INSIGHTS: ELUCIDATION OF SQUALENE MONOOXYGENASE INHIBITORS FOR LOWERING CHOLESTEROL IN CARDIOVASCULAR DISEASES.

A thesis submitted in fulfilment of the requirements for the degree

of

MASTER OF SCIENCE IN CHEMISTRY

at

RHODES UNIVERSITY, SOUTH AFRICA

Department of chemistry /Faculty of science

By

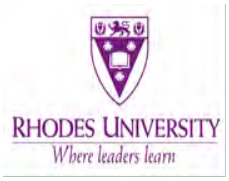
MOFELI BENEDICT LEOMA

2023

17L5037

Supervisor: Dr Tendamudzimu Tshiwawa

Co-supervisor: Prof. Kevin Lobb



“Wise words although written by my decaying hand remain imperishable through time, imbued with the medicine of immortality by the all master.”

Hermes Trismegistus.

To my mother **: Mamafohla Josephine Leoma**

O my divine Mother, may your soul continue to rest in peace. For you wished to preserve the memorial of seeing me as a master, I have journeyed over a long road for this thesis and armed with words of power. Thou divine soul who inspired me. Hail, thou who art adored. Homage to thee, mighty one of souls.

To my supervisors **: Prof K.A. Lobb and Dr T. Tshiwawa :**

Day after day, I begged for your wisdom, until at last I drank deep from your cup of wisdom. Thou worked over the heavens, being glad at heart at all times during my quest of education. Praises be unto thee.

ABSTRACT

Statins have been used to lower high cholesterol levels in the past few decades. However, several studies have shown that some people taking statins experience side effects over time, especially elderly patients, women of childbirth possibility, and children. Several studies have shown that the majority of people with underlying cardiovascular complications caused by high cholesterol are at a greater risk of fatality due to COVID-19, regardless of age and sex. The literature suggests that antimycotic squalene monooxygenase inhibitors, terbinafine and its derivatives, and anticholesterolemic squalene monooxygenase (SM) inhibitors could be another option and a safer remedy for lowering cholesterol in mammals. Molecular docking calculations, molecular dynamics (MD) simulations, molecular mechanics generalized born surface area (MM-GBSA) calculations, quantum mechanics/molecular mechanics calculations (QM/MM), and density functional theory (DFT) calculations were used in this study. An early stage in drug discovery, in which small molecular hits from high- throughput screening (HTS) are evaluated and undergo limited optimization to identify promising lead compounds, is referred to as lead generation. To address the first step of lead generation, the number of compounds to be tested was narrowed down, and the hit compounds that could be taken for further tests were obtained. Thus, the molecular docking technique was taken advantage of, which assisted us in identifying the antimycotic ligand SDZ 18, which had a good binding affinity of about -8,4 kcal mol^{-1}. Another widely employed strategy, the molecular mechanics-generalized born surface area (MM-GBSA), was used to investigate the binding free energies of the protein-ligand complexes to validate the binding affinities obtained from molecular docking. Despite the excellent docking results, it must be emphasized that the stability of the ligand in the binding pocket must be investigated. To address this, the protein-ligand complexes were then taken through molecular dynamics for 100 ns simulations calculations which showed that the inhibitors stayed in the binding pocket with the RMSD values below 3.5 Å for most systems. This provided insight into a realistic model because the docked complexes were placed in conditions closer to the physiological environment at 300 K and 1.01325 bar, and in an explicitly solvated dynamic environment. Density functional theory (DFT) at the B3LPY level of theory using the standard 6-31G(d,p) basis set was used to assess the reactivity and other properties of the SM inhibitors. ONIOM calculations were performed to explain what was happening at the microscopic level by calculating the total energy of the complex. The aim of this project was to efficiently uncover the non-physical aspects of SM inhibitors with the help of computational techniques to identify new drugs that can lower high cholesterol levels. From a theoretical perspective, the results obtained from docking indicated that the antimycotic ligands SDZ SBA 586 18 and TNSA 84 (trisnor-squalene alcohol) have good binding affinities, and the MM-GBSA method provided free energy calculations. MD results indicated that the stability of the ligand in the binding pocket was achieved during the 100 ns simulations. The HOMO-LUMO energy gaps obtained from DFT calculations provided information on the reactivity of the ligands. Other insights into the protein-ligand complexes were obtained from a hybrid ONIOM QM/MM study.

ABBREVIATIONS

SM Squalene monooxygenase

SDZ SBA 586 SDZ

TNSA Trisnor-squalene alcohol

HA Heavy atoms

LE Ligand efficiency

LLE Ligand Lipophilicity efficiency

LELP Ligand efficiency dependent lipophilicity index

MD Molecular dynamics simulations

CSBDD computational structure-based drug discovery

MEP Molecular electrostatic potentials

MM-GBSA Molecular mechanics-generalized born surface area

MM-PBSA Molecular mechanics Poisson-Boltzmann and generalized born surface area

QM/MM Quantum Mechanics and Molecular Mechanics

ADMET Adsorption, Distribution, Metabolism, Excretion and Toxicity

HTVS High-throughput screening

GPU The graphics processing unit

TIP3P Three-site transferrable intermolecular potential

RB Rotatable bond

AFE Alchemical free energy

FEP Free energy perturbation

TI Thermodynamic integration

BFE Binding free energy

RMSD Root mean square deviation

RMSF Root mean square fluctuation

L-RMSF Ligand Root mean square fluctuation

LJ Lennard Jones potential

PBC Periodic boundary conditions

QFT Quantum field theory

BO Born Oppenheimer approximation

DFT Density functional theory

SCF Self consistent field

LCAO Linear combination of atomic orbitals

HOMO Highest occupied molecular orbital

LUMO Lowest unoccupied molecular orbital

EA Electron affinity

MEP Molecular electrostatic potential

IRC Intrinsic Reaction Coordinate

DEDICATION

I fully dedicate this thesis to my late mother **Mamafohla Josephine Leoma**, whom I lost in the middle of my project. Being fatigued and overwhelmed by the multitude of my thoughts for many months, I was cast back on the project by the will of God.

ACKNOWLEDGEMENTS

First and foremost, I praise and honor God, the creator of all the elements of nature. I equipped myself with prayer throughout my project and he showed me the full interpretation of compounds that can perfect an imperfect body.

I would also like to thank my supervisor Dr Tendamudzimu Tshiwawa and co-supervisor Prof Kevin Lobb, who blessed me with the knowledge of atom and showed me the laws of nature that govern the universe, which in turn are the laws of God that establish health and harmony when applied.

The author would also like to thank Rhodes university and Henderson Scholarship for undertaking the massive expenses in taking care of my financial needs throughout the project. I am also grateful to the Centre of High-Performance Computing in Cape Town (CHPC) for the high computational calculations.

The author acknowledges gratefully his indebtedness to his main colleague Kabelo Phuti Mokgopa through whose valuable assistance and encouragement made it possible for me to reach the hilltop of superior achievement. I also honorably salute my research team CMCDD and Mr Bienfait Isamura.

Special recognition and gratitude also go to my whole family, for being there for me always in everything and inducing me to persevere to the end, to the beauty and completeness of the work.

RESEARCH OUTPUTS

List of publications

1. Abstract from thesis published in the 56th Annual (2023) SASBCP Conference of the South African Society for Basic and Clinical Pharmacology booklet.

List upcoming publications from thesis.

2. **Mofeli Benedict Leoma**, K. Lobb, T. Tshiwawa. Pharmacophore modelling, 3D-QSAR and ADMET prediction for drug design in cardiovascular biology (2023). Royal society of chemistry (RSC).
3. **Mofeli Benedict Leoma**, K Lobb, T. Tshiwawa, Kabelo Mokgopa. New insights into the squalene monooxygenase inhibitors for lowering cholesterol in cardiovascular biology with the aid of computational methods (2023). Royal society of chemistry (RSC).

TABLE OF CONTENTS

ABSTRACT **4**

ABBREVIATIONS **5**

DEDICATION **7**

ACKNOWLEDGEMENTS **8**

RESEARCH OUTPUTS **9**

LIST OF PUBLICATIONS **9**

Chapter 1 : INTRODUCTION AND LITERATURE REVIEW **15**

Introduction to the chapter **15**

1.1 Cardiovascular disease **15**

1.2 Computational approaches utilized in this study **24**

(a) Molecular docking **24**

(b) Molecular mechanics-generalized born surface area (MM-GBSA) and Molecular mechanics Poisson-Boltzmann and generalized born surface area (MM-PB/SA) **28**

(c) Molecular Dynamics (MD) simulations **31**

(d) Quantum Mechanics/Molecular Mechanics **34**

(e) QM with Density functional theory (DFT). **38**

1.3 Summary of the chapter **42**

Chapter 2 : COMPUTATIONAL BACKGROUND AND METHODOLOGY **43**

Introduction to the chapter **43**

2.1 Molecular docking **43**

2.2 Molecular mechanics-generalized born surface area (MM-GBSA) **47**

2.3 Molecular Dynamics (MD) simulations **47**

2.4 QM with Density functional theory (DFT) 49

2.5 Combined Quantum mechanics and molecular mechanics (ONIOM QM/MM)..... 50

Chapter 3 : MOLECULAR DOCKING..... 51

Introduction to the chapter..... 51

3.1 Protein PDB crystal structures used for the study..... 51

3.2 The Ramachandran Plots..... 52

3.3 Molecular docking results..... 54

3.4 Molecular mechanics-generalized born surface area (MM-GBSA)..... 70

3.5 Summary of the chapter..... 72

Chapter 4 : MOLECULAR DYNAMICS (MD) SIMULATIONS..... 74

Introduction to the chapter 74

4.1 The Protein and Ligand RMSD results..... 74

4.2 P-RMSF calculations 77

4.3 The Ligand Root Mean Square Fluctuation (L-RMSF)..... 81

4.4 Ligand Torsion Profile 83

4.5 Protein-Ligand Contacts and contacts timeline..... 85

4.6 Summary of the chapter..... 91

Chapter 5 : QM WITH DENSITY FUNCTIONAL THEORY (DFT)..... 92

Introduction to chapter..... 92

5.1 Geometry optimization 92

5.2 Energy profiles of the frontier molecular orbitals 93

5.3 Visualized frontier molecular orbitals..... 95

5.4 Molecular electrostatic potentials (MEP) 98

..... 99

Chapter 6 : HYBRID QUANTUM MECHANICS / MOLECULAR MECHANICS (ONIOM QM/MM)..... 100

Introduction to the chapter 100

6.1 The ONIOM models with respect to B3LYP/6-31G:UFF level..... 100

6.2 Total intrinsic Reaction Coordinate (IRC) .....

6.2 Total intrinsic Reaction Coordinate (IRC) 102

6.3 RMS Gradient norm along intrinsic Reaction Coordinate (IRC) 103

6.4 Summary of chapter..... 104

Chapter 7 : Conclusion and future work..... 105

References..... 108

..... 133

1. Pharmacophore modelling results 138

2. 3D-QSAR results 149

3. ADME results 154

LIST OF FIGURES

Figure 1.1 The general structure of cholesterol. 17

Figure 1.2 The diagram illustrating the cholesterol biosynthesis and the enzymes involved. The enzymes are in blue, the molecules in black and the proteins in green. adapted from [4], [16], [21], [9], [16], [26]..... 18

Figure 1.3 The 2D and 3D figure of O-methly GCG showing the position, poses and interaction with the protein. Used with permission from Kabelo Mokgopa. 25

Figure 2.1 The KNIME workflow. 44

Figure 3.1 Protein PDB crystal structures. 52
Figure 3.2 The Ramachandran plots with phi and psi degrees for proteins. 54
Figure 3.3 Bar chart of binding affinities in kcal mol − 1for the ligands. 56
Figure 3.4 The 2D representation of protein interaction for (a) SDZ18 and 2ZCS and (b) The 2D representation of TNSA84 and 3COX. 58
Figure 3.5 The 2D representation of protein interaction for (a) SDZ18 and 6GQF and (b) SDZ18 and 3D4S. 58
Figure 3.6 The 2D representation of protein interaction for (a) SDZ18 and 3W7F and (b) SDZ18 and 1M17 59
Figure 3.7 The correlation heatmap of the proteins based on the binding affinities. 60
Figure 3.8 Protein-3D docked poses for (a) 3COX protein and TNSA84 (b) 2ZCS protein and SDZ18. 61
Figure 3.9 The ligand efficiency plots for the proteins. 68
Figure 3.10 The illustration of the native ligand and docked ligand in the active site of protein. The green ligand is the docked ligand and the yellow ligand is the native ligand of protein. (a) 3COX protein and TNSA84 (b) 2ZCS protein and SDZ18. 69
Figure 4.1 The RMSD plots of the proteins for 100 ns simulations. 76
Figure 4.2 The RMSD plots of the proteins for 100 ns simulations. 77
Figure 4.3 The RMSF plot for a simulation of 100 ns for the proteins. 80
Figure 4.4 The RMSF plot for a simulation of 100 ns for the proteins. 82
Figure 4.5 Ligand Torsion Profile for 100 ns simulation for the proteins. 84
Figure 4.6 The Protein-ligand contacts and contacts timeline for (a) 2ZCS protein SDZ18 and (b) 1M17 protein and SDZ18. 88
Figure 4.7 The Protein-ligand contacts and contacts timeline for (a) 6GQF protein SDZ18 and (b) 3D4S protein and SDZ18. 89
Figure 4.8 The Protein-ligand contacts and contacts timeline for (a) 3COX protein TNSA84 and (b) 3COX protein and SDZ18. 90
Figure 5.1 The optimized geometrical structures of the Ligands. 93
Figure 5.2 The HOMO and LUMO energies of the ligands. 97
Figure 6.1The intrinsic reaction coordinate (IRC) for SDZ 18 and 2ZCS. 103

Figure 6.2 The RMS gradient norm (RMS) along IRC for SDZ18 and 2ZCS. 104

Chapter 1 : INTRODUCTION AND LITERATURE REVIEW

Introduction to the chapter

This study entailed a search for new drugs with the potential to lower high cholesterol. Inspired by recent studies, computational chemistry approaches were utilized to explore the properties of squalene monooxygenase inhibitors (SM) as they have the potential to lower high cholesterol in the body [1], [2]. Despite much interest in SM inhibitors, new insights have not yet been established, specifically with finding potential drug candidates. This chapter provides an overview of the theoretical foundation which is applicable to the study.

1.1 Cardiovascular disease

On 11 March 2020, the World Health Organization (WHO) declared a pandemic caused by the corona virus disease (COVID-19). High mortality rate due to COVID-19 was attributed to the presence of underlying diseases such as cardiovascular disease (CVD) [3], [4], [5]. It is important to point out that many patients with COVID-19 have underlying CVD or develop acute cardiac injury during the course of the illness [6]. According to the WHO, cardiovascular diseases (CVDs) are also the leading causes of death globally with an annual estimation of 17.9 million lives. About 102,595 people have died so far in South Africa. According to the research work of Jennifer G. Robinson, careful examination reveals that cardiovascular disease is thus far the leading cause of death and disability in person with age greater than 70 years in the United states and other Westernized countries [7]. This looming disaster has lead us to the quest of finding alternative drugs that can be used to lower high cholesterol in the body.

In the human body, there are many arteries that are responsible for the physiological functions such as transportation of oxygen to the heart. [2], [8]–[10]. The build-up of high cholesterol in this arteries near the heart increases the risk of CVD such as coronary heart disease and peripheral arterial disease as it blocks the flow of blood [11]. When the concentration of low density lipoprotein (LDL) is high and the high density lipoprotein (HDL) is low, cholesterol in the body is dangerous [12], [13], [14].

The most efficient treatment against atherosclerosis is the treatment with statins, which has shown to decrease the risk both of stroke and coronary artery disease in all age groups [15], [2], [7], [8], [16], [17]. Several studies have proved that regardless of their efficiency, stains are associated with a lot of side effects. In a study of Peter H. Langsjoen et al, fifty consecutive new cardiology clinic patients who were on statin therapy (for an average of 28 months) on their initial visit were evaluated for possible adverse statin side effects (myalgia, fatigue, dyspnea, memory loss and peripheral neuropathy) [18]. In another study by Jennifer G. Robinson et al, there is evidence that the use of statins with moderate dose is of little benefit to a population aged 70 to 82 years when given for 3 years and some evidence also suggests that the elderly experience more mortality with the treatment [7]. The use of statins is poorly tolerated by some patients and myopathy is a well known side effect [9], [8], [15], [16].

Despite these side effects of statins, the alternative drugs that have been developed still show lower effectiveness against CVD. The development of new drugs that can balance critically cholesterol across the body is a matter of great health and commercial urgency. Through an economic lens, perhaps it might also be easier to see the importance of the development of new alternative drugs. According to Yahoo finance records, the global market in pharmaceutical industry for statins is estimated at $13.6 Billion in the year 2020 and is projected to reach $15.2 Billion by 2027.

(a) Cholesterol

In order to understand the causes of cardiovascular disease and establish remedies to combat this disease, the origin of cholesterol must first be understood. In 1784, cholesterol (figure 1.1) was first isolated from gallstones by the French physician- chemist François Poulletier and later the cholesterol structure discovered by Heinrich. O Wieland and Adolf Windaus [8], [19], [8]. Cholesterol is a major sterol present in the animal tissues and in mammals it plays a vital role in life as it is necessary for the cells to be able to function [12],[2], [9], [20], [21]. It is essential in cell membrane synthesis and cell growth, a vital lipid in animals but can be toxic in excess [22], [23]. The cells of the body acquire cholesterol by de novo synthesis or uptake of low density lipoprotein particles that have esterified cholesterol [24], [23], [25]. The food products such as processed eggs, meat and their derivatives that are exposed to high temperatures and oxygen during manufacturing or heat processing are the main causes of cholesterol oxidation [11]. Processed food

products cause cholesterol oxidation because cholesterol is a reactive molecule that undergo oxidation reactions and thus resulting in oxysterols that present atherogenic, cytotoxic, mutagenic and carcinogenic effects [11]. Figure 1.1, shows the cholesterol. It is a polycyclic monounsaturated alcohol with a molecular formula $C_{27}H_{46}O$.

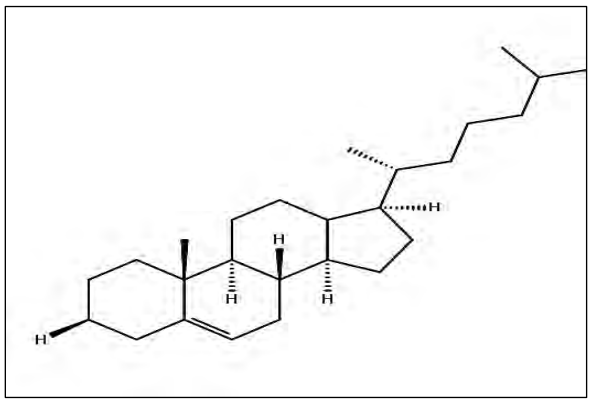


Figure 1.1 The general structure of cholesterol.

(b) Cholesterol biosynthesis

Acetyl-CoA is the main substrate for the biosynthesis of cholesterol [11], [12].

Figure 1.2 shows the bioorganic transformation of Acetyl-CoA to cholesterol. Cholesterol biosynthesis begins with the formation of mevalonate from acetate, followed by the condensation of two acetyl-coenzyme A (Acety-CoA) molecules form acetoacetyl-CoA [12]. HMG-CoA synthase catalyzes the reaction between the two molecules, Acetyl-CoA and Acetoacetyl-CoA to form HMG-CoA [12], [22]. The HMG-CoA enzyme makes a molecule that is used to make cholesterol and other similar sterol molecules [26]. It is the first step of the synthesis of mevalonate. The 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR) is the rate-limiting enzyme of mevalonate pathway [27], [8], [23], [28].

Commercially available statin drugs work by blocking the conversion HMG-CoA reductase to mevalonate acid [27], [22], [23]. As a result of the interruption of cholesterol synthesis in the liver, the activation of microsomal HMGR and cell surface LDL receptor occurs [27]. Therefore, the cholesterol levels decrease by the range of 20% to 50% [27]. The inhibition of HMGR also inhibits other processes and formation of other sterols which are essential for the bodies, this results in a

lot of side effects discussed in section 1.1 of this chapter. The defects in the cholesterol synthesis are lethal in mice, thus this shows the importance of cholesterol synthesis for survival [29].

Mevalonate pathway is an important pathway which plays a key role in multiple cellular processes, and it provides isoprenoid building blocks such as cholesterol, steroids, bile acids, vitamin D, dolichols, Haem A, ubiquinone and isopentenyl adenine [21], [23]. The use of statins block the synthetic pathway for the formation of all these essential metabolites [14], [17], [24]. In light of this predicament, our search project focused on the search for new drugs targeting the squalene pathway as the starting point beyond the mevalonate pathway.

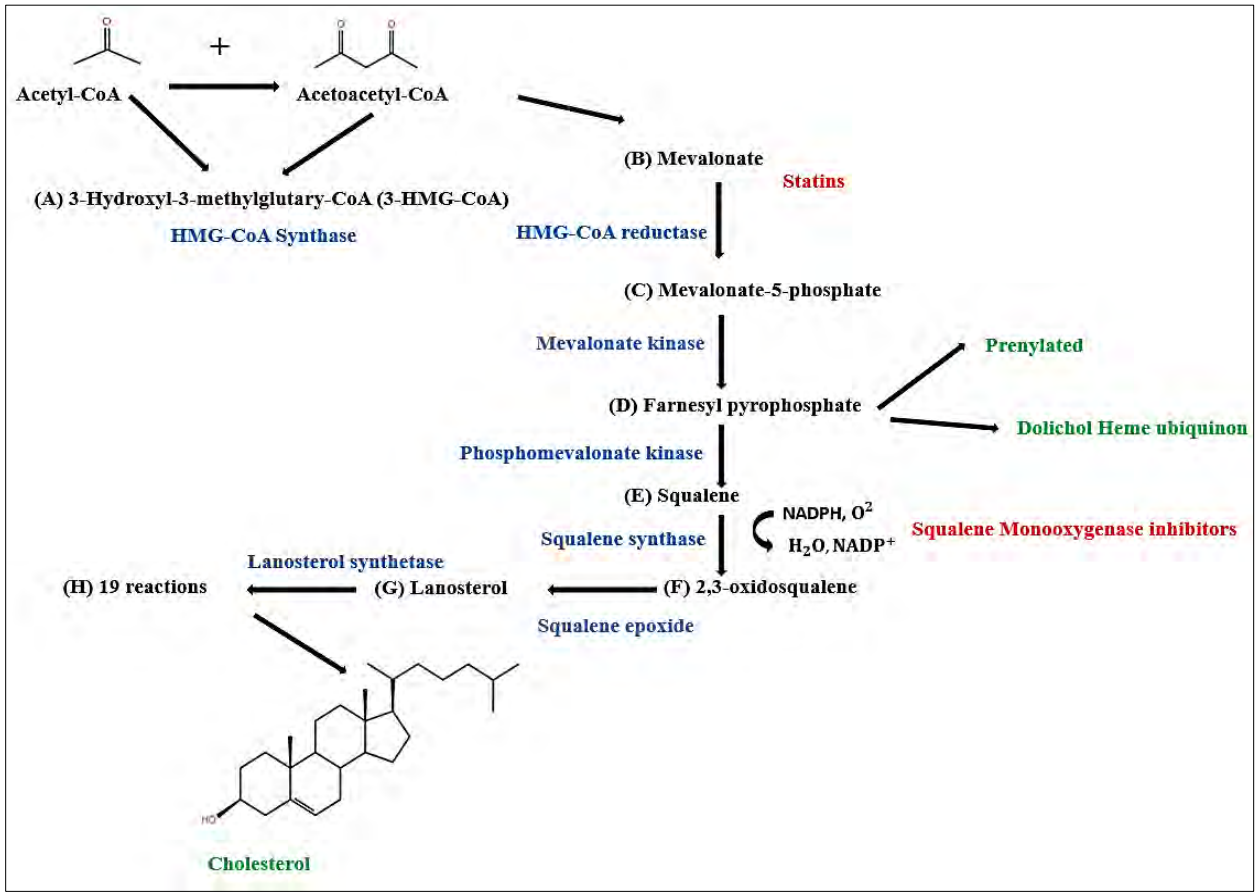


Figure 1.2 The diagram illustrating the cholesterol biosynthesis and the enzymes involved. The enzymes are in blue, the molecules in black and the proteins in green. adapted from [4], [16], [21], [9], [16], [26].

(c) Squalene monooxygenase (SM) inhibitors

In 1969 SM, the enzyme of cholesterol synthesis was first detected in the livers of a rat [19], [30]. It is interesting that it is a therapeutic target for treating certain fungal infections and in mammals it is a rate limiting enzyme that represents a key control point in the cholesterol synthesis pathway [9], [31], [32]. Squalene potentiates the cholesterol lowering effects of statins and it has an excellent safety profile. The injected squalene of 21 mg into over million cases including infants has shown a very low incidence rate of side effects, usually just pain at the injection side [19]. The inhibition of SM due to squalene accumulation at high concentrations is non toxic for cells even at high concentrations [19]. Indeed, among the enzymes from cholesterol synthetic pathway, SM seems to be the best candidate for lowering cholesterol.

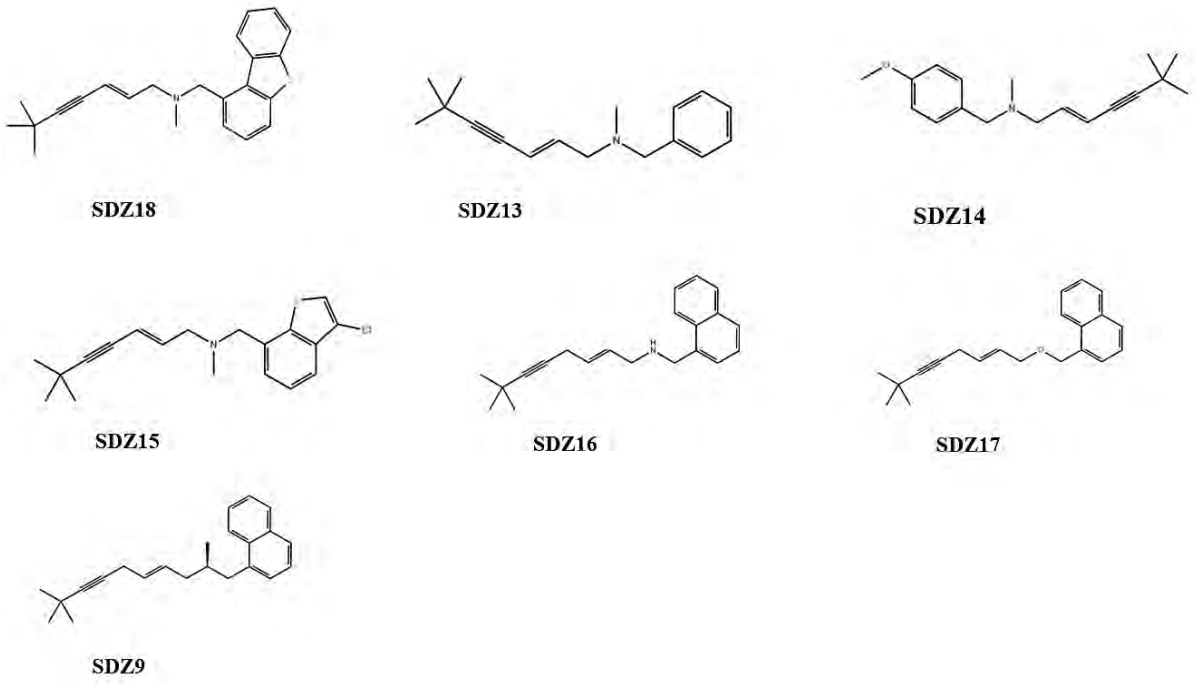
Thus, it is at this juncture that we may give a detailed description of SM inhibitors. SM catalyzes the epoxidation of C=C double bond of squalene to yield 2,3-oxidosqualene. It catalyzes the first oxygenation step in cholesterol synthesis, the conversion of squalene to nonsterol precursor for lanosterol, 2,3-oxidosqualene [22],[25], [32]–[34]. It has been reported that in dogs, some signs of dermatitis-like toxicity are presented as a side effect associated with the intracellular accumulation of squalene in skin cells. Another observation of side effects associated with the intracellular accumulation of squalene is neuropathy weanling in rats [2], [18], [35], [34], The computational investigation of SM inhibitors is significant as it paves the road in the pipeline of drug discovery.

