UNIVERSITY OF TURKU
Faculty of Science

In silico & In vitro study to estimate Plasma Protein Binding of anti-parasitic compounds for Sleeping sickness (Human African trypanosomiasis)

Akash James
Master's Thesis
Master's Degree Programme in Biomedical Science:
Medicinal and Radiopharmaceutical Chemistry
Department of Chemistry
Faculty of Science
University of Turku
July 2022

Supervisors:
Adj.Prof. Outi Salo-Ahen
Dr. Parthiban Marimuthu

ABSTRACT

UNIVERSITY OF TURKU

Department of Chemistry/Faculty of Science

JAMES, AKASH: In silico & In vitro study to estimate Plasma Protein Binding of anti-parasitic compounds for Sleeping sickness (Human African trypanosomiasis).
Master's thesis, 71 pages
Computational Chemistry
July 2022

Human African trypanosomiasis (HAT), also known as sleeping sickness, is a disease caused by a group of parasites called Trypanosoma brucei (Tb). The two main types causing HAT are T. brucei gambiense and T. brucei rhodesiense. T. brucei gambiense is the most common form of HAT, accounting for ninety seven percent of all reported cases of sleeping sickness. According to WHO, HAT is endemic in 36 sub-Saharan African countries. The disease can lead to death during the second stage if left untreated. Several drugs have been developed for the first stage such as pentamidine and suramin, and for the second stage such as melarsoprol, nifurtimox-eflornithine combination therapy (NECT). In 2019, fexinidazole was introduced as an oral treatment for the first stage and non-severe second stage of HAT. Several antiparasitic compounds prepared by our collaborator's research group at the University of Graz, Austria showed varying levels of activity against Tb in vitro, whereas the compounds had only a moderate in vivo effect if at all. The suggested reason for the poor in vivo activities is that the compounds may bind tightly to plasma proteins, or they are metabolized before reaching the target sites for therapeutic effect. The prediction of plasma protein binding is of paramount importance in the pharmacokinetics characterization of drugs, as it causes significant changes in volume of distribution, clearance and drug half-life. Human serum albumin (HSA), an abundant plasma protein, can bind a remarkable variety of drugs impacting their delivery and efficacy and ultimately altering the drug's pharmacokinetic and pharmacodynamic properties. In this current investigation, the overall aim was to investigate whether a strong HSA binding could be a probable reason for the poor in vivo activity of the provided antiparasitic compounds. The interaction of the antiparasitic compounds with HSA was studied computationally by docking them in the HSA drug binding site I and II. The compounds with the highest docking score were additionally studied using molecular dynamics simulations to evaluate the stability of the binding interactions. Moreover, the HSA binding affinity of the compounds was estimated by calculating the binding free energies using the MM-GBSA approach. In addition, experimental HSA binding studies using Microscale thermophoresis (MST) were conducted for some of the compounds. The results of the in silico studies suggest that majority of the investigated compounds may bind to HSA with varying affinity whereas a few of them did not show favorable binding interactions with HSA. Further, none of the compounds studied in vitro by MST showed HSA binding. In sum, plasma protein binding may be the reason for the in vivo inactivity for some of the investigated antiparasitic compounds.

Key words:

Sleeping Sickness, Human Serum Albumin, Molecular Dynamics Simulation, binding free energy estimation, Microscale thermophoresis.

ACKNOWLEDGMENT

The current work was carried out in Department of Pharmacy, Åbo Akademi University, Turku, Finland in 2021-2022. Computational resources by CSC- Scientific Computing, Ltd., is greatly acknowledged.

First and foremost, I am deeply and truly grateful for my principal supervisor Adj. Prof. Outi Salo-Ahen, who opened the first door on my journey to computational drug design and guided me throughout the research. Without her, I would not have been able to explore the magnificent field of Computer-Aided Drug Design (CADD).

Further, I owe my sincere gratitude to my co-supervisor, Dr. Parthiban Marimuthu, for his guidance and support during my project work.

I would like to extend my thanks to Prof. Pasi Virta for his invaluable knowledge, to Dr. Rajendra Bhadane and Pekka Roivas for their support in running the *in vitro* experiments (MST), and to the CADD research group members and fellow colleagues for all the discussions and support.

Last but not least, I express my deepest gratitude to my parents James and Amtul, my brothers Romman and Farhan for their love and source of support in all times, and my friends for countless coffee cups and encouragement.

Thank you!

Akash James

Turku, Finland

July 2022

TABLE OF CONTENTS

LIST OF ABREVIATIONS

ADME	Absorption, distribution, metabolism and excretion
Å	Ångström, 10^{-10}m
BFE	Binding free energy
CADD	Computer Aided Drug Design
CG	Coarse grained
CNS	Central nervous system
CPU	Central Processing unit
DND*i*	Drugs for Neglected Diseases *initiative*
E_{ele}	Sum of electrostatic energies
E_{lipo}	Lipophilic energy
E_{VDW}	Van der Waals energy
FDA	Food and Drug Administration (U.S.)
FDT	Free drug theory
GPU	Graphic processing unit
HAT	Human African trypanosomiasis
H-bond	Hydrogen bond
HTS	High-Throughput Screening
IC50	Half maximal inhibitory concentration
kDa	Kilo dalton
LBDD	Ligand-based drug design
MD	Molecular Dynamics
MDS	Molecular Dynamics Simulation
MM	Molecular mechanics
MST	Microscale thermophoresis
NECT	Nifurtimox-eflornithine combination therapy
NMR	Nuclear Magnetic Resonance
PBC	Periodic boundary condition
PCR	Polymerase chain reaction
PDB	Protein Data Bank
Pf	*Plasmodium falciparum*
PK-PD	Pharmacokinetics-Pharmacodynamics
PLS	Partial least squares

PPB	Plasma protein binding
Prime/MM-GBSA	Prime/Molecular Mechanics-Generalized Born Surface Area
QM	Quantum Mechanics
QM-MM	Quantum Mechanics-Molecular Mechanics
QSAR	Quantitative Structure-Activity Relationship
QSPR	Quantitative Structure-Property Relationship
RMSD	Root-Mean-Square Deviation
RMSF	Root-Mean-Square Fluctuation
SBDD	Structure-Based Drug Design
SMILES	Simplified molecular input line entry system
Tb	*Trypanosoma brucei*
Tbg	*Trypanosoma brucei gambiense*
Tbr	*Trypanosoma brucei rhodesiense*
Trp	Tryptophan
WHO	World Health Organization
2D	Two dimensional
3D	Three dimensional
ΔG_{bind}	Binding free energy
$\Delta G_{coulomb}$	Coulomb energy
$\Delta G_{covalent}$	Covalent binding energy
ΔG_{solv}	Solvation free energy
ΔG_{VdW}	Van der Waals energy

1 INTRODUCTION

1.1 Trypanosoma brucei

Human African trypanosomiasis, also recognized as sleeping sickness is a disease, which is initiated by a group of parasites called *Trypanosoma brucei* (Tb). The two main types of Tb causing human disease are *Trypanosoma brucei gambiense* (Tbg) and *Trypanosoma brucei rhodesiense* (Tbr) (Büscher et al., 2017). There are other types of *Trypanosoma brucei*, which are the reason for animal disease and very rarely infect humans. Chagas disease, also known as American trypanosomiasis, is another disease caused by *Trypanosoma cruzi* (Pérez-Molina & Molina, 2018). *Trypanosoma brucei gambiense* is the most common cause of African trypanosomiasis accounting for about ninety seven percent of all cases. It is found in Central and West African countries and tends to cause the chronic form of the disease. *Trypanosoma brucei rhodesiense* accounts for about three percent of the cases and is found in the Eastern and Southern African countries and causes a more acute infection (Büscher et al., 2017). The parasite is transferred to humans through the bite of an infested tsetse fly. These flies are most commonly found in the rural areas of sub-Saharan Africa. Both the female and male flies can spread the disease and tend to bite during daytime. Flies can get infested by Tb, when they bite a human or an animal already infected with the Tb parasite. Both the wild and domestic animals can transmit the parasite and are significant storage of the infection. Human being are the leading reservoir for *Trypanosoma brucei gambiense*. It is unusually that the infection can be transferred from a mother to a baby or over blood transfusions, infected medical tool such as needles, or during sexual intercourse (Brun et al., 2010; Malvy & Chappuis, 2011). There are two main phases of the Tb. During the initial stage, the parasite is located away from central nervous system (CNS), and mostly in the organs, lymphatic system, and blood. During the latter stage, the parasite infects the CNS. Primarily, there can be agonizing sore at the site of the bite by tsetse fly, also known a chancre. The common indications of the initial stage involve itching, headaches, fever, pain in joints and muscle, and lymph nodes inflammation. It can cause ailment in the affected organs such as swelling of the heart muscle, inflammation of the liver and spleen. During the latter stage, several neurological symptoms can occur such as mood or personality change, psychiatric issues, instability and weakness during movement. The disease can upset sleep due to change in sleep circadian rhythm. This can gradually get dreadful as the patients is sleeping majority of the time. That is why the disease is called sleeping sickness (Brun et al., 2010).

The lack of treatment can cause the death of the patient. The intensity of the disease and how rapidly it advances from the initial to the latter stage depends on the form of the parasite. In *Trypanosoma brucei rhodesiense* infections, it is quite quick and usually require couple of weeks. However, the infection by *Trypanosoma brucei gambiense* may take numerous years as the infected person does not shows symptoms for many months or years. Once they found out, they are usually in the latter or last stage. The *Trypanosoma brucei* life cycle in tsetse fly and human being is shown below;

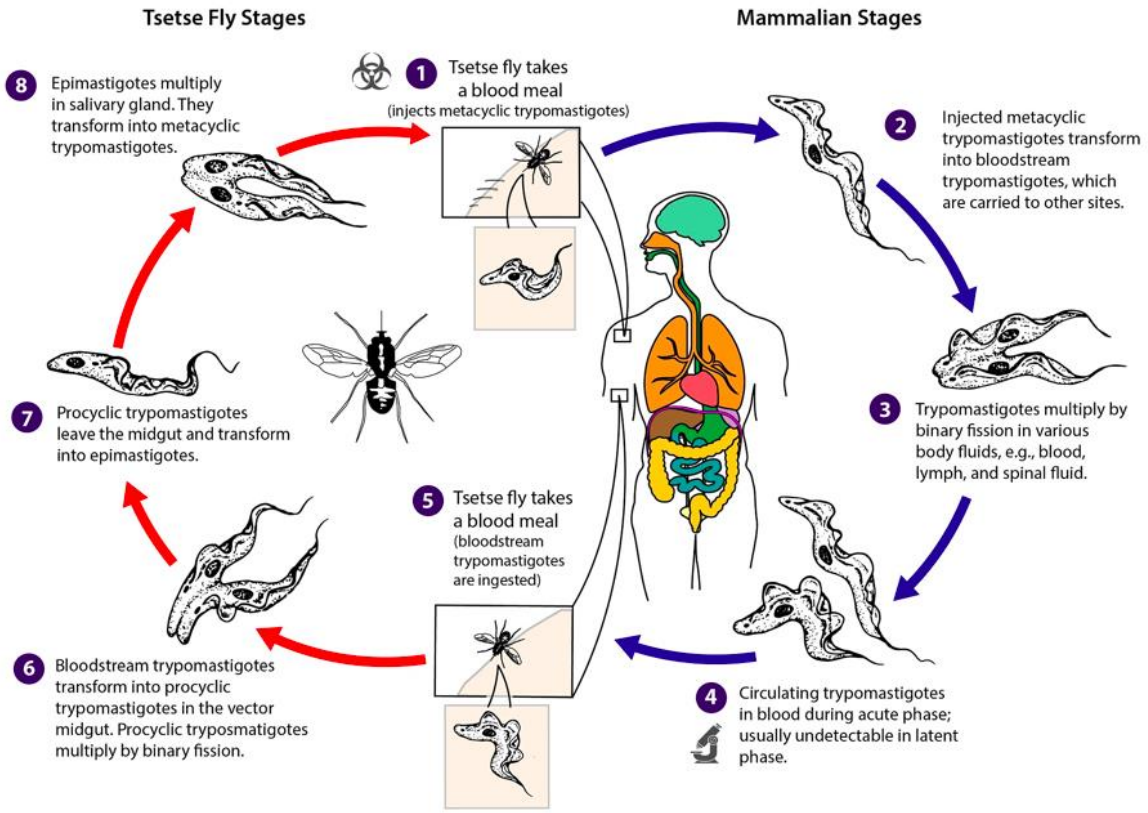


Figure 1.1: A schematic representation of a life cycle of Trypanosoma brucei in tsetse fly and human being. (Image courtesy of DPDx, Centers for Disease Control and Prevention: https://www.cdc.gov/dpdx)

The disease is detected in laboratory by analysing the blood under the microscope and looking for the parasite. This works good for *Trypanosoma brucei rhodesiense* infections, but it is tougher to do in case of *Trypanosoma brucei gambiense* infections. *Trypanosoma brucei gambiense* infections are diagnosed by studying the fluid taken from a lymph node under a microscope. The latter stage infections can be identified by probing the cerebrospinal fluid (WHO, 2019). The antibodies used against the parasite, can be used during the blood test. Card Agglutination test is used for screening for people with *Trypanosoma brucei gambiense*

infections. PCR tests that look for the parasites genetic fingerprint can be used too. There are a inadequate variety of efficient drugs available to treat the sickness. The drug used, will depend on the parasite type and the disease stage. It is harder to treat the disease once it has advanced to the latter stage and the drug used is more toxic and complicated to manage.

The precautionary measures focus on preventing the interaction with the tsetse fly, which involves wearing suitable clothing that covers the bare parts of the body and keep away from bushes in the region that tsetse flies are well-known to live in (WHO, 2019). So far, there are no available vaccines or precautionary drugs that can be taken to stop people from receiving the disease. Extreme works to regulate the disease have been headed by the World Health Organization (WHO) national sleeping sickness control programs and several pharmaceutical companies and non-governmental organizations (Malvy & Chappuis, 2011; WHO, 2019).

1.2 Previous research

A novel bicyclic compound prepared by our collaborator's research group at University of Graz, Austria, showed antiparasitic activities on screening. The compound prepared was further derivatized for activity enhancement. Out of these compounds, some showed moderate *in vivo* activity against *Plasmodium falciparum* (Pf) and *Trypanosoma brucei* (Tb) while some were ineffective *in vivo*. In case of *Trypanosoma brucei*, possible reasons for the poor (or lack of) *in vivo* activity may be their extensive binding to plasma proteins or metabolism of the compounds before they reach the target sites for therapeutic effect.

1.3 Plasma protein binding

In human and animals, blood manages the distribution of several substances inside the body. Blood can be separated into its components, a dense section of erythrocytes or red blood cells (RBC), and the less dense region of the yellowish plasma.They are divided by the buffy coat, a white layer having platelets and leukocytes, also well-known as white blood cells (WBC) (Lambrinidis et al., 2015).The plasma makes upto 55% of blood´s volume as depicted in Figure 1.2. It is the fluid portion of the blood that consist of 91 % of water, 7 % of proteins such as albumin, fibrinogen and immunoglobin, nutrients such as glucose, fatty acids and aminoacids, electrolytes such as sodium, calcium, magnesium and bicarbonates, waste products such as lactic acid, carbon dioxide and urea, and hormones (Lambrinidis et al., 2015; Toma et al., 2019).

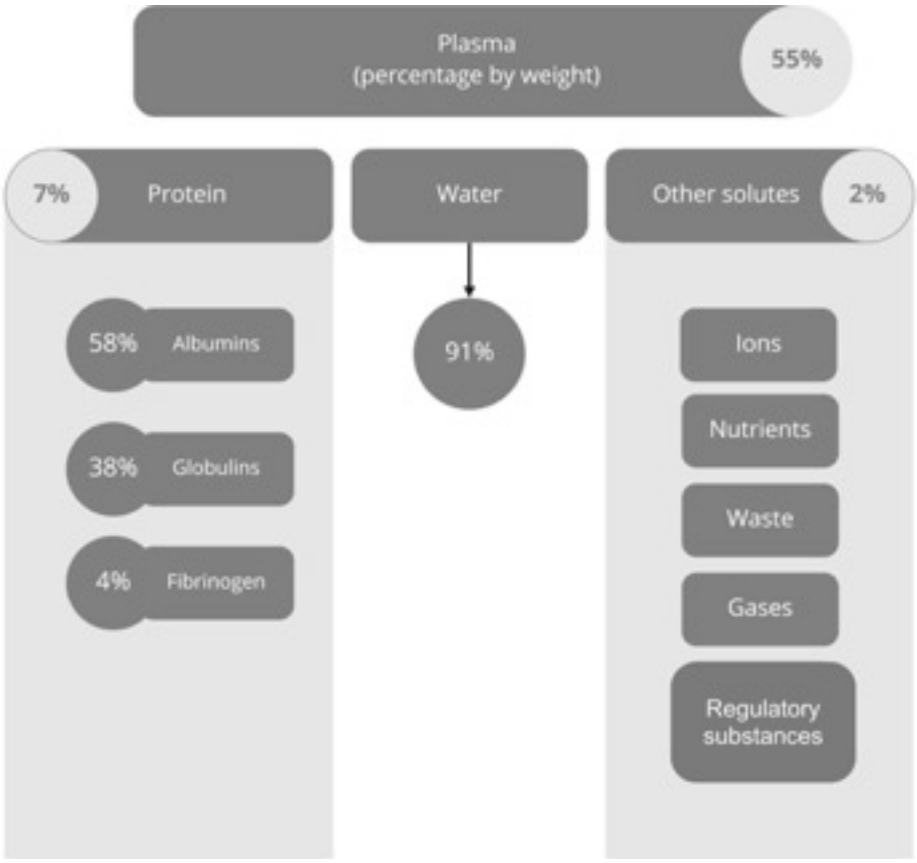


Figure 1.2: A schematic representation of the composition of plasma in human being.

The affinity of substances for plasma proteins and tissues plays a significant role in drug delivery. After a drug is absorbed into the bloodstream, it may bind to plasma proteins. However, this binding is nonspecific and reversible, meaning that multiple drugs will bind to the same protein. The bound and unbound drugs reach a dynamic balance with one another (Kumar et al., 2017).

Small variations in the fraction bound to proteins may have a big impact on the drug's bioavailability, e.g. a difference between 98 percent and 99 percent bound drug results in twice the amount of drug present in plasma (Toma et al., 2019).

Free Drug Theory (FDT) is being progressively utilized to classify pharmacokinetics/ pharmacodynamics (PK/PD) connections. According to this theory, a lack of energy-dependent processes (uptake and efflux transporters, pH gradient), free or unbound concentration of a drug in plasma (extracellular space) equals to concentration of free drug in tissues (intracellular space), and merely the free drug in the tissues is accessible for target binding sites and thus having a pharmacologic activity, as illustrated in Figure 1.3. As a result, the concentration of unbound drug in plasma should represent the pharmacologically important concentration at the target site in the tissues (Bohnert & Gan, 2013; Lambrinidis et al., 2015).

While entire drug concentration in plasma and tissues can vary greatly, for a drug with a quick rate of permeation, concentration of free drug intracellularly equals the systemic free drug in circulation. It is critical to consider that plasma proteins are not target tissues, and that drug binding to plasma proteins has little physiological impact. Drug-plasma protein binding makes a drug reservoir, but just the free drug is obtainable to tissues for therapeutic effect (Bohnert & Gan, 2013; Lambrinidis et al., 2015).

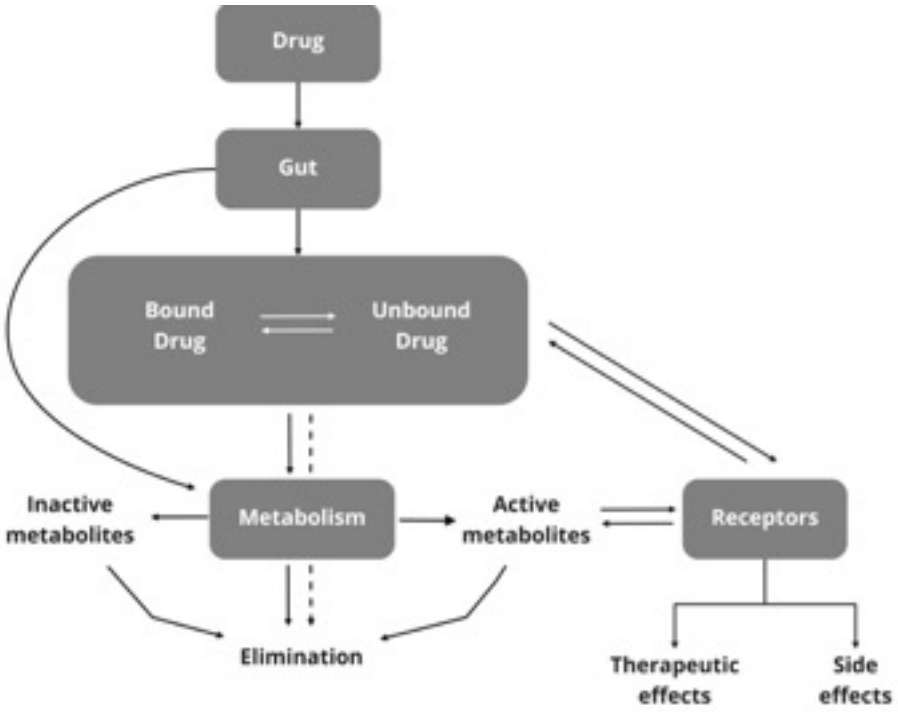


Figure 1.3: A schematic representation of a drug's pathway in the body.

Plasma protein binding (PPB) is a significant criterion for a drug's effectiveness and safety that should be explored amid each drug development program. In both animals and humans, drug molecules are either bound to proteins and lipids or are readily available. Since it is commonly assumed that the free drug concentration drives the clinical effect, the degree of protein binding can have an influence on a drug's effectiveness and toxicology (Buscher, Brigitte, 2014). PPB in animal and human plasma should be determined before starting Phase I trials, and it should be used as a supporting data set when assessing drug–drug interactions, as recommended by many regulatory bodies. As a result, PPB is regarded as a critical parameter of any current drug discovery program (Buscher, Brigitte, 2014; Kumar et al., 2017). Also, in the systemic circulation, higher protein binding is commonly due to high lipophilicity and thus, low free-drug concentration (Buscher, Brigitte, 2014).

The association between the drug and proteins in plasma is reversible. The main forces involved in these associations are hydrogen bonding, hydrophobic effect and van der Waals interactions. (Ghafourian & Amin, 2013).

PPB may adversely affect drug clearance, but in particular, only the free drug is subjected to hepatic clearance. Similarly, molecules with a high PPB have a limited renal clearance. Protein binding can also affect the amount of delivery of drug at steady state and, as a result impact half-life of drug. Different drugs may compete over the same binding pocket, resulting in competitive displacement interaction This will be especially problematic for heavily protein bound drugs with a narrow therapeutic index, such as tizoxanide, which is an active metabolite of nitazoxanide (Ghuman et al., 2005; Mullokandov et al., 2014).

1.4 Human serum albumin (HSA)

Human serum albumin (HSA) is the major drug binding protein in human blood plasma, accounting for about 60% of the mass of plasma proteins, with a typical concentration of 5 g/100 mL in the bloodstream (Artali et al., 2005). It is a highly soluble, positively charged monomeric polypeptide with 585 amino acid residues. It is produced in the liver and has a molecular weight of 69 kDa. It has a high ligand-binding potential and can bind both acidic and neutral compounds reversibly at different sites. HSA binds to basic drugs as well, but to a minor extent (Bohnert & Gan, 2013; Ghuman et al., 2005). The HSA is mainly involved in the transport of fatty acids, maintaining colloidal osmotic blood pressure, fluid exchange between body compartments, organism defense against toxic metabolites, and acts as a nitric oxide storage. It has pseudo-enzymatic properties and accounts for the majority of the antioxidant potential of human plasma (Lambrinidis et al., 2015).

Several experimental structures of HSA, both bound and unbound with endogenous compounds or drugs, are available in the Protein Data Bank (PDB) database (Moroy et al., 2012). The general structure of HSA (PDB ID:1H9Z) is shown in Figure 1.4.

The amino acid distribution of HSA shows low content of methionine (Met), glycine (Gly) and isoleucine (Ile), while charged amino acids, both acidic amino acids: aspartates (Asp), glutamates (Glu) and basic amino acids: arginine (Arg), lysine (Lys), histidine (His) are very abundant (Artali et al., 2005; Aureli et al., 2005). This facilitates the interaction of HSA with cationic and anionic ligands. Also, one tryptophan (Trp) amino acid residue in the HSA protein has been especially beneficial in fluorescence binding *in vitro* experiments (Artali et al., 2005). Carter and Ho (1994) determined the first three-dimensional crystal structure of HSA at a resolution of 2.8 Å. It is comprised of three similar helical domains (1–3) arranged

asymmetrically in the form of a heart, with approximate dimensions of 80 Å on the side, an average thickness of 30 Å and a volume of 882493 $Å^3$ (Artali et al., 2005). Out of 585 amino acid residues, residues 1-195 are in domain-1, 196-383 in domain-2 and 384-585 in domain-3 (Pragna Lakshmi et al., 2017).