(d) Molecules that are object of the study

The process of drug repurposing is a method that is used in medicinal chemistry to discover new therapeutic use of drugs which have already been approved for clinical use [36], [37] [38]. This section introduces the SM inhibitors which attracted our attention. They were obtained from the research paper by A. Belter and others [22], who proposed the repurposing of antimycotic compounds, anticholesterolemic, and anticholesterolemic from garlic extracts as SM inhibitors. The compounds are discussed in subsections below.

(i) Antimycotic compounds

Naftifine compound is the first antifungal agent that inhibits the epoxidation of squalene, subsequently blocking the biosynthesis of ergosterol [19]. Presently, it is used in the treatment of infections caused by the pathogenic fungi known as Trichophyton rubrum and C. albicans. Terbinafine (2), butenafine (3), SDZ SBA 586 (4) were obtained by substitution of cinnamyl group in naftifine with 6,6-dimethyl-2-hepten-4-ynyl, 4-t-butylbenzyl and a [4-1-methly-1-phentylethyl)phenlyl]methyl [19].

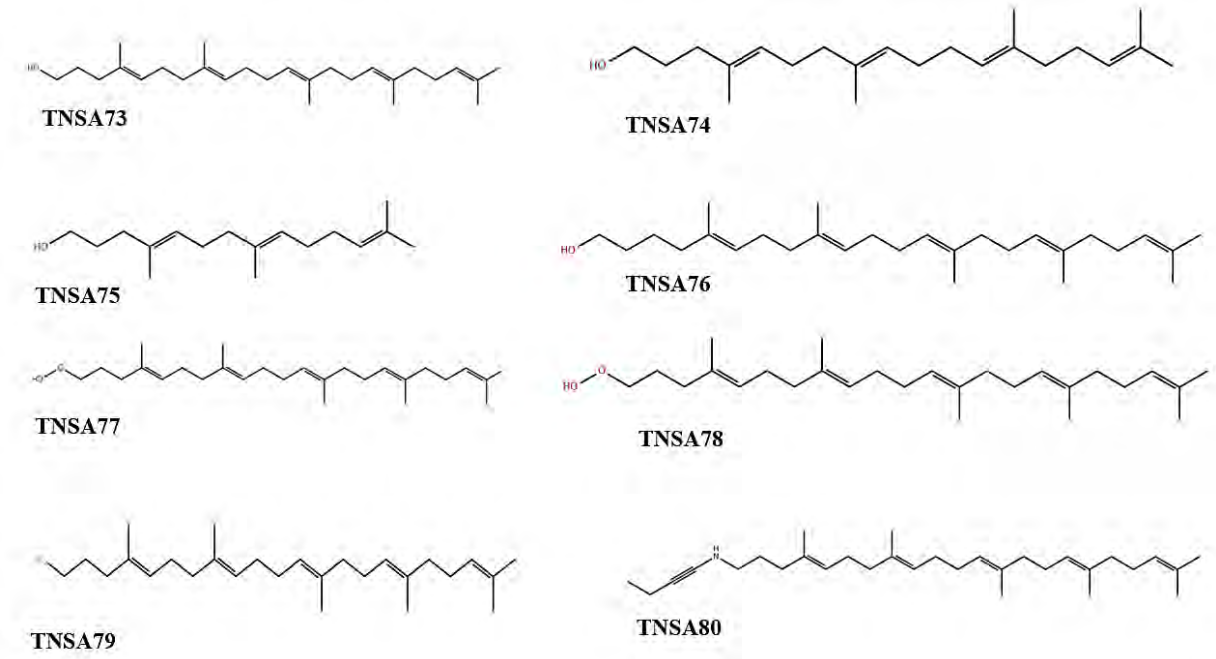


Structures 1 Antimycotic compounds from Naftifine.

(ii) Anticholesterolemic squalene monooxygenase inhibitors

The new potent competitive inhibitors of SM were obtained using squalene analogues as a starting point. The inhibitory effect was achieved by the trisnorsqualenoid moiety after adding and removing isoprene units and removal of its terminal isopropylidene group resulted in decrease in inhibitory potency [19]. The terminal isoprene unit is important in substrate binding to an enzyme.

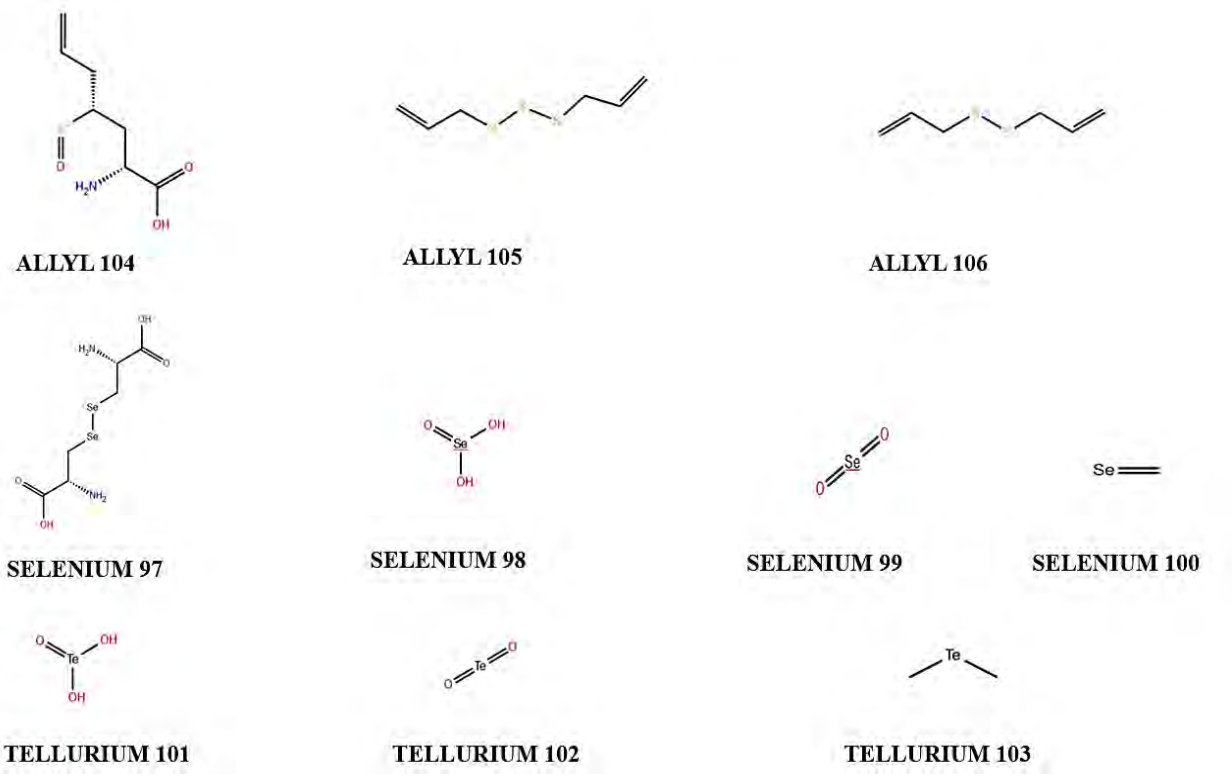
Trisnorsqualene alcohol (TNSA) was the first effective inhibitor of vertebrate squalene monooxygenase [19].



Structures 2 Anticholesterolemic compounds from Trisnorsqualene alcohol.

(iii) Anticholesterolemic squalene monooxygenase inhibitors from garlic extract

Apart from chemically synthesized compounds, there are natural inhibitors which notably lowers the cholesterol level in plasma. As such, it has been found that garlic compounds also inhibit the squalene monooxygenase (SM) activity. The garlic compounds have three groups which can be divided into selenium compounds, tellurium compounds and allyl compound [19], [35]. Tellurium compounds inhibits SM by the reaction with sulfhydryl groups on the enzyme [39]. Tellurium disrupts the cholesterol synthesis in the Schwann cells, through the inhibition of the SM [35]. However, dimethyl tellurium has toxic effects in rats, it causes neuropathy [19], [35]. Selenium is an essential micronutrient which inhibits SM although in lower concentrations. In the same manner, selenium also induces neuropathies in the animals when consumed in large amount with toxic effects occurring at levels just 10-fold greater than the recommended daily allowance of 70µg [35].



Structures 3 Anticholesterolemic compounds from garlic.

This study was a continuation of project that begun in 2021. Our previous study demonstrated with ADMET, Pharmacophore modelling and 3D-QSAR prediction that these SM inhibitors exhibit the acceptable drug-like properties, the compounds followed the Lipinski's rule of five and the results are shown in the appendix. In modern drug discovery, ADMET properties are always carried out to find a safer and better compound that will eventually enter the clinical trials if other tests are passed and reducing the chances of risk during the clinical trials [40], [41].

(e) Importance of computational chemistry in the study

In recent years computational chemistry has emerged as a remarkable tool that helps modern chemists to study matter. Computational chemistry uses mathematical equations that originates from physics to model chemistry [42]. Thus, many atomic structures, bond energies and other chemical properties are well understood due to these equations from classical physics and quantum

mechanics. The equations obtained from classical physics are built primarily from Newtonian laws. Quantum chemistry focuses mainly on the energy of particles in order to explain and rationalize subatomic particles [43]–[45] .

The time-independent Schrödinger wave equation (equation 1) is the theoretical foundation of quantum computational chemistry and quantum physics, it helps in measuring the total energy in matter, revealing many mysteries to us [46].

Equation 1: Schrödinger wave equation.

$$\hat{H}\Psi = E\Psi$$

Where Ψ is the wavefunction, which is the function of positions of particles in a system, Ĥ is the Hamiltonian operator, which is associated with total kinetic and potential energy, it is responsible for an operation on the wave function Ψ and E is the total energy of the system [43], [45] .

Nowadays, modern chemists are extensively conducting the in-depth calculations of properties of atoms and molecules. Many molecular properties can be computed computationally, namely, bond energies, reaction energies, structure of ground states, vibrational frequencies (IR and Raman), atomic charges, electrostatic potentials, NMR chemical shifts, dipole moments, polarizabilities, hyperpolarizabilities, transition energies and intensities for UV and IR spectra and reaction pathways and mechanisms [43], [45].

The limitations of experimental work have been achieved as computational chemistry helps chemists to investigate the chemical reactions and molecular properties using calculations rather than using chemicals which may be expensive and hazardous to the environment. The use of computers to study chemical properties is impressively safer and less costly financially and in terms of time as compared to experimental investigations.

It is true that the costs of synthesis of many compounds have been reduced as the calculations in computational chemistry assists us to predict the compounds that have the desired properties and finding the hit compounds from a large library of compounds, however, the advancement of

computational chemistry does not eliminate the need for experimental validation of the results obtained computationally.

1.2 Computational approaches utilized in this study

This project focused on finding squalene monooxygenase inhibitors with the potential of solving the problem of high cholesterol through computational approaches. Six computational techniques that have proved in literature to predict accurately the inhibition properties of molecules were considered, namely, molecular docking, molecular mechanics born generalized surface area (MM-GBSA), molecular dynamics simulations (MD), Density functional theory (DFT) and Quantum mechanics and molecular mechanics (QM/MM) :

(a) Molecular docking

Drug discovery has always been the biggest challenge facing the scientific community and one of the techniques which is a game changer and has solved the obstacles is molecular docking[47], [48],[47]. In recent years, molecular docking has gained a lot of awareness in modern drug design for the investigation of protein-ligand interaction at the molecular level [49]–[51]. Molecular docking is an *in silico* method for recognizing the correct binding pose of a protein-ligand complex and evaluates the strength of the protein-ligand interaction using various scoring functions for selecting the best pose generated by each molecule to a rank order [47], [52]. Molecular docking is composed of three objectives, namely, virtual screening, pose prediction and binding affinity estimation.

During molecular docking, the mechanistic models explain the interaction between a small molecule and protein at the atomic level. This aids us to characterize the behavior of small molecules in the binding pocket of the target protein and moreover to clarify the fundamental biochemical processes [52]. In biomolecular system, the first step in protein-ligand binding process is diffusion followed by vibrations and collisions between the protein and ligand molecules in the solvent [52], [53]. The collisions between the two molecules are essential for further interactions to occur, this causes the kinetic energy to be released from water molecules which causes Brownian motions of individuals water molecules [54].

Furthermore, the motion of can lead to entropy maximation of water solvent. The driving forces of protein-ligand binding are vibratory energies such entropy, free energy, and enthalpy. The higher the concentrations, the higher the probability of protein-ligand contact [52], [53]. Figure 1.3 shows an example of the representation of molecular docking 2D interaction and the 3D interaction of the ligand and protein.

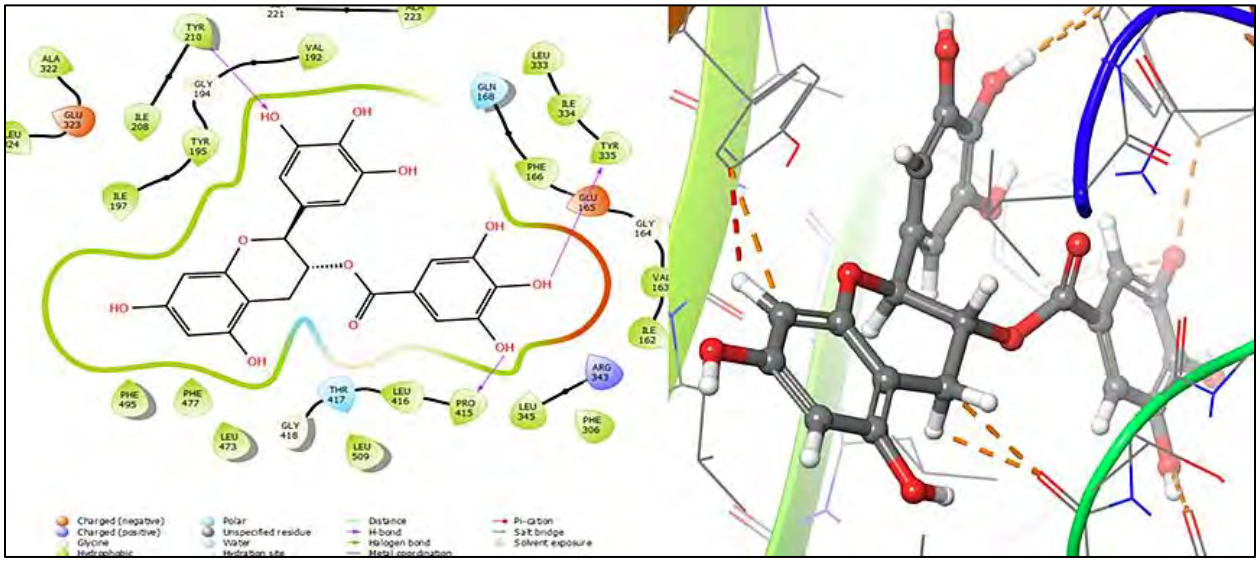


Figure 1.3 The 2D and 3D figure of O-methly GCG showing the position, poses and interaction with the protein. Used with permission from Kabelo Mokgopa.

(i) The molecular docking processes

The proteins are the essential components of the living organisms involved in the biological processes and they act as enzymes, signalling components, structural and transport or storage constituents [55]. They bind to other molecules with great affinity and specificity of the binding site with the aid of the shape of their binding pocket and physio-chemical properties of the amino acids' residues forming the binding cavity [56]. This binding is essential for many biological functions in the body. In this research project, the technique of molecular docking was employed to analyze at the atomic level the interactions between the proposed inhibitors in section 1.5 and SM.

In blind docking where the binding site of the protein is not known there are programs or online servers that help to identify the active site with proteins [47], [57]**.** They are namely GRID [20,21], POCKET SURFER, AFT [27], Surfnet [23,24], PASS [25], MMC [26] and PATCH-SURFER [47], [52]. In this study, the GRID [2021] was used. Discovering the binding site and calculation of binding affinity between the ligand and protein plays a very important role in the understanding the mechanisms and therefore designing therapeutic structure-based drugs [58] [21].

Some models explain the protein-ligand binding mechanism with the lock-and-key model proposed by Fischer and induced-fit models created by Koshland. It is insisted that the protein is the lock, and the ligand is the key [52], [57]. They are used to idealize binding process at a single molecule level which simplify the binding of ligand to rigid and flexible proteins, however not at a population level. The binding-driving force of the lock-and-key is the difference between the product of entropy change and the absolute temperature and enthalpy change which is known as the Gibbs free energy [50]. Molecular docking prediction of ligand-receptor complex is done through sampling the conformations of ligand in active site of protein then ranking them via a scoring function [52], [58], [59].

(ii) The scoring functions

Scoring functions are the algorithms designed to compute the binding affinity of a protein-ligand complex and its purpose is to delineate the correct poses from incorrect poses, or binders from inactive compounds in a reasonable computation time [52], [57]. Nevertheless, the scoring functions estimate rather than calculate the binding affinity between the protein and ligand, and through these functions accepting assumptions and simplifications [52]. The force fields used in scoring functions range from molecular mechanics force fields such as AMBER, OPLS or CHARMM, [25] [57],[60], [61]. In this work, we have employed OPLS force field. The force fields play a very important role in computational chemistry as they help in determining the interactions in a molecular simulations system. The scoring functions are divided into three types namely, classical force field based, empirical based and knowledge-based scoring functions [52], [57].

The classical force-field based scoring function is used to assess the binding energy by running calculations for the sum of the non-bonded interactions (electrostatics and van der Waals) [52], [57]. It considers the hydrogen bonds, solvation, and entropy contributions. Coulombic function helps in calculating the electrostatic terms [52]. However, there are problems in the modelling of the protein's real environment due to point charge calculations, a distant dependent dielectric function helps in modulation of the charge-charge interactions contribution. The other problem that is encountered is how solvents are treated during the ligand binding [57]. Another challenge faced by classical force field is that it does not predict accurately the entropic effect which makes the de-solvation effect very difficult [57]. The major problem is the slow computational speed, the cut-off distance that is used to handle the non-bonded interaction although this decreases the accuracy of the long range effects involved in the binding [57].

The empirical scoring function break down the binding energy into several energy components such as binding energy, hydrogen bond, ionic interaction, entropy, electrostatic, de-solvation and hydrophobic effect [52], [57]. The binding energy obtained by empirical scoring function is evaluated by the addition of a set of weighted empirical terms [57]. The knowledge-based function obtains the interatomic contact frequencies and distances between the ligand and protein by applying the statistical analysis of the ligand-protein system [52], [57]. They are obtained from experimental methods by determination of structural formation of protein-ligand complexes [57]. The final score is obtained by the great frequency of occurrence in the interaction and rejection of the repulsive interaction [52]. The major problem is the interaction that occur which are underrepresented in the limited training set of the crystal structure and the inherent in the selection of the proteins for successful structure determination [52]. In the current work, the empirical scoring function was used.

The assessed scoring functions take into account the contributions from van der Waals contacts, electrostatic interactions, de-solvation effects and entropy [52], [57], [62]. During molecular docking technique, the main focus is on the calculation of the interactions of the ligand and protein, we analyze the position of the ligand in the protein binding site. Molecular docking can predict

and differentiate interactions such as electrostatic forces due to the charges existing in the system, electrodynamic forces such as the van der Waals interactions, steric forces caused by the entropy and solvent-related forces such as hydrogen bonding and hydrophobic interactions due to the structural changes of the solvent [56].

(iii) Limitations of molecular docking

During molecular docking, the features of the ligand such as conformational preferences, protein flexibility and the important role played by explicit water molecules in the binding sites are generally ignored. The binding process is driven by the solvent entropy gained and system enthalpy loss which occurs deep inside the physical matter. The scoring functions are usually optimized for speed and efficiency by the use of empirical functions which lack clear theoretical basis [63]. Although molecular docking is computationally efficient and low-cost, it is important to remark, that the prediction of binding poses measured are not of high accuracy and it has been difficult to reliably differentiate the ligands with comparable binding affinities [64]

Thus, in drug-discover other techniques are employed to remedy the shortcomings of molecular docking. Further examination of the docked poses will do as the highest scoring poses cannot be undoubtedly trusted due to the need to consider factors such as the flexibility, explicit water and entropy [62]. Concertedly, some of the problems faced by the typical scoring function due to limited treatment of solvation effect can be solved by the physics-based scoring methods such as Molecular mechanics-generalized born surface area (MM-GBSA) and Molecular mechanics Poisson-Boltzmann and generalized born surface area (MM-PB/SA [52].

(b) Molecular mechanics-generalized born surface area (MM-GBSA) and Molecular mechanics Poisson-Boltzmann and generalized born surface area (MM-PB/SA)

The MMGBSA method is used to calculate the binding free energy for the protein-ligand complexes, thus restoring the final docking poses by improving accuracy of virtual screening and evaluating false positive hits [65], [66]. The free energy (ΔG) defines the binding energy between the two interacting molecules in the cell which initiates several biological processes [50], [56],

[67]. The kinetic and thermodynamic properties of the system are described by change in free energy [50], [59], [68] [10]. It may be argued that the MMGBSA and MMPBSA are popular methods used for calculation of free energy prediction since they are more accurate than most scoring functions of molecular docking and less computationally costly than the alchemical free energy methods [64]. At present, the calculations of the absolute binding free energies remain a challenge for the Alchemical free energy (AFE) methods [64].

The computational techniques for calculating of the binding free energy can be divided into pathway and end-point methods. In the AFE methods which are also known as pathway methods, there are free energy perturbation (FEP) and thermodynamic integration (TI) [64]. These are computational techniques used in the alchemical free energy computations. The main problem is that they are of high computation cost and slow in the convergence of structural transitions or large environmental reorganization. Fortunately, the graphics processing unit (GPUs) has emerged, and it is helping the TI and FEP calculations [69] .

MM/PBSA and MM/GBSA allow analysis of the contributions from individual residues or energy terms by free energy decomposition analysis that gives detailed residue-specific energetic contributions to the system binding, identifies dominant interactions in the binding process [64]. To the best of our knowledge, many chemical reactions and biological reactions are carried out in solvents such as water [64]. The MMGBSA method divides the solvation energy into polar and non-polar contributions [64]. Molecular docking has a problem of limited solvation effect and it is critical to investigate the solvation effects of the structures and functions of biological systems such as proteins, DNAs, RNAs and the interactions between them [64]. To tackle the problem of the limited treatment of solvation effect and the affinity prediction, molecular mechanics-generalized born surface area have been used and it is a physics-based scoring that involves the rescoring or lead optimization to improve the accuracy of binding affinity prediction [52].

Several biological processes in the cell are initiated by the binding between two interacting molecules and the significance of these specific interactions can be judged through the free energy of binding (ΔG_{bind}) [56], [70]. The amount of energy released or absorbed during a chemical reaction is known as the change in free energy (ΔG_{bind}). Moreover, it describes the

thermodynamics and the kinetic properties in a solvent specifically the binding of the ligand to a protein [56], [71].

Another interesting aspect is that the binding free energy (BFE) calculations are the key to understanding the in-depth binding mechanism between a protein and ligand, which involves both entropic and enthalpic contributions [72]. As demonstrated in thermodynamics branch of science, free energy is known as the amount of internal energy of a system that is required to do work and determine the direction of the process. In drug design, the binding strength between the protein and a drug molecule is characterized by the free energy [56], [73], [74] [64]. The final value of ΔG_{bind} is positive if the chemical process is non-spontaneous and it is negative if the chemical process is spontaneous. The method was developed in the late 90s by Kollma et al and used to calculate the free energy of a state by decomposing the energy into both entropic and enthalpic terms using the free energy equation (equation 2) [64].

Equation 2 : Free energy equations.

$$\Delta G_{bind} = \Delta H - T\Delta S \approx \Delta E_{MM} + G_{sol} - T\Delta S$$

$$\Delta E_{MM} = \Delta E_{internal} + \Delta E_{electrostatic} + \Delta E_{vdw}$$

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

Where ΔG_{bind} represents the binding free energy of protein-ligand, G_{sol} represents the de-solvation free energy,ΔE_{MM} represents the internal energy term which is the summation of bond, angle and dihedral energies, $\Delta E_{electrostatic}$ represents the electrostatic energy term, ΔE_{vdW} is the van der Waals interaction, G_{sol} is the sum of electrostatic $\Delta G_{PB/GB}$ (the polar) and ΔG_{SA} (non-polar) and T represents absolute temperature and S entropy [75], [66], [76]. The $\Delta G_{bind}$is calculated as the summation of the gas-phase molecular mechanics interaction energy between the ligand and protein shown as E_{MM}, the solvation of the free energy (G_{sol}) and the change in the conformational entropy associated with ligand binding (- $T\Delta S$) [75], [66].

The MM-GBSA technique is used to calculate the free binding energy of the system to estimate and compare the binding affinity of the system [56], [60], [64]. It is noted that the reason for applying an implicit solvation model to calculate the free energy of a system is to avoid the large fluctuation of potential energies when explicit water molecules are used in the calculation. The MM-GBSA calculations we have employed for the study neglects the structural relaxation of the ligand in the protein binding pocket. However, several studies have reported that the snapshots from molecular dynamic trajectories make the MM-GBSA calculations comparatively accurate, but it is computationally expensive [64]. Although we have used the general sufficient method of ranking the compounds against the protein by the ΔG_{bind}, the dynamical effects are ignored, this results in the calculations based on the starting structure only. Thus, this study also employed molecular dynamics to accommodate the dynamic effect ignored by MM-GBSA.

(c) Molecular Dynamics (MD) simulations

(i) Introduction

For many centuries it has been known that every reaction in the universe has its own cause and every cause has its reaction this is known as the principle of cause and effect [77]. Undeniably, the atoms that make everything around the universe are in flux and have vibrational frequencies thus it has to be considered that the docked system is not fixed and also there are interactions between the system and solvents such as water. Therefore, molecular dynamics simulations were carried out on docked systems. The interactions of a ligand with a protein seen in docking does not guarantee the stability of the ligand in the binding pocket of proteins because molecular docking studies only give binding energy information and the orientation of the ligand in the pocket [47], [48], [78]. The nature and stability of the target protein-ligand interaction in the presence of solvent (water) molecules can be studied with molecular dynamics (MD) [47], [59], [79]. Therefore, MD have been used as the imitator of nature to cause vibrations to the system and carefully analyze the protein-ligand interactions when the system is not stationary.

(ii) Classical mechanics in Molecular dynamics Model

MD simulation is a computational technique that uses dynamical trajectory for a system composed of N particles with initial conditions and is produced by solving the integral Newtonian equations of motion [36], [61], [80], [81]. The Newton's equations of motion are called classical mechanics. They are used in Molecular dynamics simulations to derive the motion of atoms in molecules or larger objects over time [42]. Kinetic energy is added to the potential energy surface and this is followed by the motion of the system which follows the classical mechanics laws over potential energy surface [43].
The following classical mechanics equation of motion (equation 3) is solved in MD.

Equation 3: Molecular dynamics equation.

$$m_i \frac{d^2 \boldsymbol{r_i}}{dt^2} = \boldsymbol{f_i} = -\frac{d\boldsymbol{r_i}}{dr_i} \cup (r_i, r_2, \ldots, r_N)$$

U is the potential energy depending on the coordinates of the N particles. Each atom is represented by i at position r_i. The equation simply calculates the total force and acceleration experienced by every atom [61], [43].

The differential equations are applied in MD because they offer a description of how the universe works. In the present case, it models the behavior of fluids and represent any motion that occurs with the human body. The atomistic model system is composed of a collection of the interacting particles represented as atoms explaining both solute and solvent placed in a adequately large simulation box where their movements are solved by Newton classical equations [36], [45]. The atoms and molecules are permitted to interact over a period of time by running a simulation, then the potential energies and interatomic forces are calculated using molecular mechanics force fields [36].

(iii) Force fields

The body is a microcosm of the macrocosm (aggregation of infinitely smaller particles) with force fields. Thus, when physical drugs are put in the body and the effects of the force fields to the physical drugs by the use of computer modelling with force fields has to be studied. Mathematically, force fields are mathematical expression that describes the dependence of the energy of a system on the coordinates of its particles and used in a model of potential energy of a chemical system [43], [61].

Thus, force fields are very important in the model and the frequently used force fields are namely GROMACS, AMBER, CHARMM, COMPASS, OPLSAA_IMPACT_2001 and OPLSAA_IMPACT_2005 force fields [43], [45], [62]. For in this research, the OPLSAA_IMPACT_2005 force fields through Desmond was used [82]. The OPLS force fields are often employed in the simulations of biomolecules and were developed mainly to simulate condensed matter [61]. The force fields consist of the set of parameters and the interatomic potential energy which are obtained by the help of *ab-initio* or *semi-empirical* quantum mechanical calculations [62]. The general force field expression (equation 4) is shown below.

Equation 4: Force field equation.

$$U = \sum_{bonds} \frac{1}{2} k_b (r - r_0)^2 + \sum_{angles} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{torsions} \frac{V_n}{2} [1 + COS(N\emptyset - \delta)] + \sum_{improper} V_{imp} + \sum_{LJ} 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) + \sum_{elec} \frac{q_i q_j}{r_{ij}}$$

In this equation, the terms for bond stretching, angle bending, torsions (dihedral torsions and improper torsions) refer to the local or intramolecular contributions to the total energy. The last two terms, LJ (Lennard-Jones potential) and the Elec (coulombic interaction) refer to the repulsive and Van der Waals interactions and Coulombic interactions [61]. The first four terms, the intramolecular terms are represented with simple harmonic functions which controls the covalent bonds length. The dihedral or torsional terms are included for molecules that have more than four atoms in a row. The spring constant K is estimated from the infrared or Raman spectra, the r_0 can be obtained from X-ray diffraction experiments [43], [45], [61]. The last two terms, the intermolecular terms are represented with the interactions between the two atoms that arise from the balance between the repulsive and attractive forces. The induced dipoles form the attractive

forces while the overlap of the electron cloud of both atoms forms the repulsion forces [43], [45], [61].

(iv) Limitations of force fields and Molecular dynamics simulations

As mentioned previously, MD uses classical mechanics equations and the limitations for its application in drug discovery cannot predict chemical reactions and reactivity [42], [43], [61]. Due to the nature of movement of electrons which are wavelike and quantized, classical mechanics cannot work to determine reactive [42], [43], [45], [61]. The empirical force fields are unable to handle reactions such as breaking and formation of bonds, electron excitations, charge transfers and others [43]. They also cannot explain the electronic structures and also their predictive powers are only limited to systems that have the functional groups [43]. With the help of hybrid quantum mechanics – molecular mechanics (QM/MM), researchers can now develop calculations for complex chemical and biological systems such as proteins and study in depth the electronic structure of atoms and molecules and thus chemical properties [42]. For instance, atoms that are close to forming and breaking chemical bond in a reaction can be studied with QM [42]. Due to this unachieved goal of MD, QM/MM is employed to the systems.

(d) Quantum Mechanics/Molecular Mechanics

(i) Quantum mechanics (QM)

Throughout the 6th century, Pythagoras in his Pythagorean university of Krotona taught principle of hydromancy (the motion of liquids) and believed also that there is a rumbling sound inside brass bowls even when they were perfectly fitted [83]. In 1920, the discovery of quantum mechanics based on pure mathematics by Heisenberg Schrödinger , Paul Dirac and other physicists developed a theory that can explain how chemistry arises from interactions of nuclei and electrons [42]. In the QM approach, the nuclei are arranged in space and the corresponding electrons are spread all over the system in a continuous electronic density and computed by solving the Schrödinger equation [43], [45], [84]. As mentioned earlier, QM is applied to study the central part of the

system, where atoms are close to forming or breaking chemical bond in a reaction and thus chemical properties can be obtained [42].

The human body is made of cells, which are filled with chemicals made of molecules and atoms. The atoms are made of protons, electrons and neutrons, and going further hadrons and quarks are present. Intriguingly, inside the hadrons and quarks is pure energy in state of vibration forming standing waves. The motion of electrons are wavelike, quantized and can be described accurately by quantum mechanics (QM), which is capable of describing the electronic structure of atoms, molecules and chemical properties and reactions [42]. QM describes the energy of a molecule in terms of interactions among the nuclei and electrons as given by Schrödinger equation [85].

(i) Classical mechanics and Quantum mechanics' background

In 1687, Sir Isaac Newton developed the field of classical mechanics which predict trajectory for particles, allows the translational, rotational and vibrational modes of motion to be excited to any energy with the control of forces that are applied [86]. However, much later in the 1900s it was found that it has a lot of failures as it fails to explain the black body radiation, photoelectric effect, atomic and molecular spectra and heat capacities [86]. In summary, it can only predict large objects visible by the eyes but fails to predict the behavior of subatomic particles. Before the revolution in physics, it was believed that there is an indivisible particle inside the atom which forms everything else in the materialistic dream and that was the end of atomic theory.

To remedy the failures of classical mechanics, Erwin Schrödinger and other physicists developed Quantum Mechanics (QM) model which solved the problem by showing that the electrons are travelling in the quantized orbits at their lowest energy state, thus do not exist as the cloud around the nucleus. Max Planck and Niels Bohr proposed that the electron maintains a stable orbit due to its energy being quantized and the electron in its lowest energy state has the minimum possible energy. Thus, classical mechanics have been neglected in favor of quantum mechanics [84], [86].

Nikola Tesla also disregarded the idea that electrons are particles but believed that they are waves that contained energy [87]. This was later described in detail in wave functions in experimental physics that the lowest energy of the electron is not at the center but spread out in a wave form. A wave function is a mathematical function which contains information about the motion and location of the particle [86].

Attempting to acquire deeper understanding of chemical compounds quantum mechanics must be employed. The physical state of a particle at time t is described fully as possible by the wave function [84]. The nuclei are assembled in space and position of electrons with respect to occupancy around the nuclei are solved using the Schrödinger equation [45], [84], [88]. The square of the wavefunction gives the probability of finding an electron and for hydrogen atom, the familiar wavefunctions are s, p, d . . . atomic orbitals [85].

(ii) Schrödinger wave equation

The Schrödinger equation is used in determining the electron wave functions within molecules, then calculates the total energy within the wave function and predicts how fast the energy is going to evolve over time The Quantum mechanical simulations show the motion of nuclei and electrons in molecules as they change from reactants to products and also determine how much energy it takes for bonds of a molecule to be reorganized through nuclear and electron rearrangements to form other molecule (the product) [42]. The following is the 3-dimensional Schrödinger equation (equation 5) which describes the energy of a molecule in the terms of interactions among the nuclei and electrons and the wave functions [45], [84], [86].

Equation 5: Schrödinger equation.

Kinetic energy Potential energy

$$\frac{-\hbar^2}{2m}\nabla^2\psi + V\psi = E\psi \text{ where } \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

$\hbar$ represents the Planck constant $\hbar = \frac{h}{2\pi}$, ∇ is the quantity that describes the wave function ψ is the probability amplitude of Born Interpretation, it gives the shape of the wave, the squared represents the dynamic geometry, V describes the force acting on the particle, E represents the energy. $\frac{-\hbar^2}{2m}\Delta^2 + V$ represents the Hamiltonian composed of with kinetic energy and potential energy. In summary the Schrödinger equation can be written in this form in Hartree Fock method $\hat{H}\psi = E\psi$ [45], [86], [84].

(iii) Combination of quantum mechanics and molecular mechanics

Today, we know that the vibrational energy and bond energy are quantized, the electrons, protons and neutrons are excited from low orbital levels to higher orbital levels. Herein, we apply the combined Quantum mechanics and molecular mechanics (QM/MM) to model reactions in biomolecular systems. In the QM/MM system, there are two regions. The first region is Quantum mechanics (QM) model which is needed to describe the chemical reactions and other electronic processes such as charge transfer or electron excitation [89]. However, the QM model is only limited to a few 100 atoms, and it needs to treat up to several 100 000 atoms and allowing the simulation over time scales of tens of nanoseconds [89].

To address the problem of the QM model, the second model is made which is molecular mechanics (MM) which is an environment described by force fields which makes protein and solvent. Hence, the whole system is called the combined QM/MM method which provides the accurate reactive biomolecular system [89]. The MM region uses the Newtonian classical equations to explain the movement of atoms. Thus QM/MM is used the chemical regions in a system [89]

(e) QM with Density functional theory (DFT).

(i) Introduction

DFT is a computational quantum mechanical approach that is utilized in physics, chemistry and materials science to investigate the electronic structure of many-body systems, in particular atoms, molecules, and the condensed phases. The DFT model is applicable to molecules of moderate size of about 50-100 atoms [45]. It uses the electron densities which shows the location of the electrons [85]. The large values of electron densities indicate the atomic positions while the smaller values indicate the overall molecular size [85]. To shed further light where no information from experiment is obtain, the electron densities allow for the description of bonding in transitions states [85].

(ii) Born Oppenheimer approximation

The first basic approximation in quantum chemistry, The Born Oppenheimer (BO) approximation is used in the calculation of the potential energy surface as it is required in the description of the dynamics of nuclei [46], [43], [90]. BO approximation method that is used for exploring reaction pathways of a chemical system by the combination of self-consistent field (SCF) optimization procedure used in the QM part of QM/MM geometry optimizations and classical MD [43], [45], [90], [91]. BO approximation has many applications such as generation of theoretical spectra, exploration of reactivity and simulation and calculation of thermodynamics properties [43], [46], [90]. For biomolecules, BO approximation is sometimes used for computation of energy for the system and electronic density (although this is often in the context of QM/MM combined calculations, since the cost of performing of QM calculations on a biological macromolecule is prohibitive.

BO approximation isolates the motion of electrons and motion of nuclei. The nuclei have higher mass than the electrons, thus the motion of nuclei is neglected due to their relatively slow movement [92]. Furthermore, there is an electrostatic force that is acting on the nuclei and electrons and this force causes the acceleration of both particles. The magnitude acceleration is

inversely proportional to the mass, a= $\frac{F}{M}$. The acceleration of electrons is faster than the one for the nuclei, the acceleration for the nuclei is neglected [45], [46], [90], [93].

Consequently, the nuclei are considered to be stationary at each point in the DFT calculation, hence the motion of the nuclei and the mass of the nuclei are neglected while solving the Schrödinger equation using self-consistent field (SCF) calculations [43], [45]. They are done in two steps, the first computes the Hamiltonian or energy matrix in which the elements of the matrix are the integrals which are the main causes of the bonding terms between atoms, they involve the atomic orbitals and terms obtained from the Schrödinger equation [43]. The second step is the one that involves the *ab-intio* calculations that determine the linear combination of atomic orbitals (LCAO) coefficient based on the molecular orbital theory [43].

In 1964, Hohenberg and Kohn developed two theorems which are the basics of the DFT theory, and they explain the importance of the electron density and determines all ground state properties of the system. This was a great step in the development of the quantitative modelling of electronic structure which determines F_{ni}[n] such that the kinetic energy N is evaluated of non-interacting particles given only their density distribution and thus finding the corresponding potential V, which is used in the Schrödinger equation [94]. The functionals are used in DFT, and they are functions of functions, in this case of electron density [95].

The Hohenberg-Kohn theorem 1 states that the ground state energy is the unique functional of electron density and all ground-state properties of an electronic system can be determined. Theorem 2 states that the electron density that gives out the minimum energy of the overall functional is the true ground state electron density [43], [45]. The expensive method of calculating the Schrödinger equation is avoided as a result of the theorems above. The electron density is used in DFT instead of treating each electron separately like in the Hartree-Fock approximation [43], [45]. The calculations are significantly speed up when using the density. If the electron density functional is known, then the total energy of the system can be determined [45].

(iii) Molecular geometry

The main object of DFT is to obtain the most stable chemical structure, equilibrium molecular geometry and transition state. The stability of a system is directly related to the minimum energy configuration [43], [93], [96], [97]. The equilibrium molecular geometry such as bond lengths and angles describe the minimum values of the potential energy surface. Therefore, the starting chemical structure is represented by this minimum potential energy [43], [45].

(iv) Frontier molecular orbitals

The molecular orbitals (MO) energy and spatial distribution can be obtained from DFT optimized structures. MOs are the sums and differences of atomic solutions (atomic orbitals) [85]. The biological interactions between the ligand and protein involves electrons in the frontier orbitals (Lowest unoccupied molecular orbitals (LUMO) and the highest occupied molecular orbital (HOMO)). Hence, the two frontier orbitals of chemical species are very important in describing the reactivity [38], [95].

The higher the value of the HOMO energy of the molecule, the more it can donate electrons to appropriate acceptor molecule of low empty molecular orbital energy [85]. The lower value for the LUMO energy, the more it can accept the electrons. The energy gap value is shown by ΔE_{H-L} which is basically the difference between the HOMO and LUMO energies, it often represents the reactivity of the ligand or kinetic stability [38], [78]. The larger value of the energy difference shows that the ligand is less reactive and vice versa for the lower values of energy difference [38].

Koopman's theorem was developed within the Hartree-Fock theory, it gives approximations to the first ionization potential (I) which is the minimum energy that is needed to ionize an atom in its ground state and is equal to the negative of the HOMO energy (equation 6). The electron affinity (EA) is similarly equal to the negative of the LUMO energy (equation 7) [93], [95].

Equation 6: Ionization potential equation.

$$IE = - E_{HOMO}$$

Equation 7: Electron affinity.

$$EA = -E_{LUMO}$$

(v) Dipole moment (μ)

Dipole moment describes the polarity of a molecule, and the SI unit is coulomb meter (Cm) [95]. Moreover, the permanent electric dipole of molecules must be determined in order to understand the rotational excitation by radiation, where one end is $\delta+$ and the other δ – h [95]. This quantum chemical calculation assists us in obtaining the atomic charges in a molecule as it cannot be measured experimentally [45]. The language of organic chemistry relies in the atomic charges which help in understanding reaction mechanisms [45]. The dipole moment is a measure of a polar covalent bond and it is defined as the product of the charge on the atoms and the distance between the two bonded atoms. The following is the equation of dipole moment where the Q represents the charge and R is the distance (equation 8) [95].

Equation 8: Dipole moment.

$$\mu = QR$$

(vi) Molecular electrostatic potentials (MEP)

To overcome the problems of reactivity behaviour and site prediction, the molecular electrostatic potential was used in order investigate the chemical reactivity of the ligands in drug discovery process [38]. Simply, the main purpose is to identify the reactive site of a molecule [98]. The Molecular electrostatic values are used to show both the electrophilic and nucleophilic reactions. The electrostatic potential is the energy of interaction of a point positive charge (electrophile) with the nuclei and electrons of a molecule [45]. A single point calculation is used to calculate the electron density, which is simply the probability of finding an electron at a point in space [99] [100]. They were calculated at B3LYP/6-31G(d) optimized geometries using the molecular electrostatic potential equation (equation 9).

Equation 9: Molecular electrostatic potential equation.

$$V(R) = \sum_A \frac{Z_A}{|R_A - r|} - \int \frac{p(r^i)}{|r^i - r| d|r^i}$$

Where V(r) at any given point in the x, y, z component of the molecule represents the interaction energy between the electrical charge generated from the electrons and nuclei and a positive test charge (proton) located at r. The summation runs over all the nuclei A in the ligand, Z_A is the charge of nucleus A, located at R_A and $p(r^i)$ is the electron density function of the molecule [100]. In material science, V(r) is determined experimentally by diffraction and in computational chemistry, it is determined with the help of DFT [100].

1.3 Summary of the chapter

In this chapter, we have explained the theoretical background of the techniques which are applied in the current study. As we conclude this section, for more in-depth information the reader is encouraged to refer to these highly recommended books, “Quantum mechanics” by Eugen Merzabecher, vibrations and waves in physics by Iaing G Main and finally the excellent book of the whole work, “Hyperchem computational chemistry, tools for molecular modelling” by Hypercube, Inc.

Chapter 2 : COMPUTATIONAL BACKGROUND AND METHODOLOGY

Introduction to the chapter

This section provides a thorough explanation of computational methods utilized in this research. The detailed procedures outlined in this chapter are connected to the subsequent chapters where the main focus is on presenting the results obtained. The following techniques were used ; molecular docking, ligand efficiency, molecular-mechanics born generalized area (MM-GBSA), molecular dynamics simulations (MD), Density functional theory (DFT) and Quantum mechanics and molecular mechanics (QM/MM).

2.1 Molecular docking

In this section, as part of an effort to identify potential hits for squalene monooxygenase inhibitors, the main focus was to find the ligands that have high binding affinities towards squalene monooxygenase. Our approach entailed the use of squalene monooxygenase inhibitors instead of squalene synthase inhibitors and applying several computational techniques that precede molecular docking.

(a) ***Ligands and proteins preparation***

The protein 3D crystallographic structures were downloaded from the protein data bank [101]. The quality of the proteins was examined using Schrödinger Ramachandran plots prior to the screening. Molecular docking was done on the following six cholesterol proteins, 2ZCS, 3D4S, 3COX, 6GQF, 3W7T and 1M17. There were 23 compounds that were taken from A. Belter et al study about squalene monooxygenase (SM) inhibitors which were used for the high-throughput screening (HTVS) based on molecular docking in our study [19]. The 2D structures of the compounds were prepared in Schrödinger suite (v3.1, Schrödinger 2021-1) [102]. The HTVS was done through the KNIME analytics platform, and the results were visualized using Maestro from Schrödinger 2021-1 v3.1.

(b) The KNIME workflow

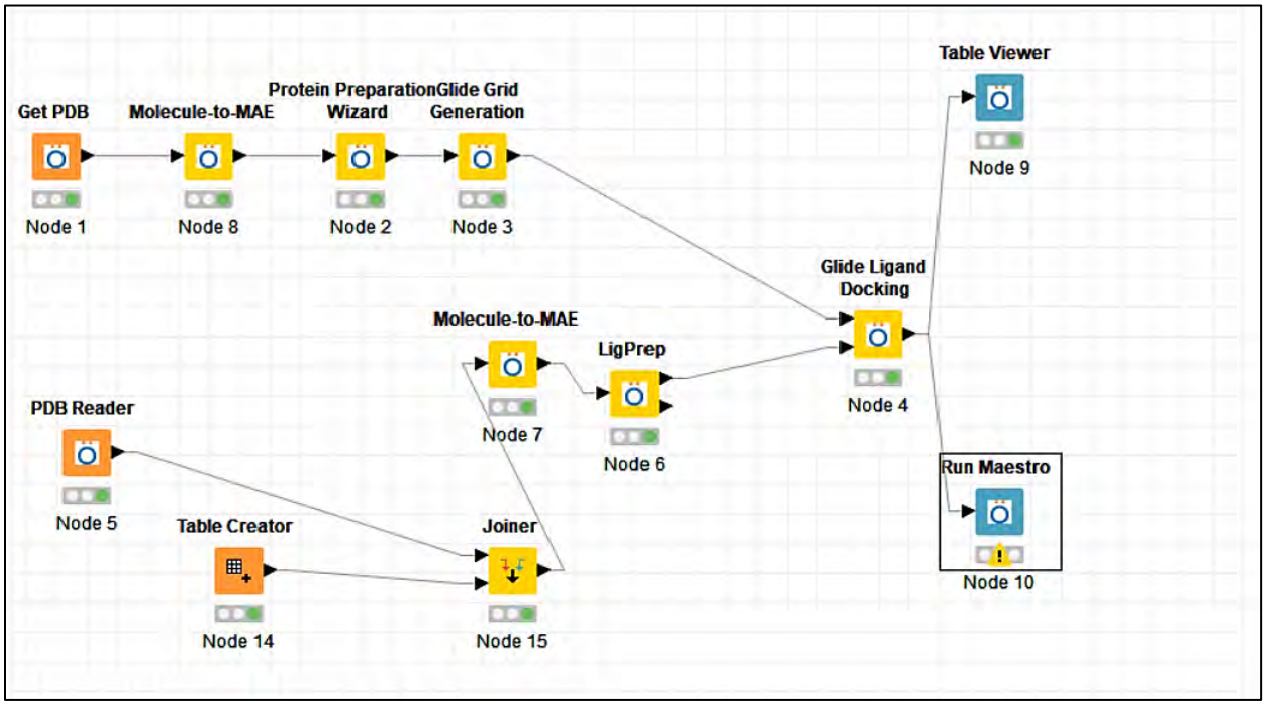


Figure 2.1 The KNIME workflow.

In figure 2.1, the PDB Reader (node 1) is used to read in the protein from the protein data bank. The protein is then converted to maestro format using the Molecule-to-MAE (node 8). Preparation Wizard (node 8) prepares the protein for docking by performing tasks such as correcting the missing hydrogens, correcting the charge states and bond orders. Subsequently, the grid box of the protein was prepared using the Glide Grid Generation (node 3). The grid box was generated using the native ligands center of mass at the active site of the protein.

The ligands were imported to the workflow using the PDB reader (node 5) connected to a table reader node (node 14) which helps in naming the ligands properly after they are read into the workflow. Then the Molecule-to-MAE (node 7) exported the ligands to maestro format. The ligprep (node 8) which is known as the ligand preparation wizard, helps in preparation of the ligands prior to glide docking. The ligand preparation is necessary tool in the Schrödinger software because it minimizes the energy, improves the incorrect valences and the protonation of the ligands.

Finally, molecular docking was done using the glide ligand docking (node 4). The docking scores were viewed using Table viewer (node 9). The 3D structure of the protein-ligand poses and the protein-ligand interactions 2D diagram were viewed in using the Maestro (node 10).

(c) Post docking calculations/matrix

Potency of the drug is often linked with lipophilicity and molecular weight, it increases as they increase and thus the idea of normalizing the biological potency value based on the descriptors of size such as heavy atom count (HA) in case of ligand efficiency (LE), lipophilicity has been explored in detail [103]. The clinical candidates which obey the rule of five with a special focus on lipophilicity are often preferred [104].

(i) Ligand efficiency (LE)

The ligand efficiency (LE) is widely used in drug discovery and it is a useful metric for lead selection [105]. It relates efficiency of potency per heavy atoms in a ligand by optimizing the binding affinity with respect to molecular size [104], [105]. It gives priority to small molecules with relatively lower potency rather than larger molecules with high potency. Potency is considerable when selecting a lead compound for further optimization into drug and it is correlated with the molecular weight (MW) [105]. Comparison of potency among the compounds within a wide range of molecular size is done as the LE measures potency with respect to molecular size using ligand efficiency equation (equation 10) [105].

Equation 10 : Ligand efficiency (LE).

$$LE = \frac{\Delta G}{Hev}$$

Where ΔG is the binding energy of the ligands and Hev is heavy atoms.

(ii) *Lipophilicity*

The lipophilicity is the ability of a compound to dissolve in fats, oil, lipids and non-polar compounds and the parameters of lipophilicity solved using lipophilicity equation (equation 11) [68], [105]. The assessment of lipophilicity in vivo which indicates the process of molecular

desolvation during the transfer from aqueous phases to cell membrane and protein binding sites [105]. The increased lipophilicity, molecular size and molecular complexity results drugs lacking the affinity to function in most chemical starting points [68].

Equation 11: Lipophilicity

$$Log\ P = log(\frac{[X]_{Octanol}}{[X]_{water}})$$

(iii) Ligand Lipophilicity efficiency (LLE)

The ligand lipophilicity efficiency (LLE) is a ligand efficiency index which links the potency and lipophilicity in an effort to estimate the drug likeness [103], [105]. It maximizes the minimum acceptable lipophilicity per unit of vitro potency while maintain the lipophilicity to be low. It is the difference between the log P and the pIC_{50} values solved using ligand lipophilicity efficiency (equation 12).

Equation 12: Ligand lipophilicity efficiency (LLE)

$$LLE = pIC_{50} - log\ P$$

(iv) Ligand efficiency dependent Lipophilicity index (LELP)

The ligand efficiency-dependent lipophilicity index (LELP) overcomes the limitation of the LLE by combining lipophilicity, molecular size and potency into one composite descriptor, it is solved using ligand efficiency dependent lipophilicity index (equation 13). [105].

Equation 13: Ligand efficiency dependent lipophilicity index (LELP).

$$LELP = \frac{logP}{LE}$$

The ligand efficiency data and plots were calculated and constructed using the open-source chemoinformatic toolkit RDK implemented on google colab, jupyter. The chemical structures were converted to smiles string using Maestro Schrödinger software 2021-1. The SMILES string

is a chemical notation system designed for modern chemical information processing and is well suited also for high-speed machine processing [106]. The related RDkit functions were imported and the data was also employed in form of a dictionary, which can be seen in the appendix.

2.2 Molecular mechanics-generalized born surface area (MM-GBSA)

The molecular mechanics-generalized born surface area (MM-GBSA) study serves as a complement to the molecular docking study. It calculates the free energy between the ligand and the protein. The binding free energy plays a vital role for studying the mechanism between the protein and ligand which encompasses both entropic and enthalpic contributions [107].

MM-GBSA method uses the molecular mechanics (MM) force fields and implicit solvation models represented by GBSA to calculated the free binding energy of the protein-ligand complex [13]. The MM-GBSA tool of prime module of Maestro Schrödinger 2021-1 v3.1 was utilized for the calculation. The docked complex was used and the solvation model VSGB and force field OPLS_2005 used.

2.3 Molecular Dynamics (MD) simulations

Molecular dynamics (MD) simulation is one of the integral parts of computational aided drug discovery approaches that is used in the investigation of the binding stability of the docked ligand in the active site of a protein [66]. Dynamical systems are a field of mathematics concerned with the motion of objects. It assess the stability of a ligand in the binding pocket of a protein by application of classical mechanics equations to calculate the position and speed of each atom of the system. The best docked compounds obtained from molecular docking were taken for the MD simulations. The MD simulations were performed using Desmond [82].

The solvated environment for the docked complex was simulated using the system builder on Schrödinger software, an orthorhombic box shape with the dimensions of 10.0 Å distances on all sides and angles of 90 degrees for alpha, beta, and gamma was generated using desmond. The system was solvated using an equilibrated three-site transferrable intermolecular potential (TIP3P)

water box which contained 45131 water molecules represented by TIPS and the solvent model was predefined as TIP3P, all the atoms were within the 10.0 Å of the orthorhombic water box. The boundary condition box volume was computed along with the Na^+ and Cl^- which were added to neutralize the system randomly in the orthorhombic box [107].