Each of the three domains (1-3) has ten helices and is further divided into two subdomains (A and B). These subdomains include six helices in subdomain A and four helices in subdomain B, and they are connected by a long-extended loop. The individual helices are linked by a conserved sequence of 17 disulfide bridges, which maintains the protein structure in all mammalian organisms. The improved thermostability of HSA is also due to the disulfide bridges. Subdomains 1 and 2 are perpendicular to one another, with hydrophobic and hydrogen bond interactions connecting subdomain 2A to the interface area between subdomains 1A and 1B. Domain 3 associates only with subdomain 2B, and domains 2 and 3 form a Y-shaped assembly. Domains 1 and 3, which are only linked by a few connections, are separated by a large channel formed by subdomains 1B, 3A, and 3B (Artali et al., 2005; Lambrinidis et al., 2015; Moroy et al., 2012).

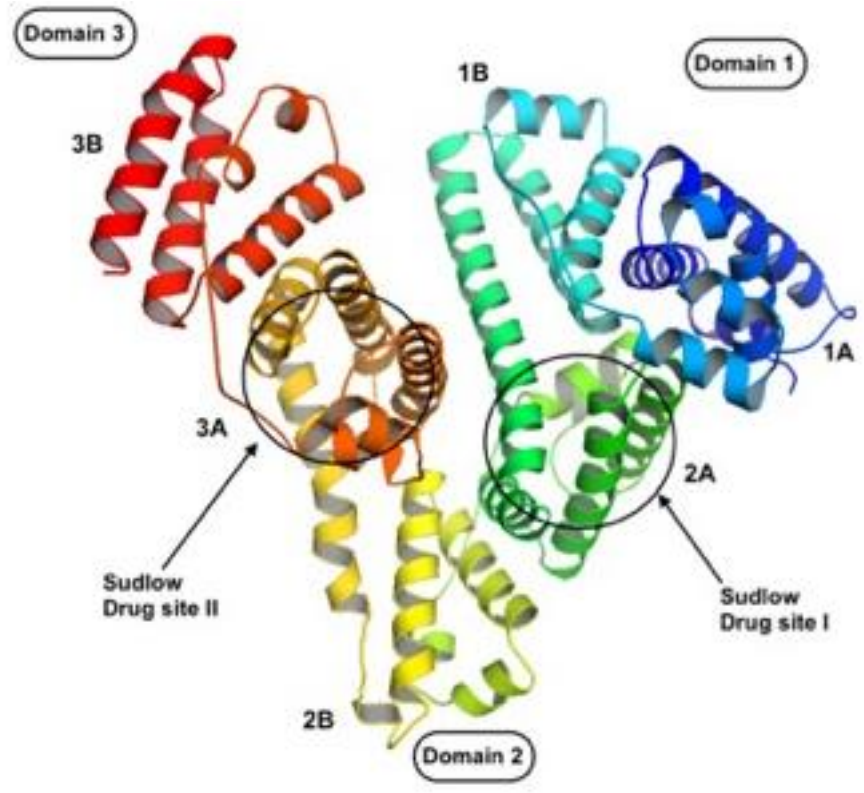


Figure 1.4: General view of the structure of human serum albumin (HSA) (PDB ID: 1H9Z). Drug binding sites in the protein are shown with circles around them.

Sudlow and Sjoholm (Rimac et al., 2017; Sjoholm et al., 1979) have proposed the presence of two HSA high-affinity binding sites (site I and site II) based on fluorescent probes that bind specifically to each of these sites. In practice, most drugs bind to one of these two primary binding sites located in the hydrophobic cavities of sub-domains 2A and 3A and able to accommodate a large range of chemical compounds (Artali et al., 2005; Moroy et al., 2012). In addition, the three homologous domains of HSA protein have seven fatty acid (oleic acid) binding sites. Several X-ray analyses of HSA-fatty acid complexes have revealed this reliable list of fatty acid binding site positions. The location of these sites are, one in sub-domain IB, one located between sub-domains 1A and 2A, two sites in 3A (constituting to the Sudlow's drug binding site II), one in sub-domain 3B, one between the 2A and 2B interface, and the last within the drug-binding cavity of sub-domain 2A (Sudlow's drug binding site I).

Drug site I, also known as the warfarin binding site, is formed by a pocket located in subdomain 2A of HSA. Ligands that have affinity to the highly flexible site I are mainly negatively charged large heterocyclic or dicarboxylic acid compounds such as azapropazone, warfarin and dansylamide. This could be due to the existence of Arg257 and Lys195/199 residues at this binding site, which bestow a partial positive character (Deeb et al., 2009; Moroy et al., 2012).

Binding site II is situated in subdomain 3A of HSA. It is also known as the benzodiazepine binding site and often binds aromatic moieties and carboxylic acids having a negatively charged carboxylate group at the termination of the molecule, far from a hydrophobic center. Binding site II takes fewer ligands than site I, which could be due to site II is positioned in a subdomain that relates to a more unstable protein conformation compared to site I. In addition, the compounds such as ibuprofen, diazepam, phenylbutazone and ketoprofen binds to this site selectively (Deeb et al., 2009).

A high degree of conformational flexibility due to the absence of beta sheets in HSA structure has been observed by a MD simulation study (Carter & Ho, 1994; Vallianatou et al., 2013). This high flexibility may explain HSA's incredible binding ability, which is expressed in several binding sites, according to the crystallographic evidence.

1.5 Computational prediction of drug binding to human serum albumin

The process of discovering and developing a novel drug is typically regarded as being exceedingly difficult and requires a significant effort and resources. Computer-aided drug design (CADD) techniques are thus now routinely employed to improve the effectiveness of the drug research and development process. Structure-based drug design (SBDD) and ligand-based drug design (LBDD) are the two main methods used in CADD. SBDD methods such as

molecular docking and molecular dynamics simulations, require understanding of the three-dimensional (3D) structure of the biological target of interest. In LBDD, the target protein's 3D structure is unknown, but the ligand or ligands that bind to the intended target location are known. LBDD includes methods such as quantitative structure-activity relationship (QSAR) and pharmacophore modeling (Jain, 2017).

Zsila et al. (2011) used a hybrid chemometrics-molecular docking approach that allows to predict whether a given ligand binds to HSA and what its likelihood is for binding to site I or site II. Based on the experimental HSA-binding data, a dataset of 163 small compounds were used, of which 62 compounds in the collection binds to HSA site I, 38 binds to site II, and 63 compounds have no or poor affinity for HSA. For each compound, a total of 3250 descriptors were generated, and 45 descriptors were selected based on the physiochemical properties such as logP, compound aromaticity and number of carboxylic acid moieties, and 3D descriptors to make a distinction in the shapes of site I and site II. These selected descriptors contributed to HSA binding affinity and used in support vector machine (SVM) model building. The 100 compounds from the dataset were randomly split into 60 % training and 40 % test set, and the remaining 63 compounds were chosen as the external validation set. They reported that the model showed an accuracy of 100% for training set and a test set accuracy of 78 % was achieved. The validation of the model on external dataset predicted compounds with an accuracy of 81%. The available X-ray crystal structures of ligand-HSA complexes were carefully examined to identify potential variability at the binding sites since choosing the right protein geometry is essential for precise complex geometry estimation. There has been some side-chain variability at the binding sites, based on the comparison of X-ray data of the ligand-HSA complexes. The variations at HSA binding site I comprises of the rotation of residues such as aromatic (Tyr150, Trp214) and basic (Lys199, Arg218, Arg222), depending on the size and chemical structure of the ligand. The HSA site II, which binds both negatively charged (such as ibuprofen, indoxyl sulphate) and neutral substances (such as propofol, diazepam), the HSA side-chain conformational changes can also be seen due to the ligand. The binding of ibuprofen and diazepam at site II differ significantly, revealing significant side-chain alterations due to ligand. The carboxylic moiety of Ibuprofen forms a salt bridges with HSA residues Lys414 and Arg410. In contrast, the binding of diazepam, which lacks an acidic functional group, occupy alternative steric position does not involve binding to of basic residues such as Lys414 and Arg410. Furthermore, diazepam binding causes substantial rotation of HSA binding pocket residue such as Leu387 and Leu453, enabling the phenyl group of diazepam to enter the binding pocket from rear.

Aureli et al. (2005) suggested a site II specific 3D-QSPR (three-dimensional quantitative structure-property relationship) model for predicting HSA binding of structurally related compounds. The 37 compounds in the training set included 10 novel interleukine-8 (CXCL8) inhibitors, 17 structural analogs of diflunisal and 10 classical non-steroidal anti-inflammatory drugs (NSAIDs) (Aureli et al., 2005; Lambrinidis et al., 2015). The test set contains seven compounds having varied biological activities but a structural similarity close to the training set compounds. The flexible protein-ligand molecular docking was used to study whether the training set compounds preferred to bind to Sudlow's drug site I or II. The results suggested that the majority of the compounds would preferentially bind to site II, which is consistent with the experimental studies that showed that despite lacking carboxylic group, these novel class of interleukine-8 inhibitors had an extraordinarily high affinity for HSA and that they resemble structurally the site II binding NSAIDs. The training set compounds interacted extensively with the hydrophobic residues (Leu453, Ile388, Asn391, and Leu407) in site II binding pocket. Additionally, the hydrogen bond and electrostatic interactions with site II amino acids, such as Tyr411, Arg410, Ser489 and backbone of Leu430 appear to be particularly important in defining binding affinity to HSA site II. It also showed that the physicochemical descriptors, for instance, hydrophilic and hydrophobic regions of the compounds are inversely and directly correlated with HSA binding. The substantial influence of acceptor groups on the compounds shows that the binding to HSA site II takes place through hydrogen bonding with the residues Arg410, Tyr411, Ser489 and backbone of Leu430. The specific residues Arg410 and Leu430 of HSA site II form hydrogen bonds with high affinity ligands containing oxygen atom of carbonylic and sulfonylic groups.

1.5.1 Molecular docking-based prediction

Molecular docking is a widely used structure-based drug design (SBDD) tool for predicting protein-ligand binding interactions. It has become a widely used drug discovery tool especially due to the relatively low costs associated with its use and the perceived simplicity of use. It involves identification of a low-energy binding position of a small molecule or a ligand within the active site of a biological macromolecule or receptor. Molecular docking in drug discovery is usually applied for virtual screening where large databases of 3D structures of small molecules are searched through to find those fitting the binding pocket of the receptor by using an appropriate docking program (Halperin et al., 2002).

Ligands used in molecular docking can be small molecules, peptides, peptidomimetics or DNA etc., but the larger the ligand the more challenging it is to achieve a reliable and accurate

docking result. Receptors on the other hand are 3D structures of proteins that are usually published in the PDB database. Docking can be categorized into flexible docking and rigid docking. In rigid docking, both the receptor's and the ligand's internal geometries are considered as rigid, like in case of a lock and a key. On the other hand, in flexible docking, an enumeration on the rotations of the fragments in one of the molecules (usually the smaller one, i.e. the ligand) is performed, while energy is calculated for every rotation and the most optimum pose is selected (Raval & Ganatra, 2022). Some docking methods can also take into account the flexibility of the receptor to some extent, e.g. induced-fit docking.

Typically, ligand-protein docking software consists of two main components that work together to find the most stable binding conformation of the two partners (Kitchen et al., 2004; Sousa et al., 2006). A search algorithm is responsible for exploring the conformational space by making systematic changes to the molecule's orientation and conformation, in order to generate a large number of poses of a molecule in the binding site. There are many searching algorithms that have been developed. Some are more precise than others but all of them are meant to search within the conformational space for the most stable complex between the two partners.

The strength of the binding affinity of the docked poses is assessed using scoring functions (SF) that estimate the ligand-protein binding free energy (Kitchen et al., 2004). It merely quantifies the energy of protein-ligand interactions by summing basic bonding and non-bonding (hydrogen-bonds, electrostatics, van der Waals, hydrophobic, pi-pi) parameters of the two entries composing the complex. Most scoring functions are very similar in concept but have different levels of simplifications and different variation of molecular mechanics force fields, that are used to estimate the energy of different poses in the binding site (Patrick Walters et al., 1998; Taylor et al., 2002).

The scoring methods can be categorized into empirical, knowledge-based and field-based. They mostly employ the linear regression model. Empirical SF evaluate a ligand-protein complex binding affinity by combining the essential energetic parameters for complex binding, such as hydrogen bonds, ionic interactions, hydrophobic effects, and steric effect (Guedes et al., 2018). The knowledge-based SF, based on statistical principle of inverse Boltzmann, acquire the relevant pairwise potentials from 3D structures of a vast set of ligand-protein complexes. The scores are calculated by penalizing the repulsive and favoring the preferred interaction among each atom of protein and ligand within a specified cutoff. The force field-based SF calculate the total of the non-bonded interactions (van der Waals and electrostatics) between ligand-protein atom pairs, which take into account the significance of enthalpy to energy. The electrostatic and van der Waals terms are represented by Coulombic and Lennard-Jones

potentials, respectively. However, it ignores entropy and the solvent effects, which are enhanced by including ligand torsion entropy and the solvation/desolvation as defined by implicit and explicit solvent models (Li et al., 2019; Meng Zhang, H. X., Mezei, M., & Cui, M., 2011).

The complexes calculated using molecular docking techniques can be more or less stable depending on various factors, but the binding affinity calculated with the scoring function can be a good indicator of the binding strength and the stability. Usually, a high negative value indicates a strong binding energy between the investigated ligand and the receptor. On the other hand, a positive value would mean an artifact clash or repulsion between the atoms of the participating molecules in the docking. However, docking programs are fairly robust and they will eliminate any unrealistic poses from the final list of ligands.

After calculating the binding position between the ligand and the receptor, the main purpose is to study the interaction between the molecules. In this way, the molecular docking method allows screening of entire libraries with a much lower cost than experimental methods. Regardless of the advantages, docking-based virtual screening is not the substitute of the actual laboratory experimental screening. In fact, these two approaches are highly complementary.

1.5.2 Molecular dynamics simulation

Molecular dynamics simulation (MDS) is a significant theoretical tool to visualize detailed physical motion and interactions of atoms of biological macromolecules over time at a specified temperature. It can also be useful to refine models coming from docking experiments or generated by experimental techniques or to complement biological experiments. Everything known about biomolecules at the atomistic level are built with models. Biological structures are all refined using molecular modeling. Sometimes structural biology tools cannot be used, for example a structure cannot be crystallized by X-ray crystallography, a protein is not soluble for an NMR experiment, or the molecule is very small and does not has enough contrast for cryo-EM. However, using homology model to construct the atomic-resolution model of the desired protein structure, followed by molecular dynamics simulation can capture the behavior of the biological macromolecules and provide necessary information in full atomic detail (Hollingsworth & Dror, 2018; Patodia, 2014).

Based on the number of degrees of freedom, MD simulations can be described into different approaches. An atomistic approach where one particle describes one atom can be used for analyzing the atomic-level behavior of a molecular system. The atomistic approach belong to

the category of classical molecular mechanics (Joshi & Deshmukh, 2020; Salo-Ahen et al., 2021).

Generally, one mole of a macroscopic sample in the experimental environment has Avogadro's number (6 x10^{23}) of molecules and the average taken in the measurement is an average of the total number of molecules. That is, the laboratory experiment measures an average property, averaged over a large number of molecules in a macroscopic sample usually over the time of measurement. However, in MD simulations conformations are generated for one molecule in time. The ergodic hypothesis states that a single system developing in time is replaced by a large number of relative replications of the same system that are considered at the same moment. That is, an average of a large number of conformations is equal to the average of the same property in time. Hence, the time set for MDS should be sizeable to allow the generation of enough conformations to represent the system.

Therefore, the ultimate goal of MDS is to generate a large enough ensemble of conformations of a system as depicted in Figure 1.5. In such a way, an accurate value of a property of interest can be calculated by applying the Newton's second equation of motion as presented in the following equation:

$$\text{Force} = \text{mass} \times \text{acceleration}$$

$$F = ma$$

According to Newton's equation of motion, the particle of mass 'm', experiences an acceleration 'a' in the path of the applied force 'F'.

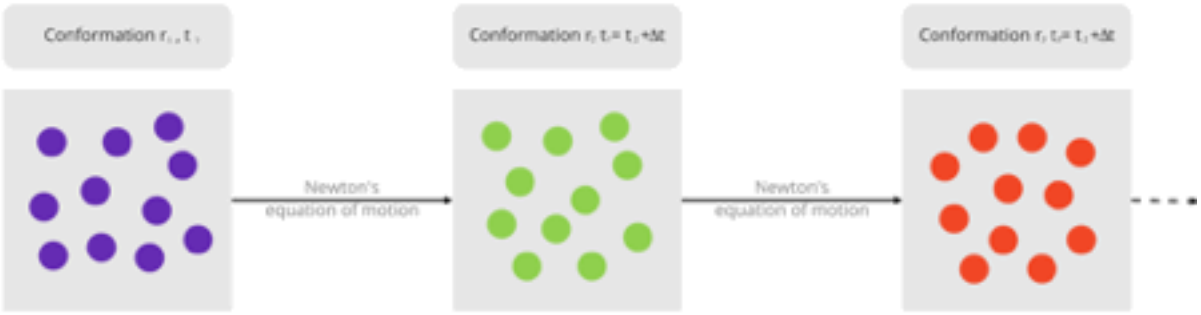


Figure 1.5: MD simulations ensemble of conformations generated with time.

The total force F acting on an individual particle at each MD simulation step can be obtained from a potential energy function *V(total),* also called a force-field (Lin & Mackerell, 2019; Ponder & Case, 2003). There are several force fields accessible, with varying degrees of

complexity and aimed at treating various types of systems. However, a general force field (González, 2011) can be expressed as follow:

$$V(total) = \sum_{Bond} \frac{1}{2} k_b(r - r_o)^2 + \sum_{Angles} \frac{1}{2} k_a(\theta - \theta_o)^2 + \sum_{Dihedral} \frac{V_n}{2} (1 + \cos(\eta\emptyset - \delta)) + \sum_{electrostatic} \frac{q_i q_j}{r_{ij}} + \sum_{Lennard-Jones} 4\varepsilon_{ij} \{ (\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^{6} \}$$

As indicated in the above equation, the parameters contributing to the total potential energy function V(r) are split into bonded interactions and non-bonded interactions, as shown in Figure 1.6. Bonded interactions include bond length, bond angle and dihedral angles to keep the chemical groups sterically stable during the MD simulation. Bond stretching is computed by a harmonic potential, which depends on the square of the separation distance between the atoms. The term k_b is a force constant for bond stretching. Angle bending is also described by a harmonic potential for the reference bond angle θ_O. The term k_a is a force constant for angle bending. This parameter is defined for a group of three bonded atoms. Dihedral angles are expressed by a torsion or 1-4 interaction function between a group of four bonded atoms and are computed by a Fourier potential such as cos function. The parameters η is the number of minima when the bond is rotated, δ is the phase factor that determines the position of minima. The term ϕ is the angle formed by the two planes. The last two expressions represents coulombic and van der Waals interaction (Levitt et al., 1995).

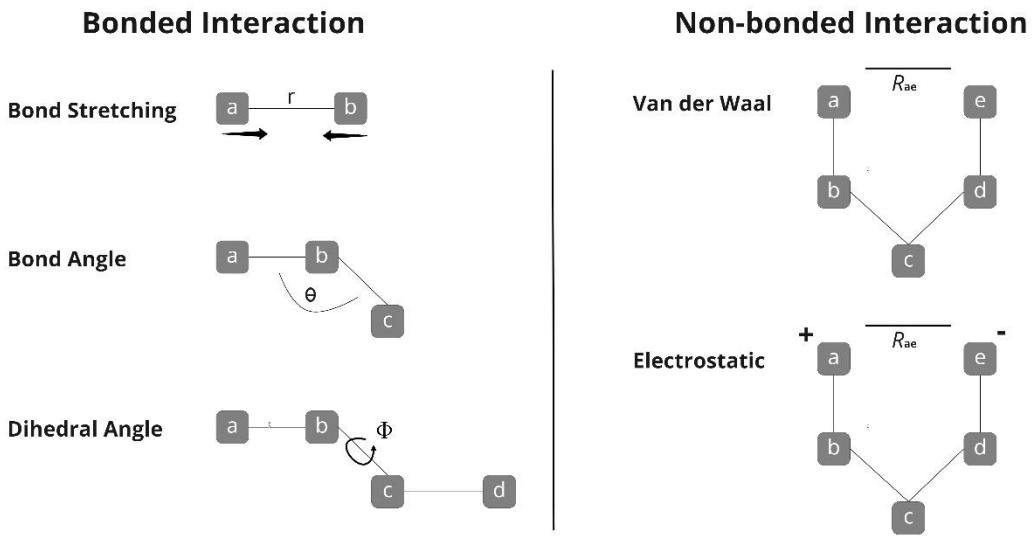


Figure 1.6: Bonded and non-bonded interaction parameters.

The non-bonded interactions are long-range interactions between a pair of atoms in the macromolecule. These are usually calculated over every pair of atoms in the system except for the atoms that are bonded to each other or separated by two or three bonds (Sagui & Darden, 1999). These interactions include both van der Waals and electrostatic interactions. The van der Waals interactions are described by the Lennard-Jones 12-6 potential (Heinz et al., 2008; Wang et al., 2020) which contains the depth of the potential well ε, finite inter-particle distance σ at which the potential is zero, and distance r between particles. The electrostatic interactions are estimated among the pairs of point charges q using Coulomb's law. Non-bonded interactions are the most time-consuming to calculate, as for N number of atoms in a system, N^2 interactions are required to be calculated. However, it is known that non-bonded interactions also decay very fast with the distance. The van der Waals interaction decays faster than charge-charge Coulombic interaction ($1/r$) and thus the long-range contributions are ignored. This can be done by applying a cut-off value at which non-bonded potential and forces are not any more calculated. In order not to completely neglect the electrostatic interactions coming from a distance larger than the cutoff radius r_c, usually the Ewald summation method (Darden et al., 1993) is used to sum up electrostatic interactions in real space (short range) and in reciprocal space (long range), which results in a faster convergence (Essmann et al., 1995; York et al., 1993).

However, Predescu et.al. (2020) established U-series, a new and more effective decompositions for coulomb potential method with an improved efficiency and performance with around half

of the computational requirements, compared to conventional (Ewald) decomposition method. It decomposes Coulombic interaction ($1/r$) into a near and far components, resulting in a variety of desirable properties for the resultant approximation. It has a consistently limited relative error for $r > r_c$ and is uniform along the entire real space. However, the majority of the inaccuracy occurs from the interactions with particles located immediately outward the radius $r \leq rc$. The u-series approach is less accurate for systems containing lipid bilayer, where some charge structure is conserved on scales greater than the cutoff radius r_c.

Subsequently the total force F acting on the system is computed as a negative gradient of the potential energy function V(total), which allows to calculate the acceleration of particles based on the equations below;

$$F = -\nabla V(total)$$

$$F = \frac{dP}{dt} = m\frac{d^2r}{dt^2}$$

Along with the initial positions of particles acquired from the simulated structures and initial velocities produced according to atom types and simulation temperature, the acceleration is used to compute for next positions and velocities in the simulation run. The algorithms used for the incorporation of classical Newton's equation of motion apply an approximation known as Taylor Series expansion for the positions, velocities and accelerations. Some of the well-known algorithms for this purpose are Verlet algorithm, Leapfrog algorithm and Beeman's algorithm (van Gunsteren & Berendsen, 1990).

The most common force fields used for biomolecular simulations are AMBER (Hornak et al., 2006), CHARMM (Best et al., 2012), GROMACS (Oostenbrink et al., 2004), OPLS (Beveridge & Jorgensen, 1988) and for the coarse-grained setting a Martini force field (Marrink et al., 2007). It is suggested not to mix and match force field parameters. The parameters in the force field come from experimental or ab initio studies of small molecular compounds, e.g., bond lengths and bond angles come from crystal data, force constants for bonds and angles come from spectroscopic data (IR, Raman) and the charges might come from a quantum mechanics electrostatic potential. The parameters developed for small compounds should be transferable to a new set of molecules (Lin & Mackerell, 2019).

The span of time step Δt, used in a simulation is needed to be considerably smaller than the fastest movement in the system, for instance the bond vibrations. The integration of an appropriate time step is critical. With an overly small-time step, the events might never happen

as it will take longer to simulate whereas with a too large a time step, the system will get unstable. The time step is usually set to be around 1-5 femtoseconds.

Another aspect that governs the outcome of a simulation is keeping the temperature and pressure close to the experimental conditions. To regulate (maintain) these physical properties during a MD simulation run, several statistical ensembles are employed. Ensembles such as NVE, NVT and NPT are the most common settings. A simulation in the NVE ensemble, also known as the microcanonical ensemble, is performed at a constant number of particles N, constant volume V and energy E. This ensemble does not control the temperature or the pressure of the system, and thus, is not suggested for equilibration. The NVT ensemble (NosÉ, 2002), referred as canonical ensemble, maintains the number of particles N, volume V and temperature T constant during the MD simulation. In this ensemble the energy of endothermic and exothermic processes is exchanged with a thermostat. The role of a thermostat is to guarantee a suitable average temperature and an appropriate extent of temperature fluctuation. Most commonly used thermostats are the Berendsen weak-coupling thermostat (Berendsen, 1981), velocity-rescaling temperature (Bussi et al., 2007) and Nose-Hoover temperature coupling (Nose, 1984; Hoover, 1985). The NPT or isothermal-isobaric ensemble (Yamauchi et al., 2019) keeps the number of particles N, pressure P and temperature T conserved. The pressure is associated to the system's volume and the particles interaction with each other. This allows the pressure to be adjusted by varying the volume of the system and scaling the position of the molecules (Victor, 2007). Examples of barostats are the Berendsen barostat (Berendsen, 1984), stochastic cell rescaling (Bernetti & Bussi, 2020) and Parrinello-Rahman (Parrinello, 1981; Nose, 1984).

To sum up the MD simulations workflow, the starting configuration R_0 can be attained from experimental data from structural biology. The initial velocities v_0 are usually taken from random distribution (Maxwell distribution) with the kinetic energy corresponding to the desired temperature of the simulation. The force acting on each atom in the MD simulations can be calculated by the potential energy function V(r), in order to acquire the accelerations a. The new configuration can then be achieved from the previous position repetitively until a desired length trajectory is produced.

1.5.2.1 Molecular dynamics trajectory analysis

Several properties such as root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), number of hydrogen bonds, can be analyzed to extract valuable information about the

protein-ligand interaction.

RMSD estimation is the most usual way to measure a structure's stability during the timespan of the whole simulation. It shows the average variation in dislocation of the protein Cα atoms. This dislocation in each specific simulation frame is measured against the backbone atoms of the initial frame (reference structure) of the simulation. This gives insight into the structural conformation throughout the simulation and indicates if the simulation has reached equilibrium. Usually, the RMSD values of 1-3 Å are considered as standard, whereas values higher than 3 Å may show that a protein (or part of the protein) goes through large conformational changes during the course of the simulation. The equation shown below is used to calculate RMSD for the whole trajectory (Bowers et al., 2007):

$$\boldsymbol{RMSD} = \sqrt{\sum_{\boldsymbol{i=1}}^{\boldsymbol{N_{atoms}}} (\boldsymbol{r_i(t_x)} - \boldsymbol{r_i}(\mathbf{t}_{\boldsymbol{ref}}))^2 \ / \boldsymbol{N_{atoms}}}$$

where, N_{atoms} is the number of atoms whose positions are being compared, r_{i} (t) is the position for atom i at time t, t_{ref} is the initial or reference frame at time t=0 and t_x is the recorded frame at a certain time. Time-specific and averaged RMSD values are given for the atoms in the molecule.

Alike RMSD, RMSF can also be used to analyze the most and the least fluctuating residues in a protein Cα atoms. The regions that generally show high RMSF are loops and the termini. In contrast, secondary structures of proteins such as helices and beta-sheets show less fluctuation since they are structurally rigid. The equation shown below is used to calculate RMSF (Bowers et al., 2007).

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T} < (r_i'(t)) - r_i(t_{ref}))^2 >}$$

The equation presents RMSF calculated for the trajectory time T, where the reference time is *tref*, the position of a residue *i* is *ri*, and the position of atoms in residue *i* after superimposition on the reference frame is *r′*. The angle brackets (<, >) specify that the average of the square distance is taken over the selection of Cα atoms in the protein residue. RMSF is averaged over time and gives a value for each particle (residue) *i*.

Hydrogen bonds play a vital role in ligand binding by bringing stability and rigidity to the binding pocket.

1.5.2.2 Binding free energy estimation

The assessment of the binding strength among the protein and ligand is one of the fundamental components of essentially all structure-based approaches. The molecular mechanics generalized Born surface area (MM-GBSA) is the most widely used techniques developed by Kollman et al. (1998) to approximate the binding free energies (BFE) between two molecules such as a ligand and a biological macromolecule.

Through this approach, the generalized Born (GB) implicit solvent model is used to estimate ΔG_{solv}, and a classical molecular mechanics (MM) force-field is used to estimate ΔE_{MM} components. A large number of snapshots generated from an MD simulation run are used to calculate the ΔG_{solv} average. After eliminating the explicit solvent molecules, the binding free energies of these structures are estimated in the implicit solvent, followed by assigning an average free energy of binding ΔG_{Bind} as illustrated below;

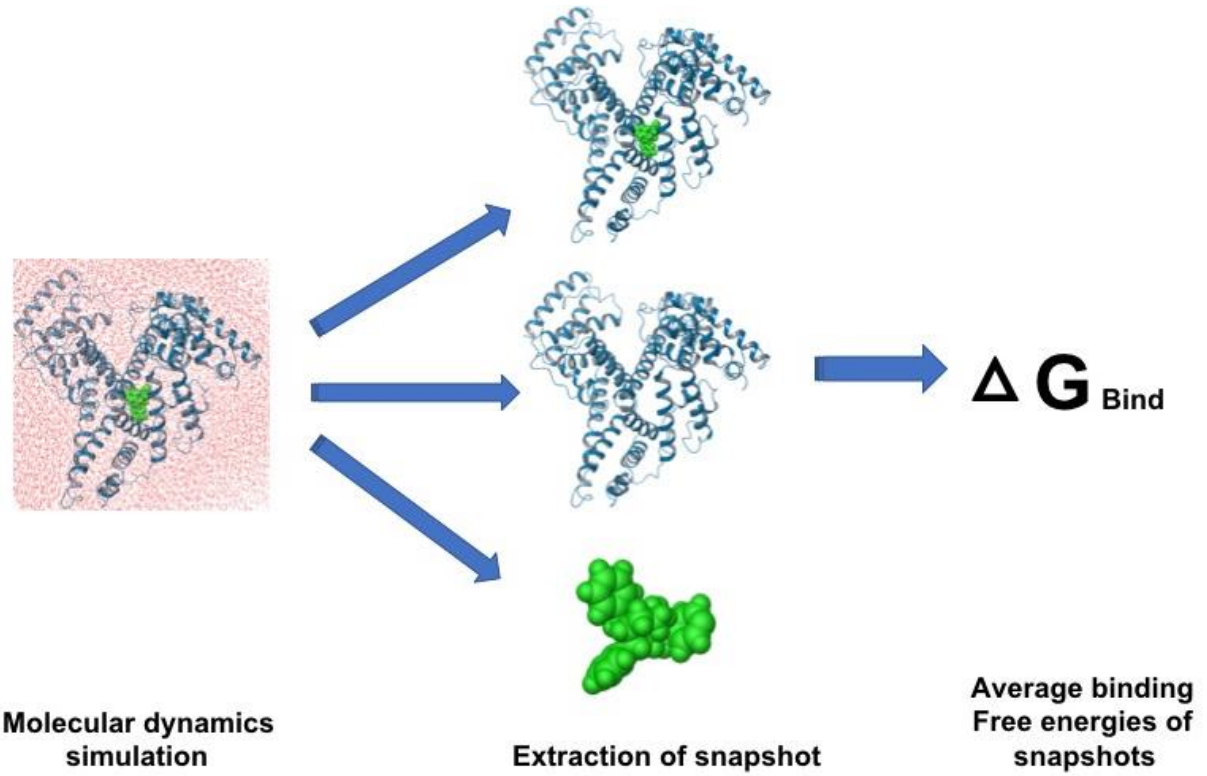


Figure 1.7: MM/GBSA flowchart.

Equation below is used to estimate the binding free energies calculated by MM-GBSA methods (Li et al., 2011):

$$\Delta\Delta G_{binding} = \Delta G_{complex} - (\Delta G_{protein} + \Delta G_{ligand})$$

where $\Delta G_{\text{complex}}$, $\Delta G_{\text{protein}}$ and ΔG_{ligand} represent the delta free energy of the complex, the protein and the ligand, respectively. Each term can be estimated according to Tuccinardi (2021) as in the following equations;

$$\Delta G_{binding} = \Delta G_{MM} + \Delta G_{sol} - T\Delta S$$

$$\Delta G_{MM} = \Delta G_{ele} + \Delta G_{vdw}$$

where the term ΔG_{MM} denotes the molecular mechanics free energy that splits into electrostatic interaction (ΔG_{ele}) and van der Waals interaction (ΔG_{vdw}) energies.

$$\Delta G_{sol} = \Delta G_{GB} + \Delta G_{SA}$$

ΔG_{solv} term in the above equation stands for solvation free energy that includes polar contributions of electrostatic solvation energy (ΔG_{GB}) and non-polar contributions of the non-electrostatic solvation (ΔG_{SA}). Additionally, several other components also contribute to the total energy such coulombic, covalent binding, van der Waals, lipophilic, generalized Born electrostatic solvation, hydrogen-bonding correction and pi-pi packing correction as stated by (Li et al., 2011).

1.5.3 Pharmacokinetic property prediction

Drugs bind to proteins in order to produce their pharmacologic effects. As a result, a drug needs to be structurally capable of interacting with the suitable protein in order to have a pharmacologic effect. However, a therapeutically effective agent needs more than just a drug's capacity to bind with a protein. The field of pharmacology known as pharmacokinetics studies the four stages that drugs go through as they move through the body. These stages include absorption, distribution, metabolism, and excretion (ADME). The delivery of a drug from its point of administration to the blood system is referred to as absorption. Distribution refers to the drug's passage through the bloodstream and into body tissues. The procedure that breaks down the drug is called metabolism. Enzymes such as cytochrome P450 (CYP450) are accountable for the biotransformation of majority of drugs. Excretion refers to the process by which a drug is eliminated from the body via kidneys (Backes, 2007).

Several in silico tools are available to predict ADME specifications from chemical structure. The web tool such as SwissADME (Daina et al., 2017) is publicly available ADME and pharmacokinetics resource. It considers several physiochemical properties, which includes lipophilicity, polarity, size, solubility, flexibility and saturation. Likewise, admetSAR (Cheng et al., 2012) is another open source tool to predict the ADMET properties of molecules. The physiochemical properties taken into account by admetSAR are amount of H-bond donor and acceptor, molecular weight, topological polar surface area and octanol-water partition (log *P*). Similarly, QikProp is a rapid and convenient ADME estimation module of Schrödinger. It predicts a total of 46 tangibly meaningful descriptors and pharmaceutically related properties of organic molecules. Additionally, it provides recommended values to correlate a certain molecule's properties with those of 95% of known drugs. The QikProp prediction method was developed by Professor William L. Jorgensen and Professor Erin M. Duffy. The Monte Carlo method was used to identify important physically descriptors for predicting aqueous solubility, solvation free energies, and log P octanol/water. This method averaged several descriptors such as solute-water Coulomb and Lennard-Jones interaction energies, solvent-accessible surface area (SASA) and its components (hydrophobic, hydrophilic and aromatic), total number of solutes as H-bond donor and H-bond acceptor. This method, however, has the practical drawback of being slow due to use of more computational resources. As a result, they created algorithms for quick estimation of key descriptors, particularly hydrogen-bond counts. This take account of electronic and steric surroundings of the H-bonding sites, as well as the possibility of intramolecular H-bonding. The input requires a 3D structure of the compound, followed by performing SASA yielding its components, such as solvent-accessible volume, weakly polar surface area for sulphur, halogens and phosphorous, hydrophobic surface area, hydrophilic surface area and carbon-pi surface area. Multiple Linear Regression model was performed for aqueous solubility on a training dataset of 317 wide-ranging organic molecules, with a special emphasis on heterocyclic compounds and drugs. The overall MLR model has an r^2 and rms value of 0.90 and 0.63 log unit respectively, with a standard deviation error of 0.48 log unit.

The anticipated log S (Solubility) values from QikProp (QP log S) are typically within a few tenths of a log unit for the vast majority of conformers, with the largest ranges very seldom covering more than one log unit. Even though there are variations, extended structures usually produce outcomes that are more consistent with the experimental data. The QikProp regressions were created with extended conformers, which are typically generated by standard 2D to 3D structure conversion programs.

Out of the 46 properties, the only property of interest in the current thesis work was the prediction of a molecule's binding to HSA (QPlogKhsa) (William L. Jorgensen & Duffy, 2002). As discussed earlier, the effectiveness of a drug molecule may be disturbed by the level to which it binds to the blood plasma proteins such as HSA. This significantly lowers the magnitude of the free drug in general blood circulation and the amount of the drug reaching the target. Hence, the more bound a drug is to blood plasma proteins, the less efficiently it can pass through cell membranes or diffuse to reach the target to produce its effect. Since HSA is the most abundant plasma protein and because it has a high capacity to bind various molecules, the prediction of drug binding to HSA has been estimated to represent the overall plasma protein binding. The logarithm of HSA predicted binding constant $\log K_{HSA}$ parameter has a recommended value range between −1.5 to 1.5 for 95% of known drugs. The more positive value represent strong binding of the ligand to HSA due to the reaction equilibrium constant toward the complex, whereas lower values indicate weak binding between ligand and HSA (Colmenarejo et al., 2001).

1.5.4 Examples on computational studies of drug-HSA binding

As discussed earlier in the section 1.4., the HSA protein has several ligand binding sites. Out of them, drug site I and II are the most common for ligand binding. The pharmacokinetics and pharmacodynamic characteristics of chiral compounds can be significantly impacted by enantioselectivity in protein binding. Among the plasma proteins, HSA has the ability to demonstrate the greatest possible enantioselectivity and the capacity to bind to a range of drugs. Therefore, using rigid-receptor docking techniques for flexible proteins (such as HSA) may result in an inaccurate pose because of conformational changes in the binding process between the receptor and the ligand. The binding of warfarin enantiomers (R-, S-) cause HSA to experience gradual conformational modification, resulting in HSA-warfarin complex. According to Gumede et al. (2012) the R-enantiomer of warfarin indicated a considerable change in the conformation of the rotatable groups in binding pocket of HSA site I. Therefore, the lock and key method (flexible ligand-rigid protein) appears to be inaccurate. However, induced fit docking (IFD) using flexible ligand-protein, consider the movement of protein side chain and back bone during binding to the ligand. The HSA side chains Lys199, Arg257, and His242 were shown to have intermolecular interaction with the ligand, whereas, Tyr150 plays a key role in hydrogen bonding at site I, followed by His242, Lys199, and Arg222. They also reported that the affinity and enantioselectivity was observed in favor of the S-pose over R-pose due to the presence of an additional π−π interaction and a strong hydrogen bond between

the Arg218 N-H group (hydrogen bond donor) and a negatively charged oxygen group of the coumarin ring (hydrogen bond acceptor). The achieved binding affinity and enantioselectivity estimates were reported to be comparable with experimental data.

Warfarin ((3-(α-acetonylbenzyl)-4-hydroxycoumarin) binds to subdomain IIA of drug site I within HSA with high affinity (Kd ≈3 μM). Additionally, the affinity of warfarin to HSA may be enhanced up to 3-fold in the presence of low levels of fatty acids. Range of other drugs like indomethacin, tolbutamide and phenylbutazone compete with warfarin for the same binding site. Despite the fact that different drugs bind to site I, the structure of these drugs generally exhibits bulky heterocyclic features with a negative charge located toward the molecule center. Therefore, warfarin exemplifies this type of compound and consequently, the study of the HSA-warfarin complex structure could explain this broad selectivity. Petitpas et al. have reported the crystal structures of HSA-myristate complexed with both the R- and S- enantiomers of warfarin and elucidated the molecular details of the interaction between warfarin and HSA. Warfarin enantiomers were both bound to the protein in the same way which accounts for albumin poor stereoselectivity. The coumarin and benzyl moieties of warfarin could be seen accommodating two adjacent but separate sub-chambers. The interactions between HSA-warfarin complex were shown to be dominated by hydrophobic contacts, however, specific electrostatic interactions could also be noticed. Phe211, Trp214, Leu219, and Leu238 forms the sub-chamber that fits the benzyl moiety with additional aliphatic contacts with Arg218 and His242. The coumarin moiety accommodates the main chamber located further from the entrance and interacts primarily with the surrounding side chains via hydrophobic contacts. Ala291 and Leu238 delimit the roof and floor of the pocket, respectively. Meanwhile, Ile260, Ile264, Ile290, and the aliphatic parts of Arg257 and Ser287 make contact with the coumarin ring back end. Further hydrophobic contacts were seen with the right side from Val241 and on the left from the aliphatic part of Arg222. Moreover, the coumarin moiety lies close to the fatty acid in the adjacent binding site, however, it doesn't make direct contact with it. Interestingly, the coumarin chamber shows an additional side pocket that formed by Leu219, Arg222, Phe223, Leu234, Ile264, Leu257 and Ile290. Although this side chamber is not occupied by warfarin, it can fit further hydrophobic portions of other site I drugs. More importantly, two of the three oxygen atoms of the coumarin moiety are involved in electrostatic interactions. The $O_{(a)}$ atom of warfarin is close to N_{ϵ} of Arg222 while the acetonyl group $O_{(c)}$ lies close to NH_2 of Arg222 closer to the pocket entrance. The hydroxyl group $O_{(d)}$ atom makes hydrogen bonds to the side chain of His242 and also to a bound water molecule. The authors suggested that the electrostatic interactions made by $O_{(b)}$, $O_{(c)}$, and $O_{(d)}$ altogether help maintain the orientation of warfarin

within its binding pocket. However, the authors also suggest that the shape complementarity between the drug and the pocket plays in an additional major role in its binding (Petitpas et al., 2001).

Meenu Narwal et al. (2018) used short MD simulations run to distinguish ester and non-ester drugs binding with HSA. HSA has esterase activity, which allows it to break drugs with ester bonds and so modulate their effects. The initial poses of the compounds at the drug sites I and II were created by molecular docking. The preference for the particular binding site was determined by the greatest number of hydrogen bonds and other close interactions between the protein and the ligand. The docking analysis demonstrates that some drugs prefer binding at HSA site I while others preferred binding at HSA site II, regardless of the kind of drug (ester or non-ester). The best pose of each ligand with the most favorable calculated binding energy at the identified binding site was submitted for a 2 ns MD simulation. After running the MD simulations, they observed that the protein backbone RMSD in case of both ester and non-ester drugs remained between 1-3.4 Å. Also, the ester drugs showed constantly 1 to 2 hydrogen bonds and non-ester drugs 1-4 hydrogen bonds during the simulations. They concluded that the presence of an ester bond cannot determine which site the drug binds to since both ester and non-ester drugs were detected bound in the two HSA binding pockets, either at drug site I or site II.

Another study by Pragna Lakshmi et al. (2017) investigated the interaction of a compound 2,4-diacetylphloroglucinol (DAPG) with HSA drug site I and site II through 50-ns MD simulations. The antibacterial, antiviral, antihelminthic, and anticancer activities of 2,4-diacetylphloroglucinol (DAPG) have long been recognized. The obtained RMSD plot of DAPG in site I illustrates that the simulation system converged during the first 1000 ps of the simulation time and stayed generally stable for the remaining of the simulation, with minor fluctuation between 30 ns and 40 ns. The complex containing DAPG in site II, on the other hand, was more fluctuating. The findings of this investigation suggest that DAPG binds to HSA site I.

1.6 Microscale thermophoresis (MST) in studying biomolecular interactions

Biomolecular interactions between drug compounds and serum albumin are important for understanding drug pharmacokinetics. Microscale thermophoresis (MST) can be used to perform biomolecular interaction studies (Jerabek-Willemsen et al., 2011b). This is based on the basic principle of thermophoresis, which is motion of molecules in temperature gradient. A temperature field in an aqueous solution of biomolecules stimulates both the flow of heat and

molecules. This coupling effect among a heat flow and mass flow was initially detected by Carl Ludwig in 1856 (Jerabek-Willemsen et al., 2011a) and is thus known as Ludwig-Soret effect or thermophoretic effect. This effect is very sensitive to the solvent-molecule interface. It allows the quantification of biomolecular interactions by the thermophoretic changes in size, charge and hydration shell of a molecule. For example, the size changes when a protein binds to another protein, the charge changes when a DNA binds to a protein. When a protein binds with a small molecule, the charge or size may never change, but still one water molecule is displaced upon binding. Therefore, any change in these three parameters (size, charge or hydration shell) will change the thermophoresis behavior (Duhr & Braun, 2006). As a result, the approach is appropriate for quantification of a broad variety of interactions, ranging from small molecules binding to proteins to affinities in multi-subunit complexes and liposomes.

The MST experiment is easy to setup and handle; it has a low sample requirement (6 µl per sample), a short analysis time allowing high throughput, a wide temperature range (20°C - 45°C) for the analysis. In addition, MST measures interactions with essentially no limitation on molecule size or molecular weight (100 Da to 1 MDa), and it allows instantaneous quality control (aggregation, precipitation, sticking) and the sample does not need any immobilization (Jerabek-Willemsen et al., 2011b). The MST instrument couples infrared (IR)-laser into the direction of fluorescence emission with the help of an IR-mirror. The IR-laser is directed onto the sample via the same optics that are used for detection of fluorescence. This allows a stable and accurate observation of thermophoresis (Jerabek-Willemsen et al., 2011b).

In the beginning, the molecules are dispersed homogeneously at an ambient temperature without laser heating in order to measure a constant initial fluorescence. The instant an IR-laser is turned on, a quick temperature jump is witnessed, due to the molecules moving away from the heat, as illustrated in Figure 1.8. This temperature jump is associated with temperature dependent variations in fluorescence that are sensitive to the local environment of the fluorophore. This allows the affinity of the interaction to be quantified by the intrinsic temperature dependent changes in fluorescence. This is followed by fluorescence variation induced by thermophoretic molecule motion. The decrease in fluorescence is measured for around 30 seconds until it reaches a steady state. Once the laser source is turned off after 30 seconds, an inverse temperature jump or a steep rise in signal is observed caused by the sample cooling. Lastly, back diffusion of molecules, which is solely driven by mass diffusion, allowing to derive information on the size of molecule (Jerabek-Willemsen et al., 2011b).

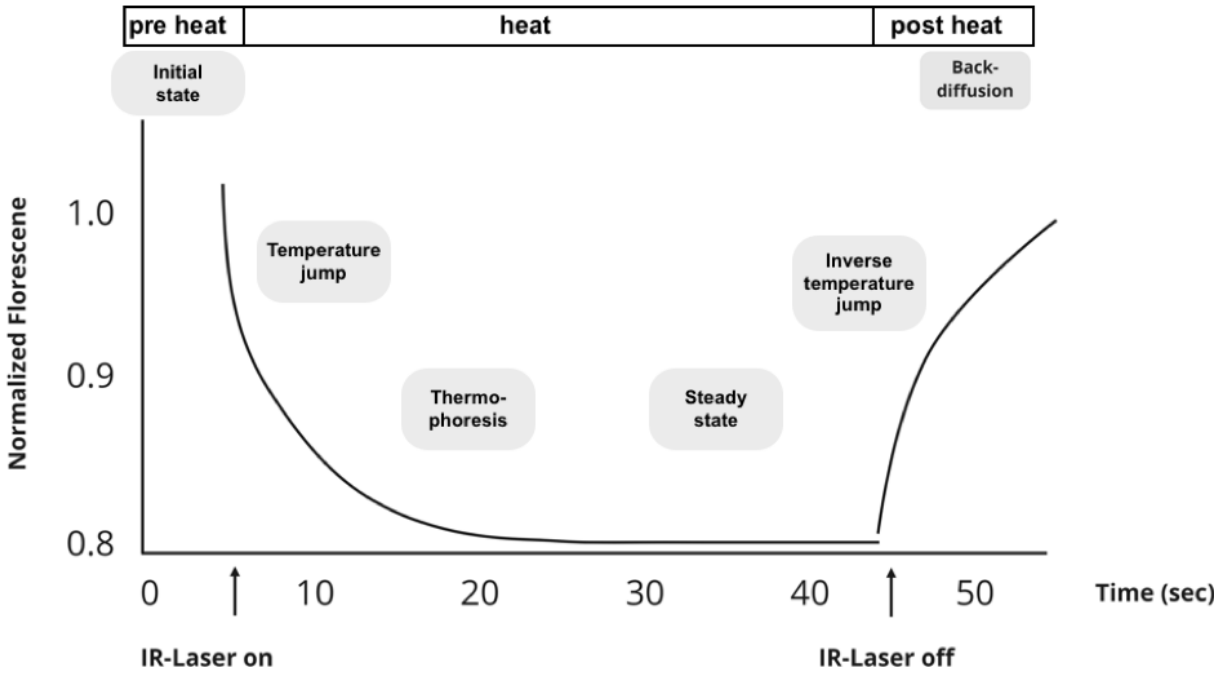


Figure 1.8: The schematic diagram of MST time trace. Different phases in fluorescence over time are distinguished by their time scale.