The OPLS_2005 force fields were used for the model with the Brownian dynamics NPT ensembles followed by relaxation. The NPT is employed to mimic systems in realistic conditions and they are conditions of constant pressure (P), constant number of particles (N) and constant temperature (T). The water molecules were subjected to energy of 100 ns was performed for the simulations of the complexes, approximate number of frames 1000, recording interval trajectory of 100 ps and energy 1.2. The system was heated to temperature from 0 to 300k and Pressure of 1.01325 bar gradually for 100ps while maintaining a fixed number of atoms and volume [107]. The host used was graphics processing unit gpu_1(36,1). Thereafter, the MD simulation was run on a high-performance computer cluster through the Centre for high performance computing (CHPC) in Cape Town, the calculations usually take 10 hours to finish after submission. For inspection of the trajectories, the frames were automatically assembled and inspected using the simulation interaction diagram on Schrödinger software. The diagrams of the molecular dynamics simulations and analysis of the graphs were visualized using the Desmond on the Schrödinger software 2021-1 [108]. Figure 2.2 shows an illustration of the initial water boxes that were created.

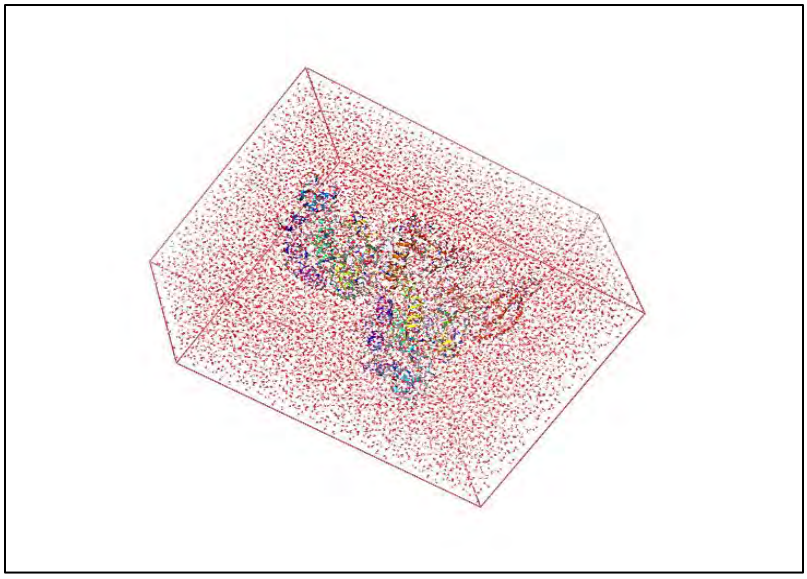


Figure 2.2 The initial water box with the 2ZCS protein and SDZ18 ligand before MD simulation.

2.4 QM with Density functional theory (DFT)

The Density functional theory (DFT) studies provide insights into the compounds by using the Born-Oppenheimer Approximation to calculate the properties such as geometries, energies and spectroscopy properties [79], [97], [109]. The GAUSSIAN-09 software was used to refine and optimize the geometries of the ligands at B3LYP/6-31G(d,p) level of theory and the DFT calculations run on the cluster at Rhodes university and the results were then visualized with Gauss view [110].

Specifically, the selection was made in favouur of the B3LYP because many studies have proved that it has been more accurate energetic predictions when compared to the gradient approximation (GGA) functional and it provide the better bond length results [45], [111], [112], [113]. The solvation model of SCI-PCM model was used and the molecular geometry optimizations and vibrational frequency calculations were performed.

2.5 Combined Quantum mechanics and molecular mechanics (ONIOM QM/MM)

In this study, the Our-own N-layered Integrated molecular Orbital Molecular mechanics (ONIOM) method was used to calculate the energy of our complexes. The ONIOM QM/MM was developed by Morokuma and co-workers, the concept is to partition a large molecular system into onion-skin layers using different combinations of the quantum chemical and molecular mechanics method for different layers of the system [91]. To reduce the computation cost, the two layered hybrid ONIOM method as implemented in the GUASSIAN-09 program was used in this section. The inner and total real parts of the ONIOM system remained the same throughout the calculations [91]. The Becke's three-parameters Lee-Yang-Parr (B3LPY) hybrid density functional method was used for all the higher level calculations of the ONIOM model [114]. The basis set used for geometry optimization and frequency was B3LYP/6-31G for higher layer, and the lower layer was calculated with the semi-empirical UFF and the remaining atoms.

The amino acids of the protein that were close to the ligand were described as part of the higher layer and the other amino acids remained as the lower layer. Therefore, the inner model part was treated with high level method. It is important to also obtain the graphical representation of chemical reaction which is called the intrinsic reaction coordinate (IRC), instead of struggling with a Schrödinger many-body problem [115]. Therefore, special attention was also paid to the IRC plot of the interaction energy of ligand and protein.

Chapter 3 : MOLECULAR DOCKING

Introduction to the chapter

In computational chemistry, molecular docking is a profound tool that is used in the prediction of the binding energy of the ligands and to discover novel drugs as well as optimize the already available drugs [53]. The computational approaches of molecular docking are outlined in chapter 2, section 2.1.

3.1 Protein PDB crystal structures used for the study

Table 1 shows the proteins used in this study, their full names and descriptions; the experimental protein crystallographic structures are placed in the Protein Data Bank (PDB) and denoted with different PDB codes. The starting step in molecular docking was to download PDB crystallographic structures.

Table 1: Protein PDB crystal structures.

PDB crystal structure code	Name of protein	Organism
2ZCS	The structure of C (30) carotenoid dehydro-squalene synthase.	Staphylococcus aureus.
3D4S	The structure of cholesterol bound form of bound beta2 adrenergic receptor.	Homo sapeins, Tequatrovirus T4.
6GQF	The structure of mouse Aster A (GramD1a) with 25-hydroxy cholesterol.	Mus musculus.
3COX	The structure of cholesterol oxidase complexed with a steroid substrate.	Brevibacterium sterolicum.
3W7F	The structure of C (30) carotenoid dehydro-squalene.	Staphylococcus aureus complexed with farnesyl thiopyrophosphate
1M17	The structure of the epidermal growth factor re.	Homo sapiens.

It is more convenient to also discuss the crystal structures of the proteins used in this study. In figure 3.1 shows the protein 2ZCS from *staphylococcus aureus* with no mutations and sequence length of 293. The protein 3D4S from the *Homo sapeins, Tequatrovirus* T4 organism with sequence length 490. The protein 6GQF is shown and it is from Mus musculus organism, with chains A,B,C and D with the sequence length of 231. The protein 3COX obtained from

Brevibacterium sterolicum with sequence length 507. The protein 3W7F obtained from organism *staphylococcus aureus* with chains A and B with has sequence length 293. Lastly, the protein 1M17 obtained from *Homo sapiens organism*, sequence length 333.

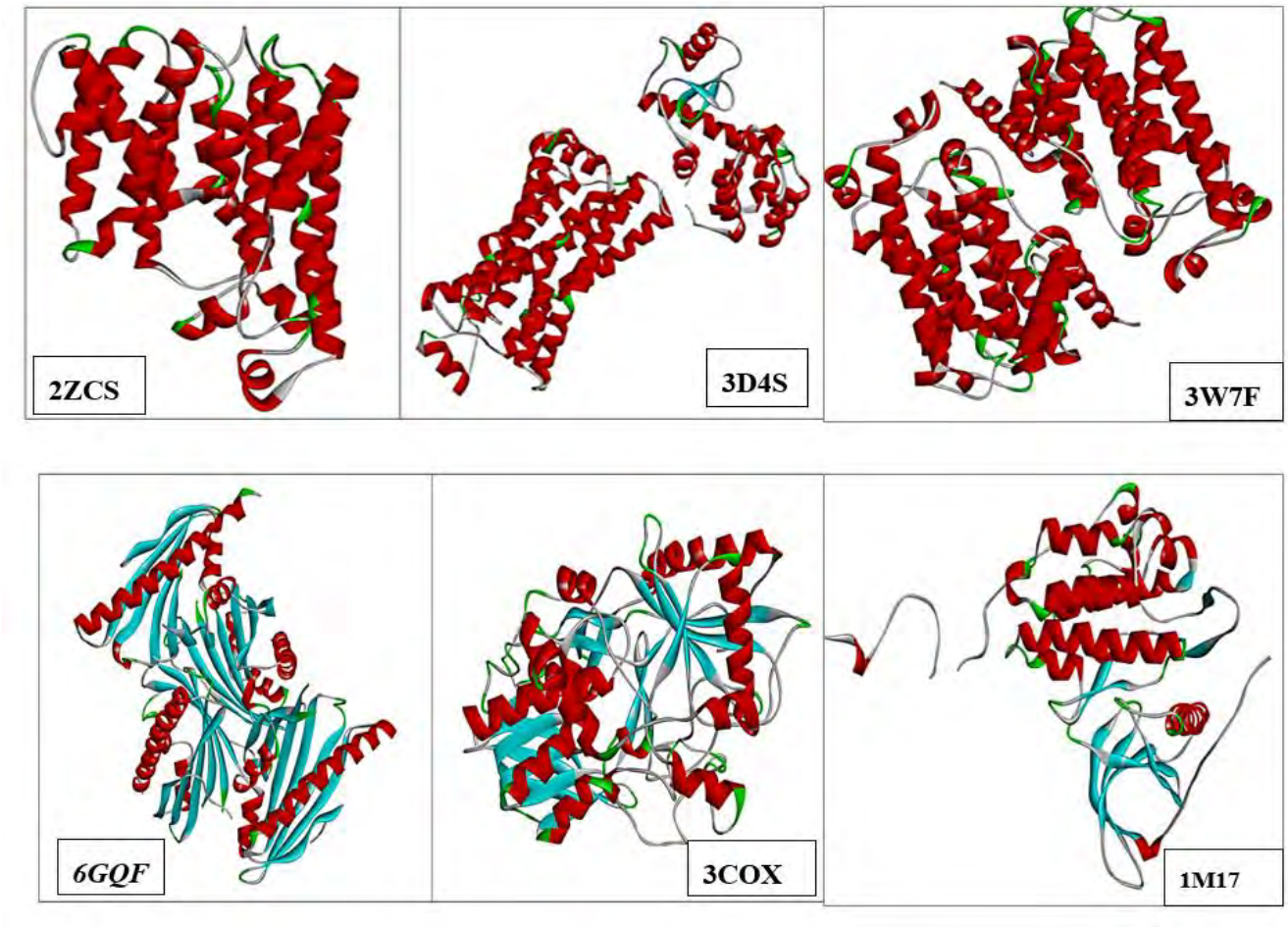


Figure 3.1 Protein PDB crystal structures.

3.2 The Ramachandran Plots

The Ramachandran plots were utilized to predict the quality of the proteins in this study. It is apparent in literature that the quality of the protein structure can also be judged by visual inspection [116]. In 1997, Rob W. Hooft et al published an intriguing research about objectively judging the quality of the protein structure from a Ramachandran plot [116]. The Ramachandra plot was originally developed by G. N Ramachandran, C. Ramakrishnan, and V. Sasisekharan. To visualize

the quality of the proteins, the 2-D graphical representation is provided in the terms of the torsion angles [117], [118]. It is a useful tool for validation of the accuracy of the protein structures which have been determined experimental by methods such as NMR, X-ray crystallography and Cryo-EM [117], [118]. It also checks the protein structures that are produced by the homology modelling or ab-initio methods.

In the figure 3.2 below, the points in the Ramachandran plot give indication to the torsional angles and the plot also shows the combination of $\psi \; and \; \phi$ which represents the amino acids residues that were obtained using the structure determination methods [117]. The symbols $\psi \; and \; \phi$ define the polypeptide or protein conformations, consequently showing the representation of the amino acid residues as points on the Ramachandran plot which can be used to identify secondary structure of the protein [117]. While these points may appear to be common feature for all proteins, closer inspection shows that there is some discrepancy. To begin with, we note that in figure 3.2 (a) the dots represent the amino acids, the red region is the completely favored one, the yellow region is the acceptable one and the white region is an unacceptable region [117], [118]. Nevertheless, the points that are shown on the unfavored region represent the sterically disallowed conformations of the polypeptide chain due to the steric collisions between the main chain and side chains [117], [118]. Now looking at the points in the red area, they are in the favorable region, and they represent the conformations that do not have the steric collisions [117], [118].

It appears, therefore, that protein 2ZCS crystallographic structure is of good quality, as it contains a lot of the torsional angles in the favored region represented by the dots. In this perspective, there are only two amino acid residues found in the unacceptable region. Correspondingly, in (c) and (d) for protein 3COX and 3D4S, the same characteristics were observed with just about two amino acid residues in the unacceptable region. The obtained results thus show that the generated protein structure is a better conformer than 1M17 in (e). For a bad quality protein, a lot of numbers of torsional angles will be shown in the unacceptable region [116]–[118].

Indeed, in contrast to previous structures, it is observed in (e), that protein 1M17 had several amino acids in the unacceptable region. However, the overall Ramachandran plots show that the proteins that were used for this research project had good quality structure as they contain a lot of the torsional angles in the favored region.

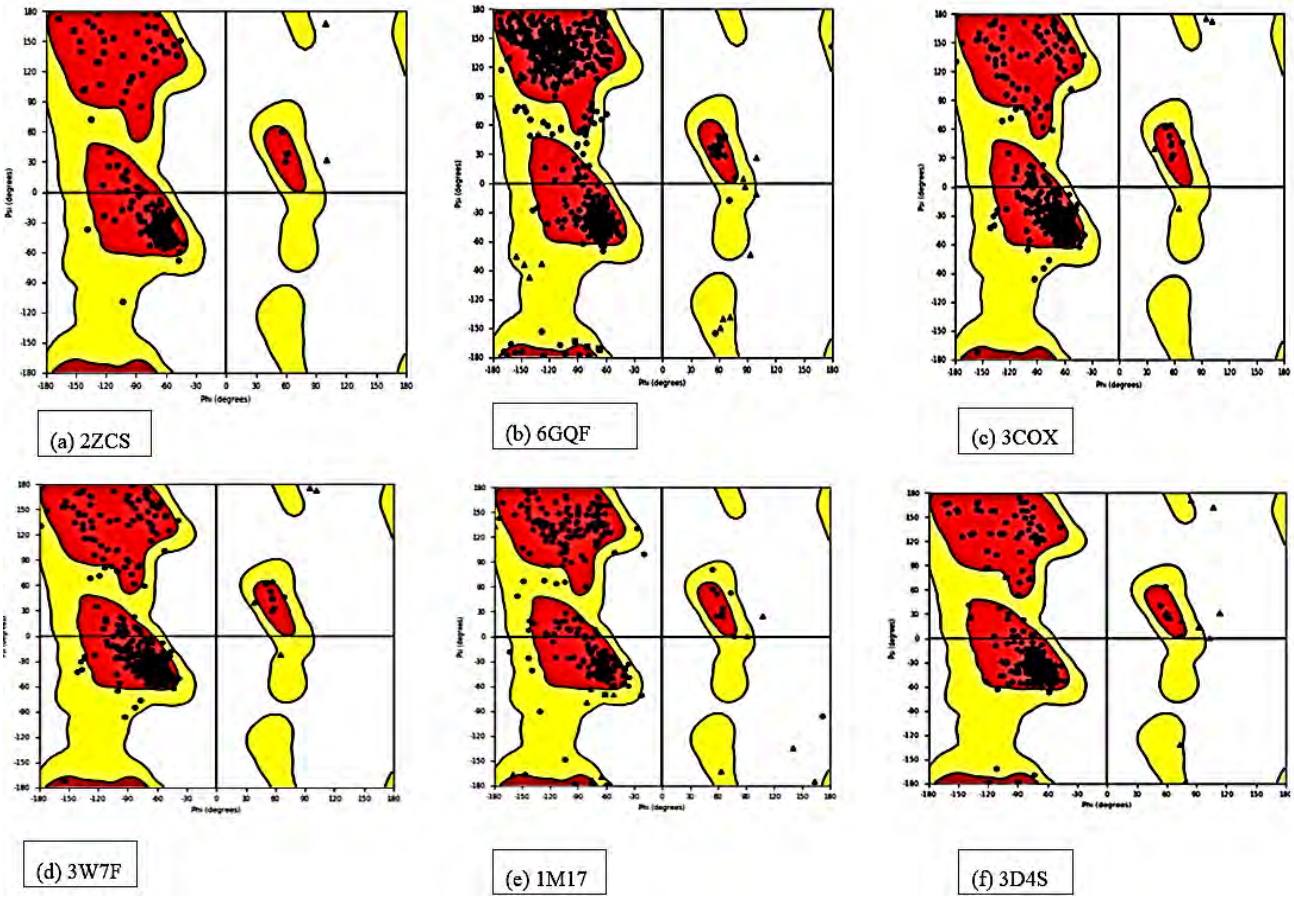


Figure 3.2 The Ramachandran plots with phi and psi degrees for proteins.

3.3 Molecular docking results

(a) Molecular docking scores

The data in table 2 shows the colour coded molecular docking scores in kcal mol^{-1} (good docking scores range from green colour, the intermediate in orange and the least docking scores are in red colour). The ligand SDZ 18 showed the best binding affinity towards all the proteins considered in this study, with the lowest bocking score being -6.0 kcal mol^{-1}. The ligands SDZ 9, SDZ 18 and SDZ 13 also showed good binding affinities, and this suggests that these molecules have a high potential for drug repurposing. Nevertheless, the least binding ligand was Allyl 105, with all docking scores above -1 kcal mol^{-1}. Based on the present data some general observations were made, the binding activity was mostly greater than -5 kcal $mol^{-1}$for most compounds indicating

that the binding force between them and these SM proteins was very good [119]. The empty spaces in the table represents the ligands that showed no binding affinity with certain proteins, they failed to dock in the active site.

Table 2 : Molecular docking binding affinities in kcal mol^{-1}.

Ligands	6GQF	3COX	3D4S	2ZCS	3W7F	1M17
SDZ 18	-7,5	-6,5	-6,0	-6,8	-8,4	-6,3
SDZ 13	-6,5	-6,4	-5,2	-6,2	-5,2	-4,2
SDZ 14	-5,1	-5,0	-3,8	-5,0	-4,7	-3,2
SDZ 16	-6,4	-5,4	-4,0	-5,3	-6,5	-4,4
SDZ 15	-7,9	-6,7	-5,8	-5,8	-ND	-4,9
SDZ 17	-6,8	-5,6	-4,0	-5,3	-6,4	-4,6
SDZ 9	-7,7	-6,9	-6,1	-5,3	-7,3	-5,0
TNSA73	-6,4	-6,7	-4,8	-ND	-6,7	-4,5
TNSA74	-7,1	-5,6	-5,7	-ND	-6,8	-4,4
TNSA75	-3,2	-3,0	-2,6	-4,2	-3,8	-2,7
TNSA76	-7,0	-6,8	-5,3	-ND	-6,2	-5,6
TNSA77	-7,2	-5,6	-5,1	-ND	-6,7	-4,6
TNSA78	-7,2	-5,6	-5,1	-5,4	-6,7	-4,6
TNSA79	-6,8	-5,5	-5,3	-ND	-5,2	-3,7
TNSA80	-4,7	-7,2	-4,8	-4,9	-ND	-2,8
TNSA81	-8,1	-5,9	-5,2	-5,4	-6,6	-4,3
TNSA82	-7,2	-6,1	-5,6	-5,6	-5,6	-5,7
TNSA83	-7,2	-6,1	-5,6	-5,6	-5,6	-5,7
TNSA84	-5,2	-8,1	-5,7	-ND	-6,1	-4,1
TNSA85	-5,0	-7,6	-5,4	-ND	-ND	-4,3
Allyl 104	-4,7	-4,3	-4,6	-4,3	-6,0	-3,7
Allyl 106	-2,5	-2,8	-3,1	-2,8	-2,8	-2,0
Allyl 105	-ND	-0,5	-0,6	-0,6	-0,5	-ND

-ND – Not Docking.

(b) The molecular docking bar chart and proteins

Based on the molecular docking binding affinities, the bar chart was plotted in order to visualize the data better. Complementary to the data in table 2, a complete picture of bar chart is shown in figure 3.3, the bar chart of binding affinities in kcal mol^{-1} for the ligands. The docking scores are shown on the y-axis and the ligands on the x-axis. It is observed that the ligand SDZ18 had better binding affinities overall than the other ligands as shown by the long bars which were all above -6.0 kcal mol^{-1}. On the other hand, the ligand Ally 105 had the low binding affinities as shown by the low bars which were less than -1.0 kcal mol^{-1}. It is reasonable to critical establish that the SZD18 is an optimal option for further investigation with more computational techniques as it has frequently occurring good binding affinities. For further analysis, we notice that SDZ15, SDZ9, TNSA74, TNSA80, TNSA81, TNSA82, TNSA84 and TNSA85 had a mixture of results, with fairly good binding affinities with some proteins, as good as -7.0 kcal mol^{-1}, not so good docking scores with other proteins.

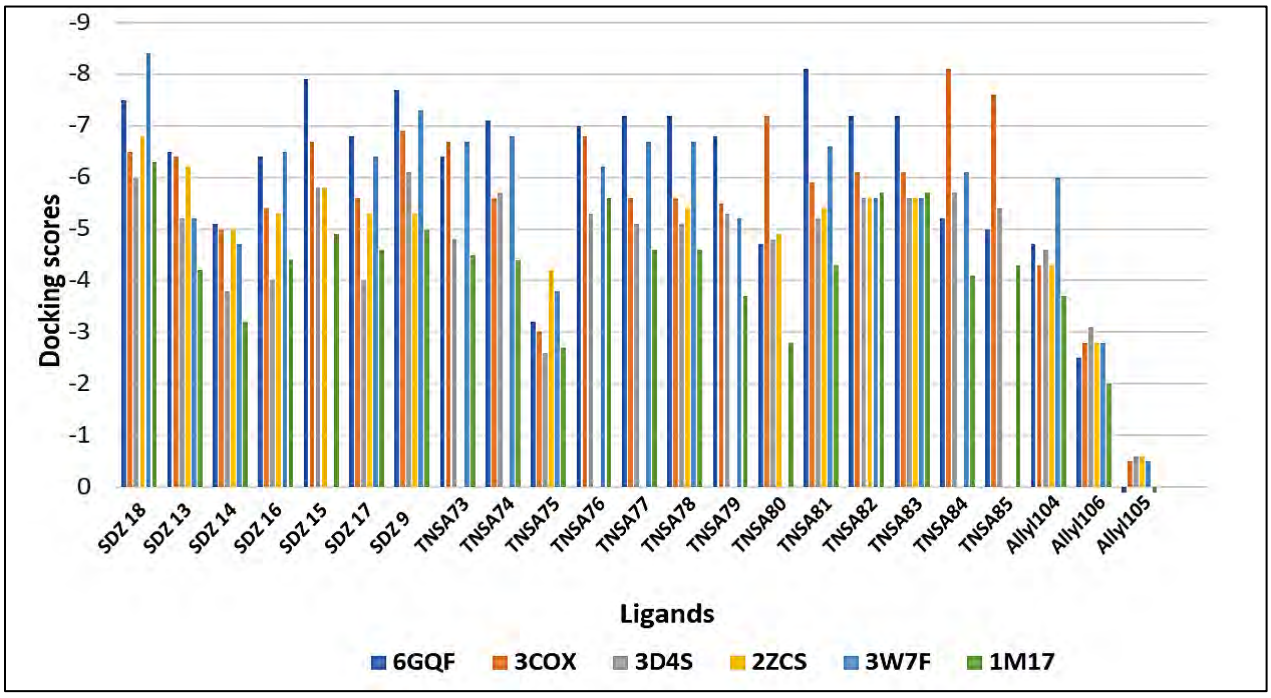


Figure 3.3 Bar chart of binding affinities in kcal mol^{-1} for the ligands.

(b) Protein-ligand interactions

To get the full understanding about the binding strength and types of interactions of ligand with protein residues, the 2D representations were analyzed. Figure 3.4 (a) shows the representation of interaction of SDZ18 with protein ID 2ZCS. The compound SDZ18 had the best binding affinity towards 2ZCS of -6.8 kcal mol^{-1}. The favorable interactions were observed between SDZ18 and protein 2ZCS. The interactions involved pi-pi stacking with PHE26. The protein-ligand diagrams shows that the amino acids that dominated in the active site were the hydrophobic ones (green), preceding ones in dominance were the polar (blue), glycine(yellow) and lastly the positively charged (violet) and negatively charged (brown). The compound had the hydrophobic bonding interaction with amino acid PHE26, had good docking scores.

In figure 3.4 (b), there is intramolecular hydrogen binding interaction between the ligand TNSA84 and LYS296, this might contribute to the stability of the ligand in the pocket and in order to be confident of this assumption we have to employ other techniques. Closer examination of figure 3.5 (a), (b) and figure 3.6 (a) reveals that the favored intramolecular interactions were pi-pi stacking. In figure 3.6 (b) shows no interactions of the ligand SDZ18 and 1M17. The main reason is that the quality of protein was not good in contrast to others as analyzed in the Ramachandran plot. Moreover, it had moderately low docking scores with other ligands. There might be clues that the crystal structures had variations in the docking scores and binding residues due to the differences in the molecular conformations of the enzymes [120]. To sum up, there is certainty that the molecular docking interactions revealed the ligands were successfully stable in the complex. However, further work needs to be done to demonstrate that the ligands stay in the active side of the protein during a simulation.

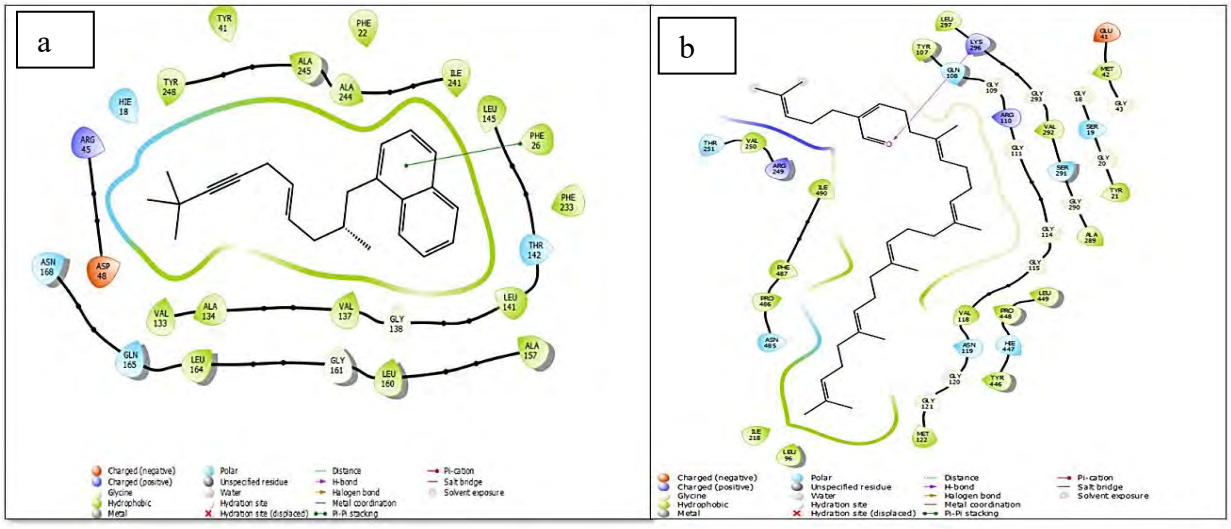


Figure 3.4 The 2D representation of protein interaction for (a) SDZ18 and 2ZCS and (b) The 2D representation of TNSA84 and 3COX.

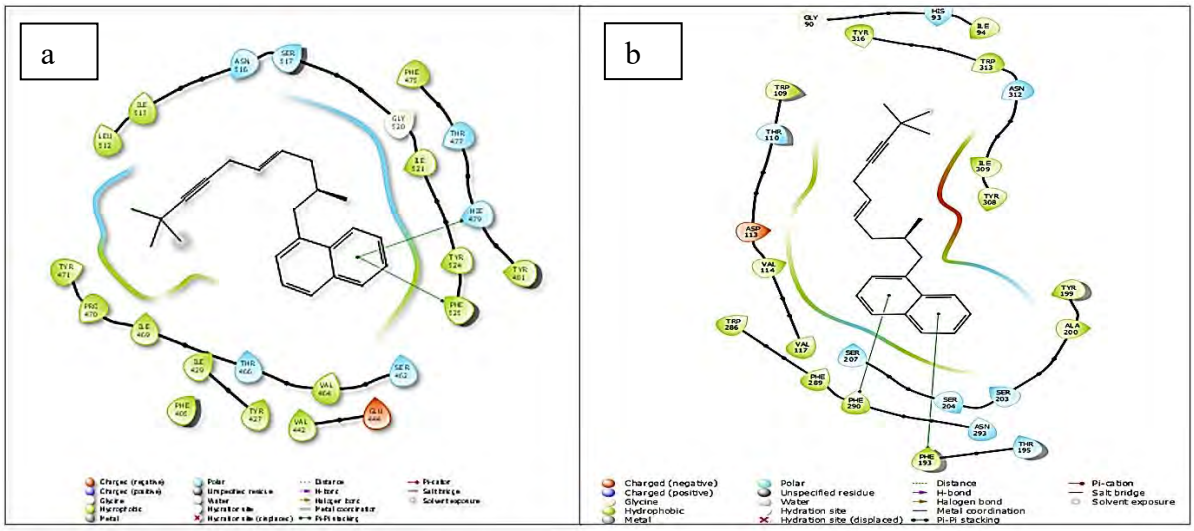


Figure 3.5 The 2D representation of protein interaction for (a) SDZ18 and 6GQF and (b) SDZ18 and 3D4S.

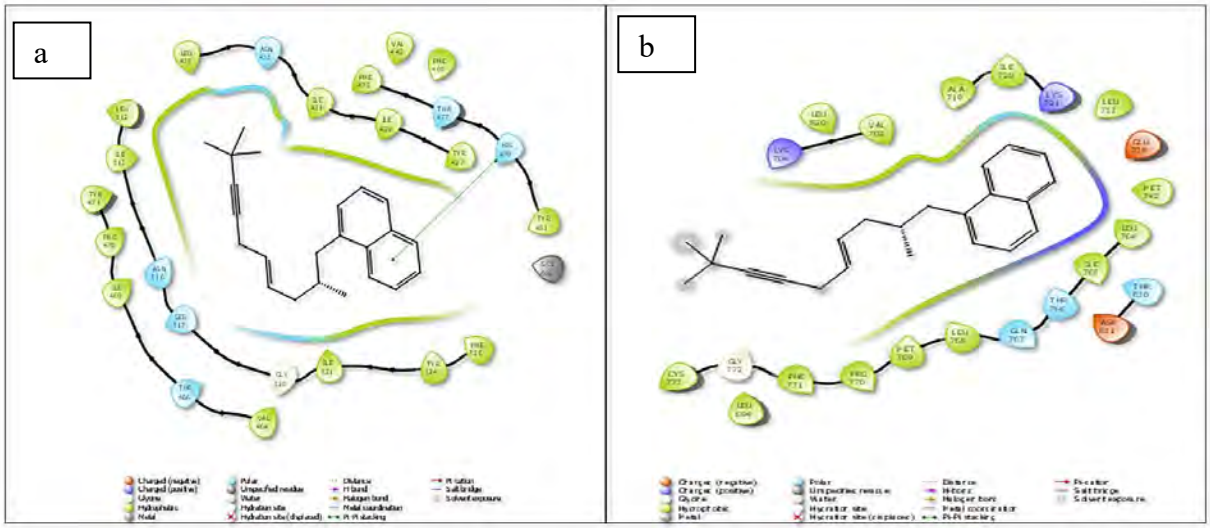


Figure 3.6 The 2D representation of protein interaction for (a) SDZ18 and 3W7F and (b) SDZ18 and 1M17.

(c) The correlation heatmap of the proteins

The correlation analysis is a statistical method which is used to find the strength and direction of the relationship between two variables [121]. It is expressed using the correlation coefficient which varies between -1.0 and +1. A value close to 0 signifies no relationship between the variables while the one close to 1 represents a close relationship [121]. Proteins control the cellular functions in the body [121]. In this study we compute the correlation heatmap in order to find the cholesterol proteins that may have high correlations. In figure 3.7, the heatmap was determined based on the binding affinities for all the ligands and proteins. The x-axis and y-axis shows the six proteins that were used in the study.

A glance at the plot shows several colours, yellow represents high correlations, green represents medium correlations and blue represents low correlations. Strong positive correlations of 0.88 was observed between 3COX and 3D4S, and both proteins were dominated by hydrophobic amino acids in the active side. Protein 3D4S and 6GQF also had high correlation of a 0.86. As compared to the results of the 2D representation previously discussed, the two proteins had similar amino acids in the active side which were polar and hydrophobic aminos. It also appeared that the PHE amino acid was interacting with directly with the ligand with a pi-pi stacking bond. This might be the reasons for the high correlation between 3D4S and 6GQF.

The protein 2ZCS and 3COX had low correlation of 0.09. Both proteins had hydrophobic and polar amino acids in the active side, however the best docked ligands were different for the proteins. Under the circumstances just described, we must admit that this might be the reason for the low correlation. The protein 2ZCS and 6GQF had low correlation of 0.33, 3W7T and 6GQF had a medium correlation of 0.56. These findings indicate that some of our proteins may have similar results as we try to obtain more insight into the behavior of the ligand in the pocket. We may assume that the results will not differ much.

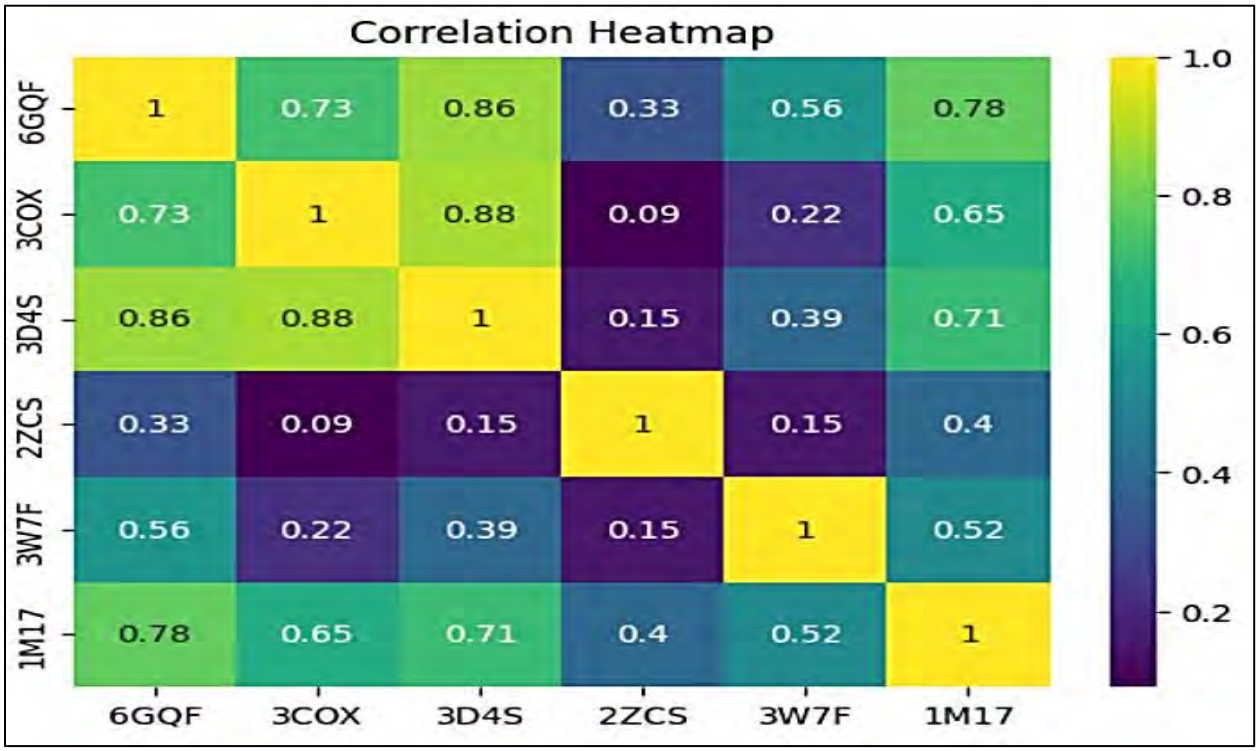


Figure 3.7 The correlation heatmap of the proteins based on the binding affinities.

(d) Protein-ligand 3D docked poses

In this section, figure 3.8 shows illustrate 3D docked poses of the ligands with high binding affinity, TNSA84 and SDZ18. The ligand is in the binding pocket in all the 3D diagrams represented. The ligands form at least one bonded interactions with the key residues which were described section above and also the majority of non-bonded interactions are observable [122].

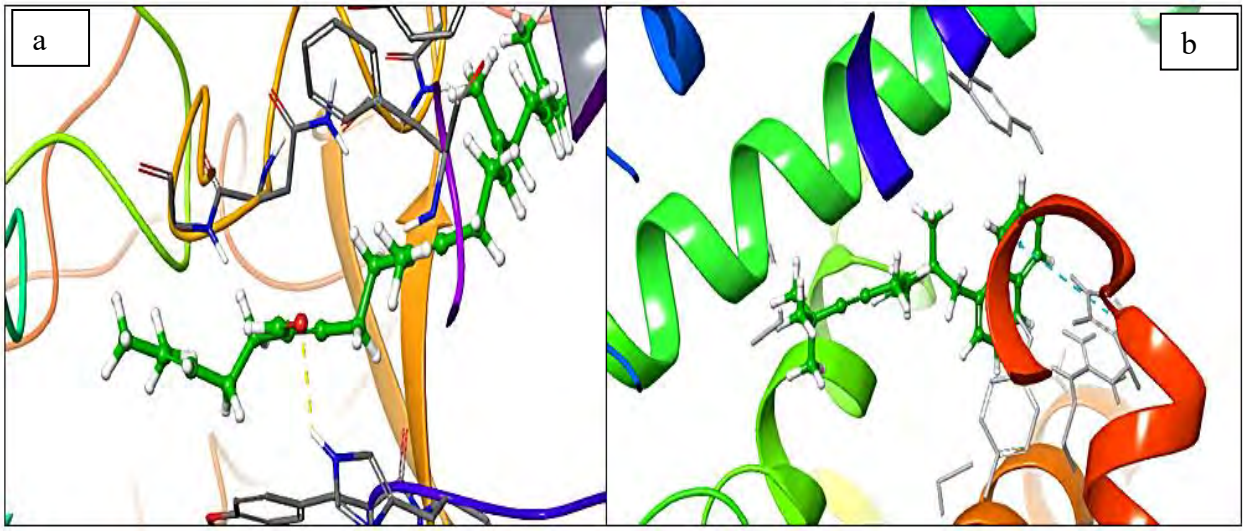


Figure 3.8 Protein-3D docked poses for (a) 3COX protein and TNSA84 (b) 2ZCS protein and SDZ18.

(e) The ligand efficiency and other post docking matric results

Several papers have been published based on the exploration of molecular docking data using ligand efficiency calculations, what makes these papers interesting is that they discuss how heavy atoms and binding affinities correlate. Motivated by the work of Charles H. Reynolds et al, we have also explored deeply into the results of molecular docking [123]. We assembled the same database of 23 ligands with their pIC50 values and target proteins.

Ultimately, drug design is the process aimed at developing new molecules with pharmacological activity, and it is difficult to predict reliable models of action of a drug in the model [71]. Ligand efficiency is used in the assessment of whether a ligand derives its potency from optimal fit with the protein target or simply by the virtue of making contact [123]. To this end, we have calculated post docking matrices for the six proteins which include Ligand efficiency (LE), Ligand lipophilicity efficiency (LLE), ligand efficiency dependent lipophilicity index (LELP), ligand lipophilicity efficiency (LLE), Binding efficiency index (BEI) and surface binding efficiency index (SEI).

(i) The ligand efficiency tables for the proteins

In this section, we have table 3 to table 7 one table for proteins which were calculated. A glimpse at table 3 for 3D4S, the pIC50 values are only shown in this table for the ligands range from 3,96 to 6,96 nM. The heavy atoms ranged from 18 to 36. This is also the same for the other tables. The lowest value for the ligand efficiency was found to be -0,37 in table 4 for 3W7F and the lowest value 0,01 for table 6 for 1M17. However, it is usually more easier to analyse the data obtained from the tables by the visual representation of data.

Table 3 : The ligand efficiencies of 3D4S.

	Ligands	PIC50	HA	LE	LE(pIc50/HA)	LLEP	LLE	BEI	SEI
0	SDZ13	5,41	18,00	-0,29	0,30	-12,84	1,69	0,02	160,49
1	SDZ14	4,31	20,00	-0,19	0,22	-19,65	0,59	0,01	30,47
2	SDZ15	6,43	25,00	-0,23	0,26	-26,49	0,34	0,02	179,01
3	SDZ16	5,66	22,00	-0,18	0,26	-27,36	0,74	0,01	33,25
4	SDZ17	6,58	22,00	-0,26	0,30	-20,59	1,23	0,02	62,84
5	SDZ18	6,49	23,00	-0,26	0,28	-24,63	0,09	0,02	0,00
6	SDZ9	4,30	22,00	-0,28	0,20	-19,97	-1,29	0,02	188,27
7	TNSA73	5,40	28,00	-0,17	0,19	-50,77	-3,23	0,01	23,73
8	TNSA74	4,00	23,00	-0,25	0,17	-27,62	-2,90	0,02	28,18
9	TNSA75	5,39	18,00	-0,17	0,19	-30,46	-1,78	0,01	14,83
10	TNSA76	4,39	29,00	-0,18	0,15	-50,12	-4,62	0,01	26,20
11	TNSA77	5,39	29,00	-0,19	0,19	-48,04	-3,73	0,01	19,01
12	TNSA78	5,39	29,00	-0,18	0,19	-50,71	-3,73	0,01	17,31
13	TNSA79	4,52	28,00	-0,19	0,16	-50,36	-5,05	0,01	0,00
14	TNSA80	5,70	32,00	-0,15	0,18	-64,00	-3,90	0,01	39,90
15	TNSA81	5,27	25,00	-0,21	0,21	-41,39	-3,43	0,01	0,00
16	TNSA82	5,34	29,00	-0,19	0,18	-51,81	-4,50	0,01	0,00
17	TNSA83	5,00	29,00	-0,19	0,17	-51,81	-4,84	0,01	0,00
18	TNSA84	4,15	36,00	-0,16	0,12	-71,94	-7,36	0,01	33,39
19	TNSA85	3,96	36,00	-0,15	0,11	-75,14	-7,31	0,01	20,75
20	Allyl 104	6,96	12,00	-0,38	0,58	-0,42	6,80	0,02	5,72
21	Allyl 105	6,71	9,00	-0,34	0,75	-9,96	3,32	0,02	0,00
22	Allyl 106	6,40	8,00	-0,07	0,80	-39,14	3,66	0,00	0,00

Table 4: The ligand efficiencies of 3W7F.

	Ligands	HA	LE	LE(pIc50/HA)	LLEP	LLE	BEI	SEI
0	SDZ13	18,00	-0,29	0,30	-12,84	1,69	0,02	160,49
1	SDZ14	20,00	-0,24	0,22	-15,55	0,59	0,02	37,69
2	SDZ15	25,00	0,00	0,26	0,00	0,34	0,00	0,00
3	SDZ16	22,00	-0,30	0,26	-16,42	0,74	0,02	54,03
4	SDZ17	22,00	-0,29	0,30	-18,46	1,23	0,02	69,34
5	SDZ18	23,00	-0,37	0,28	-17,31	0,09	0,03	0,00
6	SDZ9	22,00	-0,33	0,20	-16,95	-1,29	0,02	225,31
7	TNSA73	28,00	-0,24	0,19	-35,96	-3,23	0,02	33,12
8	TNSA74	23,00	-0,30	0,17	-23,02	-2,90	0,02	33,61
9	TNSA75	18,00	-0,21	0,19	-24,66	-1,78	0,02	18,78
10	TNSA76	29,00	-0,21	0,15	-42,96	-4,62	0,02	30,65
11	TNSA77	29,00	-0,23	0,19	-39,69	-3,73	0,02	22,74
12	TNSA78	29,00	-0,23	0,19	-39,69	-3,73	0,02	22,74
13	TNSA79	28,00	-0,19	0,16	-50,36	-5,05	0,01	0,00
14	TNSA80	32,00	0,00	0,18	0,00	-3,90	0,00	0,00
15	TNSA81	25,00	-0,26	0,21	-33,43	-3,43	0,02	0,00
16	TNSA82	29,00	-0,19	0,18	-51,81	-4,50	0,01	0,00
17	TNSA83	29,00	-0,19	0,17	-51,81	-4,84	0,01	0,00
18	TNSA84	36,00	-0,17	0,12	-67,71	-7,36	0,01	35,74
19	TNSA85	36,00	0,00	0,11	0,00	-7,31	0,00	0,00
20	Allyl 104	12,00	-0,50	0,58	-0,32	6,80	0,03	7,46
21	Allyl 105	9,00	-0,31	0,75	-10,93	3,32	0,02	0,00
22	Allyl 106	8,00	-0,06	0,80	-45,66	3,66	0,00	0,00

Table 5:The ligand efficiencies of 2ZCS.

	ligands	HA	LE	LE(pIc50/HA)	LLEP	LLE	BEI	SEI
0	SDZ13	18,00	-0,34	0,30	-10,95	1,69	0,03	191,36
1	SDZ14	20,00	-0,25	0,22	-14,93	0,59	0,02	40,10
2	SDZ15	25,00	-0,23	0,26	-26,49	0,34	0,02	179,01
3	SDZ16	22,00	-0,24	0,26	-20,52	0,74	0,02	44,06
4	SDZ17	22,00	-0,24	0,30	-22,30	1,23	0,02	57,42
5	SDZ18	23,00	-0,30	0,28	-21,35	0,09	0,02	0,00
6	SDZ9	22,00	-0,24	0,20	-23,30	-1,29	0,02	163,58
7	TNSA73	28,00	0,00	0,19	0,00	-3,23	0,00	0,00
8	TNSA74	23,00	0,00	0,17	0,00	-2,90	0,00	0,00
9	TNSA75	18,00	-0,23	0,19	-22,51	-1,78	0,02	20,76
10	TNSA76	29,00	0,00	0,15	0,00	-4,62	0,00	0,00
11	TNSA77	29,00	0,00	0,19	0,00	-3,73	0,00	0,00
12	TNSA78	29,00	-0,19	0,19	-48,04	-3,73	0,01	18,33
13	TNSA79	28,00	0,00	0,16	0,00	-5,05	0,00	0,00
14	TNSA80	32,00	-0,15	0,18	-64,00	-3,90	0,01	40,73
15	TNSA81	25,00	-0,22	0,21	-39,51	-3,43	0,02	0,00
16	TNSA82	29,00	-0,19	0,18	-51,81	-4,50	0,01	0,00
17	TNSA83	29,00	-0,19	0,17	-51,81	-4,84	0,01	0,00
18	TNSA84	36,00	0,00	0,12	0,00	-7,36	0,00	0,00
19	TNSA85	36,00	0,00	0,11	0,00	-7,31	0,00	0,00
20	Allyl 104	12,00	-0,36	0,58	-0,45	6,80	0,02	5,35
21	Allyl 105	9,00	-0,31	0,75	-10,93	3,32	0,02	0,00
22	Allyl 106	8,00	-0,07	0,80	-39,14	3,66	0,00	0,00

Table 6: The ligand efficiencies of 1M17.

	ligands	HA	LE	LE(pIc50/HA)	LLEP	LLE	BEI	SEI
0	SDZ13	18,00	-0,23	0,30	-16,19	1,69	0,02	129,63
1	SDZ14	20,00	-0,16	0,22	-23,33	0,59	0,01	25,66
2	SDZ15	25,00	-0,20	0,26	-30,46	0,34	0,01	151,23
3	SDZ16	22,00	-0,20	0,26	-24,63	0,74	0,02	36,58
4	SDZ17	22,00	-0,20	0,30	-26,76	1,23	0,02	47,67
5	SDZ18	23,00	-0,27	0,28	-23,72	0,09	0,02	0,00
6	SDZ9	22,00	-0,23	0,20	-24,31	-1,29	0,02	154,32
7	TNSA73	28,00	-0,16	0,19	-53,94	-3,23	0,01	22,24
8	TNSA74	23,00	-0,19	0,17	-36,34	-2,90	0,01	21,75
9	TNSA75	18,00	-0,15	0,19	-34,52	-1,78	0,01	13,35
10	TNSA76	29,00	-0,19	0,15	-47,48	-4,62	0,01	27,68
11	TNSA77	29,00	-0,16	0,19	-57,05	-3,73	0,01	15,61
12	TNSA78	29,00	-0,16	0,19	-57,05	-3,73	0,01	15,61
13	TNSA79	28,00	-0,13	0,16	-73,60	-5,05	0,01	0,00
14	TNSA80	32,00	-0,09	0,18	-106,66	-3,90	0,01	23,28
15	TNSA81	25,00	-0,17	0,21	-51,13	-3,43	0,01	0,00
16	TNSA82	29,00	-0,20	0,18	-49,22	-4,50	0,01	0,00
17	TNSA83	29,00	-0,20	0,17	-49,22	-4,84	0,01	0,00
18	TNSA84	36,00	-0,11	0,12	-104,64	-7,36	0,01	24,02
19	TNSA85	36,00	-0,12	0,11	-93,92	-7,31	0,01	16,53
20	Allyl 104	12,00	-0,31	0,58	-0,52	6,80	0,02	4,60
21	Allyl 105	9,00	-0,22	0,75	-15,40	3,32	0,01	0,00
22	Allyl 106	8,00	0,01	0,80	273,98	3,66	0,00	0,00

Table 7 : The ligand efficiencies of 6GQF.

	ligands	HA	LE	LE(pIc50/HA)	LLEP	LLE	BEI	SEI
0	SDZ13	18,00	-0,36	0,30	-10,34	1,69	0,03	200,62
1	SDZ14	20,00	-0,26	0,22	-14,36	0,59	0,02	40,90
2	SDZ15	25,00	-0,32	0,26	-19,04	0,34	0,02	243,83
3	SDZ16	22,00	-0,29	0,26	-16,98	0,74	0,02	53,20
4	SDZ17	22,00	-0,31	0,30	-17,27	1,23	0,02	73,67
5	SDZ18	23,00	-0,33	0,28	-19,41	0,09	0,02	0,00
6	SDZ9	22,00	-0,35	0,20	-15,98	-1,29	0,02	237,65
7	TNSA73	28,00	-0,23	0,19	-37,53	-3,23	0,02	31,64
8	TNSA74	23,00	-0,31	0,17	-22,27	-2,90	0,02	35,10
9	TNSA75	18,00	-0,18	0,19	-28,77	-1,78	0,01	15,82
10	TNSA76	29,00	-0,24	0,15	-37,59	-4,62	0,02	34,60
11	TNSA77	29,00	-0,25	0,19	-36,51	-3,73	0,02	24,44
12	TNSA78	29,00	-0,25	0,19	-36,51	-3,73	0,02	24,44
13	TNSA79	28,00	-0,24	0,16	-39,87	-5,05	0,02	0,00
14	TNSA80	32,00	-0,15	0,18	-64,00	-3,90	0,01	39,07
15	TNSA81	25,00	-0,32	0,21	-27,16	-3,43	0,02	0,00
16	TNSA82	29,00	-0,25	0,18	-39,37	-4,50	0,02	0,00
17	TNSA83	29,00	-0,25	0,17	-39,37	-4,84	0,02	0,00
18	TNSA84	36,00	-0,14	0,12	-82,22	-7,36	0,01	30,46
19	TNSA85	36,00	-0,14	0,11	-80,50	-7,31	0,01	19,22
20	Allyl 104	12,00	-0,39	0,58	-0,41	6,80	0,02	5,85
21	Allyl 105	9,00	-0,28	0,75	-12,10	3,32	0,01	0,00
22	Allyl 106	8,00	0,01	0,80	273,98	3,66	0,00	0,00

(ii) The ligand efficiency plots for the proteins

The question to be addressed in this section is does the ligand binding affinities show a strong influence due to molecular size? To address these question, the number of heavy atoms was utilized to calculate the ligand efficiencies using the equations which were previously elaborated on chapter 2, section 2.1. To find the answers of the ligand efficiency calculations and the trends, we have analyzed the plots. The initial aim was to provide more insights into the quality of the compounds being developed as potential drug candidates for lowering high cholesterol. After further work using Jupiter notebook service on google colab, the non-binders which were

previously assigned 0 where eliminated and the plots were generated. Since the intent was to focus on the binders only, the non-binders were completely omitted.

In particular, analysis of figure 3.9 (a) shows the plot of Ligand efficiency (LE) against the heavy atoms (HA) and the legend shows different protein targets. The LE falls dramatically when looking HA between 5 and 15 , then there is a minor increase throughout for all the proteins as heavy atoms increase. In the protein active site, the small ligands have surface area which is available for making favorable contact with the proteins because it is much large on per atom basis than the surface area of the large ligands [123]. The observations drawn from the figure 3.9 (d) are that the LE was much higher around 10 heavy atoms than at 30 heavy atoms.

Perhaps it is easier to see the concept on figure 3.9 (b) where the LE was calculated based on the pIC50 ("LE (pIC50/HA)") and heavy atoms then plotted against the heavy atoms. This was another way to illustrate the phenomena described above. In this plot, also there is a sharp decline in the LE as the number of heavy atoms increase. The LE was 0.8 for heavy atoms of 8 and 0.1 for heavy atoms of 40. A further illustration is shown figure 3.9 (c), in the plot of LLEP against HA (heavy atoms). The R-squared calculated were as follows for data of each protein 1M17 : -0,723, 3COX : -0,687, 6GQF : -0.681, 2ZCS : -0,696, 3D4S : -0,824 and 3W7F : -0,672.

Moreover, to gain more confidence the correlation coefficient was calculated in figure 3.9 (d) and determined to be -0.909. However, we would like to assert that the data presented here is fairly limited due to the fact that affinities can vary widely for a given heavy atom since it is possible to have ligands of any size that have low binding affinities simply because they do not have correct geometry or functionality that bind to the active site [123]. And as one may expect, they are some outliners on the plots especially in figure 3.9 (e).

Analyzing the impact of lipophilic efficiency on compound quality, Ákos Tarcsy et al. conducted an elegant study to explore the plot of LLE and LELP in which was increasing and they concluded that it suggest balanced optimization of potency and ADMET profile [124]. However, the plot neglects the effect of ligand size and in this section other alternative metrics are used [124]. On figure 3.9 (e), it is apparent that there is an increase in trend in the plot of LLE vs LLEP. In fact, the latter has the same increasing plot as illustrated by Ákos Tarcsy and other authors for their research work [124].

In 2007, Charles H Reynold et al attempted to show this decreasing trend from smaller ligands to larger ligands in hopes of finding the best prospective drug [125]. Despite that, it is usually expected that larger, more complex ligands to have less optimal binding than the smaller less complicated ligands [123]. Our plots are in agreement with the plots discussed by Charles H Reynold. According to I.D. Kuntz, careful examination of the best ligands survey done experimentally showed that the strongest binding ligands indicates that the free energy of the binding increases with the nonhydrogen atoms with an initial slope of -1.5 kcal $mol^{-1}$and the free energy of binding increases with little relative molecular mass [126]. In conclusion, the ligand efficiencies for smaller ligands are much higher than the large ligands due to the available surface area [105], [123].