1.6.1 MST studies on drug binding to HSA

Several studies have incorporated MST for the purpose of analyzing binding affinity of compounds with HSA. Joanna Wasko et al. (2020) employed the MST approach, in order to determine if HSA attaches to native human insulin and its A13-A19 and B12-B17 fragments (hot spots), and whether binding with HSA helps to decrease the ability of insulin and its hot spots to aggregate in human body. These hot spots regions are structural components that regulate the aggregation of native polypeptides. Insulin can misfold, resulting in the formation of organized fibrillar amyloid aggregates inside the body. The conformational changes in the structure of metabolic hormones (insulin) may be detected during aggregation, with a considerable rise in the quantity of B-sheet with time. Recent research has revealed that small fragments (hot spots) of insulin are involved for the instigation and advancement of soluble peptide aggregation into fibrous deposition in numerous tissues and organs. Furthermore, calculating the dissociation constant (Kd) of HSA-fragments complexes revealed the binding of both insulin fragments (A13–A19 and B12–B17) with HSA, minimizing the possibility of fragments to aggregate in the presence of HSA. The utilization of HSA as a drug delivery mechanism for peptidic inhibitors of the insulin aggregation development, may result in cutting edge diabetic treatment strategies.

Another study by Christoph J. Wienken et al. (2010) also reported MST being an excellent

method for understanding the effect of HSA on the binding of small molecule inhibitor quercetin to its target kinase, protein kinase A (PKA). HSA, a universal transport protein is being well-known for having a hydrophobic pocket that can bind flavonoids such as quercetin. The experiment showed that HSA binds to quercetin and thus reduced the affinity of quercetin for PKA.

Chushak et al. (2015) investigated the binding of toluene to HSA using MST. The affinity chemicals, drugs and metabolites to HSA is proportional to their distribution throughout the body as well as the percentage of unbound compounds accessible for biological activities and toxic effects. The benzene and its metabolites interact with a cysteine residue in HSA to form a protein adduct, which is utilized as an indicator of benzene exposure. Similarly, toluene binding to HSA causes nonspecific binding and promotes aggregation, which occurs commonly with HSA and small molecules. The authors concluded that HSA certainly interacts with toluene.

2 AIM OF THE STUDY

The overall aim of the study was to investigate whether a strong HSA binding could be a probable reason for the poor *in vivo* activity of the provided antiparasitic compounds.

The detailed objectives of the study were to:

- Computationally study the binding interaction of the provided antiparasitic compounds and fexinidazole with HSA by docking them in the HSA drug binding sites I and II.
- Evaluate the stability of binding interactions of the compounds with the best docking scores using MD simulations.
- Estimate the HSA binding affinity of the compounds by calculating the binding free energies using the MM-GBSA approach.
- Predict the binding interaction of the compounds with HSA using an ADME property prediction program QikProp.
- Conduct *in vitro* HSA binding studies using MST for some of the provided compounds.

3 MATERIALS AND METHODS

3.1 Computational protocols

Molecular modeling and visualization of the molecular structure was carried out with the Maestro molecular modeling suite (Maestro v2021.2 by Schrödinger LLC graphical user interface) and PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. were used for visual analysis and structural investigation of the protein, ligand and protein-ligand complexes. The graphs for the MD structure stability analysis (RMSD, RMSF, and H-bonds) were plotted using the Grace-5.1.25 plotting tool. The generous computational resources for MD simulation were provided by CSC – IT Center for high performance Scientific Computing, Finland.

3.1.1 Ligand preparation

An initial library of 418 antiparasitic compounds was provided by our collaborator Dr. Werner Seebacher from the University of Graz, Austria. The compound structures were provided in an Excel sheet, in 1D SMILES (Simplified Molecular Input Line Entry System) format. In addition, *in vitro* activities and toxicity values as IC50 (nM) were given for these antiparasitic compounds. The *in vitro* inhibitory activities of the compounds against the growth of trypanosome (Tb *rhodesiense* STIB 900) had been studied using a fluorescence-based assay and the toxicity assay had been performed in L-6 cells acquired from rat skeletal myoblasts (Ahmad et al., 2016; Mohsin et al., 2018). The antiparasitic *in vivo* activities of some of the compounds had also been evaluated at Swiss Tropical and Public Health Institute (Basel) in Naval Medical Research Institute (NMRI) mice.

All of the 418 compounds and their biological data have been published previously (Ahmad et al., 2012, 2015, 2016; Berger et al., 2009; Faist et al., 2009; Hoffelner et al., 2019; Seebacher et al., 2005, 2006, 2007b, 2007a; Weis, Berger, et al., 2008; Weis, Kaiser, et al., 2008; Wolkinger et al., 2016).

Additionally, fexinidazole, an oral drug used for sleeping sickness and the co-crystallized warfarin R-(+)-enantiomer, present at the HSA drug site I of the HSA crystal structure (PDB ID: 1H9Z), were also used for the computational study.

To prepare the above-mentioned compounds, the LigPrep tool of Maestro (Schrödinger Release 2021-2: LigPrep, Schrödinger, LLC, New York, NY, 2021) was used to produce low-energy 3D structures from the initial SMILES strings. The ionization states were calculated at pH of

7.0+/-2.0 with Epik (Greenwood et al., 2010; Shelley et al., 2007) (Schrödinger Release 2021-2: Epik, Schrödinger, LLC, New York, NY, 2021). All possible tautomers and a maximum of two stereoisomers per compound were generated using the OPLS4 (Lu et al., 2021) force field.

3.1.2 Protein preparation

The 3D structure of the target protein HSA was retrieved from the RCSB Protein Data Bank (PDB) (Berman et al., 2000) (http://www.rcsb.org/) that is a freely available database containing more than 190,000 macromolecular structures as of June 2022 (https://www.rcsb.org/stats/growth/growth-released-structures). The techniques used to solve 3D structures of macromolecules are mainly röntgen (X-ray) crystallography, nuclear magnetic resonance (NMR) and cryo-electron microscope (EM). Out of these techniques, X-ray crystallography is the most widely used. The retrieved X-ray crystal structure of HSA (PDB ID: 1H9Z, resolution 2.50 Å), contains a single protein chain including 583 modelled amino acids with a structural weight of 68.25 kDa. It has been co-crystallized with the R-(+) enantiomer of warfarin contains also six myristic acid molecules and several water molecules (Petitpas et al., 2001).

The HSA experimental structure obtained from the PDB database cannot be used straightaway for molecular modeling purposes. To prepare the protein, the Protein Preparation Wizard tool of Maestro (Madhavi Sastry et al., 2013) was used. First of all, HSA was preprocessed by adding hydrogen atoms, filling in missing side chain residues and loops using Prime (Schrödinger Release 2021-2: Prime, Schrödinger, LLC, New York, NY, 2021). Water molecules beyond 3 Å of heteroatom (non-protein) groups were removed and hydrogen bonds (H-bonds) networks were optimized. The only water molecule retained within the 3Å is H_2O-2028, as it mediates the interaction between the co-crystallized ligand warfarin and HSA residue Arg257. This specific water molecule (H_2O-2028) was retained in docking of other compounds too. Finally, the whole protein structure was minimized to eliminate strain using the OPLS4 force field and letting convergence criteria for the heavy atoms to be 0.30 Å RMSD.

3.1.3 Molecular docking

For the purpose of the current investigation, both drug site I and drug site II were used for ligand binding studies, to find whether the provided antiparasitic compounds would bind to either of these sites.

The HSA co-crystallized ligand (warfarin) was self-docked by Glide extra precision (XP) mode to reconstruct the pose in drug site I of HSA, and the resulting poses were scored and visually assessed. The poses generated from the self-docking of co-crystallized warfarin were evaluated for their binding free energy by MM-GBSA approach. The free energy of binding values estimated for all the generated poses were compared with the HSA co-crystallized ligand.

Before docking the library of ligands into the putative target protein HSA, the docking sites were defined using the Receptor Grid Generation tool of Maestro. For the drug site I, the grid box was generated around the co-crystallized ligand; the internal box size was 15x15x15$Å^3$ and the external box 28.58x28.58x28.58$Å^3$. In case of the drug site II, the already present fatty acid was removed from the pre-processed HSA and the Site Map tool of Maestro (T. Halgren, 2007; T. A. Halgren, 2009) (Schrödinger Release 2021-2: SiteMap, Schrödinger, LLC, New York, NY, 2021) was used to explore the binding sites on the protein surface. SiteMap mainly focuses on the expected pockets according to their size, exposure levels, hydrophobic and hydrophilic regions to occupy hydrophobic and hydrophilic ligands groups, compactness, and hydrogen-bond donors and acceptor descriptions. The predicted binding site by SiteMap tool around sub-domain IIIA (drug site II) was selected for grid generation. The grid box size for the drug site II was defined as the internal box 17x17x17$Å^3$ and the external box 37x37x37$Å^3$.

Once the docking grids were generated around the putative binding sites (site I and site II), the entire prepared library of ligands (altogether 1690, including their tautomers and stereoisomers) was docked into these sites with Schrödinger's Grid-based Ligand Docking with Energetics (Glide) program (Friesner et al., 2004; T. A. Halgren et al., 2004) (Glide, Schrödinger, LLC, New York, NY, 2021) using the Glide extra precision (XP) mode.

The energy terms in XP scoring function (Friesner et al., 2006) are more complex than those in classic molecular mechanics. It includes terms, such as hydrophobic interaction ($E_{hyd\_enclosure}$), hydrogen bond neutral-neutral motifs ($E_{hbond\_nn\_motif}$), hydrogen bond charged-charged motif ($E_{hbond\_cc\_motif}$), pi-stacking and pi-cation (E_{PI}) interactions. It also uses standard ChemScore function such as lipophilic-lipophilic pair (E_{phoic_pair}) and hydrogen bonds term ($E_{hbond\ \_pair}$). The hydrogen-bonding expression ($E_{hbond\ \_pair}$), like the standard the ChemScore function, is divided into multiple weighted components depending on whether the donor and acceptor are both neutral, one of them is neutral and the other is charged, or both of them are charged. The binding penalties in the XP scoring function includes desolvation penalties (E_{desolv}) and contact penalties in rigid-protein docking (E_{lig_strain}). The equation below illustrates the parameters in Glide XP scoring function.

$$\text{XP\_Glide score} = E_{coul} + E_{vdW} + E_{bind} + E_{penalty}$$

$$Ebind = E_{hyd\_enclosure} + E_{hbond\_nn\_motif} + E_{hbond\_cc\_motif} + E_{PI} + E_{hbond\ \_pair} + E_{phoic\_pair}$$

$$E_{penalty} = E_{desolv} + E_{lig\_strain}$$

3.1.4 Molecular dynamics simulation

Based on the molecular docking results, best-ranked docking poses of the protein-ligand docking complexes at both drug site I and II were further investigated by MD simulation. For comparison, a docking complex of one of the ligands tested experimentally for HSA binding (150MFC6c) and the apo (ligand free) HSA were also simulated. MD simulation was carried out by using Desmond module (Schrödinger Release 2021-2: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2021. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2021) (Bowers et al., 2007) to study the stability and dynamic behavior of these complexes. The total of four parallel MD simulation runs with four randomly generated seeds were used and the simulation outcome of the seed with the best RMSD value was chosen. A cubic simulation box with distance 10 x 10 x 10 Å and a three-point TIP3P explicit water model (Jorgensen et al., 1983; Mark & Nilsson, 2001) were used to prepared a solvated system employing the OPLS4 (Lu et al., 2021) force field. The sodium ions (Na^+) as counter ions were added to neutralize the system charge and a physiological salt concentration of 0.15 M. The NPT ensemble was employed for the 400 ns production simulation; temperature of 300 K was maintained with the Nose-Hoover chain coupling scheme (Martyna et al., 1992) and 1 bar pressure controlled using the Martyna-Tuckerman–Klein barostat (Martyna et al., 1994). High performance computational (HPC) recourses from IT Center for Science Finland - CSC were used to run all MD simulations.

3.1.5 MD simulation analysis

The outcomes of MD simulations are collected in the form of trajectories, which are a series of snapshots (called as frames) collected over the period of the simulation time. Several aspects of the structural stability of the complexes were extracted from the generated trajectories. The script $SCHRODINGER/run analyze_simulation was used followed by an automated *Simulation Interactions Diagram (SID)* tool (Bowers et al., 2007). The values estimated by the SID tool were RMSD and RMSF. Additionally, a script $SCHRODINGER/run trajectory_analyze_hbonds.py was employed to calculate the number of protein-ligand hydrogen bonds. The mentioned simulation analysis was implemented for the full span of the

MDS trajectories.
Lastly, the number of hydrogen bonds between the protein and the ligands in the complexes over a simulation time was measured. Apart from this, a visual inspection was done for all protein-ligand complexes.

3.1.6 Binding free energy calculation

The MM-GBSA method incorporated in Maestro's Prime module (Bowers et al., 2007), accessible via the Prime/MM-GBSA module of Maestro (Jacobson et al., 2002, 2004) (Schrödinger Release 2021-3: Prime, Schrödinger, LLC, New York, NY, 2021.) was applied to calculate BFE of the docked protein-ligand complexes.
The protein-ligand complexes binding free energies of MD simulation were estimated by MM-GBSA approach, accessible in Schrödinger Maestro suite by running the *thermal_mmgbsa.py* script. The 50 evenly spaced snapshots for the last 4000 trajectory frames (200 ns) were taken into consideration for BFE calculation.

3.1.7 Computational prediction of plasma protein binding by QikProp

Plasma protein binding property of the compounds was analyzed by using the QikProp module from the Maestro suite (Schrödinger Release 2021-4: QikProp, Schrödinger, LLC, New York, NY, 2021). The prediction by QikProp module was applied to the LigPrep generated compounds.

3.2 Experimental protocol

3.2.1 Microscale thermophoresis

The MST experiment was conducted on Monolith NT.115 (NanoTemper Technologies, model: Monolith NT.115). Human Serum Albumin ≥98% (SDS-PAGE) SRP6182-1MG was purchased from Sigma-Aldrich. The HSA was labelled with MO-L011 Protein Labeling Kit RED-NHS 2nd Generation (Amine Reactive). High precision glass capillaries from Monolith™ NT and Automated Capillaries Chips (MO-AK002) were used for the samples. The four ligands for an *in vitro* MST experiment (150MFC6c, 39APR4b, 150MFC11 and Berr5) were obtained from our collaborator Dr. Seebacher. The three first are published structures and were also provided in the Excel sheet of the 418 compounds. MST assays were carried out as described in Jerabek-Willemsen et al. (Jerabek-Willemsen et al., 2011b).

4 Results and Discussion

The molecular interactions of the ligands with HSA were computationally studied by docking them into the crystal structure of HSA (PDB ID: 1H9Z) drug site I and drug site II. Initially, a library of 418 ligands, a co-crystallized ligand (warfarin) and fexinidazole were prepared with Maestro's LigPrep tool to convert mentioned ligands from 1D SMILES to low energy 3D structure format. The LigPrep generate tautomers and at most 2 stereoisomers per ligand. The number of ligands prepared in 3D format after applying LigPrep are compiled in Table 4.1.

Table 4.1: The number of compounds prepared by the LigPrep tool for Molecular docking

	Before LigPrep (2D format)	***After LigPrep (3D format)***
Ligand library	418	1690
Warfarin	1	4
Fexinidazole	1	1
150MFC6c*	1	4

* Example molecule from the library, was also tested in the MST assay

4.1 Molecular docking

Molecular docking is an effective method for the prediction of protein-ligands binding and strength of molecular interaction at the binding site of protein. For the current thesis work, the prepared co-crystallized ligand warfarin was self-docked to replicate the poses in the binding site and to evaluate the ability of Glide docking software in replicating the poses similar to that of reference pose, into the HSA drug site I. The extra precision (XP) scoring function of Glide was used to dock co-crystallized ligand warfarin, taken from X-ray crystal structure of HSA (PDB ID: 1H9Z). A total of four poses were generated (Table 4.2.), their BFE were estimated using MM-GBSA approach incorporated in Maestro Prime module. The generated poses were also visually inspected and compared with the reference co-crystallized ligand. All the poses showed precise overlapping with a minimal variation. The similarity of re-docked pose-1 and co-crystallized warfarin is illustrated in Figure 4.1. The evaluation of binding free energies showed similar values for pose-2, pose-3, pose-4 and co-crystallized ligand. The BFE for pose-1 showed a value higher than the co-crystallized ligand.

Table 4.2: The docking score and binding free energies of co-crystallized ligand and re-docked poses.

Warfarin	Docking score (kcal/mol)	BFE (kcal/mol)
co-crystallized ligand (1H9Z)	**-**	**-48.28**
Pose-1	-7.80	-51.91
Pose-2	-7.51	-48.97
Pose-3	-7.32	-48.26
Pose-4	-7.28	-48.62

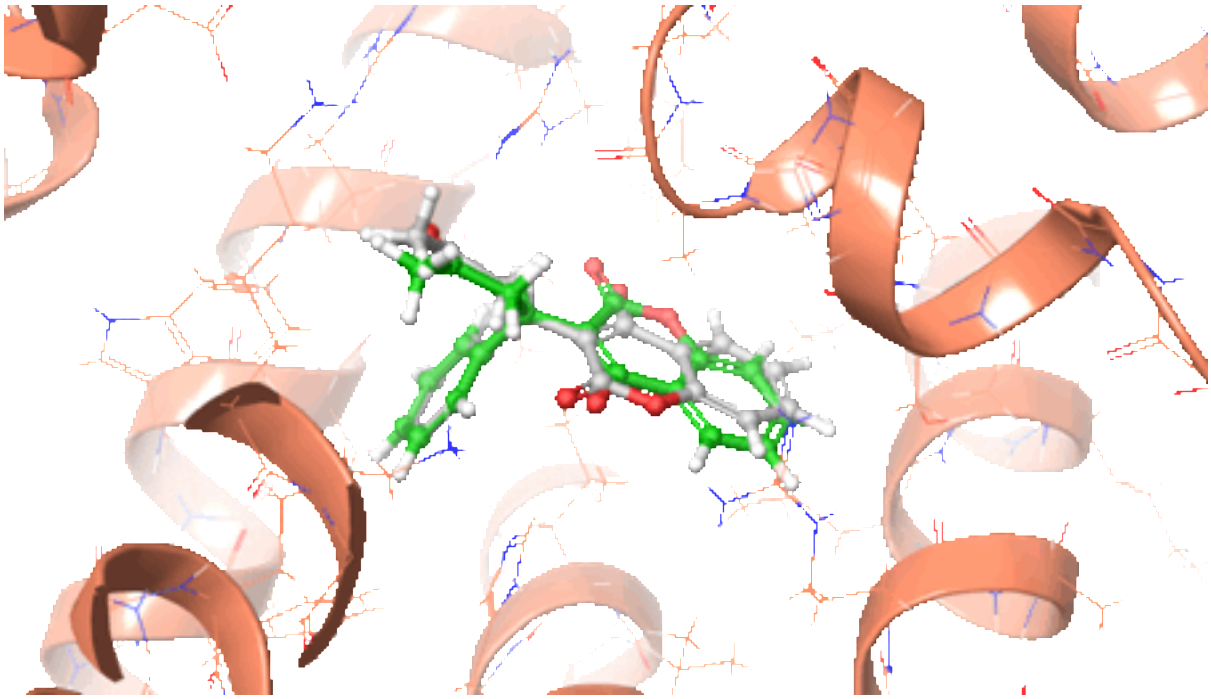


Figure 4.1: Self docking of co-crystallized ligand warfarin at HSA drug site I (PDB ID: 1H9Z), green show co-crystallized ligand, grey show pose 1.

Fexinidazole, an oral drug for the treatment of human African trypanosomiasis, was docked into HSA drug site I. In addition, the entire prepared library of compounds from our collaborator (altogether 1690, including their tautomers and stereoisomers) was docked into both the drug site I and drug site II of HSA using Schrödinger's Glide with XP (extra precision) mode. The docking score values for these compounds according to the putative drug binding sites are compiled in Table 4.3.

Table 4.3: The number of compounds docked at HSA site I and site II and their docking score range.

Compound	Number of compounds	Docked compounds	Docking Score range (kcal/mol)
HSA Drug Site-I			
Ligand library	1690	1572	-6.761 → 10.609
Warfarin poses	4	4	-7.80 → -7.28
Fexinidazole	1	1	-5.030
150MFC6c	4	2	-2.803 → -2.700
HSA Drug Site-II			
Ligand library	1690	950	-8.376 → 6.578
150MFC6c	4	4	-6.091 →-3.248

Based on Table 4.3, the generated poses of warfarin had a better docking score ranging from -7.780 to -5.836 kcal/mol (depending of the tautomer/steroisomer), compared to fexinidazole (-5.030 kcal/mol) and 150MFC6c (2.803 to -2.700), when docked to HSA Sudlow's drug site I.

Warfarin (Figure 4.2. a) makes two hydrogen bonds, one with HSA residue Arg222 and other with Arg257, mediated by a water molecule (H_2O-2028). Also, a salt bridge was observed with residue Arg222 and a pi-pi interaction with residue Trp214. In case of fexinidazole (Figure 4.2. b), the hydrogen bond was observed with the water molecule (H_2O-2028), and salt bridge between fexinidazole and Arg222. The compound 150MFC6c (Figure 4.2. c), used for *in vitro* study, showed pi-cation interaction with Lys195 and an H-bond with Lys199. The water molecule (H_2O-2028) did not show any interaction with 150MFC6c.

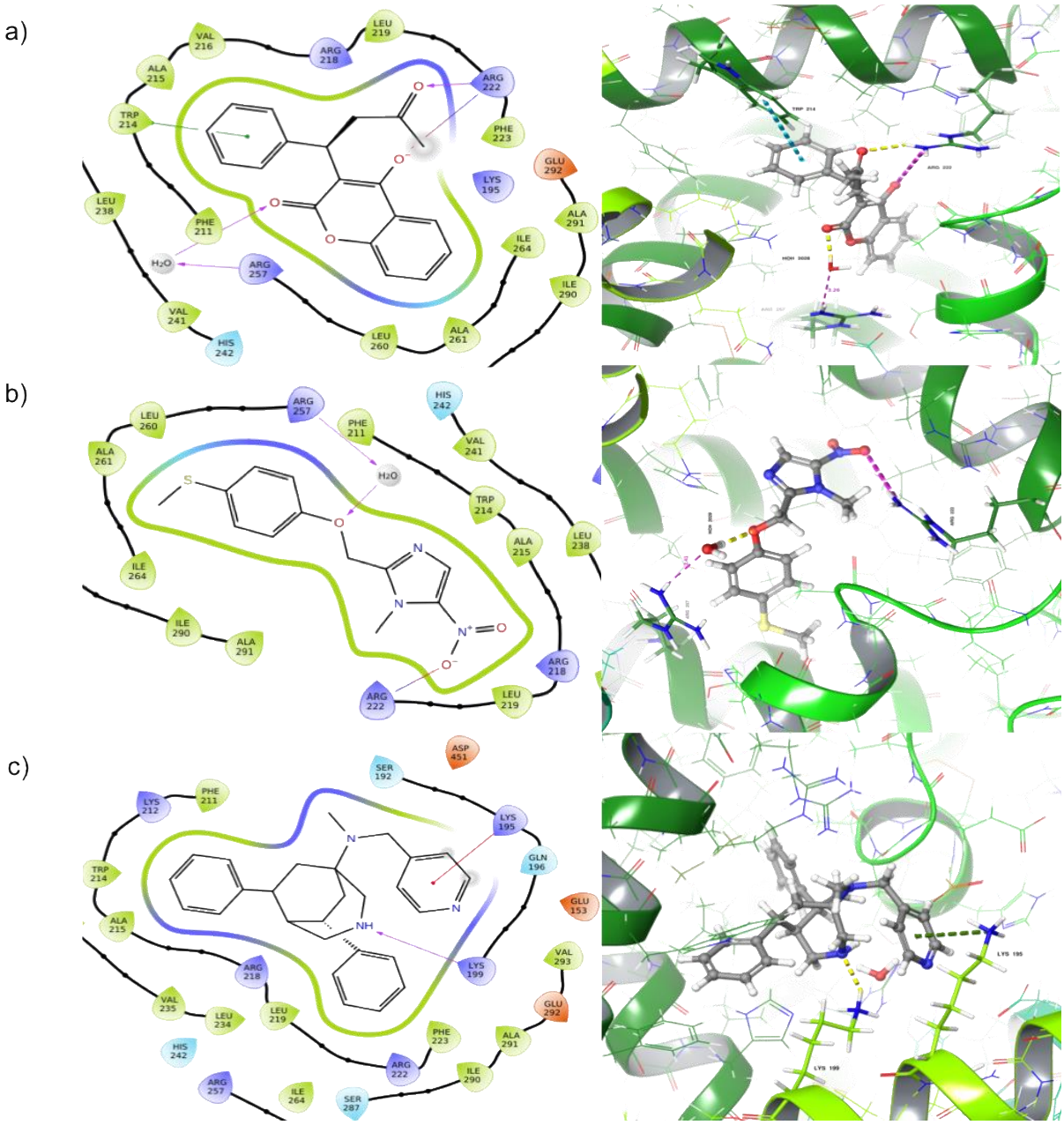
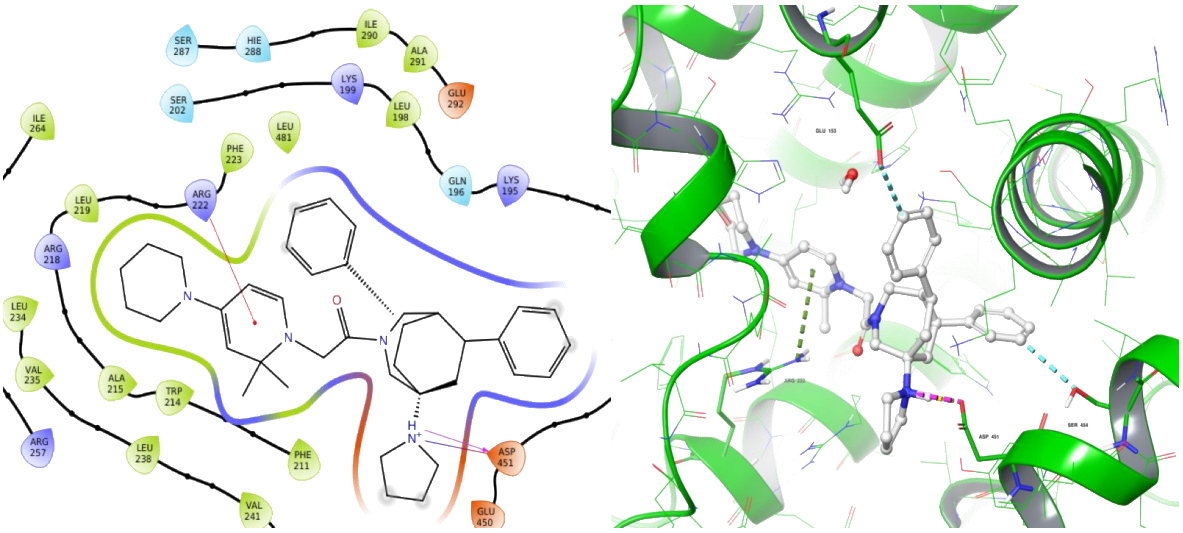