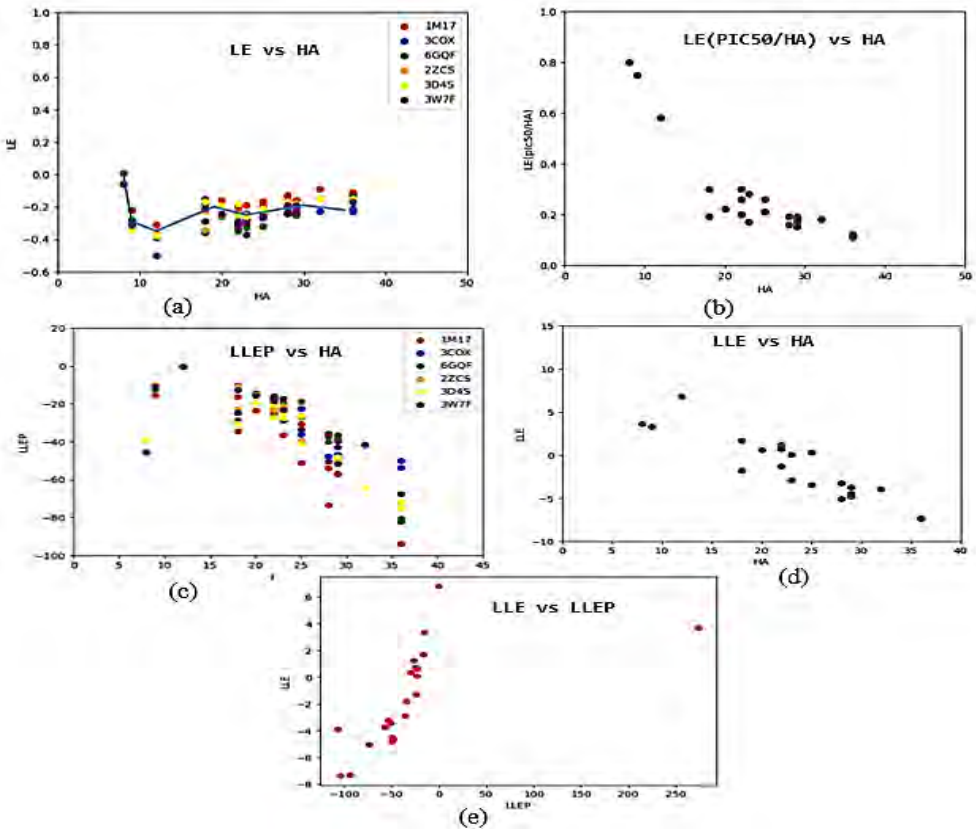


Figure 3.9 The ligand efficiency plots for the proteins.

(f) Molecular docking validation

There are various methods for validating docking programs and scoring functions. The most commonly used method is the pose selection, whereby the docking programs are used to redock the native ligand from the protein into the active site with a known conformation and orientation [127]. The results obtained showed that the molecular docking was valid collectively for all the results as the native ligand and the best ligand obtained from the results superimposed on one another in the active site. The figure 3.10 (a), shows the ligand TNSA84 superimposed on to the native ligand in the active site of the protein. Somewhat interestingly, we observe that the ligand TNSA84 was in the binding pocket and entirely superimposed on the native ligand of the protein. In similar vein, we also observe that the figure 3.10 (b) SDZ 18 ligand was also superimposed on the native ligand, thus showing the validation of molecular docking.

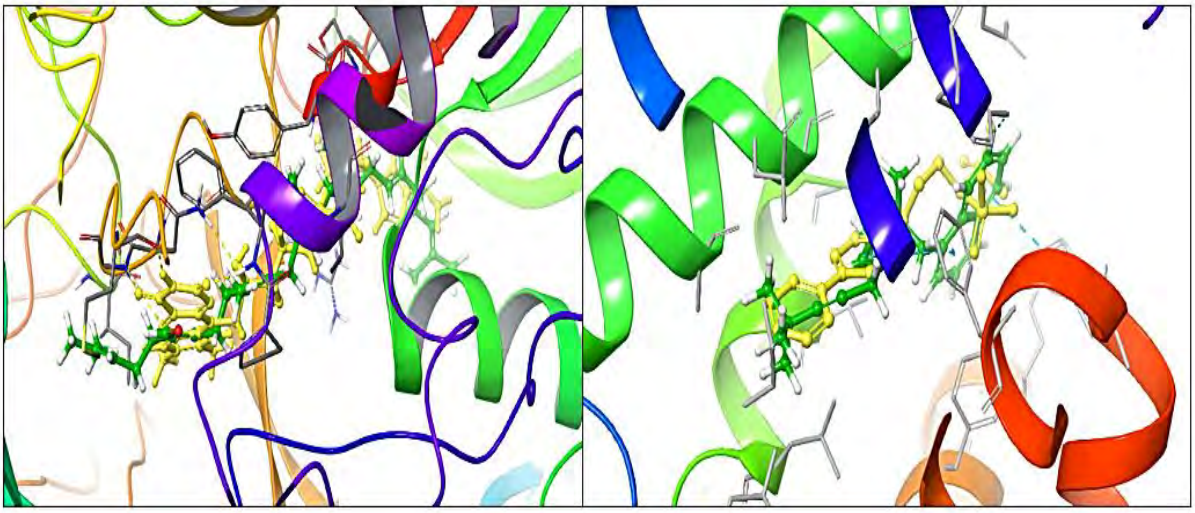


Figure 3.10 The illustration of the native ligand and docked ligand in the active site of protein. The green ligand is the docked ligand and the yellow ligand is the native ligand of protein. (a) 3COX protein and TNSA84 (b) 2ZCS protein and SDZ18.

3.4 Molecular mechanics-generalized born surface area (MM-GBSA)

This is a very short section for the calculation of free energy with a modest computational effort using the Molecular Mechanics-Generalized Born Surface Area (MM-GBSA). The binding free energy calculation is relevant in drug discovery processes because the ability of a ligand to elicit a therapeutic effect is dependent on the degree of the binding to the target [47], [128]. Moreover, the binding free energy is evaluated as the difference between maximum and minimum values of free energy change which is determined by using the potential of mean force (PMF) [47]. Thus, to correctly validate the docking process with free energy calculation, it is worth the investment to employ other techniques in order to investigate more aspects of the complex. This reports the results of the computation of free energy of the docked complex and other properties.

(a) Results

The components of the total binding free energy such as ΔG_{Bind}, $\Delta G_{Bind\ Coulomb}$,$\Delta G_{Bind\ Covalent}$, $\Delta G_{Bind\ Hbond}$,$\Delta G_{Bind\ Packing}$, $\Delta G_{Bind\ vdW}$, $\Delta G_{Bind\ Solv\ GB}$, $\Delta G_{Bind\ Lipo}$ were computed in the MM-GBSA method. Among the types of interactions $\Delta G_{Bind\ vdW}$, $\Delta G_{Bind\ Lipo}$, $\Delta G_{Bind\ Coulomb}$ and $\Delta G_{Bind\ Lipo}$ contribute more to the average binding energy and contrastingly, sx$\Delta G_{Bind\ Solv\ GB}$ and ΔG_{Bind} covalent energies contribute the lowest to attain the final average binding energies [129]. In fact, free energy (ΔG) is the amount of energy released or required during the chemical process [66]. The negative ΔG value is associated with a spontaneous process, the ligand binding to its target while the positive ΔG value represents a non-spontaneous process. Generally, the free energy describes the thermodynamics and kinetic properties of a system, in this case it is the ligand binding to a protein in the presence of a solvent [56].

More importantly, the thermodynamic free energy is the energy in a physical system that can be converted to do work. In general, the ligand that binds more strongly to the protein compared to the others has more negative free energy of binding than the others. This further indicates that the ligand is likely to remain stable within the binding pocket of the protein even during simulation. The free energy is called the Gibbs free energy in a protein-solvent system [54].

The ΔG_{Bind} determines the binding strength [128]. Not surprisingly, it is observed that in table 4 the ligand TNSA 84 docked on the 3COX protein had more strength with a value of -148.63 kcal mol^{-1} than the other ones while the SDZ18 docked on the 3W7F has the lowest strength of -41,85 kcal mol^{-1}. Previously, in section 3.3, figure 3.4 there was a hydrogen bond interaction between the ligand TNSA 84 and the protein amino acid LYS 296 and we assume that might be he reason for high ΔG_{Bind} value. To further clarify the impact of the TNSA 84 to protein 3COX, the calculated $\Delta G_{Bind\ Hbond}$ was -0,05 while it is 0,00 for the rest of proteins. A final point of interest is that these results are consistent with our early model computations.

The most notable difference between the ligands is that the free energy of the binding from the coulomb ($\Delta G_{Bind\ Coulomb}$) was observed to be more on the SDZ18 docked on 2ZCS with a value of -4,48 while 3D4S had the lowest value of 0,75. The $\Delta G_{Bind\ Solv\ GB}$ represents the solvation effect which is responsible for reactions more favorable both kinetically and thermodynamically [73], [76]. To understand and gain insights into the reactivity of a system, it is important to study the solvation effect. The protein 3COX had the highest value of 43,92 relative to the others.

The $\Delta G_{Bind\ vdW}$ and $\Delta G_{Bind\ Lipo}$ were noted to be higher for the protein 3COX while lower for protein 1M17. Based on previous analysis, the ligand SDZ 18 docked on protein 1M17 did show any interactions with the amino acids, this might be the reason for lower values. In summary, the findings in table 4 shows that the ligand TNSA 84 docked on protein 3COX had more higher values than the other ones. However, looking at the future it would be satisfactory for the system to agree with the other computations in the investigation for stability of the ligand in the binding pocket in a non-stationary environment.

Table 4 : The binding free energy analysis using MM-GBSA approach calculated in kcal mol^{-1}.

Components (kcal mol^{-1})	**SDZ18 docked on 6GQF**	**TNSA84 docked on 3COX**	**SDZ87469 docked on 3D4S**	**SDZ18 Docked on 2ZCS**	**SDZ18 Docked on 3W7F**	**SDZ18 docked on 1M17**
ΔG_{Bind}	-82,36	-148,63	-44,08	-21,29	-41,85	-44,21
$\Delta G_{Bind\ Coulomb}$	-4,48	-1,89	0,75	-5,85	-3,47	-3,12
$\Delta G_{Bind\ Covalent}$	5,24	10,38	3,60	17,09	3,26	0,64
$\Delta G_{Bind\ Hbond}$	0,00	-0,05	0,00	0,00	0,00	0,00
$\Delta G_{Bind\ Lipo}$	-58,78	-124,54	-31,94	-36,37	-30,19	-21,67
$\Delta G_{Bind\ vdW}$	-49,72	-76,45	-45,17	-40,99	-44,78	-40,50
$\Delta G_{Bind\ packing}$	-2,31	0,00	-3,38	-2,13	-0,27	0,00
$\Delta G_{Bind\ Solv\ GB}$	27,69	43,92	32,07	30.94	33,61	20,44
Ligand Strain Energy	6,03	-16,63	11,69	-2,94	3,97	4,57

3.5 Summary of the chapter

In summary, the results obtained in these chapters indicates that the antimycotic compounds namely SDZ18 and the anticholesterolemic compounds namely TNSA84 had the best binding energies. The study of the correlation of the proteins utilized for the study shows that 3D4S, 3COX and 6GQF have high correlations above 0.85. From the ligand efficiency plots, it was concluded that the ligand efficiencies are lower for larger ligands and vice versa due to the atomic surface area. On the other hand, the garlic compounds showed no docking potential at all while the others such as Allyl compounds only had bad docking scores. Hence, we did not give more insight into these compounds as they failed the first test. The best binders gave clues that the development of new drugs could be made from them, thus a starting point in the field of drug discovery was achieved.

The discussion would not be complete without mentioning that these results must be compared with the ones performed by Molecular mechanics Poisson-Boltzmann and generalized born surface area (MM-PBSA) for better accuracy. However, the work of Huiyong Sun and colleagues have examined that both methodologies achieve approximately equal accuracies at the interior dielectric constant of 4 (r_p= 0.408 $\pm$ 0.006 of MM-GBSA and r_p= 0.388 $\pm$ 0.006 of MM-PBSA) [76]. These methods are alternative approaches to improving the scoring

function errors in molecular docking and also for calculation of free energy. In summary, the MM-GBSA method was to calculate the free energy of the ligands in the binding pockets. This method was successfully implemented as we have obtained the results that remarkable free energies. These further indicates the future work to be done is MM-PSA on these ligands to investigate if there is a good agreement between the results. Thus, the most accurate results will be obtained.

Despite the good results obtained, the combined computational approaches need to be used that will give us further insights into the compounds. Disappointingly, molecular docking also assumes that the docked complex is stationary, and this violates the universal law of vibration. Next, we attempted to eliminate errors that might have occurred due to scoring functions of molecular docking and the nature of the system being stationary by computationally simulating our systems in a solvated dynamic environment.

Chapter 4 : MOLECULAR DYNAMICS (MD) SIMULATIONS

Introduction to the chapter

In this chapter, the molecules that had the best binding affinity were investigated with Molecular dynamics (MD) simulations. MD was done to investigate whether the best docked ligands remain stable in the binding pocket during the 100 ns simulations of the system at constant pressure and temperature while maintaining a fixed number of atoms. In the field of drug design, a realistic model is an objective that is highly desired when the computational chemistry techniques are employed. Molecular dynamics (MD) is a useful tool for investigating the stability of the ligand in the binding pocket of protein and with the inclusion of heat, pressure and water molecules to the system.

MD simulations are run on the model which resemble the conditions closer to physiological environment to investigate that the protein-ligand system can withstand the fiery principle of the human body. Therefore, 300k Temperature and pressure of 1.01325 bar is applied to the MD simulation box (a more detailed computational approach is given in chapter 2, section 2.3). The OPLS_2005 force field was utilized for the description of the interaction between atoms of the system regarding the covalent bound atoms and non-bonded parameters [130].

4.1 The Protein and Ligand RMSD results

The Root Mean Square Deviation (RMSD) is a statistic measure that is used in molecular dynamics usually to measure the average change in the displacement of a selection of atoms for a specific frame regarding a reference frame, calculated for all the frames in the trajectory [82], [41]. The RMSD for frame x is shown by the equation below where N is the number of atoms in the atom selection, t_{ref} refers to reference time(usually the first frame is used as the reference and it is considered as being at time t=0) and r' is the position of the selected atoms in frame x after superimposing on the reference frame, where frame x is recorded at time t_x [41], [82]. In order to elucidate the stability of the system during the course of a simulation, the

root mean square deviation (RMSD) equation (equation 14) below is used to calculate an RMSD value for every frame in the simulation trajectory.

Equation 14 : The root mean square deviation (RMSD)

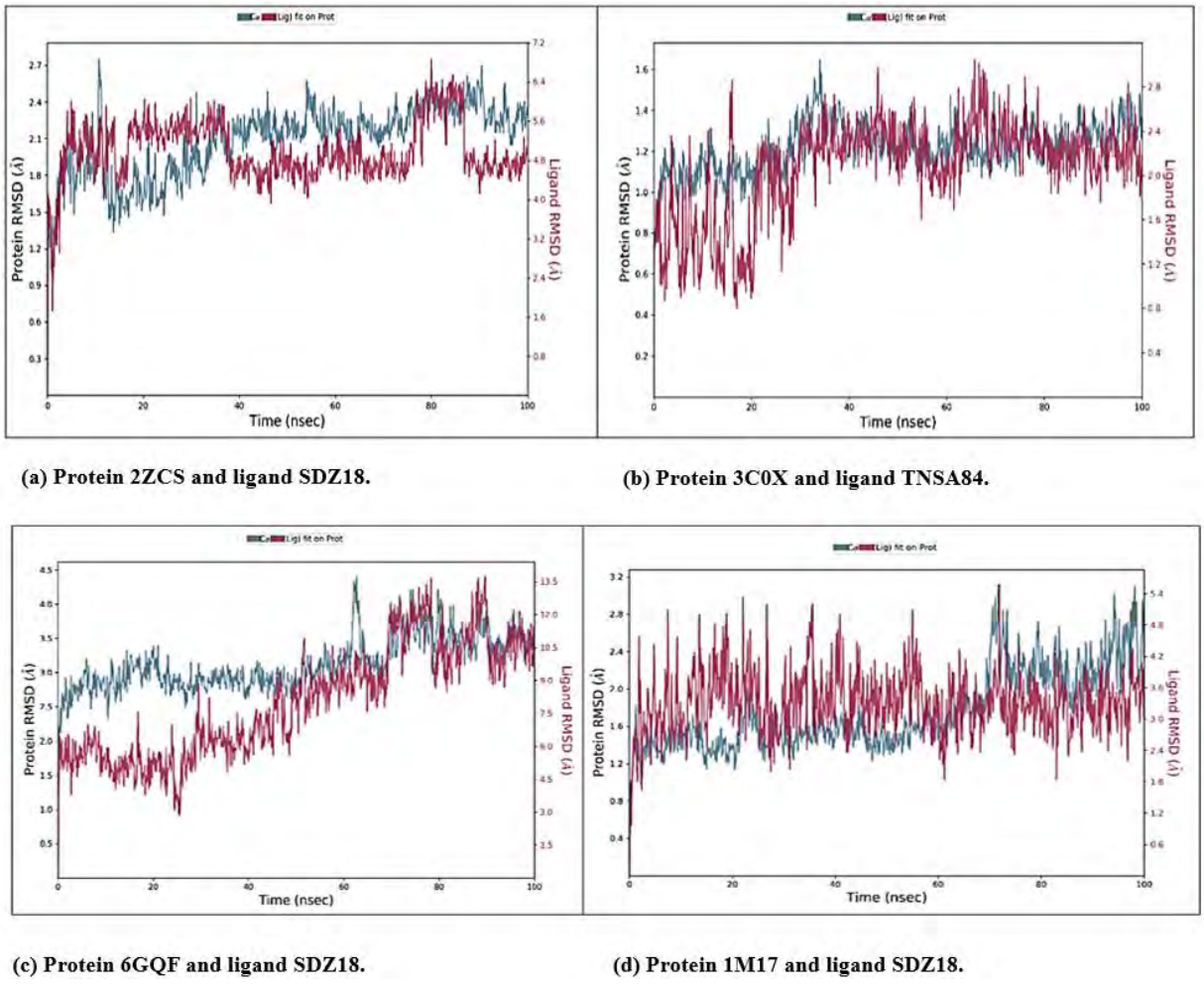
$$RMSD_X = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i'(t_x)) - r_i(t_{ref}))^2}$$

The analysis of the RMSD of the protein Cα atoms gave the insights into the structural conformation from the beginning to the end of the simulation and it provided the information about the equilibration of the system. The RMSD was used to estimate the Cα (alpha carbon) atoms of the protein residues have indicated the degree of protein stability [81]. The Cα RMSD plots were analyzed to check the stability of the ligand in the binding pocket of the protein. The t is often used to identify whether a structure is stable in a timescale of the simulations or if it is diverging from the initial coordinate [41], [131]. In addition, particular emphasis is that if the RMSD plot is still changing and not stable towards the end of the simulation, this suggests that a longer simulation is required. Notably, the results show that the 100 ns simulation was the appropriate one for the complexes.

The divergence from the initial coordinates is interpreted as a sign that the simulation has not equilibrated, and it is observed that in figure 4.1 (a) 2ZCS, there was minor divergences throughout the simulation, but stability was achieved around 2.4 Å. The plots measure the scale distance between the ligand and the protein through-out the trajectory. Moreover, particular significance is that the plots measure the scale distance between the ligand and the protein through-out the trajectory. By inspection, the RMSD evolution of a protein shown on the left Y-axis assumes a stability at a range of 0.6 to 2.7 Å for 100 ns simulation. The analysis show if the simulation is at equilibrium, the fluctuations towards the end of the simulation are around some thermal average structure. Notably, in figure 4.1 (b) 3COX, the stability of the ligand in the binding pocket of protein has equilibrated when the fluctuation occurs around a stable average conformation which is below 1.6 Å. It has the lowest RMSD in comparison to the other plots, it is at this junction that we may assume that the hydrogen bond interaction with amino acid LSY 296 which was previously observed in chapter 3, section 3.3, figure 3.4 contributed to the stability. In figure 4.1 (c) 6GQF, there was a significant drifting of the ligand in the

binding pocket for the first 50 ns as it is also clearly by the distance between the two plots and during the last 50ns the ligand finally stabilized around 4.0 Å.

Generally, the majority of the average RMSD values calculated were below the acceptable value as they were less than 3.6 Å with exceptions to figure 4.1 (c) which stabilized slightly above 3.6 Å. In Figure 4.1 (d) shows that the stability was reached at 0.6 to 2.4 Å with some minor fluctuations to 2.7 Å at 80-90ns. The RMSD plot show a stability below 2.7 Å, this indicates that the system has equilibrated , and simulation is long enough for rigorous analysis. To sum up, the distance between the ligand and protein further shows that it did not leave the active site despite the minor fluctuations that are observed throughout the simulation.



(a) Protein 2ZCS and ligand SDZ18.

(b) Protein 3C0X and ligand TNSA84.

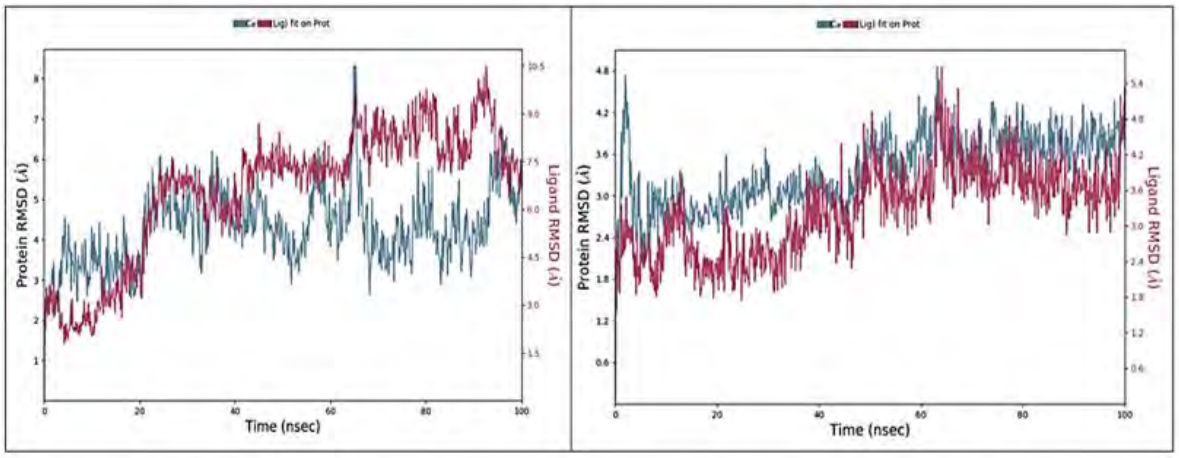
(c) Protein 6GQF and ligand SDZ18.

(d) Protein 1M17 and ligand SDZ18.

Figure 4.1 The RMSD plots of the proteins for 100 ns simulations.

Unexpectedly, figure 4.2 (a) shows that the ligand stabilized for the first 20 ns and the plot goes above 4.0 Å and this indicates that the results were not good. The increase in RMSD values

during the simulation shows that the molecule has escaped the bind pocket of the protein. In fact, we cannot always have a purely polarized system in science as we will always have results with some errors. Previously in chapter 3, section 3.2, in figure 3.5 we investigated the Ramachandran plots and the one obtained for protein 3W7F had minor points in regions not allowed. This, in turn might be the main reason for the results obtained. Convincingly, one can see that the systems had low values of RMS and it was observed that they remained stable during the 100 ns simulation. Our inspection indicates that the analysis of the RMSD of the protein gave the insights into the structural conformation from the beginning to the end of the simulation and also it provided the information about the equilibration of the system. The results obtained in this section are in agreement with the previous observations as they show that the ligand TNSA 84 is more stable in the binding pocket of protein 3COX with remarkable RMSD plot.



(a) Protein 3W7F and ligand SDZ18. **(b) Protein 3D4S and ligand SDZ18.**

Figure 4.2 The RMSD plots of the proteins for 100 ns simulations.

4.2 P-RMSF calculations

The Protein Root Mean Square fluctuation (P-RMSF) is a measure of the displace of a particular atom or group of atoms, relative to the reference structure, averaged over the number of atoms [107], [41], [82]. The RMSF for residue is shown by the equation 15.

Equation 15 : The protein root mean square fluctuation (P-RMSF)

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

where T is the trajectory time over which the RMSF is calculated, t_{ref} is the reference time, r_i is the position of residue I, r' is the position of atoms in residue I after the superposition on the reference [41], [82]. The angle of the brackets indicate that the average of the square distance taken over the selection of atoms in the residue [41], [82].

The investigation of the RMSF plot provides the insights into the dynamic residue behavior within the active side of the protein and the flexibility of the amino acids' residues upon binding to the ligand. Given the nature of assembly of the RMSF plot, it shows the areas of the protein that fluctuate more during the simulation. The x-axis shows the index number of the residues.

Generally, monitoring the tails (N- and C-terminal) fluctuate more than any other part of the protein. The alpha helices and beta strands are the secondary structure elements which are anticipated to be more rigid than the unstructured regions of protein [41]. The secondary structure regions indicated by pink colour have a low value of RMSF while the unstructured regions have higher values of RMSF. The green lines indicate the protein residues that are interacting more with the ligand.

In figure 4.3 (a), the C-alpha of protein 2ZCS peaks indicate the areas that fluctuate the most. The noteworthy observation is that amino acid ASP had less interaction with the ligand as it fluctuated around 3.6 Å. To validate this, we herein look previously at molecular docking in chapter 3, section 3.3, figure 3.4 that the amino ASP 48 had no direct interaction with the ligand. However, in overall, the protein and ligand were interacting as it is also shown by the green lines and majority of fluctuations below 2.4 Å.

Interestingly, in figure 4.3 (b) shows the ligand and the protein 1M17 are not interacting because the green lines were not observed. In similar vein, the 2D interaction diagram of molecular docking observed in chapter 3 also showed that the ligand had no interactions with the amino acids. Thus, the molecular details of these events validate the results observed in molecular docking, the ligand SDZ18 and 1M17 protein showed no interaction at all. As expected, in figure 4.3 (a) it is observed that the protein 3W7F was fluctuating more as it has many peaks above 2.4 Å in contrast to figure 4.3 (a), 2ZCS. In this context, it is worth

remembering that the Ramachandran plots showed that the protein 2ZCS was better in quality than 3W7F. Nearly identical results were obtained except for figure 4.3 (b) and (d).

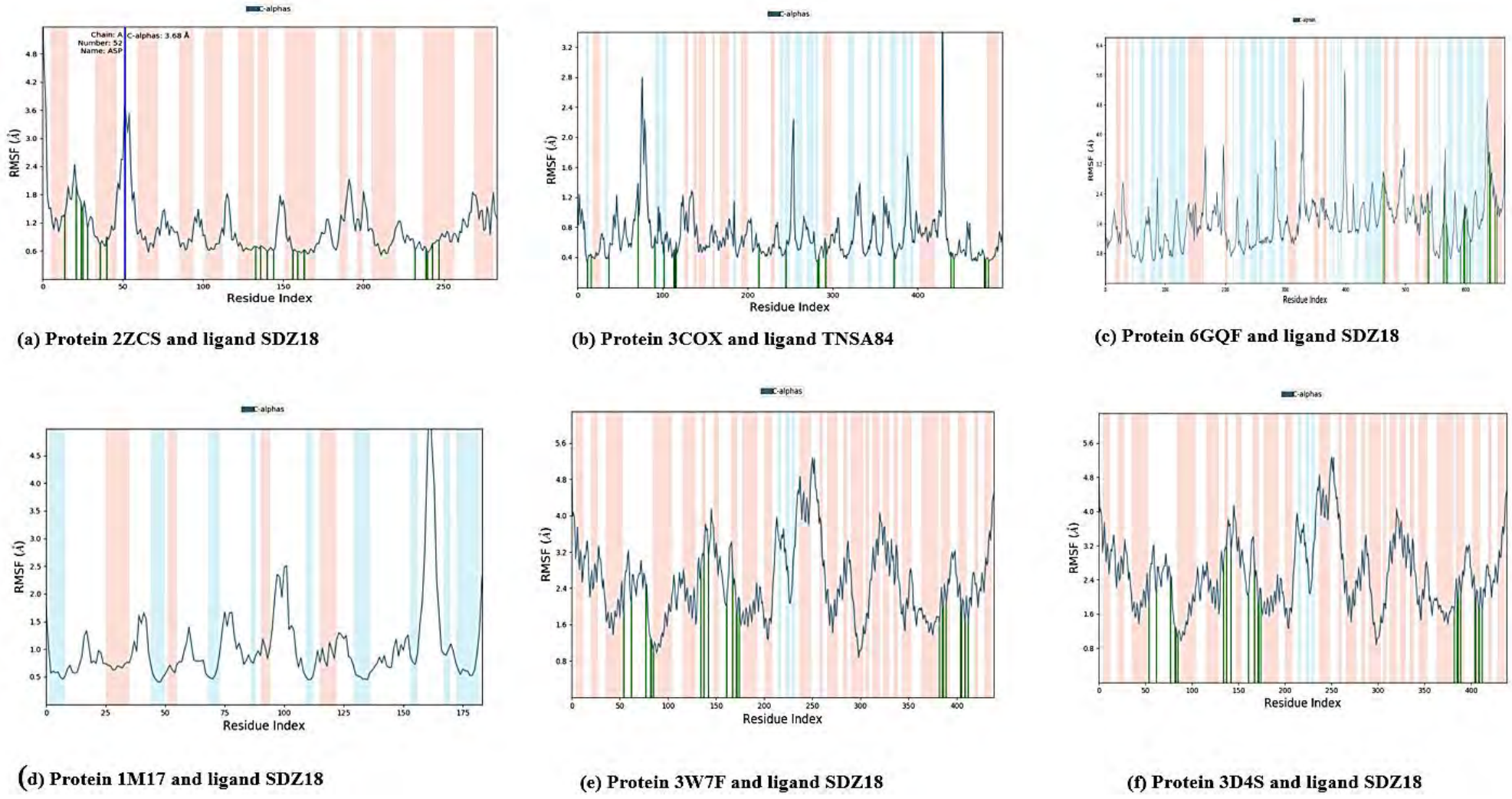
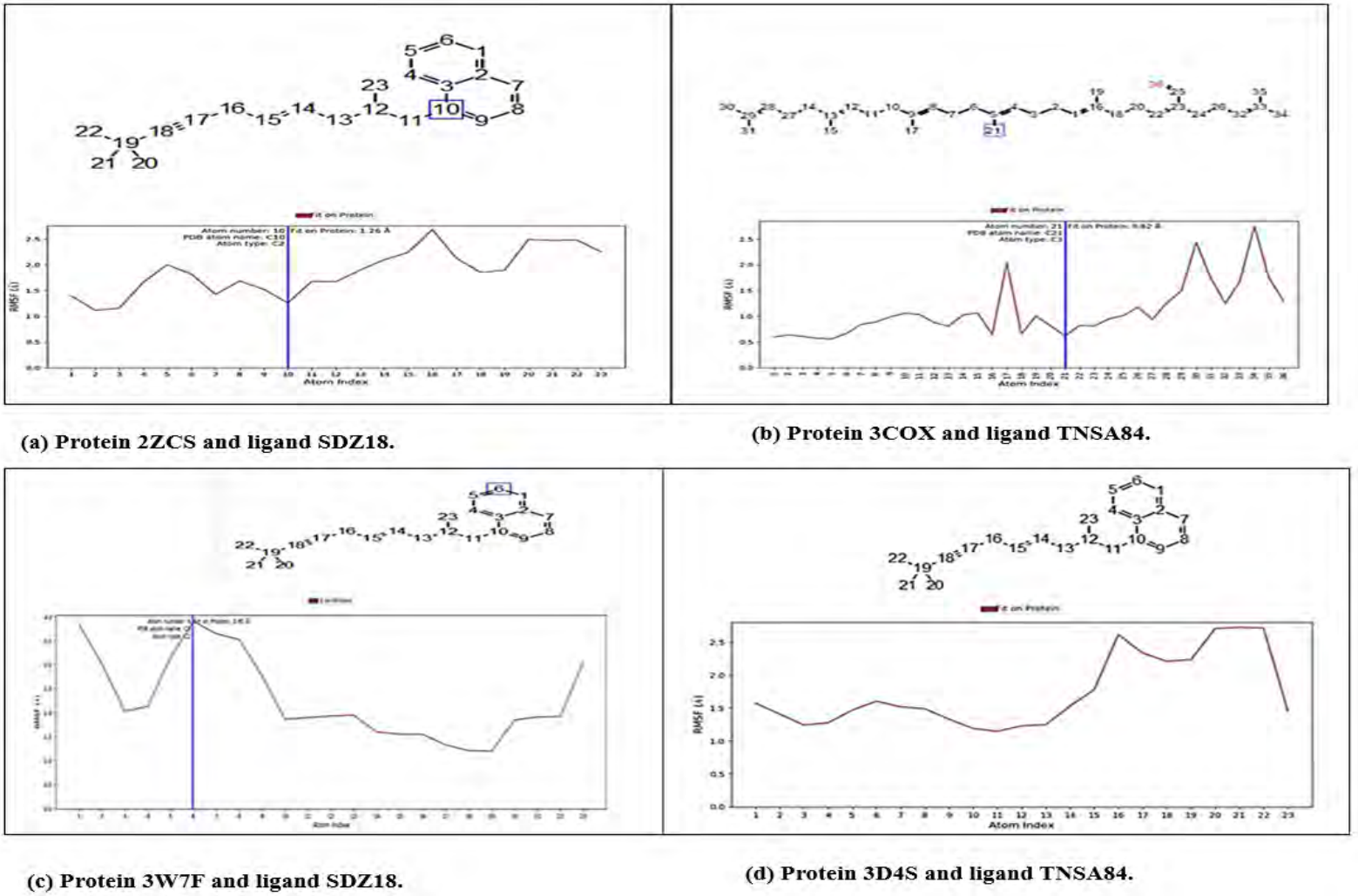


Figure 4.3 The RMSF plot for a simulation of 100 ns for the proteins.

4.3 The Ligand Root Mean Square Fluctuation (L-RMSF)

The Ligand Root Mean Square Fluctuation (L-RMSF) is used for distinguishing changes in the ligand atom positions. In the same vein, it is calculated the same way as the P-RMSF. The most notable difference between the RMSF and L-RMSF is that the L-RMSF indicates the fluctuations of ligand broken down by the atom which correlates with the 2D structure shown. To understand of the interaction of the protein and the ligand fragments corresponds in the binding process, the L-RMSF have to be studied.

In figure 4.4 (a) the peak intensities give an indication that shows atom 10 is deep in the binding pocket at L-RMSF value of 1.26 Å as it is shown by the least fluctuation. The atoms 16, 20, and 21 are the most fluctuating ones as they have L-RMSF values above 2.5 Å. They are single bonds which are more exposed, and they can easily rotate and interact with the waters. Indeed, the arrangement of the structure indicates that the atom 16, 20 and 21 are distant from the aromatic ring of the molecule. This suggest that this might be the least reactivity part of the ligand and further investigation is required. On the other hand, in figure 4.4 (b) the atoms of ligand TNSA84 that were deep in the binding pocket are found to be 1, 5, 16,18 and 21 with the L-RMSF value of 0.5 Å. The atoms that were fluctuating more than the others are atom 17 at 2.0 Å L-RMSF value, atom 30 at 2.3 Å and atom 34 at 2.6 Å. In contrary to figure 4.4 (a), figure 4.4 (c) shows that the atoms that are not in the aromatic ring are deep in the binding pocket as shown by atom 6 at 3.90 Å. The other atoms are deep in the binding pocket. Hence, they do not show high peaks, they have lower fluctuations. For the further analysis all plots show that the fluctuations were below 2.6 Å except for (c).



(a) Protein 2ZCS and ligand SDZ18.

(b) Protein 3COX and ligand TNSA84.

(c) Protein 3W7F and ligand SDZ18.

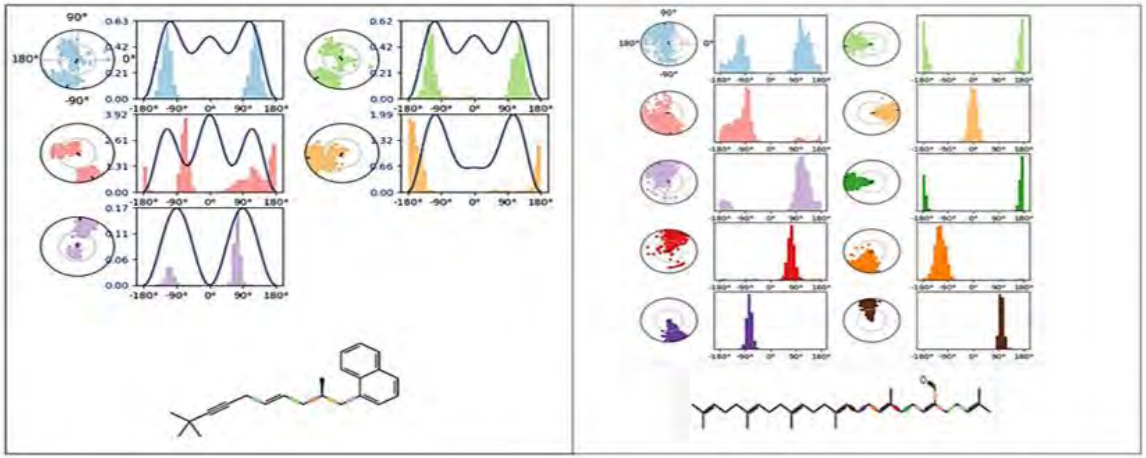
(d) Protein 3D4S and ligand TNSA84.

Figure 4.4 The RMSF plot for a simulation of 100 ns for the proteins.

4.4 Ligand Torsion Profile

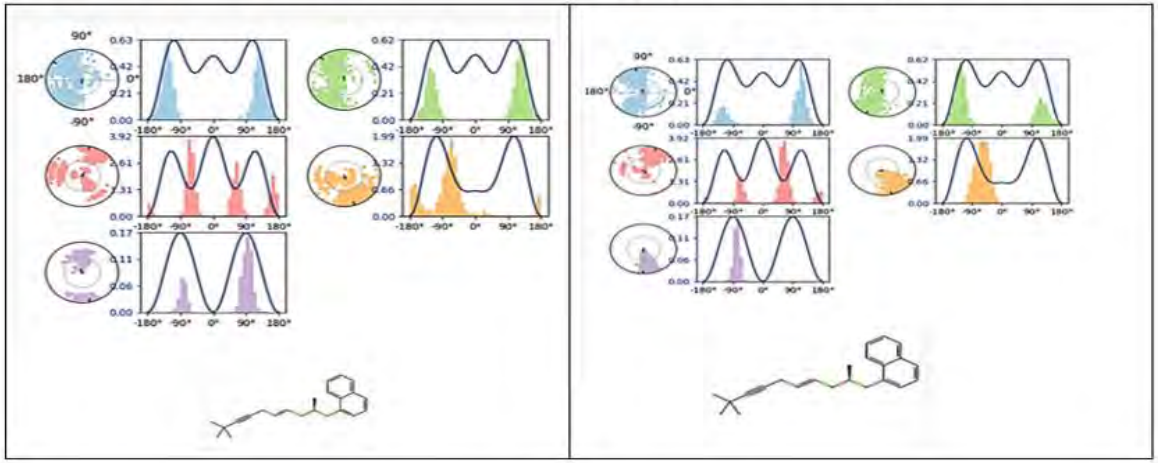
The aim of discussing the ligand torsions plot was to provide the summary of the conformational evolution of every rotatable bond (RB) in the ligand throughout the simulation trajectory from 0.00 to 100.00 ns. Alongside is a 2D schematic of the Ligand with colour-coded rotatable bonds corresponding to the dial or radial plot as well as the bar plot of the same colour. The design and the key factors considered in developing these plots are discussed below. In figure 4.5, the radial plot describes the confirmation of the rotatable bond throughout the course of the simulation where the start of the simulation is at the center and as time progresses the torsion evolution radiates outwards. In figure 4.5 (a) the blue rotatable bond seems to be spinning around more than the other ones with the value of about 0.52 Å. In fact, as indicated previously by the L-RMSF plot atom 16 and 15 that were shown to have large fluctuations, we observed that they were not deep in the binding pocket. Herein, the pink and purple bond on the ligand have less spread of the torsion angle shown on the diagram.

More interestingly, the bar plot summarizes the data by pinning the probability density of the torsions from the simulation, the plot contains the potential of the rotatable bond, and the values are on the left y-axis of the chart in blue colour, expressed in kcal mol^{-1}. By looking at the blue bond, torsion angles with a higher potential energy ranging from 0.42 – 0.63 and accompanied by the bar chart. Likewise, in figure 4.5 (c) and (d) the behavior of the ligand SDZ18 which was observed in protein 2ZCS is also observed for 6GQF, 3D4S and 3W7F. IN overall, the plots provide the insight into the conformational strain that the ligand undergoes in maintaining the protein bound conformation.



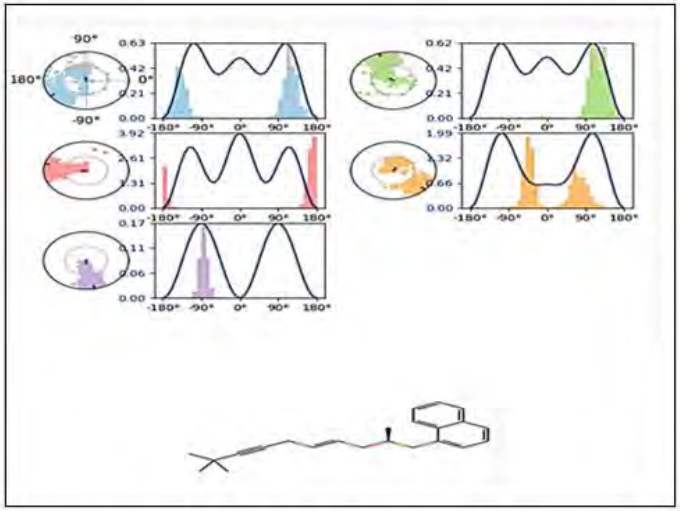
(a) Protein 2ZCS and ligand SDZ18.

(b) Protein 3COX and ligand TNSA84.



(c) Protein 6GQF and ligand SDZ18.

(d) Protein 3D4S and ligand TNSA84.



(e) Protein 3W7F and ligand SDZ18.

Figure 4.5 Ligand Torsion Profile for 100 ns simulation for the proteins.

4.5 Protein-Ligand Contacts and contacts timeline

The protein interactions give a detailed information about the interactions of the residues with the ligand during the simulation. The interactions are namely, hydrogen bonds, hydrophobic, ionic and water bridges. They all contain the more specific subtypes which can be explored through the simulation interaction diagram panel. The values are understood easily if expressed in percentages, a typical example for such expression is when the value of 0.5 suggests that the 50% of the simulation time, the specific interaction is maintained. However, some proteins may have values that are above 100% or 1.0 because some of them may make the multiple contacts of the same subtype with the ligand.

The following maestro default settings were used for the assessment of the interactions, in hydrogen bonds, the bonds to the protein were broken down into backbone and side chain, donor and donor acceptor [82]. Firstly D—H···A the distance cut off of less than 2.5 Å between the donor H and the acceptor atom A. Secondly, D—H···A was greater than 120°, it represents the two angles between Donor H and the acceptor atom A. H···A— X was greater than 90° [82].

In ionic, the protein-ligand interactions between the oppositely charged atoms were within 3.7 Å of each other and no hydrogen bonds. On the other hand, the protein-metal-ligand interactions that are mediated by a metal atom where within 3.4 Å of a protein and ligand heavy atoms in exception of carbons. In hydrophobic, the interactions are broken down into pi-pi stacking, pi-cation staking and other general interactions. In pi-pi, the two aromatic groups stacked face-to-face and face-to-edge [82], [132]. In pi-cation, the aromatic and charged groups where within 4.5 Å. The other general interactions, hydrophobic sidechain within 3.6 Å of a ligand's aromatic or aliphatic carbons. In water bridge, the interactions bridged by a hydrogen bonded water molecule. The hydrogen bonds geometry was slightly relaxed from the standard criteria D—H···A the distance cut off of less than 2.8 Å, D—H···A angle greater than 110° and acceptor angle greater than 90° [82], [132].

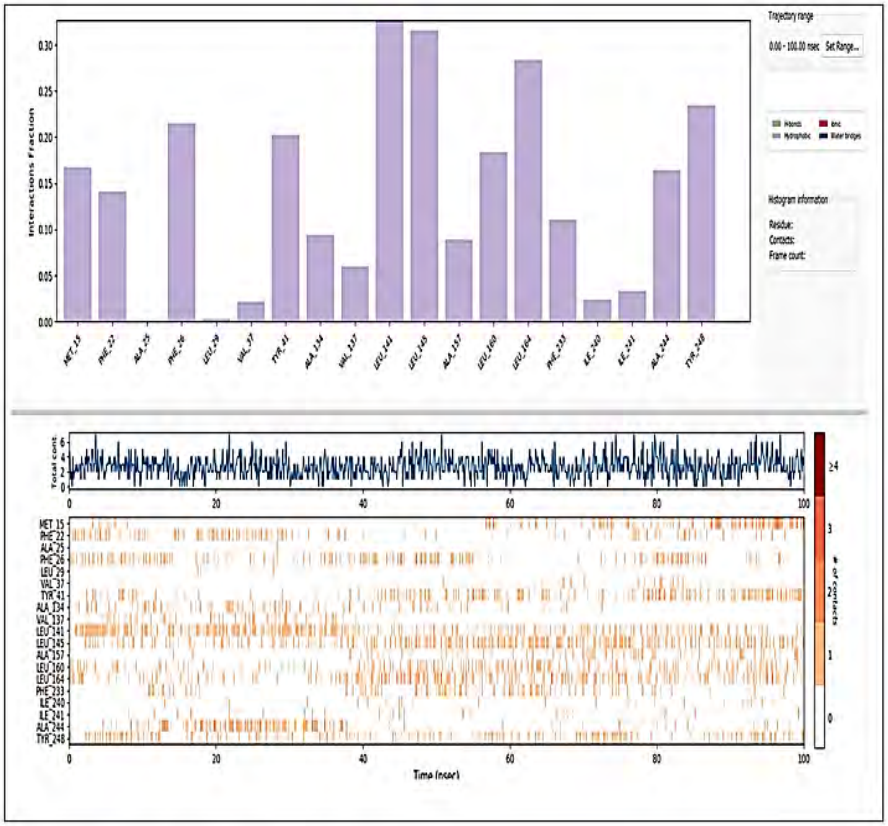
In figure 4.6, the interaction of the protein 2ZCS amino acids can be observed throughout the 100 ns simulation. It shows interaction fraction on the y-axis and the amino acids on the x-axis. The results observed from molecular docking in chapter 3, figure 3.8 (a), showed that amino acid PHE 26 was interacting with the aromatic ring of the ligand. The amino acids such as LEU

141, LEU 145, ALA 244, etc. were in the binding pocket. Closer analysis shows that the amino acid LEU 141 and LEU 145 are interacting with ligand by hydrophobic interaction over the course of 100 ns. They are both making 0.35 or 35 % contacts of the hydrophobic interactions. The amino acid PHE 26, which had pi-pi stacking bond interaction with the ligand, here in Molecular dynamics is also showing a contact of 0.30 with the ligand. Furthermore, hydrogen bonds interaction are shown for other amino acids. Below the contact plot is the timeline plot that shows the total number of contacts that are being made from 0 to 100 ns. It shows which amino acids that were making contacts at certain times. The strength of the contact is shown by the more colour intensity represented by pink. The amino acid LEU 141 had more interactions and they only start to show from 0 to 40ns. However, the interaction becomes less intense from 40ns to 100 ns. The evidence is shown by the pink colour becoming less intensity on the diagram. Furthermore, TRY 248 also contributes to the amino acid that had more interactions as it shows more intense pink colour. However, the amino acid TRY 248 did not show any direct contact with the ligand in molecular docking results. Some amino acids show very small interactions, ILE 240 shows less interactions throughout the simulation Contrastingly, there are some amino acids that do not show any interactions, no pink colour is shown namely interactions are LEU 29 and ALA 25.

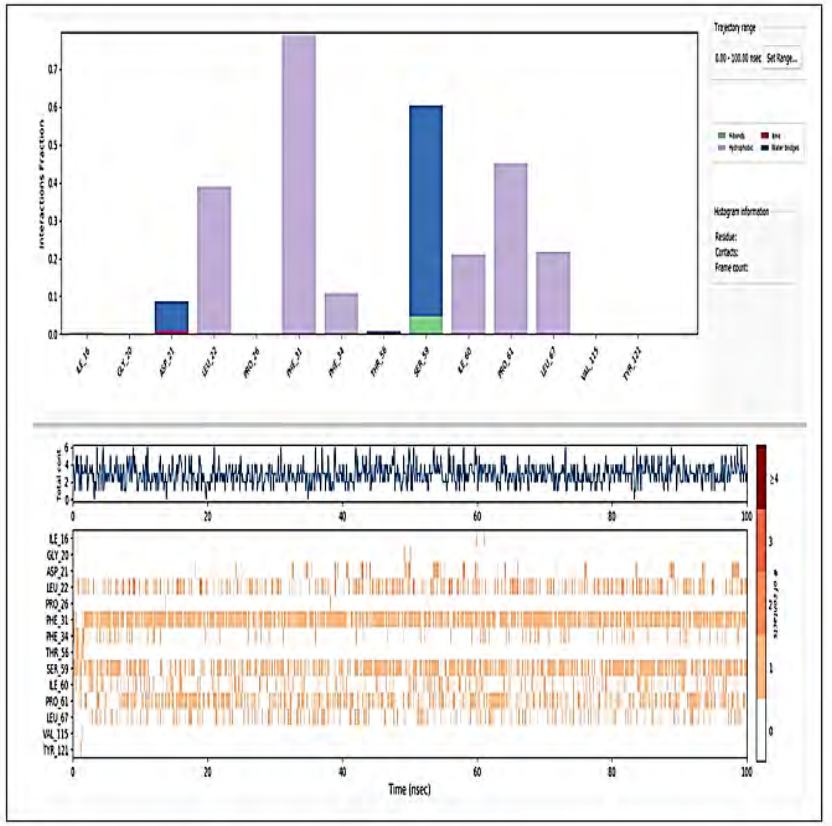
Analogous to protein 2ZCS, in figure 4.6, we observe that in figure 4.7, for protein 3D4S the amino acids that dominated were hydrophobic and here the amino acid PHE 193 showed more interaction above 0.70 throughout the 100 ns simulation. Interestingly, this is consistent with the results previously analyzed in molecular docking chapter, figure 3.9 (d) where PHE 193 and 290 were showing the direct contact with the ligand. However, the amino acid PHE 290 showed less contact with the ligand than PHE 193 ILE 390 showed minor interactions while PHE 290, PHE 289, TRP 109 and VAL 114 showed intermediate interactions.

According to figure 4.8, it is observed that protein 6GQF did not show any interaction and this is corresponding to the RMSD plot explained previously in section 4.1, figure 4.1 which showed the ligand not stable in the first 50ns simulation. The results of ligand torsional profile show the interactions becoming more intense during the last 50 ns simulation. More interestingly, we observe that in figure 4.16, there is a hydrogen bond interaction with high interaction of 0.78 with inclusion of water bridges contact of about 0.02. Unexpectedly, the hydrogen bond is in contact with the ASN 485 instead of LYS 296 amino acid as previously depicted in molecular docking, section 3.3. We may assume that during the MD simulation,

the ligand rotated and interacted with the other amino acids. At this point, the reasoning is somewhat limited and another study is needed in other to calculate the electron density of the ligand in order to observe what is happening at the quantum mechanical (QM) realm as it was previously mentioned that MD is limited due to classical mechanics. In general, we have systemically investigated the protein-ligand interaction in this work and observed the amino acids which contributes to the stability of our systems.

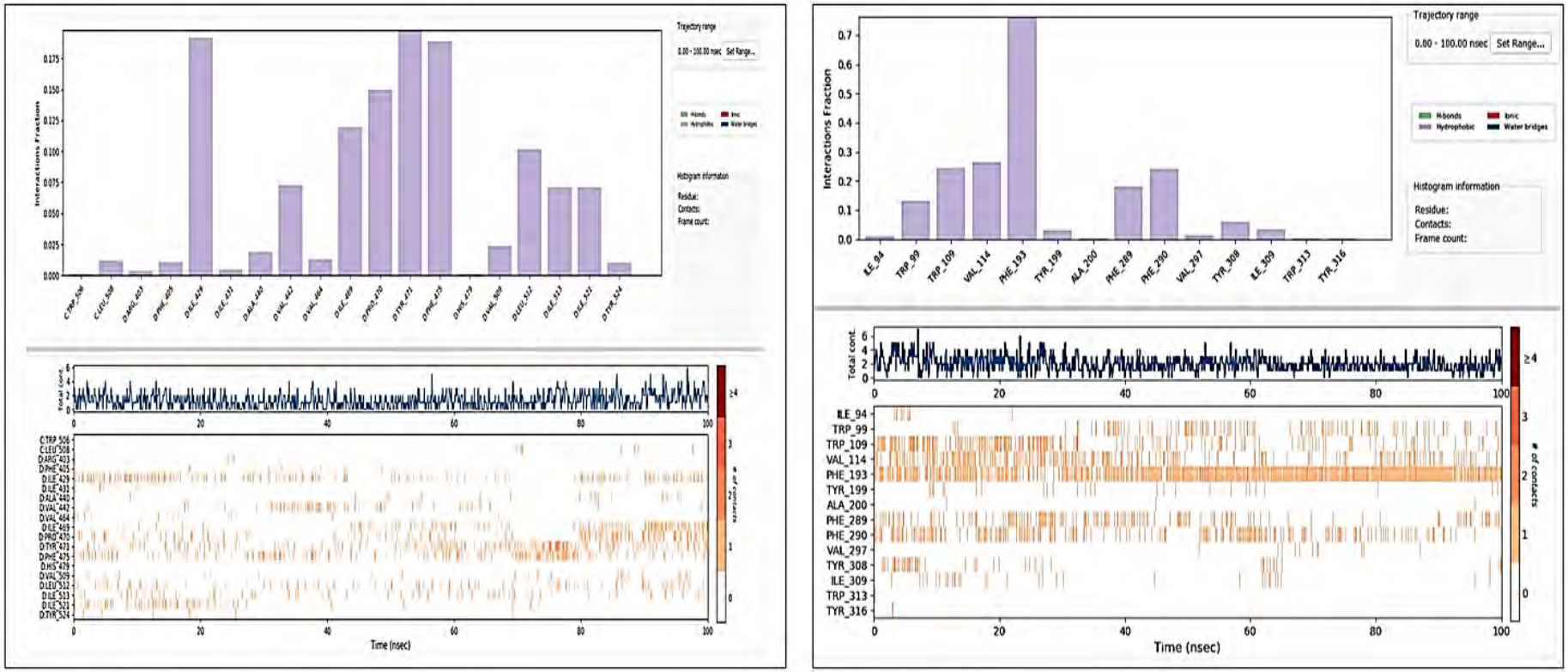


(a) Protein 2ZCS and ligand SDZ18.