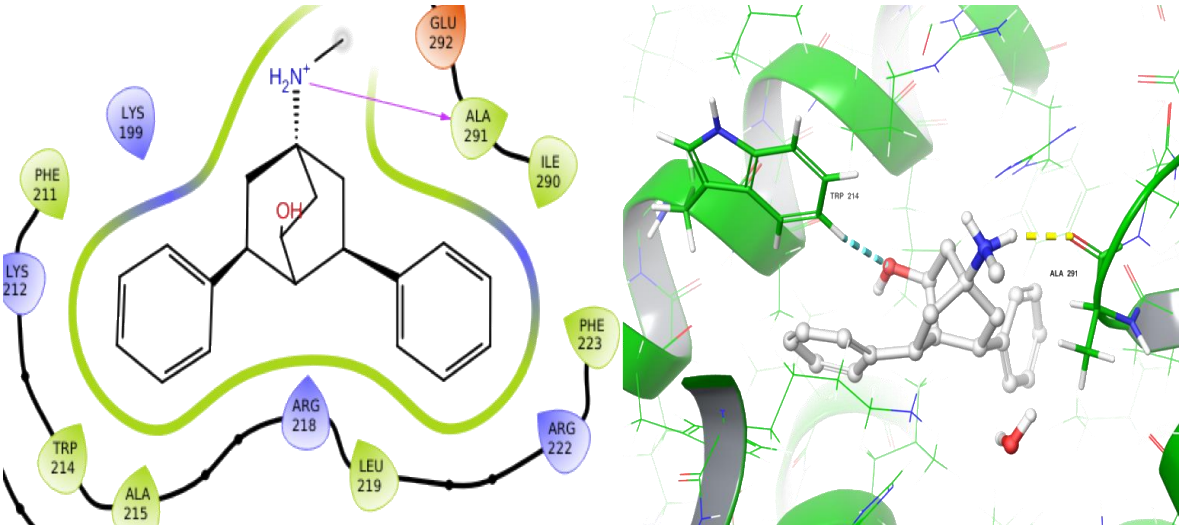
Figure 4.2: Docking pose of a) Warfarin, b) Fexinidazole and c) 150MFC6c at the HSA drug site I. Left panel: A 2D illustration of the binding interactions at the site. Right panel: 3D docking pose. HSA protein is depicted in green cartoon, Ligand: ball-and stick, atom color code: hydrogen (white), nitrogen (blue), oxygen (red), carbon (grey) and sulphur (yellow), bonds color code: h-bond (yellow), salt bridge (pink), aromatic h-bond (green) and pi-pi interaction (cyan) dashed lines. For 2D structures: h-bond-pink, salt bridge-blue-red, aromatic h-bond-red, pi-pi stacking-green.

Molecular docking of the ligand library to site I and site II of HSA showed that most of the molecules were preferential ligands for the HSA drug site I, with 1572 and 950 compounds succesfully docked, respectively. However, majority of these ligands does not seem to have a strong binding affinity to either of the target HSA sites. From the ligand library, four best docking poses of the ligand molecules that fit into the binding region (site I, site II) of the target protein HSA were kept for the MD simulation study. These compounds were selected primarily by considering the docking score but the visual inspection of these docked poses in Maestro was also taken into account. The docking poses of the chosen compounds are shown in Figures 4.3 and 4.4 for site I and site II, respectively.

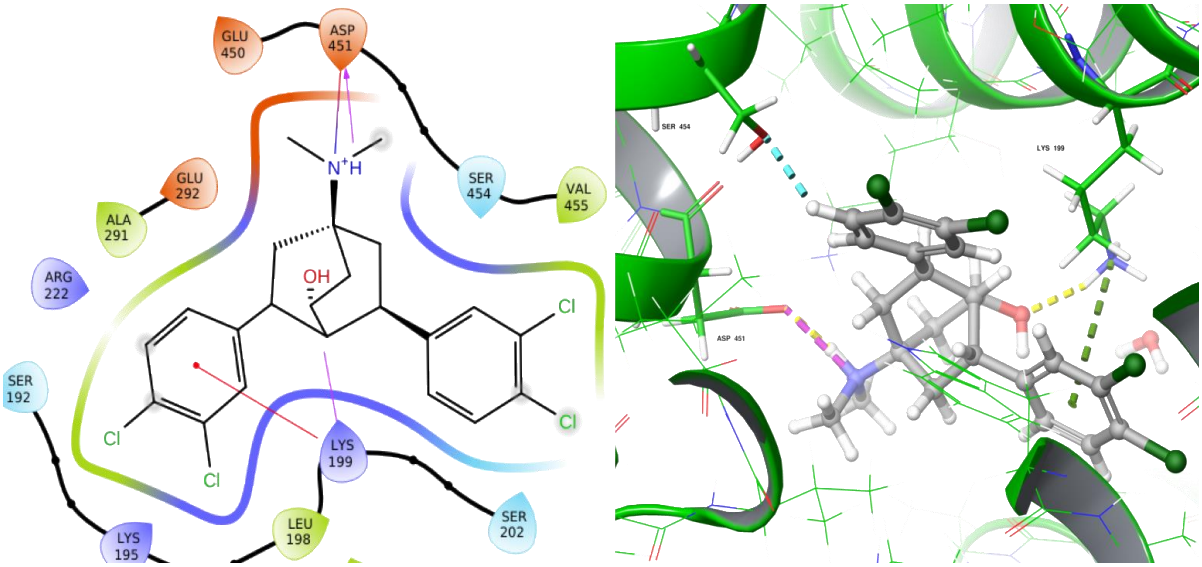
a)



b)



c)



d)

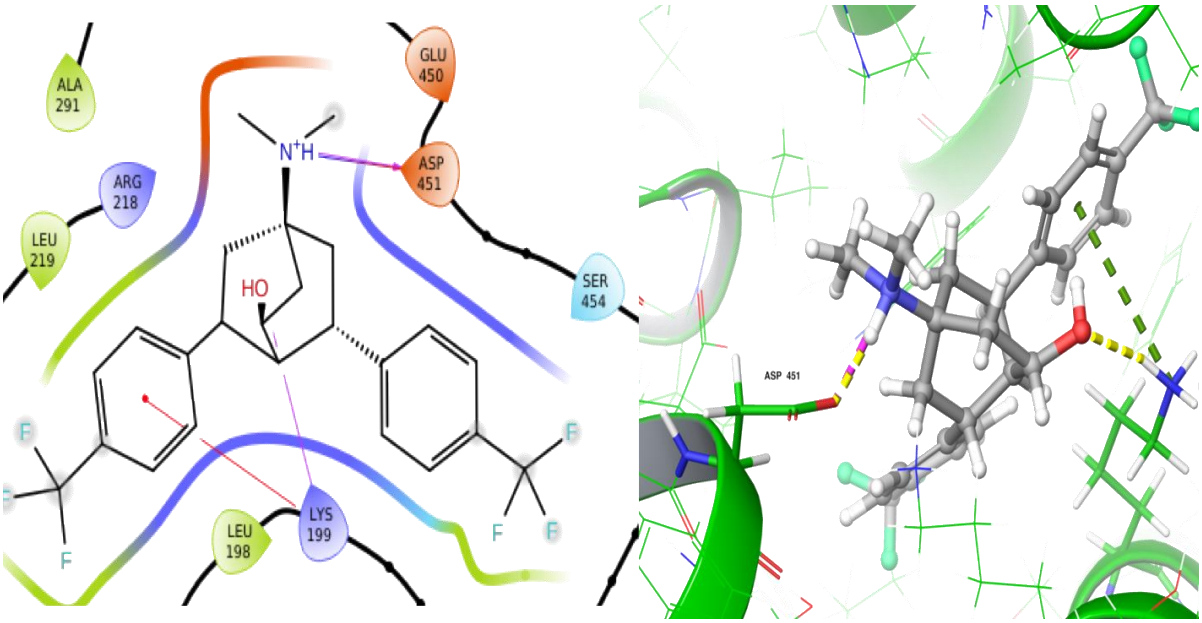
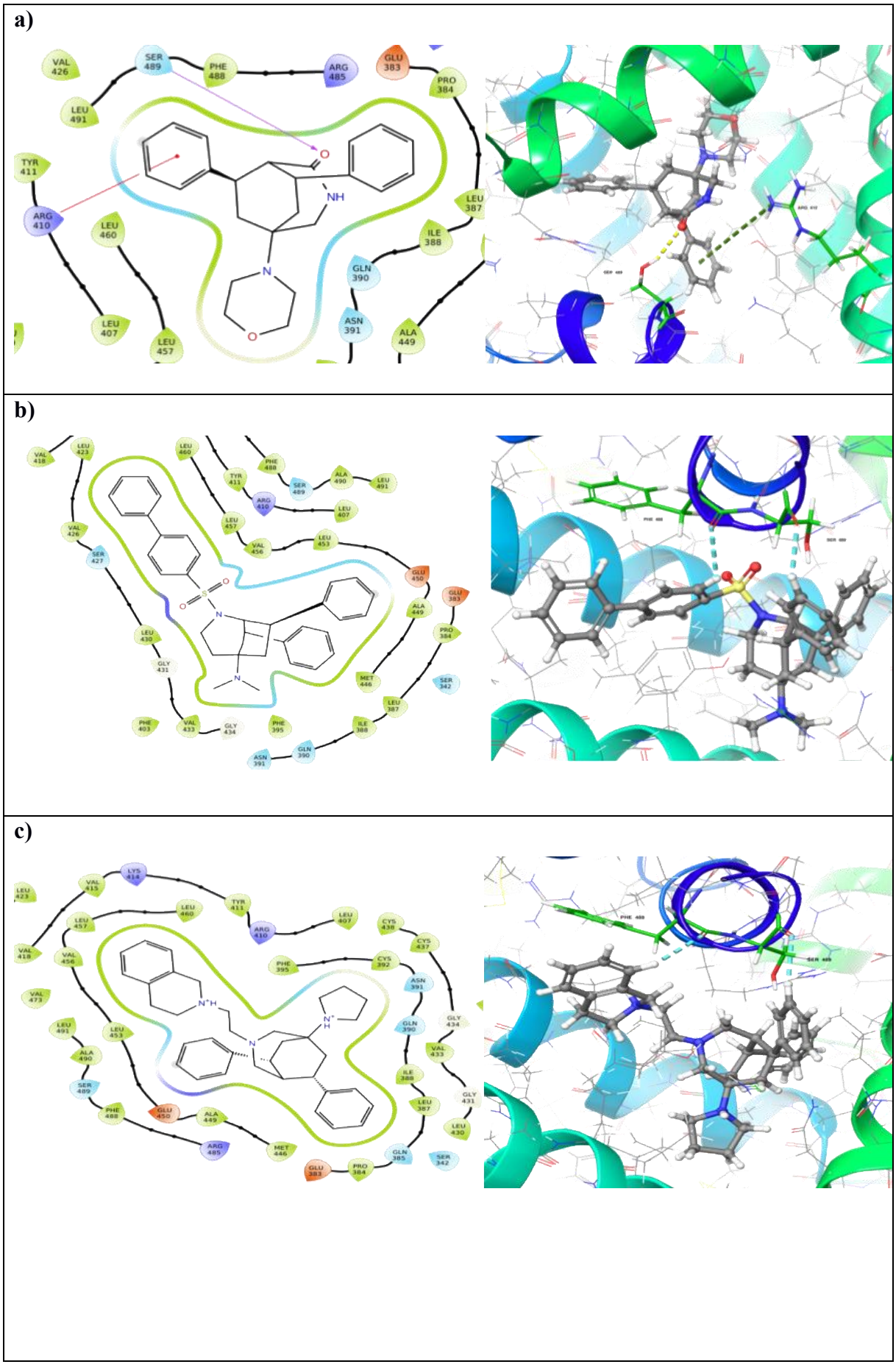


Figure 4.3: Docking pose of the best-ranked compounds at the HSA drug site I. a) 150MFC18b, b) 136MFC2e, c)140MCF4a and d) 140MCF8a. Right panel: 3D docking pose. HSA protein is depicted in green cartoon, Ligand: ball-and stick, atom color code: hydrogen (white), nitrogen (blue), oxygen (red), carbon (light grey), chlorine (green), fluorine (cyan) and sulphur (yellow), dashed bonds color code: H-bond (yellow), salt bridge (pink), aromatic H-bond (green), pi-cation interactions (cyan). For 2D diagrams: H-bond-pink, salt bridge-blue-red, pi-pi stacking-green, pi-cation interactions-red.


a)
VAL 426
SER 489
PHE 488
ARG 485
GLU 383
PRO 384
LEU 491
TYR 411
ARG 410
LEU 460
LEU 387
ILE 388
GLN 390
LEU 407
ASN 391
ALA 449
LEU 457
O
NH
N
O
b)
c)

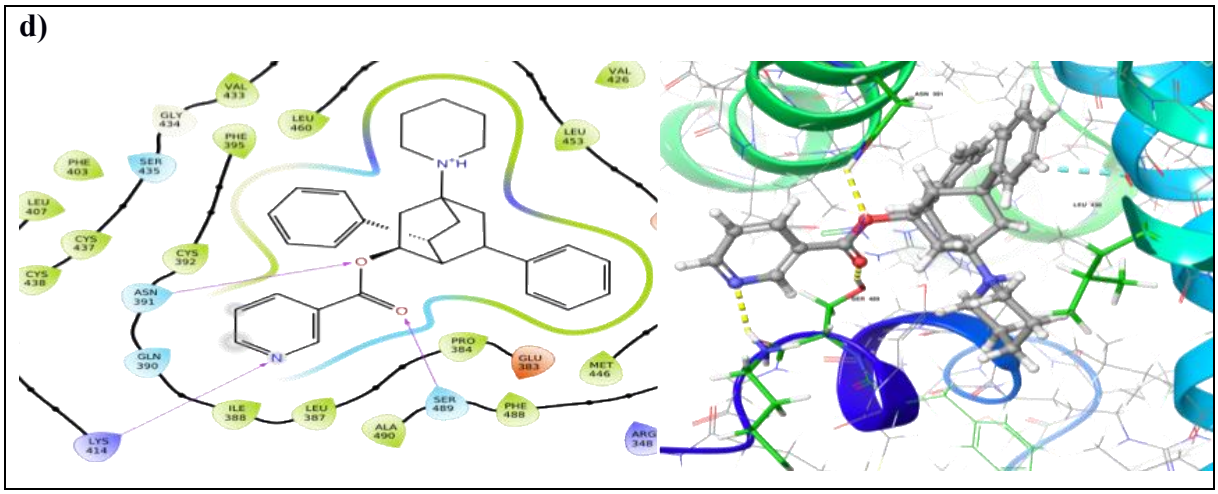


Figure 4.4: Docking pose of selected compounds at the HSA site II. a) CHEMBL3401421, b) 138MFC9a, c) 39APR7i and d) CHEMBL459424S. Right panel: 3D docking pose. HSA protein is depicted in blue/green cartoon, Ligand: ball-and stick, atom color code: hydrogen (white), nitrogen (blue), oxygen (red), carbon (grey) and sulphur (yellow), bonds color code: h-bond (yellow), aromatic h-bond (green) and pi-cation (cyan) dashed lines. For 2D structures: H-bond-pink, aromatic H-bond-cyan and pi-cations-red.

For site I, the compound 150MFC18b (Figure 4.3.a) had a highest docking score of -6.761 kcal/mol. It forms an aromatic H-bond with the binding pocket residues Ser454 and Glu153, a pi-cation interaction with Arg222, a H-bond and a salt bridge with Asp451. The second compound 136MFC2e (Figure 4.3.b) forms a H-bond with Ala291 and an aromatic H-bond with Trp214. The third compound 140bMCF4a (Figure 4.3.c) forms a H-bond with Lys199 and Asp451, an aromatic H-bond with Ser454, a salt bridge with Asp451 and a pi-cation interaction with Lys199. The fourth promising compound studied at site I was 140MCF8a (Figure 4.3. d). It formed a H-bond with Lys199 and Asp451, a pi-cation interaction with Lys199 and a salt bridge with Asp451.

Similarly, for site II of HSA, the compound CHEMBL3401421 (Figure 4.4.a) with a docking score of -8.376 kcal/mol, formed a H-bond with Ser489 and a pi-cation interaction with Arg410. The compound 138MFC9a and 39APR7i (Figure 4.4.b, c) had a docking score of -8.243 kcal/mol and -8.125 kcal/mol, and it formed two aromatic H-bonds with Ser489 and Phe488. The compound CHEMBL459424S (Figure 4.4. d) with a docking score of -7.979 kcal/mol forms three H-bonds with Asn391, Ser489 and Lys414, and an aromatic H-bond with Leu430.

4.2 MD trajectory analysis

MD simulation of 400 ns was carried out using the Desmond module of Schrödinger to optimize the best docking complexes and to evaluate the compound's relative binding affinities to HSA drug site I or II. As a reference, complexes of HSA with warfarin, fexinidazole and 150MFC6c (the MST-tested compound in the library) were subjected to MD analysis. Altogehter, the number of trajectories obtained is listed below:

a. 4 for ligands at HSA drug site I;
b. 4 for ligands at HSA drug site II;
c. 1 for warfarin at HSA drug site I;
d. 1 for fexinidazole at HSA drug site I;
e. 1 for MST tested compound (150MFC6c) at HSA drug site I;
f. 1 for HSA (apo).

4.2.1 Root-Mean-Square Deviation (RMSD)

Calculating RMSD is the most popular method for determining how stable a protein structure is during the MD simulation. By measuring the RMSD of the HSA Cα atoms with respect to the initial (before MD) protein conformation, the overall structural fluctuation of HSA with a bound-ligand and without a bound ligand (apo) were assessed over the simulation time. It was done by superimposing the trajectory frames over the reference (initial) frame versus time. The whole MD simulation time frame of 400 ns was considered for the RMSD calculation. The high RMSD value indicates that the protein experienced significant conformational change during the MD simulation time frame. Ideally, the trajectories of stable binding compounds would have an average RMSD value ranging from 1.5 to 3.0 Å (or comparable to the apo), whereas large RMSD values can result from the movement of a loop region or termini or unfolding of the protein.

The RMSD plots (Figure 4.5) for the HSA drug site I, indicated that the RMSD for the majority of the HSA-compound complexes varies between 2 to 5 Å, with the exception of HSA-136MFC2e and HSA-fexinidazole complexes. The complex HSA-150MFC18b and HSA-140MCF8a reached equilibrium during the last 230 and 200 ns, respectively. The HSA-136MFC2e complex showed the highest RMSD value of approx. 6.0 Å. It raised from the 2 to 6 Å within less than 50 ns, after which it reached a plateau. After about 120 ns of simulation, the RMSD of the HSA-140MCF4a complex appears to be quite stable, with the RMSD reaching a plateau of approximately 3.0 Å. The HSA-warfarin complex showed RMSD of 2 to 4 Å. The

complex of HSA-fexinidazole showed quite high deviation for the initial 200 ns, then started to maintain stability until deviated again after 350 ns. The HSA-150MFC6c system showed an elevation from 2 to 5 Å within the first 25 ns. The structural transformation is visible in the RMSD plot of the apo Cα atoms (Figure 4.5) during the 400 ns of the MD simulation. This is sustained with slight fluctuation for the first 50 ns, after which the RMSD Cα atoms is increased from 3 to 4.5 Å, eventually reaching approx. 5.0 Å until the end of simulations. This could be due to the absence of ligand interaction in the binding pocket. Most of the drug site I HSA-ligand complexes showed better convergence in comparison to apo. The RMSD results imply that most of the complex at HSA drug site I reached equilibrium during the last 200 ns of the simulation and thus can be used for the binding free energy analysis.

In case of HSA drug site II (Figure 4.6.), the apo showed better convergence compared to the HSA-ligand complexes. The RMSD value of apo was found to be lower than that of the complexes. The HSA- CHEMBL3401421 and HSA-39APR7i complexes showed RMSD close to apo. The HSA-138MFC9a system fluctuated from 2 to 7 Å during the whole simulation time frame and indicates that the system did not evolve into a stable state. Similarly, the HSA-CHEMBL459424S complex showed the highest RMSD value of approx. 8 Å. It raised from the 2 to 8 Å within 100 ns, after which it reached the plateau. The HSA-CHEMBL3401421 has the lowest and HSA-CHEMBL459424S has the highest fluctuation at HSA drug site II. The RMSD values higher than 3 Å may show that a protein goes through large conformational changes during the course of the simulation. In sum, the complexes at drug site II did not showed better convergence as compared to apo. Nevertheless, the complexes HSA-CHEMBL3401421, HSA-39APR7i and HSA-CHEMBL459424S, showed equilibrium during the last 200ns, and can be used for binding free energy analysis.

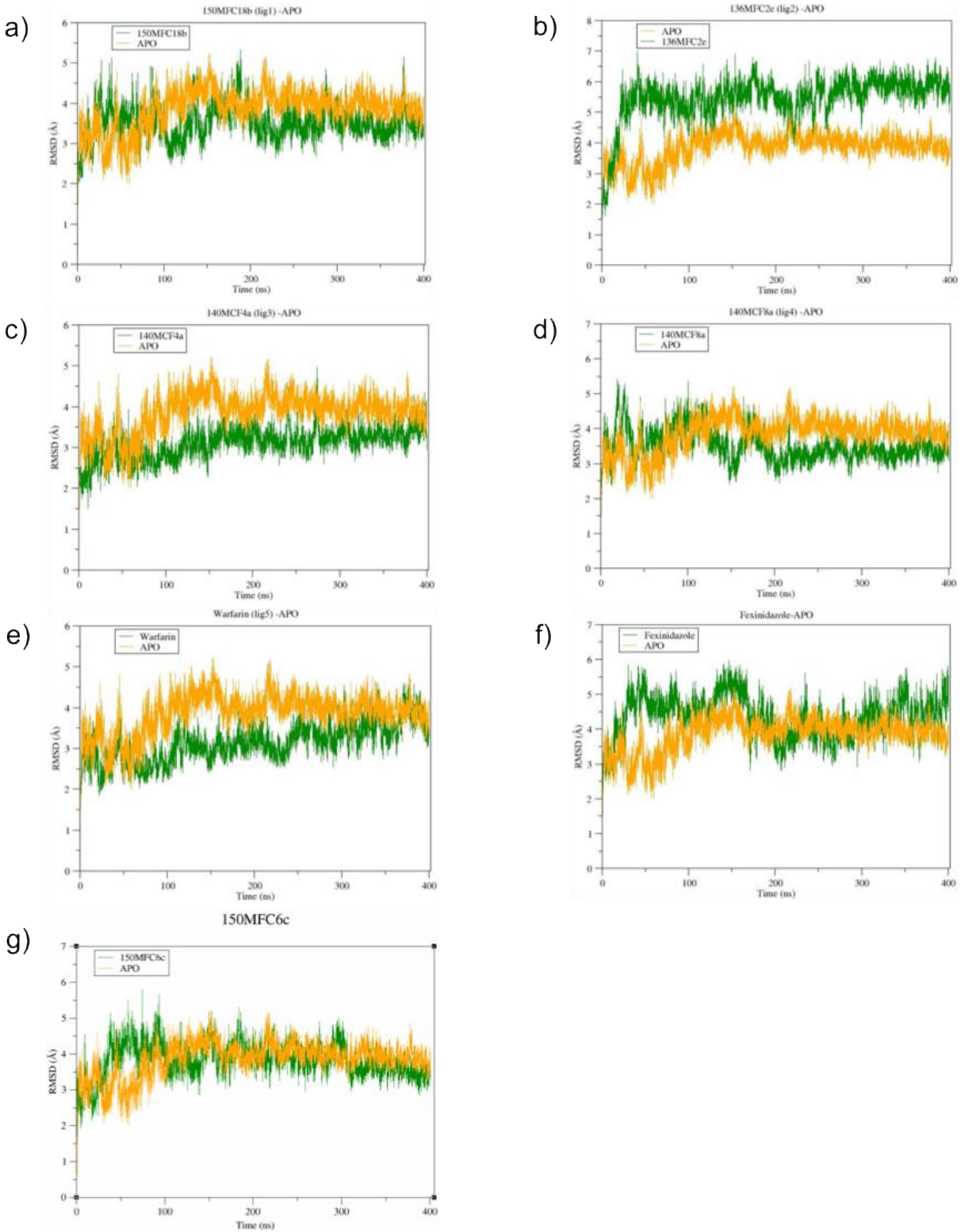


Figure 4.5: Comparison of RMSDs of Cα atoms of apo HSA and ligand-HSA complexes at HSA drug site I during the 400 ns of molecular dynamics simulation. a) 150MFC18b, b) 136MFC2e, c) 140bMCF4a, d) 140MCF8a, e) Warfarin, f) Fexinidazole and g) 150MFC6c.

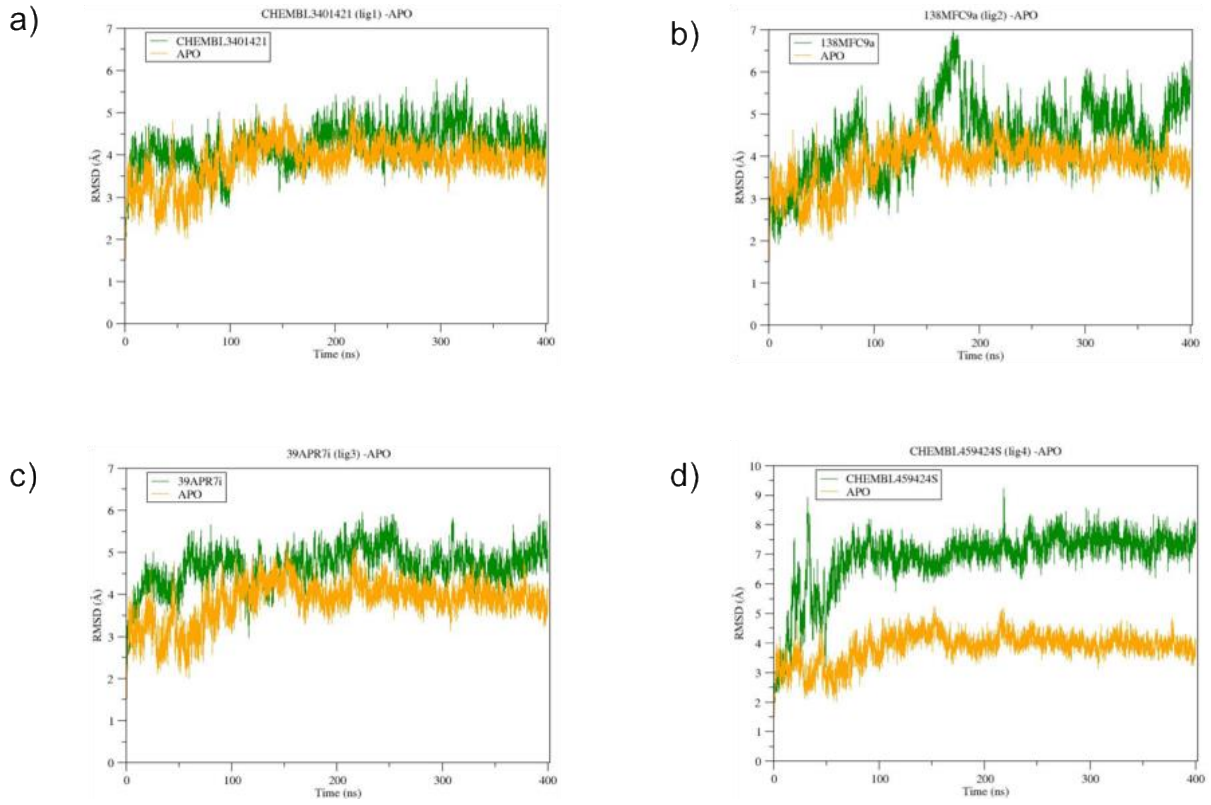


Figure 4.6: RMSD of apo HSA and ligand-HSA complexes for compounds docked at HSA drug site II during the 400 ns molecular dynamics simulation. a) CHEMBL3401421, b) 138MFC9a, c) 39APR7i and d) CHEMBL459424S.

4.2.2 Root-Mean-Square Fluctuation (RMSF)

In order to investigate the movement of HSA protein residues through the MD simulation, the RMSF plots were generated to quantify the average atomic fluctuations of Cα atoms of the protein. The RMSF plots (Figures 4.7-4.8), show that residues that constitute part of the loop regions are flexible throughout the simulation. As anticipated, the abrupt fluctuation at the start and the end of the RMSF plots shows how the protein's N-terminus and C-terminus reach a higher fluctuation value.

The residues forming the Sudlow's drug site I and drug site II are present in the range of 197-297 and 384-497, respectively. The most fluctuating residues in apo are 100-123 which are present in the region connecting subdomain 1A and 1B, 168-173 in subdomain 1B which belong to the drug site III, 289-315 and 360-373 in subdomain 2B where there are loops and no helices present, 431-444 in subdomain 3A where the Sudlow's drug site II binding pocket locates, and 487-516 in the loop region connecting the subdomains 3A and 3B.

There are regions where the apo RMSF is higher compared to the complexes when a ligand binds at HSA drug site I pocket. These include residues from approx. 400-500 where the

subdomain 3A (drug site II) and subdomain 3B lie. This region fluctuates more in case of apo. Contrary to this, the presence of a ligand molecule in drug site I, indicates lower fluctuation of the region (400-500). The residues involved in contact with ligand molecules at this region are Arg410, Tyr411, Leu430, Val433, Ala449, Asp451, Leu453, Arg485 and Ser489. Similarly, a slight fluctuation in apo can be seen in the region of residues 200-300, which form the drug site I. This region, compared to apo, shows less fluctuation in complex except in the complex HSA-150MFC18b. The residues in contact with the ligand molecules at this region are Trp214, Arg218, Arg222, His242 and Arg257.

Similarly, the binding of a ligands (CHEMBL3401421, 138MFC9a, 39APR7i and CHEMBL459424S) at the HSA drug site II (Figure 4.8), showed more fluctuation in the complexes than in the apo structure. The more fluctuation can be seen in the beginning where the subdomains 1A and 1B are located. Similarly, the fluctuating residues 197-297 in complexes, which are present in subdomain 2A, showed more fluctuation than in the apo structure. The presence of a ligands (CHEMBL3401421, 138MFC9a and 39APR7i) bound at drug site II region (residues 384-497), showed slightly lower fluctuation than in the apo. However, it is not the case with the HSA-CHEMBL459424S, which showed even higher fluctuation when the ligand (CHEMBL459424S) is bound to drug site II. This could be because site II is located in a subdomain 3A, that corresponds to a more fluctuating region, relative to site I.

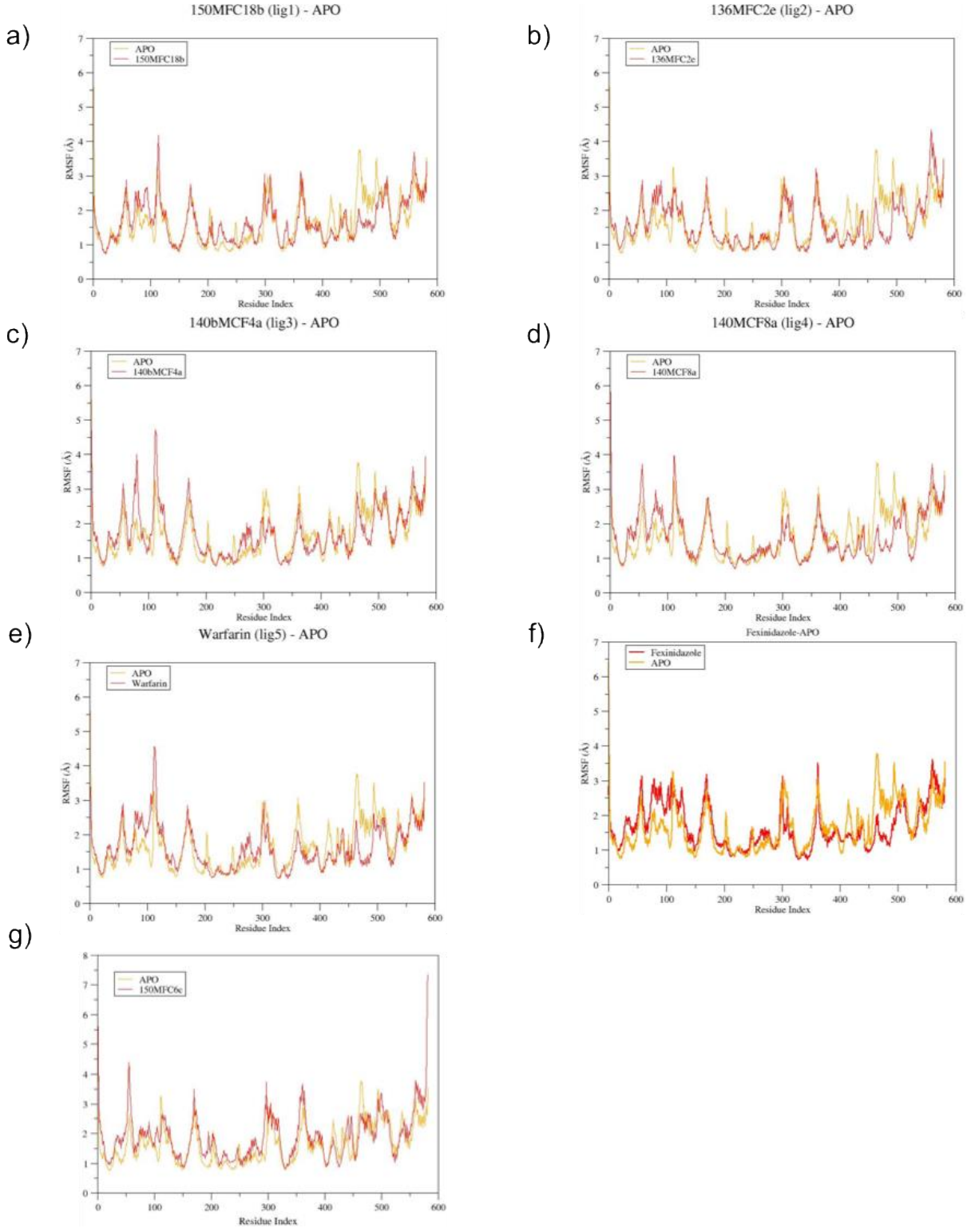


Figure 4.7: RMSF of compounds plotted against number of residues calculated from subjected to MD simulation. a) 150MFC18b, b) 136MFC2e, c) 140bMCF4a, d) 140MCF8a, e) Warfarin, f) Fexinidazole and g) 150MFC6c.

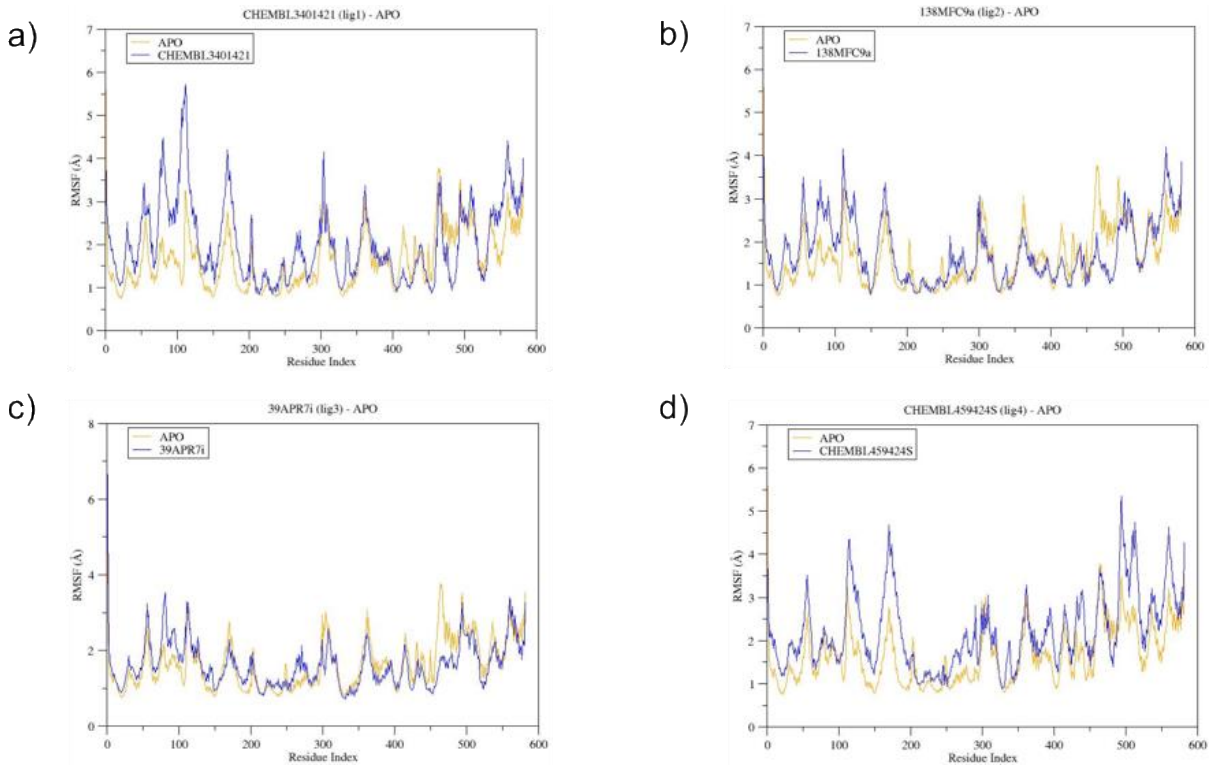


Figure 4.8: RMSF of compounds plotted against number of residues. a) CHEMBL3401421, b) 138MFC9a, c) 39APR7i and d) CHEMBL459424S.

4.2.3 Hydrogen bond count

Furthermore, the total number of hydrogen bonds formed between HSA and the docked ligands during the MD simulation was also analyzed (Figures 4.9 and 4.10). A lot of fluctuation was observed in the number of the hydrogen bonds in all the complexes.

In case of the HSA-ligand complexes at drug site I, the total number of hydrogen bonds detected ranged from 1 to 3 for the majority of complexes, with an exception of has-warfarin complex, where the hydrogen bond count was 1 to 6. The compound 150MFC18b and fexinidazole have the least (two) H-bonds, whereas 136MFC2e and warfarin have four and six H-bonds, respectively. The residues involved in H-bonds with the ligands at this drug site are Lys199, Arg218, Arg222, His242, Arg257 and Glu292.

The complexes at drug site II, the total number of hydrogen bonds detected ranged from 1 to 3. The compound CHEMBL3401421 and 39APR7i have the least (two) H-bonds, whereas three H-bonds were observed for 138MFC9a and CHEMBL459424S. The residues Arg485 and Ser489 are involved in H-bonds with the ligands at this drug site.

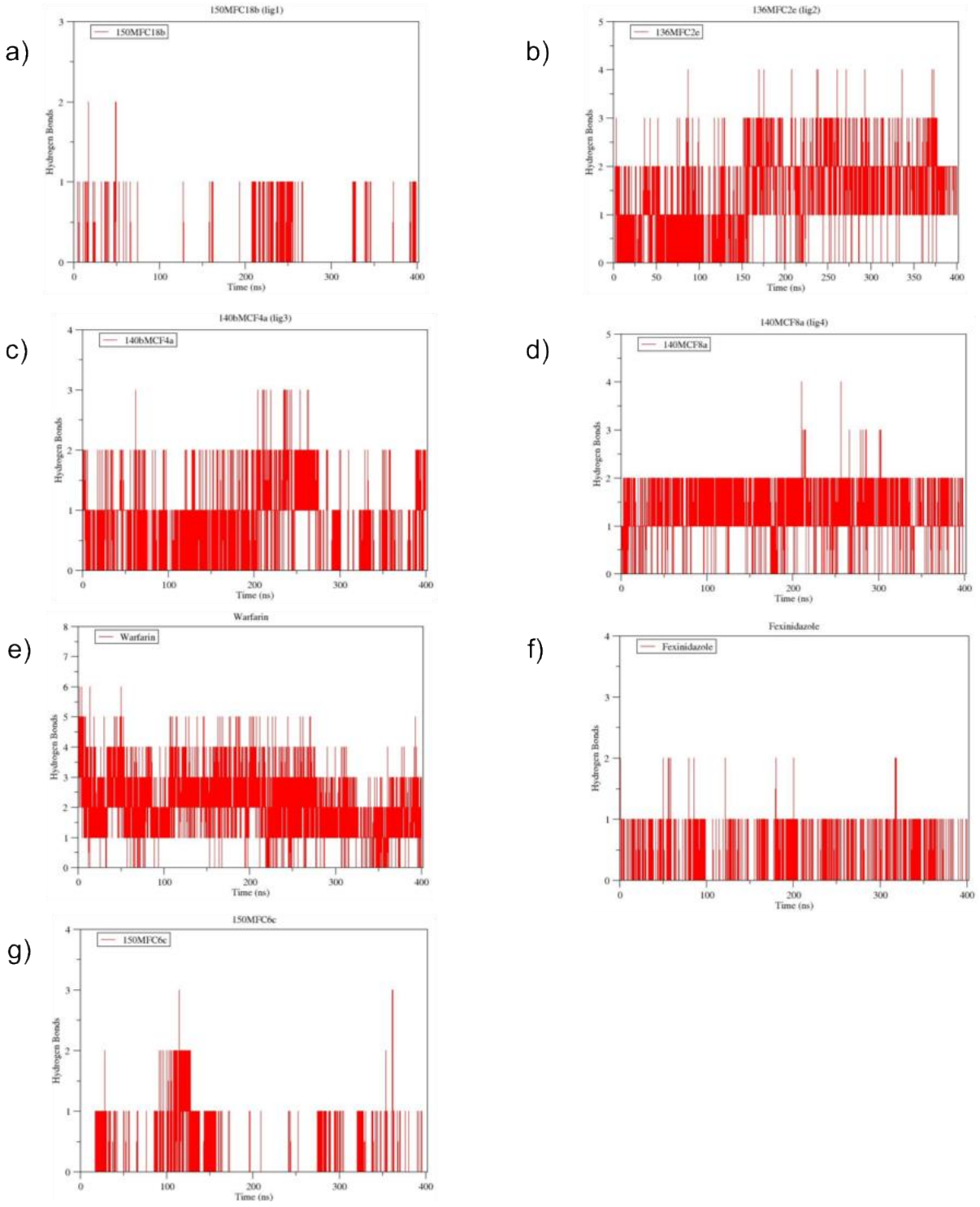


Figure 4.9: Number of ligand-protein hydrogen bonds for the docked compounds at HSA drug site I during the 400 ns. a) 150MFC18b, b) 136MFC2e, c) 140bMCF4a, d) 140MCF8a, e) Warfarin, f) Fexinidazole and g) 150MFC6c.

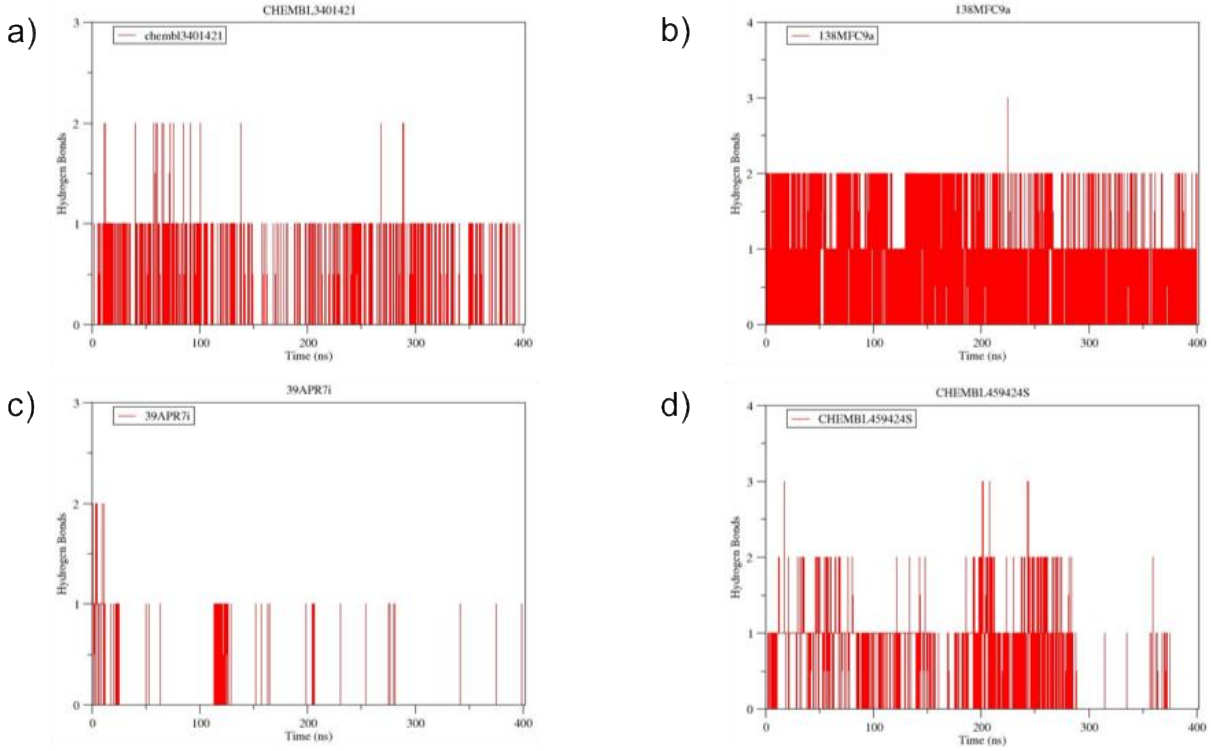


Figure 4.10: Number of ligand-protein hydrogen bonds for the docked compounds at HSA drug site II during the 400 ns. a) CHEMBL3401421, b) 138MFC9a, c) 39APR7i and d) CHEMBL459424S.

4.3 Binding free energy calculation

MM-GBSA approach was applied in order to calculate the binding free energies of the complexes. The 80 snapshots were obtained from the last 200 ns of the MD simulation with evenly spaced interval of 50 frames. The BFE estimation of the initial docked poses and their average BFE estimated from the last 200 ns of MD simulation are listed in Table 4.4.

In case of drug site I, the docking pose of warfarin showed the highest docking score of -7.780 kcal/mol. Warfarin is known for binding strongly to HSA with a plasma protein binding percentage of 99% (Valko et al., 2003). The highest docking score of warfarin correlates well with the RMSD, RMSF and H-bonds count data from the MD simulation (section 4.2). Warfarin has a maximum number (six) of H-bonds compared to the other studied ligands at drug site I. The RMSD plots of HSA-warfarin complex also showed a deviation of 2 to 4 Å, whereas the RMSF plot revealed less fluctuation of the protein Cα atoms than in the apo. The BFE estimation of warfarin-HSA docked poses showed a value of -51.91 kcal/mol, which is the second best after 150MFC6c. However, the average BFE estimated from the MD simulation (last 200 ns) revealed the opposite, a lower BFE value (-36.12 kcal/mol) compared to the initial docked pose. This is because warfarin docked pose moves in its place inside the binding pocket

during the MD simulation due to the flexibility of the HSA structure, leading to less favorable average BFE value (-36.12 kcal/mol).

Similarly, fexinidazole has a docking score of -5.03 kcal/mol, and BFE estimation for its docked poses of -46.82 kcal/mol. However, the average BFE value after MD simulation is lower (-35.01 kcal/mol), than the BFE of the docked pose. The MD trajectory analysis for fexinidazole showed RMSD convergence higher than for apo, RMSF fluctuation comparable to apo and H-bond count of two. The experimental plasma protein binding percentage value for fexinidazole is 95.4%, which does not correlate well with the average BFE value obtained by the MD simulation.

The compound 150MFC6c (also used for *in vitro* study) from the provided ligand library, showed the least favorable docking score of the simulated compounds (-3.29 kcal/mol). However, the BFE for its docked pose is calculated as the highest (-54.29 kcal/mol) among the studied compounds listed in Table 4.4. Furthermore, the average BFE calculated for this compound after MD indicated a decrease in value (-42.33 kcal/mol). This maybe because the BFE calculated after molecular docking has only one conformation, whereas the BFE calculated after MD is an average of all the conformations, with some conformations contributing more and others less.

Rest of the compounds 150MFC18b, 136MFC2e, 140bMCF4a and 140MCF8a studied at HSA drug site I, has a docking score lower than warfarin, but a very higher average BFE (after MD). Especially, the compound 150MFC18b revealed an average BFE -90.81 kcal/mol, significantly higher than that of warfarin. The highest BFE does not correlate with the H-bonds count of two for 150MFC18b, compared to warfarin having six H-bonds and a pi-pi interaction.

The compounds studied at the HSA drug site II showed docking scores similar to each other, with CHEMBL3401421 having the most favorable (-8.38 kcal/mol) and CHEMBL459424S the least favorable in this group (-7.98 kcal/mol). The compound 138MFC9a showed the least favorable BFE for docking, however the average BFE value after MD simulation is the most favorable (-72.50 kcal/mol). The favorable BFE value correlate well with its H-bonds count, which reveals that during the whole simulation time, it sustained two H-bonds, with a maximum of three H-bonds. Contrary to this, the RMSD plot shows its high fluctuation during the whole simulation, indicating that the complex did not get stabilized.

Table 4.4: Binding free energy dG (kcal/mol) estimation of docked poses and average BFE estimated from the last 200 ns of MD simulation (400 ns)

Drug site I			
Name of compound	**Docking score**	**BFE (kcal/mol)- for docked poses**	**Average BFE (kcal/mol)- *after MD**
150MFC18b	-6.76	-35.33	-90.81
136MFC2e	-6.72	-40.71	-45.78
140bMCF4a	-6.37	-35.09	-41.08
140MCF8a	-6.19	-21.99	-45.54
Warfarin	-7.80	-51.91	-36.12
Fexinidazole	-5.03	-46.82	-35.01
150MFC6c	-3.29	-54.29	-42.33
Drug site II			
CHEMBL3401421	-8.38	-45.47	-52.23
138MFC9a	-8.24	-35.27	-72.50
39APR7i	-8.12	-45.93	-45.15
CHEMBL459424S	-7.98	-42.04	-51.00

4.4 QikProp estimation

The ADME property of interest for the current work was the prediction of a compound's binding to human serum albumin (log K_{HSA}). The logarithm of HSA predicted binding constant log K_{HSA} parameter has a recommended range between −1.5 to 1.5 for 95% of known drugs. The QPlogKhsa estimates of these compounds by QikProp module of Schrödinger are shown in Table 4.5.

Table 4.5: Docking scores and QikProp predicted QPlogKhsa property of the MD simulated and the MST-tested compounds

HSA Drug Site-I			
Name of compound	**Docking score**	**QikProp prediction**	**Binding quality**
150MFC18b	-6.76	2.00	High
136MFC2e	-6.72	0.72	Intermediate
140bMCF4a	-6.37	1.41	High
140MCF8a	-6.19	1.46	High
Warfarin*	-7.80	0.09	Low
Fexinidazole	-5.03	-0.06	Low
150MFC6c	-3.29	1.23	Intermediate
HSA Drug Site-II			
CHEMBL3401421	-8.376	0.49	Intermediate
138MFC9a	-8.243	1.87	High

39APR7i	-8.125	1.99	High
CHEMBL459424S	-7.979	1.64	High

*Re-docked pose with the highest docking score.

The estimation of human serum albumin binding (QPlogKhsa) by QikProp reveals that the majority of the compounds from the provided library of our collaborator are predicted to bind strongly to HSA and thus they would not be abundantly available within the blood stream and they would not reach their target in required quantities if administered in too low doses. However, in our study, rigid docking and QPlogKhsa calculated with QikProp correlated very poorly with the experimental HSA binding affinity. Warfarin showed a QikProp value of 0.09 whereas it is very low compared to the literature value of 1.49 and a plasma protein binding affinity of 99 % (Valko et al., 2003). Similarly, fexinidazole showed a very low QikProp value -0.06. In contrary, the binding of fexinidazole to human serum protein reported *in vitro* in literature is relatively high 95.4%, however, the binding of its active metabolites showed lower binding values of 25.9 and 41.6% for M1 and M2, respectively (Gorris, 2021). Thus, QplogKhsa values predicted by QikProp may not be a reliable estimator for HSA binding affinity of ligands. QSAR and similar statistical predictive models are highly dependent on the quality and size of the training set which could limit their applicability to a certain domain. Additionally, the majority of the available crystal structures of HSA have poor resolution. Besides, the binding sites of HSA are inherently flexible which altogether render the development of a structure-based model for HSA binding challenging. Herein, Lexa et al. (2014) have developed a structure-based model of serum albumin binding while successfully incorporating albumin inherent flexibility. They applied the developed models on several series from which they observed that a combined model using logP and docking scores performed well across a broad range of test sets. However, logP alone showed only a modest correlation with HSA binding affinity while induced fit docking showed relatively good correlation but with few major outliers.

4.5 Micro Scale Thermophoresis

The microscale thermophoresis technique trace plot reveals a change in relative fluorescence as the concentration of a binding ligand increases. The binding of HSA with ligands (Figure 4.11) 150MFC6c, 39APR4b, 150MFC11 and Berr5 was experimentally evaluated using MST.

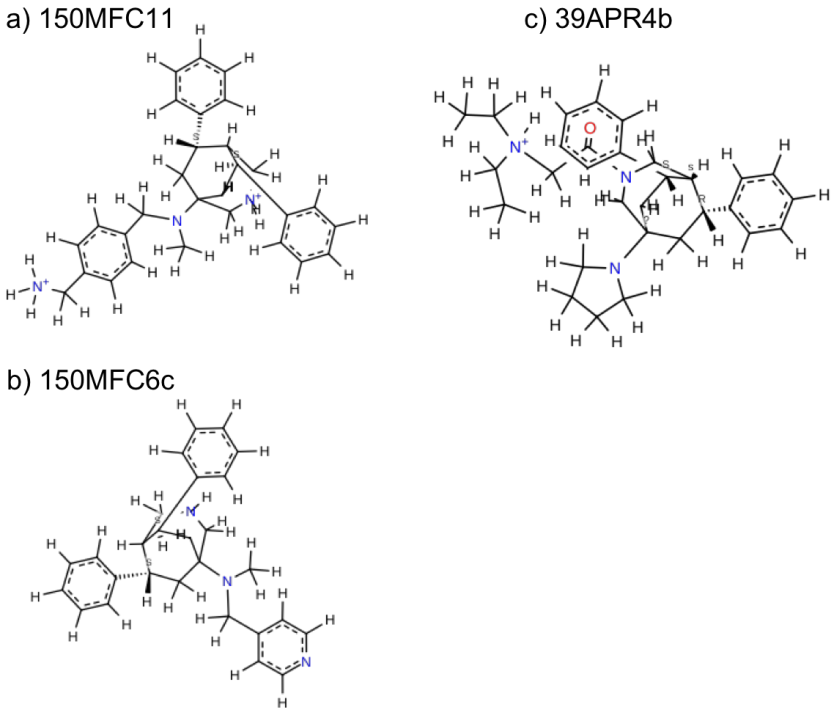


Figure 4.11: 2D illustration of the previously published ligands used for *in vitro* study. a) 150MFC11, b) 150MFC6c and c) 39APR4b.

During the *in vitro* study, none of the compounds showed binding affinity to HSA (Figure 4.12). According to the docking study (Table 4.6), the compound 150MFC6c showed some binding affinity to drug site I and drug site II with docking scores of -3.29 kcal/mol and -6.09 kcal/mol, respectively. On the other hand, the QikProp prediction for binding affinity to HSA for this compound suggests that 150MFC6c binds to HSA with a high affinity. Similarly, the *in silico* study for compound 150MFC11 suggests some binding affinity for HSA, with a docking score of -4.32 kcal/mol for site I and -6.14 kcal/mol for site II. The QikProp HSA binding prediction, however, suggests a high affinity of these compounds for HSA. The compound 39APR4b had poor docking scores of -0.49 kcal/mol and -1.64 kcal/mol for site I and site II, respectively, but a good prediction for human serum albumin binding affinity.

Table 4.6: Comparison of docking score and QikProp prediction for MST tested compounds.

Name of compound	Docking score (kcal/mol)	QikProp prediction	Binding quality according to QikProp
HSA Drug Site-I/II			
150MFC11	-4.32 / -6.14	1.27	High
150MFC6c	-3.29 / -6.09	1.20	High
39APR4b	-0.49 / -1.64	0.80	High

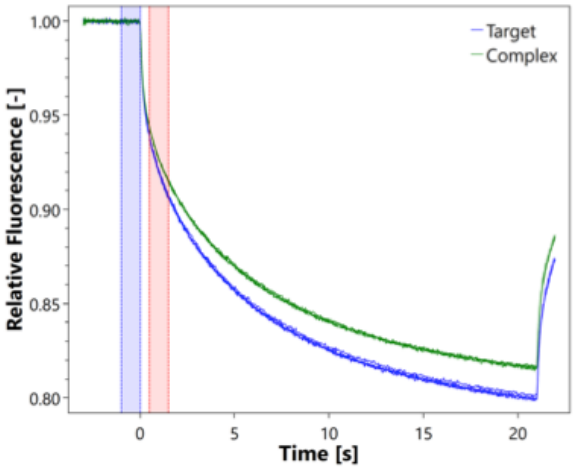
The *in vitro* study revealed that these compounds 150MFC6c, 39APR4b and 150MFC11 did not show any binding to HSA. The random signal in the Berr5 fluorescence time trace suggests nonspecific binding to HSA and aggregation induced by Berr5-HSA. The lack of these compounds binding can be due to low concentration used during the *in vitro* study. It might be possible that these compounds show more affinity for HSA if the concentration of the compounds are increased during experimental study.

The docking scores for compounds (150MFC11 and 150MFC6c) at drug site I and site II corelate with the QikProp estimation, suggesting their binding to HSA. In contrast, the experimental results show lack of binding for these compounds and thus does not coincides with docking and QikProp prediction.

Moreover, the docking score of 39APR4b at site I and site II suggested least favorable binding, and coincides well with the experimental result, which also shows no binding for 39APR4b. Contrary to this, it does not corelate with the QikProp estimation, which suggest its high affinity for HSA.

a) HSA-150MFC6c

MST Traces:



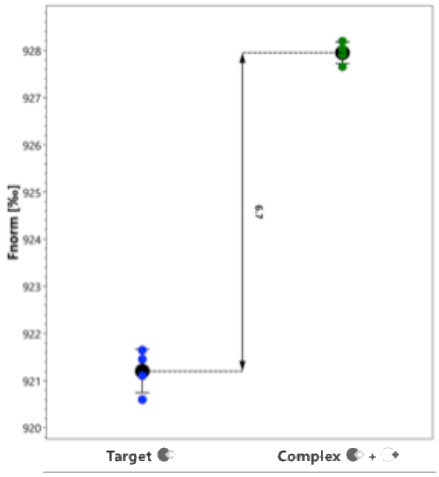
Target	Complex +
No aggregates	No aggregates
Photobleaching Rate 0.1 ‰/s	Photobleaching Rate 0.1 ‰/s

No Ligand Induced Photobleaching Rate Change

Cursor positions

Cold Region: -1s - 0s
Hot Region: 0.5s - 1.5s

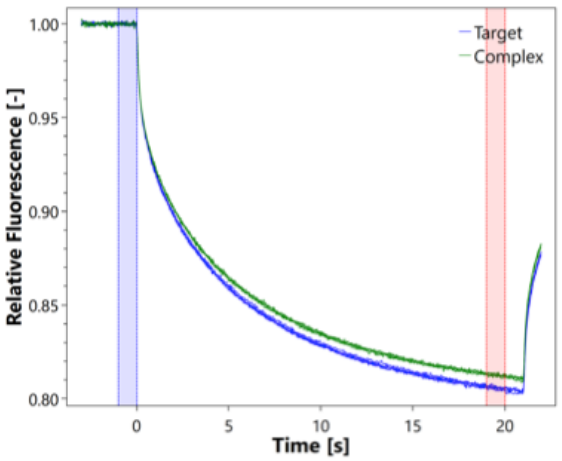
Signal to Noise Ratio:



Response Evaluation:	On Time 1.5s	
Response:	921	928
Noise:	0.5	0.2
Response Amplitude:	6.7	
Signal to Noise Ratio:	19.6	

b) HSA-39APR4b

MST Traces:



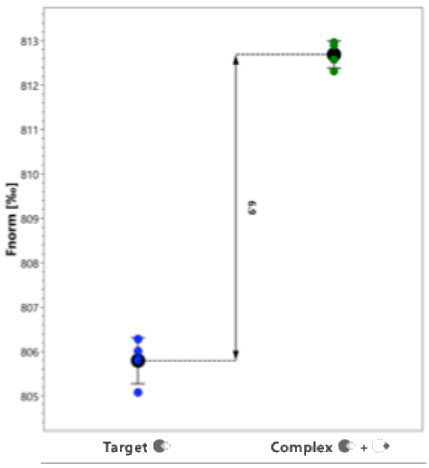
Target	Complex +
No aggregates	No aggregates
Photobleaching Rate 0.1 ‰/s	Photobleaching Rate 0.1 ‰/s

No Ligand Induced Photobleaching Rate Change

Cursor positions

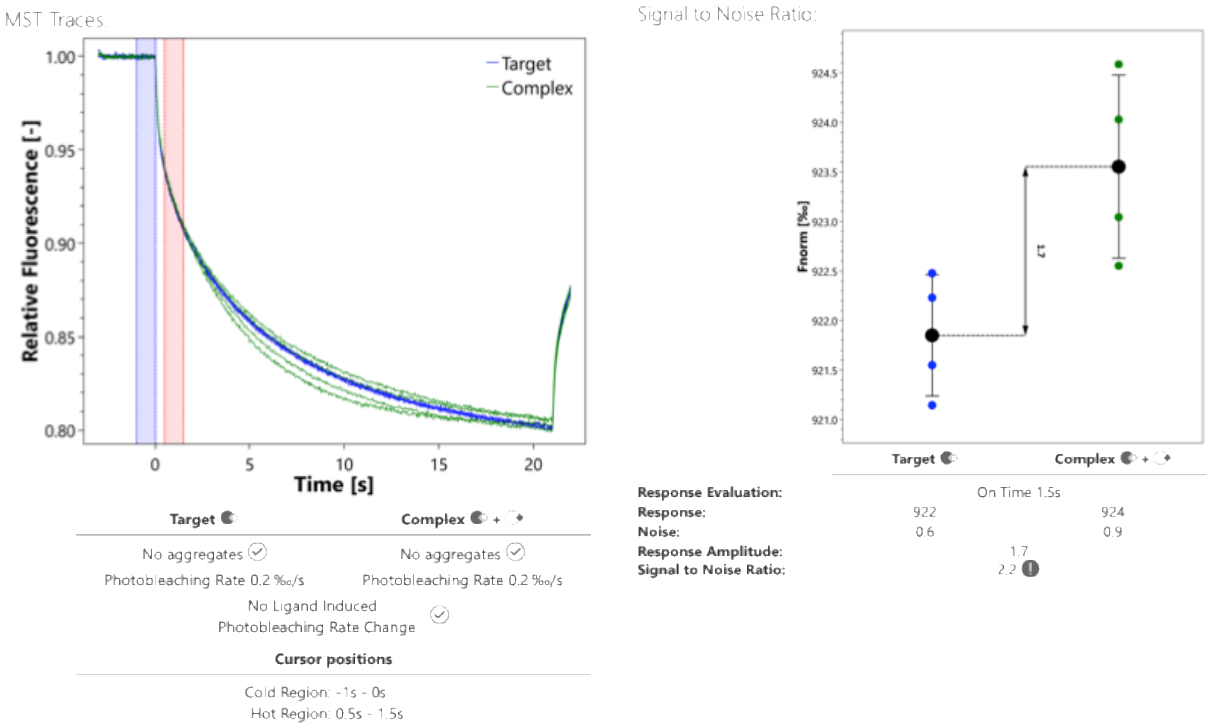
Cold Region: -1s - 0s
Hot Region: 19s - 20s

Signal to No se Ratio:



Response Evaluation:	On Time 20s	
Response:	806	813
Noise:	0.5	0.3
Response Amplitude:	6.9	
Signal to Noise Ratio:	16.9	

c) HSA-150MFC11



d) HSA-Berr5

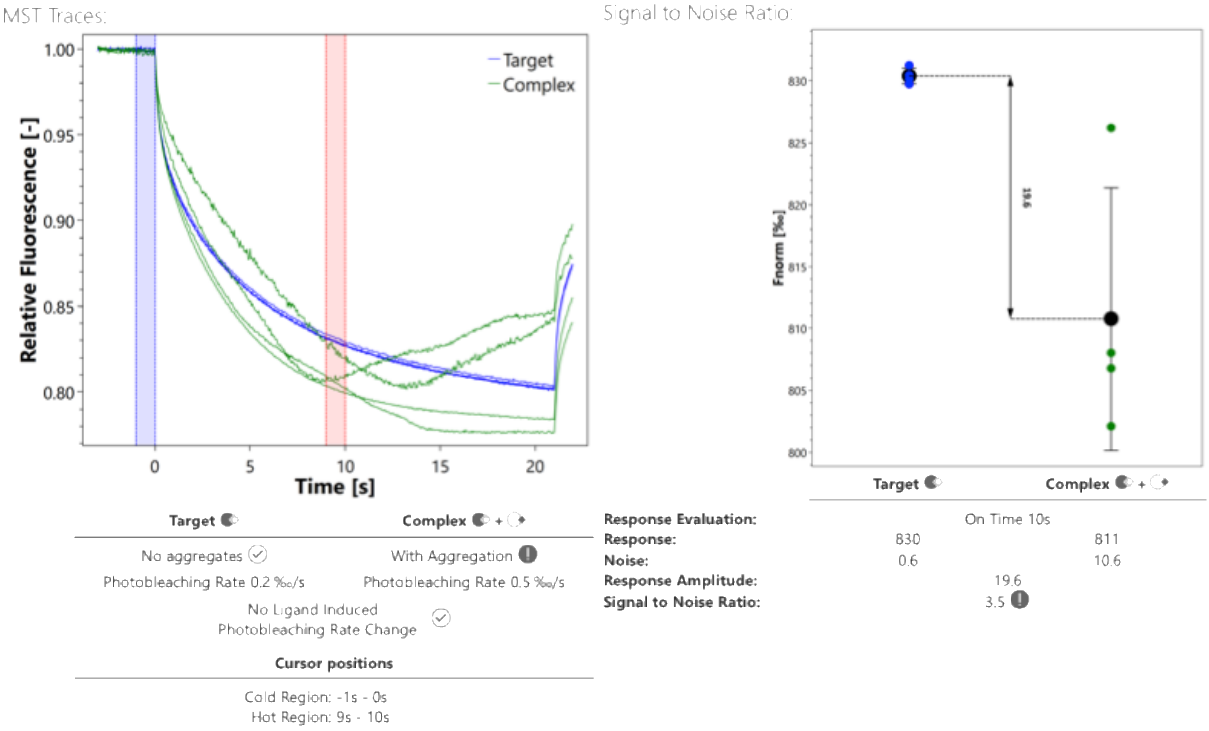


Figure 4.12: Thermophoresis raw data of HSA binding. MST trace output (left) during the binding affinity test. The blue and red vertical lines are the regions analogous to cold and hot fluorescence thermophoresis, respectively. Binding response output (right) after binding affinity test. a) 150MFC6c, b) 39APR4b, c) 150MFC11, d) Berr5, aggregation (left) of the sample during the binding affinity test.

5 Conclusions

Human African trypanosomiasis (HAT), or sleeping sickness, is an endemic disease in 36 sub-Saharan African countries caused by a group of parasites called *Trypanosoma brucei* (Tb). If left untreated, the disease could lead to death in the second stage. Entamidine and suramin are drugs developed for treating the disease at the first stage while melarsoprol, nifurtimox-eflornithine combination therapy (NECT) are approved for treating the disease at the second stage. Besides, fexinidazole is a newly approved oral drug for the treatment of the first stage and non-severe second stage of HAT.

Our collaborator's research group at the University of Graz, Austria had prepared and tested several antiparasitic compounds against Tb and *Plasmodium falciparum in vitro* for which they showed varying levels of activity. Nonetheless, the compounds tested showed moderate activity *in vivo* if at all. It was suggested that the poor *in vivo* activities might be attributed to strong binding to plasma proteins or to rapid metabolism before reaching the target sites to achieve a therapeutic effect. Human serum albumin (HSA) is an abundant plasma protein that can bind to a different variety drugs altering their pharmacokinetics and pharmacodynamics properties. Herein, the aim of this study was to explore whether a strong affinity of the provided antiparasitic compounds to HSA binding could be a plausible reason for their poor *in vivo* activity. The binding interaction between the provided antiparasitic compounds and HSA was studied *in silico.* The co-crystallized ligand warfarin was self-docked to replicate the poses in the binding site. The antiparasitic compounds were docked into HSA drug binding sites I and II. Additionally, the compounds with best docking scores were further investigated using MD simulations to assess the stability of the binding interactions. Furthermore, the HSA binding affinities of the tested compounds estimated as average binding free energies using the Prime/MM-GBSA approach were calculated throughout the second half of the MD simulations. Besides, QPlogKhsa was predicted for the tested compounds using QikProp.

The self-docking of warfarin showed that the generated poses with quite similar BFE scores and overlapping orientations and positions, with slight variation, compared to the co-crystallized ligand. The molecular docking result indicated that majority of antiparasitic compounds may bind to the drug site I with varying affinities. The MD simulations suggest more stability of the studied compounds at site I. The RMSD of these compounds at site I converged similarly or at a lower level than that of the apo, the RMSF of protein Cα atoms showed that majority of complexes fluctuated less compared to the apo structure and the H-bonds counts detected ranged from 1 to 3, with an exception of HSA-warfarin complex, where

the hydrogen bond count was 1 to 6. In case of the HSA drug site II, the HSA-ligand complexes showed lower RMSD convergence and more fluctuation of the protein Cα atoms compared to the apo, with a number of ligand-protein H-bonds ranging from 1 to 3. The higher fluctuation at drug site II, corresponds to its location in a subdomain that has a more unstable protein conformation relative to site I. Additionally, the average BFE calculation for complexes may suggest better affinities of compounds for HSA at site I. However, low estimation of the average BFE for warfarin does not correlate with the literature data, as warfarin has a plasma protein binding affinity of 99%.

Although QikProp results suggested that most tested compounds have a high to intermediate binding affinity to HSA, the results of warfarin and fexinidazole were not aligned with results reported in literature. Therefore, the reliability of the QikProp tool for the prediction of logKhsa might be questionable.

Thereafter, some of the provided compounds were tested for their experimental HSA binding affinity *in vitro* using Microscale thermophoresis (MST). However, none of the tested compounds showed binding affinity to HSA. Regardless, the concentration of the compounds might not be high enough to bind to HSA.

Lastly, the current study suggests that plasma protein binding may be the reason for the *in vivo* inactivity for some of the investigated antiparasitic compounds against *T. brucei gambiense*.

REFERENCES

Ahmad, S., Seebacher, W., Faist, J., Kaiser, M., Brun, R., Saf, R., & Weis, R. (2016). The antiprotozoal potencies of newly prepared 3-azabicyclo[3.2.2]nonanes. *Archives of Pharmacal Research*, *39*(10), 1391–1403. https://doi.org/10.1007/s12272-016-0826-5

Ahmad, S., Seebacher, W., Wolkinger, V., Presser, A., Faist, J., Kaiser, M., Brun, R., Saf, R., & Weis, R. (2015). Synthesis and antiprotozoal activities of new 3-azabicyclo[3.2.2]nonanes. *Archives of Pharmacal Research*, *38*(8), 1455–1467. https://doi.org/10.1007/s12272-014-0523-1

Ahmad, S., Weis, R., Faist, J., Wolkinger, V., Saf, R., Belaj, F., Brun, R., Kaiser, M., & Seebacher, W. (2012). Diarylcyclohexanones: Synthons for new bicyclic compounds. *Monatshefte Fur Chemie*, *143*(1), 145–152. https://doi.org/10.1007/s00706-011-0679-z

Artali, R., Bombieri, G., Calabi, L., & Del Pra, A. (2005). A molecular dynamics study of human serum albumin binding sites. *Il Farmaco*, *60*(6–7), 485–495. https://doi.org/10.1016/J.FARMAC.2005.04.010

Aureli, L., Cruciani, G., Cesta, M. C., Anacardio, R., De Simone, L., & Moriconi, A. (2005). *Predicting Human Serum Albumin Affinity of Interleukin-8 (CXCL8) Inhibitors by 3D-QSPR Approach*. https://doi.org/10.1021/jm049227l

Backes, W. L. (2007). *Pharmacokinetics*. 2007.

Berger, H., Seebacher, W., Kaiser, M., Brun, R., Saf, R., & Weis, R. (2009). SARs of the antiprotozoal action of 6,7-diaryl-bicyclo[2.2.2]octan-2-ols. *Monatshefte Fur Chemie*, *140*(5), 495–502. https://doi.org/10.1007/s00706-008-0079-1

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. In *Nucleic Acids Research* (Vol. 28, Issue 1). http://www.rcsb.org/pdb/status.html

Bernetti, M., & Bussi, G. (2020). Pressure control using stochastic cell rescaling. *Journal of Chemical Physics*, *153*(11), 1–11. https://doi.org/10.1063/5.0020514

Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., & MacKerell, A. D. (2012). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 Dihedral Angles. *Journal of Chemical Theory and Computation*, *8*(9), 3257–3273. https://doi.org/10.1021/ct300400x

Beveridge, D. L., & Jorgensen, W. L. (1988).) For reviews, see: (a) Levitt, M. In *Annu. Rev.*

Biophys. Bioeng (Vol. 110, Issue 6). UTC. https://pubs.acs.org/sharingguidelines

Bohnert, T., & Gan, L. S. (2013). Plasma protein binding: From discovery to development. *Journal of Pharmaceutical Sciences*, *102*(9), 2953–2994. https://doi.org/10.1002/jps.23614

Bowers, K. J., Chow, D. E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., Klepeis, J. L., Kolossvary, I., Moraes, M. A., Sacerdoti, F. D., Salmon, J. K., Shan, Y., & Shaw, D. E. (2007). *Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters*. 43–43. https://doi.org/10.1109/sc.2006.54

Brun, R., Blum, J., Chappuis, F., & Burri, C. (2010). Human African trypanosomiasis. *The Lancet*, *375*(9709), 148–159. https://doi.org/10.1016/S0140-6736(09)60829-1

Buscher, Brigitte, et al. (2014). W hite P aper Bioana - lysis for plasma protein binding studies in drug discovery and drug development : views and recommendations of the European Bioana - lysis Forum. *Bioanalysis*.

Büscher, P., Cecchi, G., Jamonneau, V., & Priotto, G. (2017). Human African trypanosomiasis. *The Lancet*, *390*(10110), 2397–2409. https://doi.org/10.1016/S0140-6736(17)31510-6

Bussi, G., Donadio, D., & Parrinello, M. (2007). Canonical sampling through velocity rescaling. *Journal of Chemical Physics*, *126*(1). https://doi.org/10.1063/1.2408420

Carter, D. C., & Ho, J. X. (1994). Structure of serum albumin. *Advances in Protein Chemistry*, *45*(C), 153–176. https://doi.org/10.1016/S0065-3233(08)60640-3

Cheng, F., Li, W., Zhou, Y., Shen, J., Wu, Z., Liu, G., Lee, P. W., & Tang, Y. (2012). AdmetSAR: A comprehensive source and free tool for assessment of chemical ADMET properties. *Journal of Chemical Information and Modeling*, *52*(11), 3099–3105. https://doi.org/10.1021/ci300367a

Chushak, Y. G., Chapleau, R. R., Frey, J. S., Mauzy, C. A., & Gearhart, J. M. (2015). Identifying potential protein targets for toluene using a molecular similarity search, in silico docking and in vitro validation. *Toxicology Research*, *4*(2), 519–526. https://doi.org/10.1039/c5tx00009b

Colmenarejo, G., Alvarez-Pedraglio, A., & Lavandera, J. L. (2001). Cheminformatic models to predict binding affinities to human serum albumin. *Journal of Medicinal Chemistry*, *44*(25), 4370–4378. https://doi.org/10.1021/jm010960b

Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*, *7*(October 2016), 1–13. https://doi.org/10.1038/srep42717

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, *98*(12), 10089–10092. https://doi.org/10.1063/1.464397

Deeb, O., Cecilia Rosales-Hernández, M., Gó mez-Castro, C., Garduño-Juárez, R., & Correa-Basurto, J. (2009). *Exploration of Human Serum Albumin Binding Sites by Docking and Molecular Dynamics Flexible Ligand-Protein Interactions*. https://doi.org/10.1002/bip.21314

Duhr, S., & Braun, D. (2006). Why molecules move along a temperature gradient. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(52), 19678–19682. https://doi.org/10.1073/pnas.0603873103

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, *103*(19), 8577–8593. https://doi.org/10.1063/1.470117

Faist, J., Seebacher, W., Kaiser, M., Brun, R., Saf, R., & Weis, R. (2009). Bicyclic amido compounds with antiprotozoal activity. *Monatshefte Fur Chemie*, *140*(10), 1261–1268. https://doi.org/10.1007/s00706-009-0164-0

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., & Shenkin, P. S. (2004). *Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy*. https://doi.org/10.1021/jm0306430

Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., & Mainz, D. T. (2006). Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *Journal of Medicinal Chemistry*, *49*(21), 6177–6196. https://doi.org/10.1021/jm051256o

Ghafourian, T., & Amin, Z. (2013). QSAR models for the prediction of plasma protein binding. *BioImpacts*, *3*(1), 21–27. https://doi.org/10.5681/bi.2013.011

Ghuman, J., Zunszain, P. A., Petitpas, I., Bhattacharya, A. A., Otagiri, M., & Curry, S. (2005). Structural basis of the drug-binding specificity of human serum albumin. *Journal of Molecular Biology*, *353*(1), 38–52. https://doi.org/10.1016/j.jmb.2005.07.075

González, M. . (2011). Force fields and molecular dynamics simulations. *École Thématique de La Société Française de La Neutronique*, *12*, 169–200.

Gorris, E. (2021). Annex I: *Power and Politics in the Neo-Elamite Kingdom*, 241–255. https://doi.org/10.2307/j.ctv1q26vdb.13

Greenwood, J. R., Calkins, D., Sullivan, A. P., & Shelley, J. C. (2010). Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. In *Journal of Computer-Aided Molecular Design* (Vol. 24, Issues 6–7, pp. 591–604). https://doi.org/10.1007/s10822-010-9349-1

Guedes, I. A., Pereira, F. S. S., & Dardenne, L. E. (2018). Empirical scoring functions for structure-based virtual screening: Applications, critical aspects, and challenges. *Frontiers in Pharmacology*, *9*(SEP), 1–18. https://doi.org/10.3389/fphar.2018.01089

Gumede, N. J., Singh, P., Sabela, M. I., Bisetty, K., Escuder-Gilabert, L., Medina-Hernández, M. J., & Sagrado, S. (2012). Experimental-like affinity constants and enantioselectivity estimates from flexible docking. *Journal of Chemical Information and Modeling*, *52*(10), 2754–2759. https://doi.org/10.1021/ci300335m

Halgren, T. (2007). Technology Platform [a] 2007 The Author Journal compilation [a]. *Chem Biol Drug Des*, *69*, 146–148. https://doi.org/10.1111/j.1747-0285.2007.00483.x

Halgren, T. A. (2009). Identifying and characterizing binding sites and assessing druggability. *Journal of Chemical Information and Modeling*, *49*(2), 377–389. https://doi.org/10.1021/ci800324m

Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Thomas Pollard, W., & Banks, J. L. (2004). *Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening*. https://doi.org/10.1021/jm030644s

Halperin, I., Ma, B., Wolfson, H., & Nussinov, R. (2002). Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins: Structure, Function and Genetics*, *47*(4), 409–443. https://doi.org/10.1002/prot.10115

Heinz, H., Vaia, R. A., Farmer, B. L., & Naik, R. R. (2008). Accurate simulation of surfaces and interfaces of face-centered cubic metals using 12-6 and 9-6 lennard-jones potentials. *Journal of Physical Chemistry C*, *112*(44), 17281–17290. https://doi.org/10.1021/jp801931d

Hoffelner, M., Petritsch, M., Ahmad, S., Seebacher, W., Dolensky, J., Hochegger, P., Kaiser, M., Mäser, P., Saf, R., & Weis, R. (2019). New derivatives of 3-azabicyclo[3.2.2]nonanes and their antiprotozoal activities. *Monatshefte Fur Chemie*, *150*(11), 1959–1972. https://doi.org/10.1007/s00706-019-02509-0

Hollingsworth, S. A., & Dror, R. O. (2018). Molecular Dynamics Simulation for All. *Neuron*, *99*(6), 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., & Simmerling, C. (2006).

Comparison of multiple amber force fields and development of improved protein backbone parameters. In *Proteins: Structure, Function and Genetics* (Vol. 65, Issue 3, pp. 712–725). https://doi.org/10.1002/prot.21123

Jacobson, M. P., Friesner, R. A., Xiang, Z., & Honig, B. (2002). On the role of the crystal environment in determining protein side-chain conformations. *Journal of Molecular Biology*, *320*(3), 597–608. https://doi.org/10.1016/S0022-2836(02)00470-9

Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J. F., Honig, B., Shaw, D. E., & Friesner, R. A. (2004). A Hierarchical Approach to All-Atom Protein Loop Prediction. *Proteins: Structure, Function and Genetics*, *55*(2), 351–367. https://doi.org/10.1002/prot.10613

Jain, A. (2017). Computer aided drug design. *Journal of Physics: Conference Series*, *884*(1), 504–509. https://doi.org/10.1088/1742-6596/884/1/012072

Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P., & Duhr, S. (2011a). Molecular interaction studies using microscale thermophoresis. *Assay and Drug Development Technologies*, *9*(4), 342–353. https://doi.org/10.1089/adt.2011.0380

Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P., & Duhr, S. (2011b). Molecular interaction studies using microscale thermophoresis. In *Assay and Drug Development Technologies* (Vol. 9, Issue 4, pp. 342–353). https://doi.org/10.1089/adt.2011.0380

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, *79*(2), 926–935. https://doi.org/10.1063/1.445869

Joshi, S. Y., & Deshmukh, S. A. (2020). *A review of advancements in coarse-grained molecular dynamics simulations*. https://doi.org/10.1080/08927022.2020.1828583

Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews Drug Discovery*, *3*(11), 935–949. https://doi.org/10.1038/nrd1549

Kumar, R., Sharma, A., Siddiqui, M. H., & Tiwari, R. K. (2017). Prediction of Drug-Plasma Protein Binding Using Artificial Intelligence Based Algorithms. *Combinatorial Chemistry & High Throughput Screening*, *21*(1), 57–64. https://doi.org/10.2174/1386207321666171218121557

Lambrinidis, G., Vallianatou, T., & Tsantili-Kakoulidou, A. (2015). In vitro, in silico and integrated strategies for the estimation of plasma protein binding. A review. In *Advanced Drug Delivery Reviews* (Vol. 86, pp. 27–45). Elsevier. https://doi.org/10.1016/j.addr.2015.03.011

Levitt, M., Hirshberg, M., Sharon, R., & Daggett, V. (1995). Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. *Computer Physics Communications*, *91*(1–3), 215–231. https://doi.org/10.1016/0010-4655(95)00049-L

Lexa, K. W., Dolghih, E., & Jacobson, M. P. (2014). A Structure-Based Model for Predicting Serum Albumin Binding. *PLoS ONE*, *9*(4), 93323. https://doi.org/10.1371/journal.pone.0093323

Li, J., Abel, R., Zhu, K., Cao, Y., Zhao, S., & Friesner, R. A. (2011). The VSGB 2.0 model: A next generation energy model for high resolution protein structure modeling. *Proteins: Structure, Function and Bioinformatics*, *79*(10), 2794–2812. https://doi.org/10.1002/prot.23106

Li, J., Fu, A., & Zhang, L. (2019). An Overview of Scoring Functions Used for Protein–Ligand Interactions in Molecular Docking. *Interdisciplinary Sciences – Computational Life Sciences*, *11*(2), 320–328. https://doi.org/10.1007/s12539-019-00327-w

Lin, F.-Y., & Mackerell, A. D. (2019). Force fields for small molecules HHS Public Access. *Methods Mol Biol*, *2022*, 21–54. https://doi.org/10.1007/978-1-4939-9608-7_2

Lu, C., Wu, C., Ghoreishi, D., Chen, W., Wang, L., Damm, W., Ross, G. A., Dahlgren, M. K., Russell, E., Von Bargen, C. D., Abel, R., Friesner, R. A., & Harder, E. D. (2021). OPLS4: Improving force field accuracy on challenging regimes of chemical space. *Journal of Chemical Theory and Computation*, *17*(7), 4291–4300. https://doi.org/10.1021/acs.jctc.1c00302

Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W. (2013). Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design*, *27*(3), 221–234. https://doi.org/10.1007/s10822-013-9644-8

Malvy, D., & Chappuis, F. (2011). Sleeping sickness. *Clinical Microbiology and Infection*, *17*(7), 986–995. https://doi.org/10.1111/J.1469-0691.2011.03536.X

Mark, P., & Nilsson, L. (2001). Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *Journal of Physical Chemistry A*, *105*(43), 9954–9960. https://doi.org/10.1021/jp003020w

Marrink, S. J., Risselada, H. J., Yefimov, S., Peter, D., & De Vries, A. H. (2007). *The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations*. https://doi.org/10.1021/jp071097f

Martyna, G. J., Klein, M. L., & Tuckerman, M. (1992). Nosé-Hoover chains: The canonical

ensemble via continuous dynamics. *The Journal of Chemical Physics*, *97*(4), 2635–2643. https://doi.org/10.1063/1.463940

Martyna, G. J., Tobias, D. J., & Klein, M. L. (1994). Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*, *101*(5), 4177–4189. https://doi.org/10.1063/1.467468

Meng Zhang, H. X., Mezei, M., & Cui, M., X. Y. (2011). MMeng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. 2011. "Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. Current Computer-Aided Drug Design." Current Computer Aided Drug Design 7 (2): 146–57.olecular docking: a powerful approach . *Current Computer Aided Drug Design*, *7*(2), 146–157. https://www.ingentaconnect.com/content/ben/cad/2011/00000007/00000002/art00008%0Ahttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf

Mohsin, N. U. A., Seebacher, W., Faist, J., Hochegger, P., Kaiser, M., Mäser, P., Saf, R., & Weis, R. (2018). The antiplasmodial and antitrypanosomal activities of novel piperazine derivatives of 3-azabicyclo[3.2.2]nonanes. *Monatshefte Fur Chemie*, *149*(1), 99–109. https://doi.org/10.1007/s00706-017-2109-3

Moroy, G., Martiny, V. Y., Vayer, P., Villoutreix, B. O., & Miteva, M. A. (2012). Toward in silico structure-based ADMET prediction in drug discovery. *Drug Discovery Today*, *17*(1–2), 44–55. https://doi.org/10.1016/J.DRUDIS.2011.10.023

Mullokandov, E., Ahn, J., Szalkiewicz, A., & Babayeva, M. (2014). Protein Binding Drug-Drug Interaction between Warfarin and Tizoxanide in Human Plasma. *Austin Journal of Pharmacology and Therapeutics*, *2*(7), 1–3. http://austinpublishinggroup.com/pharmacology-therapeutics/fulltext/ajpt-v2-id1038.pdf

Narwal, M., Kumar, D., Mukherjee, T. K., Bhattacharyya, R., & Banerjee, D. (2018). Molecular dynamics simulation as a tool for assessment of drug binding property of human serum albumin. *Molecular Biology Reports*, *45*(6), 1647–1652. https://doi.org/10.1007/s11033-018-4308-3

NosÉ, S. (2002). A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, *100*(1), 191–198. https://doi.org/10.1080/00268970110089108

Oostenbrink, C., Villa, A., Mark, A. E., & Van Gunsteren, W. F. (2004). A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J Comput Chem*, *25*, 1656–1676. https://doi.org/10.1002/jcc.20090

Patodia, S. (2014). Molecular Dynamics Simulation of Proteins: A Brief Overview. *Journal of*

Physical Chemistry & Biophysics, *4*(6). https://doi.org/10.4172/2161-0398.1000166

Patrick Walters, W., Stahl, M. T., & Murcko, M. A. (1998). Virtual screening—an overview. *Drug Discovery Today*, *3*(4), 160–178. https://doi.org/10.1016/S1359-6446(97)01163-X

Pérez-Molina, J. A., & Molina, I. (2018). Chagas disease. *The Lancet*, *391*(10115), 82–94. https://doi.org/10.1016/S0140-6736(17)31612-4

Petitpas, I., Bhattacharya, A. A., East, M., & Curry, S. (2001). *Crystal Structure Analysis of Warfarin Binding to Human Serum Albumin ANATOMY OF DRUG SITE I**. https://doi.org/10.1074/jbc.M100575200

Ponder, J. W., & Case, D. A. (2003). Force fields for protein simulations. *Advances in Protein Chemistry*, *66*, 27–85. https://doi.org/10.1016/S0065-3233(03)66002-X

Pragna Lakshmi, T., Mondal, M., Ramadas, K., & Natarajan, S. (2017). Molecular interaction of 2,4-diacetylphloroglucinol (DAPG) with human serum albumin (HSA): The spectroscopic, calorimetric and computational investigation. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *183*, 90–102. https://doi.org/10.1016/J.SAA.2017.04.012

Predescu, C., Lerer, A. K., Lippert, R. A., Towles, B., Grossman, J. P., Dirks, R. M., & Shaw, D. E. (2020). The u -series: A separable decomposition for electrostatics computation with improved accuracy. *Journal of Chemical Physics*, *152*(8). https://doi.org/10.1063/1.5129393

Raval, K., & Ganatra, T. (2022). Basics, types and applications of molecular docking: A review. *IP International Journal of Comprehensive and Advanced Pharmacology*, *7*(1), 12–16. https://doi.org/10.18231/j.ijcaap.2022.003

Rimac, H., Debeljak, Ž., Bojić, M., & Miller, L. (2017). Displacement of Drugs from Human Serum Albumin: From Molecular Interactions to Clinical Significance. *Current Medicinal Chemistry*, *24*(18), 1930–1947. https://doi.org/10.2174/0929867324666170202152134

Sagui, C., & Darden, T. A. (1999). MOLECULAR DYNAMICS SIMULATIONS OF BIOMOLECULES: Long-Range Electrostatic Effects. In *Annu. Rev. Biophys. Biomol. Struct* (Vol. 28). www.annualreviews.org

Salo-Ahen, O. M. H., Alanko, I., Bhadane, R., Alexandre, A. M., Honorato, R. V., Hossain, S., Juffer, A. H., Kabedev, A., Lahtela-Kakkonen, M., Larsen, A. S., Lescrinier, E., Marimuthu, P., Mirza, M. U., Mustafa, G., Nunes-Alves, A., Pantsar, T., Saadabadi, A., Singaravelu, K., & Vanmeert, M. (2021). Molecular dynamics simulations in drug discovery and pharmaceutical development. In *Processes* (Vol. 9, Issue 1, pp. 1–63).

https://doi.org/10.3390/pr9010071

Seebacher, W., Berger, H., Kaiser, M., Brun, R., Saf, R., & Weis, R. (2006). Structural requirements for the antiprotozoal activity of 4-aminobicyclo[2.2.2]octan-2-ols. *Monatshefte Fur Chemie*, *137*(4), 471–482. https://doi.org/10.1007/s00706-006-0455-7

Seebacher, W., Kaiser, M., Brun, R., Saf, R., & Weis, R. (2005). 4-Aminobicyclo[2.2.2]octan-2-ones and -ols via enamine intermediates. *Monatshefte Fur Chemie*, *136*(4), 625–634. https://doi.org/10.1007/s00706-004-0264-9

Seebacher, W., Kaiser, M., Brun, R., Saf, R., & Weis, R. (2007a). Antiprotozoal activities of epimeric aminobicycles. *Monatshefte Fur Chemie*, *138*(7), 709–714. https://doi.org/10.1007/s00706-007-0670-x

Seebacher, W., Kaiser, M., Brun, R., Saf, R., & Weis, R. (2007b). New 4-amino-2-azabicyclo[3.2.2]nonane derivatives and their antiprotozoal potencies. *Monatshefte Fur Chemie*, *138*(6), 619–625. https://doi.org/10.1007/s00706-007-0634-1

Shelley, J. C., Cholleti, A., Frye, L. L., Greenwood, J. R., Timlin, M. R., & Uchimaya, M. (2007). Epik: A software program for pKa prediction and protonation state generation for drug-like molecules. *Journal of Computer-Aided Molecular Design*, *21*(12), 681–691. https://doi.org/10.1007/s10822-007-9133-z

Sjoholm, I., Ekman, B., Kober, A., Ljungstedt-Påhlman, I., Seiving, B., & Sjödin, T. (1979). Binding of drugs to human serum albumin: XI. The specificity of three binding sites as studied with albumin immobilized in microparticles. *Molecular Pharmacology*, *16*(3), 767–777.

Sousa, S. F., Fernandes, P. A., & Ramos, M. J. (2006). Protein-ligand docking: Current status and future challenges. In *Proteins: Structure, Function and Genetics* (Vol. 65, Issue 1, pp. 15–26). https://doi.org/10.1002/prot.21082

Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., & Case, D. A. (1998). Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *Journal of the American Chemical Society*, *120*(37), 9401–9409. https://doi.org/10.1021/ja981844+

Taylor, R. D., Jewsbury, P. J., & Essex, J. W. (2002). A review of protein-small molecule docking methods. In *Journal of Computer-Aided Molecular Design* (Vol. 16).

Toma, C., Gadaleta, D., Roncaglioni, A., Toropov, A., Toropova, A., Marzo, M., & Benfenati, E. (2019). QSAR Development for Plasma Protein Binding: Influence of the Ionization State. *Pharmaceutical Research*, *36*(2). https://doi.org/10.1007/s11095-018-2561-8

Tuccinardi, T. (2021). What is the current value of MM/PBSA and MM/GBSA methods in drug discovery? *Expert Opinion on Drug Discovery*, *16*(11), 1233–1237. https://doi.org/10.1080/17460441.2021.1942836

Valko, K., Nunhuck, S., Bevan, C., Abraham, M. H., & Reynolds, D. P. (2003). Fast Gradient HPLC Method to Determine Compounds Binding to Human Serum Albumin. Relationships with Octanol/Water and Immobilized Artificial Membrane Lipophilicity. *Journal of Pharmaceutical Sciences*, *92*(11), 2236–2248. https://doi.org/10.1002/jps.10494

Vallianatou, T., Lambrinidis, G., & Tsantili-Kakoulidou, A. (2013). In silico prediction of human serum albumin binding for drug leads. *Expert Opinion on Drug Discovery*, *8*(5), 583–595. https://doi.org/10.1517/17460441.2013.777424

van Gunsteren, W. F., & Berendsen, H. J. C. (1990). Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. In *Angewandte Chemie International Edition in English* (Vol. 29, Issue 9, pp. 992–1023). https://doi.org/10.1002/anie.199009921

Victor, R. (2007). Berendsen and Nose-Hoover thermostats Temperature in MD MD at constant Temperature - NVT ensemble. *Unknown*, 1–4.

Wang, X., Ramírez-Hinestrosa, S., Dobnikar, J., & Frenkel, D. (2020). The Lennard-Jones potential: When (not) to use it. *Physical Chemistry Chemical Physics*, *22*(19), 10624–10633. https://doi.org/10.1039/c9cp05445f

Wasko, J., Wolszczak, M., Kaminski, Z. J., Steblecka, M., & Kolesinska, B. (2020). Human serum albumin binds native insulin and aggregable insulin fragments and inhibits their aggregation. *Biomolecules*, *10*(10), 1–25. https://doi.org/10.3390/biom10101366

Weis, R., Berger, H., Kaiser, M., Brun, R., Saf, R., & Seebacher, W. (2008). Synthesis of bicyclic amines and their activities against Trypanosoma brucei rhodesiense and Plasmodium falciparum K1. *Archives of Pharmacal Research*, *31*(6), 688–697. https://doi.org/10.1007/s12272-001-1214-5

Weis, R., Kaiser, M., Brun, R., Saf, R., & Seebacher, W. (2008). Acyl derivatives of 5-amino-2-azabicyclo[3.2.2]nonanes. *Monatshefte Fur Chemie*, *139*(6), 717–724. https://doi.org/10.1007/s00706-007-0815-y

WHO. (2019). WHO interim guidelines for the treatment of gambiense human African trypanosomiasis. In *Who* (Issue August).

Wienken, C. J., Baaske, P., Rothbauer, U., Braun, D., & Duhr, S. (2010). Protein-binding assays in biological liquids using microscale thermophoresis. *Nature Communications*,

1(7). https://doi.org/10.1038/ncomms1093

William L. Jorgensen, & Duffy, E. M. (2002). Prediction of drug solubility from structure. *Allergologie*, *32*(12), 453–461. https://doi.org/10.5414/ALP32453

Wolkinger, V., Ahmad, S., Seebacher, W., Faist, J., Saf, R., Kaiser, M., Brun, R., & Weis, R. (2016). New diaryl-substituted azabicyclo[3.2.2]nonanes and their antiprotozoal potencies. *Monatshefte Fur Chemie*, *147*(10), 1721–1735. https://doi.org/10.1007/s00706-016-1800-0

Yamauchi, M., Mori, Y., & Okumura, H. (2019). Molecular simulations by generalized-ensemble algorithms in isothermal–isobaric ensemble. *Biophysical Reviews*, 457–469. https://doi.org/10.1007/s12551-019-00537-y

York, D. M., Darden, T. A., & Pedersen, L. G. (1993). The effect of long-range electrostatic interactions in simulations of macromolecular crystals: A comparison of the Ewald and truncated list methods. *The Journal of Chemical Physics*, *99*(10), 8345–8348. https://doi.org/10.1063/1.465608