(a) Protein 1M17 and ligand SDZ18.

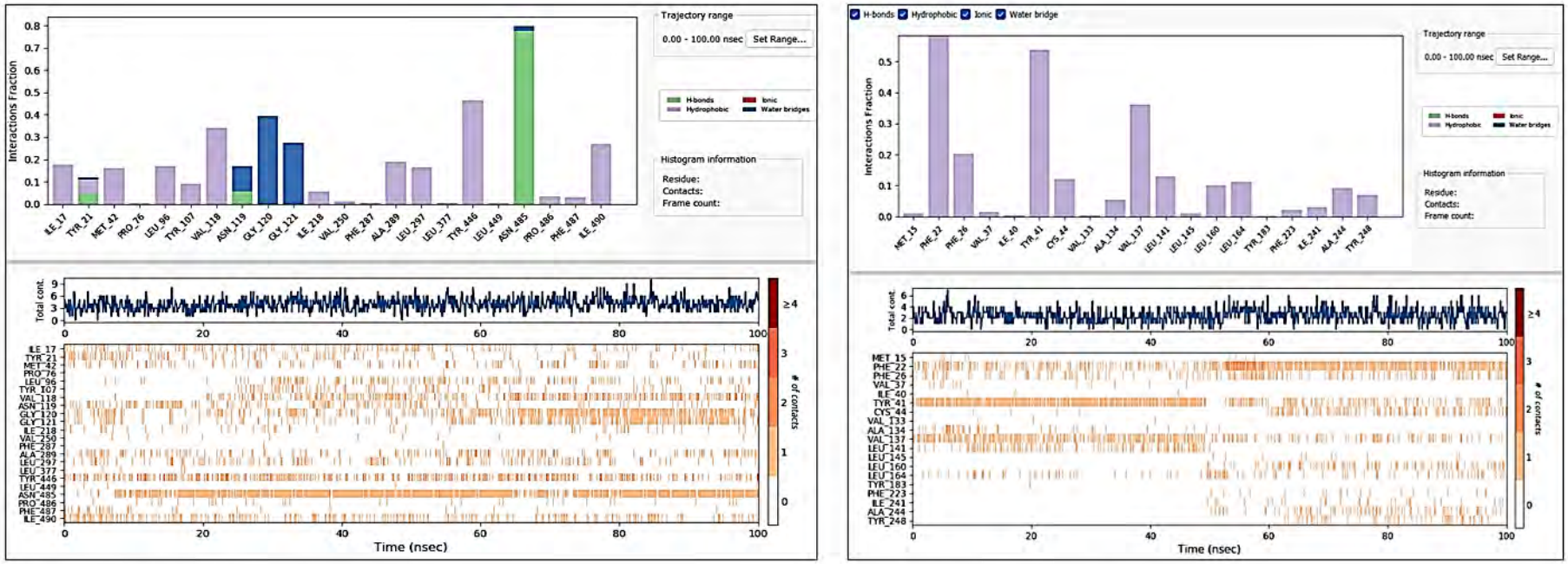
Figure 4.6 The Protein-ligand contacts and contacts timeline for (a) 2ZCS protein SDZ18 and (b) 1M17 protein and SDZ18.



(a) Protein 6GQF and ligand SDZ18.

(b) Protein 3D4S and ligand SDZ18.

Figure 4.7 The Protein-ligand contacts and contacts timeline for (a) 6GQF protein SDZ18 and (b) 3D4S protein and SDZ18.



(a) Protein 3COX and ligand TNSA84.

(b) Protein 3W7F and ligand SDZ18.

Figure 4.8 The Protein-ligand contacts and contacts timeline for (a) 3COX protein TNSA84 and (b) 3COX protein and SDZ18

4.6 Summary of the chapter

A realistic simulation was performed with molecular dynamics that investigated the stability of the ligand in the active side during a 100 ns simulation in a solvated environment. In order to explore the stability, the data obtained from MD indicated that the ligand does not leave the binding pocket at all for all the 6 proteins represented. What is evident in drug discovery is that the outcome of the good docking scores of the ligands are not sufficient to fully prove the inhibition of a protein. In fact, proteins are constantly vibrating and undergoing conformational changes. In the event that the ligand stability is not achieved or proved, the ligand may migrate to other parts of the body. Consequently, the molecules may break from its mass, then atoms disunite from one another and explode into a shower of electrons, this destruction may be dangerous to the body and very toxic to other parts of the body. Again, the molecules that are initially docked may be changed into other structures with the aid of other elements in the body in the scenario that they leave the binding pocket of the protein. To complete the picture, for the RMSD plots all the five systems had stabilized below 4.6 Å except for figure 4.2 (a).

Conclusively, the aim was to systematically investigate if ligands preserve their identity and stay in the binding pocket of the protein during the MD simulations. MD trajectories were investigated and insights into the stability and interaction between ligand and protein were presented. However, we have only studied the external forces acting on our system using classical mechanics and not yet deeply penetrated into the enzymatic reactions that happens in atoms. We end this chapter by pointing out the pitfalls of MD, it has limitations and cannot navigate to other higher dimensions quantum molecular dynamics tool which is not yet invented, would be the best tool that would use classic mechanics and quantum mechanics equations to run simulations for the whole system. Recently, a paper was published of Bienfait K. Isamura and Paul L.A. Popelier in which they develop a machine learnt force field FFLUX which uses Gaussian process regression (GPR) models to predict quantum mechanical properties in the molecular dynamics (MD) simulations [133].

Chapter 5 : QM WITH DENSITY FUNCTIONAL THEORY (DFT)

Introduction to chapter

As it was stated earlier chapter 1 section 1.2 (e), the density functional theory (DFT) is used widely in drug discovery to uncover the quantum mechanical properties of the systems with the use of the Born-Oppenheimer approximation equation [134], [44], [135]. In some cases, DFT is coupled with experimental techniques such as FT-IR and Raman spectroscopies for a proper understanding of the molecular properties of compounds [95], [100], [136]. In this chapter, DFT was utilized to illuminate the important aspects of the compounds such as the molecular orbitals, molecular electrostatic potential, ionization potential, electron affinity etc. [98]. The six ligands with high docking scores from chapter 3 were chosen for DFT studies. The model computations were carried out with DFT as implemented in Gaussian 09 package using B3LPY in conjunction with the 6-311+G(d,p) level of theory [110]. A more detailed computational approach is given in chapter 2, section 2.4.

5.1 Geometry optimization

The geometry optimization was used to determine the starting structure for the quantum mechanical calculation by finding the coordinate of the structures that represent the potential energy minimum [43]. This approach had, in fact, allows the Born-Oppenheimer Approximation to calculate the potential energy surface which is simply the specification of the classical potential energy as function of the molecular structure [43]. To find answers of the electronic properties involved in our ligands, we optimized the ligand and the structures are shown in figure 5.1 and on the left is the legend which shows the colour representation of atoms in the structures.

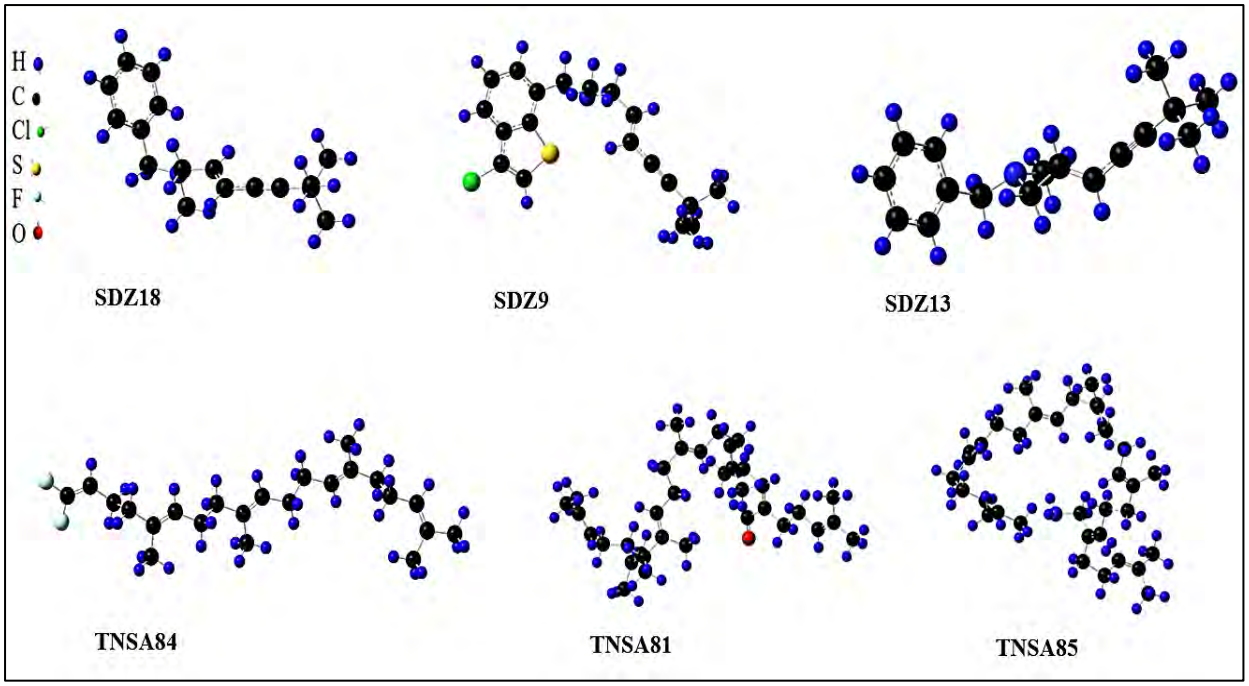


Figure 5.1 The optimized geometrical structures of the Ligands.

5.2 Energy profiles of the frontier molecular orbitals

In this study, DFT was used to investigate the reactivity of the compounds using the energy profiles of the frontier molecular orbitals. The calculated parameters were E_{HOMO} **(**highest occupied molecular orbital), E_{LUMO} (Lowest unoccupied molecular orbitals), ΔE_{H-L}, (The energy gap), the electron affinity (EA), the ionization potential (*IE*), and dipole moment (μ). The table 9 provides new insights of the molecules utilized in this study. The computed values of HOMO energies ($\boldsymbol{E_{\mathrm{HOMO}}}$) were in the range of 5,271 to 4,939 kcal mol^{-1}. The LUMO energies ($\boldsymbol{E_{\mathrm{LUMO}}}$) were in the range of -0,067 to -1,225 kcal mol^{-1}. A relatively small change in the results, this implies that there is no significant difference in the energies of the compounds as revealed by this study.

The energy gap ($\boldsymbol{\Delta E_{H-L}}$**)** relates to the reactivity of the molecule, the highest value represents the more reactive. The energy gaps of the molecules also predict the kinetic stability, which states that the large the value of energy gap, the more it is stable [98]. The ligand SDZ 18 was among the ligand with high value of $\boldsymbol{\Delta E_{H-L}}$ of 4,020 kcal mol^{-1}which indicates its high reactivity and likelihood to remain unchanged under conditions such as heat, etc. In support of the results of SDZ18, previously in chapter 3, section 3.3 figure 3.4 (a) the hydrophobic bond

interaction with the amino acid PHE was observed. It is for this reason that we may assume that is the main reason for the high energy gap value. Moreover, in chapter 4, section 4.1 figure 4.1 the RMSD plots collectively had showed stability all below 4.6 Å. This remarkably stability might also be the reason for the high energy gap value observed in this section and it might be beneficial to emphasize this assumption.

Through Koopman's theorem, there is a relation between the frontier molecular orbitals, HOMO and LUMO and the ionization potential and electron affinity [98]. Thus, we have used the HOMO and LUMO data to calculate the ionization potential and electron affinity using equations 7 and 6 in chapter 1, section 1.2 (e). Indeed, it was found that the ligand SDZ 13 had the highest ionization potential as compared to the other ligands with the value of 5,271 kcal mol^{-1}. It had the minimum energy that is needed to ionize an atom in its ground state more than the other ligands, this is surprising as the ligand had an average docking score of -6.0 kcal mol^{-1}which was lesser compared to the SDZ 18 which had average docking score of -7.0 kcal mol^{-1}as it is indicated in chapter 3, figure 3.3. SDZ 18 had the least ionization potential with the value of 4,939. SDZ 16 had the highest electron affinity of 1,225 and this means that it had high energy change that results from the addition of electrons to a gaseous atom. Perhaps, it is reasonable to assume that the docking data energies and the ionization potential and electron affinity do not have much correlation. Finally, it is evident from the calculated data of ionization potential and electron affinity that all the ligands are capable of donating electrons, they are all electron rich.

Intriguingly, as it was previously observed, that the ligand TNSA84 had higher total energy than some of the molecules. The new finding is that it also has a very high dipole moment compared to the other molecules. This might be another reason as to why previously in chapter 3, section 3.3 figure 3.4 (b) that the in binding pocket it had the hydrogen bond interaction with the amino acid LYS 296. It is worth noting that the dipole moment is a one of the fundamental properties of molecule that influences the behavior and interactions [137]. The ligand SDZ 9 had high docking scores averaging above -6 kcal mol^{-1}as it was observed in chapter 3, section 3.3 figure 3.3. The calculated dipole moment of 3.281 might have contributed to the high docking scores. We believe this to be a reasonable assumption but however would emphasize it does not correspond to the dipole moment 0.324 of SDZ18 which seem very low. In chapter 3, figure 3.3 the ligand SDZ18 was the one with mostly high docking scores than the rest of

the ligand. In 2020, a conclusion was drawn in paper published by Mohamed Hagar et al that the dipole moment and high basicity contributed to factors that enhanced the extent of the binding affinity during the investigation of some antiviral N-Heterocycles as COVID 19 drug [138].

The thermodynamic function, thermal enthalpy is the change that occurs when a molecule undergoes chemical reaction [139]. The most negative enthalpy value states that a molecule releases energy and becomes more stable while the positive value indicates that the molecule absorbed energy from surrounding and becoming less stable [139]. A glimpse at table 9 shows that molecules had negative enthalpies and they were more stable. This could be another reason as to why SDZ18 stayed in the binding pocket during the MD simulation in chapter 3.

Table 5: The HOMO and LUMO energies (kcal mol^{-1}) and the energy gap of the ligands.

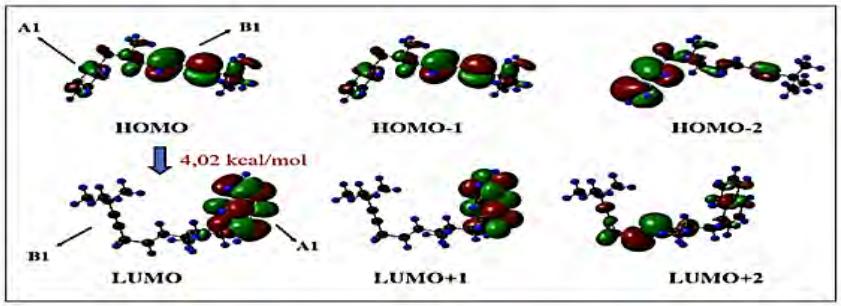
Frontier molecular orbitals	**SDZ18**	**SDZ13**	**SDZ9**	**TNSA81**	**TNSA84**	**SDZ16**
Total Energy (kcal mol^{-1})	-20581,6	-16508,6	-38010,7	-25292,5	-33237,3	-20061,2
$\boldsymbol{E_{HOMO}}$ (kcal mol^{-1})	-4,939	-5,271	-5,218	-5,245	-5,139	-5,114
$\boldsymbol{E_{LUMO}}$ (kcal mol^{-1})	-0,919	-0,827	-1,086	-0,067	-1,537	-1,225
$\Delta\boldsymbol{E_{H-L}}$ (kcal mol^{-1})	4,020	4,444	4,130	5,178	3,602	3,889
Ionization potential (I= $-\boldsymbol{E_{HOMO}}$)	4,939	5,271	5,218	5,245	5,139	5,114
Electron-affinity (EA= $\boldsymbol{E_{LUMO}}$) (kcal mol^{-1})	0,919	0,827	1,086	0,067	1,537	1,225
Dipole moment (Debye)	0,342	0,726	3,281	2,343	4,033	1,844
Thermal Enthalpies (kcal mol^{-1})	-20613,3	-16538,9	-38055,2	-25362,7	ND	-20077,3

5.3 Visualized frontier molecular orbitals

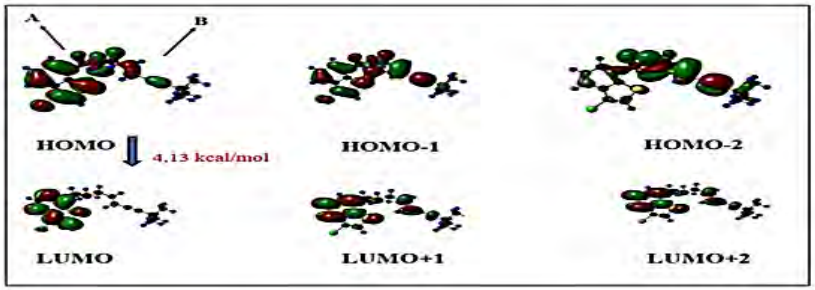
Frontier molecular orbitals (FMOs) are the HOMO and LUMO. The HOMO has electrons and it is the electron donor while the LUMO lacks electrons and it is electron acceptor. The chemical reactivity is determined by the FMOs which give insight into the electrons transfer from HOMO to LUMO. Figure 5.2 (a) reveals the HOMO of ligand SDZ 18, the electron

density is mainly distributed all over the molecule, in both part A (less bulky part of molecule) and part B (bulky part of the molecule). SDZ 18 has the energy gap of 4,02 kcal mol^{-1}. Furthermore, the HOMO, HOMO-1 and HOMO-2 that the electron density is distributed all over the ligand. Respectively, it is observed that in figure 5.2 (b) the ligand SDZ 9 showed that the electron density was localized on the aromatic ring of the ligand in both the HOMO and LUMO with energy gap value of 4.13 kcal mol^{-1}.

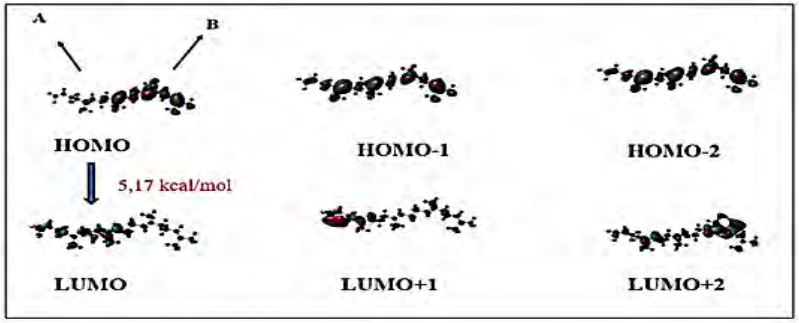
In figure 5.2 (d) the electron density is distributed all over the molecule for both the LUMO and HOMO, analogous to figure 5.2 (e) where the distribution is also all over the molecule. According to another relevant molecule is seen in figure 5.2 (c) that the electron density is distributed around the part labelled B, where the oxygen atom is present. There is an impact on this region and in order to explore this assumption we should look back at molecular docking in chapter 3, section 3.3, figure 3.4 where there was a hydrogen bond interaction with the amino acid LYS 296. However, the energy gap calculated is 3.60 kcal mol^{-1} which is smaller than the other energy gaps of the other molecules.



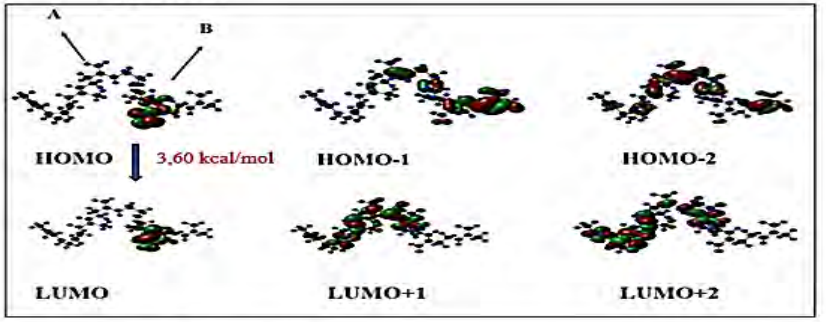
(a) SDZ18.



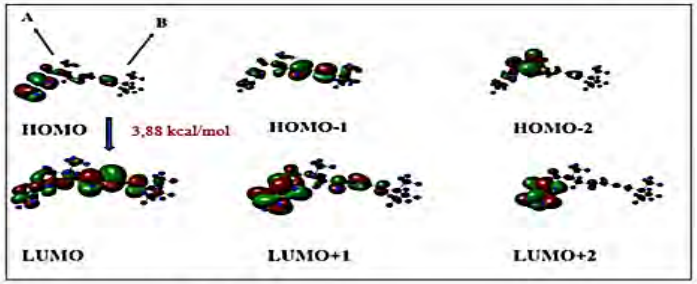
(b) SDZ9.



(c) TNSA81.



(d) TNSA84.



(e) SDZ13.

Figure 5.2 The HOMO and LUMO energies of the ligands.

5.4 Molecular electrostatic potentials (MEP)

The molecular electrostatic potential (MEP) helps us to understand the site for electrophilic attack and nucleophilic reactions and it is also related to electron density [38], [43]. It provides a visual method to understand the relative polarity of the molecule and it indicates the net electrostatic effect produced by the total charge distribution of a molecule [140]. Such information is important to understand to validate the evidence about the reactivity of the molecules [140], [138]. In this section, of interest is the Figure 5.2 shows the electrophilic regions represented with a red coloured part of the MEP cloud which corresponds to an attraction of the proton by the concentration of electron density in the molecule [140]. This electrophilic region is prone to electrophilic attack. The nucleophilic region is represented by the blue coloured part of the MEP cloud, this area is prone to nucleophilic attack and it is the positive electrostatic potential which corresponds to the repulsion of the proton by the atomic nuclei in regions with low electron density, and the nuclear charge is incompletely shielded [140], [138], [38]. The yellow and green colours represents unoccupied orbitals. Furthermore, the negative electrostatic potential represents the attraction of the proton by region of high concentration, while the positive electrostatic potential represents the repulsion of the proton by the atomic nuclei.

It is obvious that in the figure 5.2, the ligands SDZ 18, SDZ 13, SDZ 9 and SDZ 16 had both the electrophilic and nucleophilic power distributed over all the molecules. Most attention should be paid to the ligand TNSA 84, we observe that it had the red mapping of the electrostatic potential around the oxygen atom. This shows that this region had concentration of electron density. Looking back at chapter 3, section 3.3, figure 3.4 there was a hydrogen bond interaction with amino acid LYS 296 around this region.

We further observe that the electrophilic power was localized on fluorine atom for the ligand TNSA 81 indicating that these sites are more favorable for reaction. There is no sharing of the electrophilic and nucleophilic powers these two ligands TNSA 84 and TNSA 81 [78]. For the ligand SDZ 9 had the electrophilic power distributed around the aromatic ring which had chlorine and Sulphur atoms. In this regard, these region represents the molecular site suitable for an electrophilic attack [141]. To this end, it is worth stating that the mapping of electrostatic potential around the molecules could also be mainly responsible for the variation of binding

affinity with the active site of the proteins [138]. The white coloured areas on the molecules represent non-polar areas, where there is poor chemical reactivity [141].

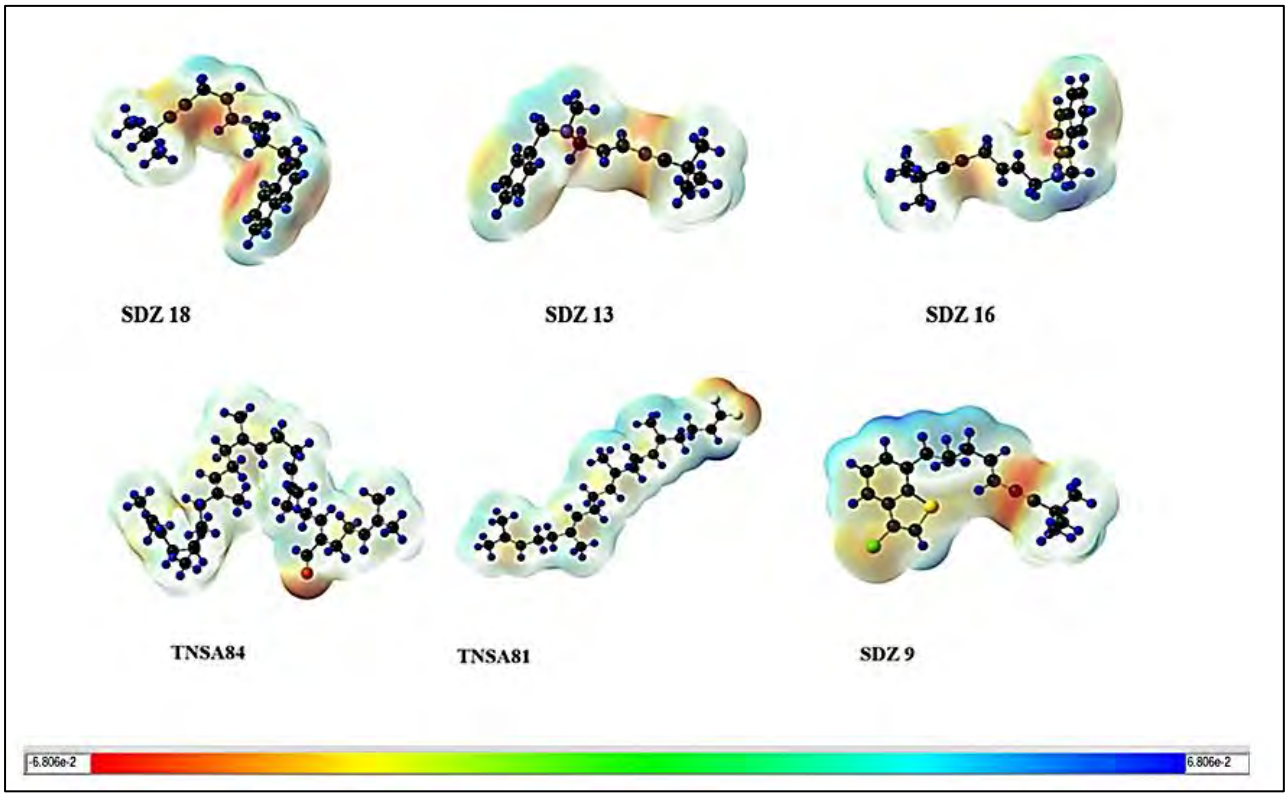


Figure 5.2 Molecular electrostatic potentials of the ligands.

Chapter 6 : HYBRID QUANTUM MECHANICS / MOLECULAR MECHANICS (ONIOM QM/MM)

Introduction to the chapter

Hybrid energy methods such as QM/MM are widely employed in describing large systems at different levels of theory [142], [143]. In this chapter, the ONIOM method is used to calculate the total energy of the complexes and to compute the intrinsic reaction coordinated (IRC) plot to understand the energetic changes associated with the interaction of the ligand to amino acids in the protein active site. MM calculates the trajectory of particle and QM calculates the wavefunction of particles [84], [86].

In 2016, Matias I. Sancho et al reported a good study where ONIOM calculations were used to obtain information about the energy and structure of the inclusion complexes, where the molecular geometries that equilibrated in MD study were optimized at QM level theory. Our approach was motivated by these previous studies. Consequently, the ONIOM Models were constructed and data extracted. Almost similar approach was done by Susanta Das et al as they presented an effective methodology for identifying probable binding modes of multiple bound states along a reaction coordinate in the protein active site for the study of docking in Terpene synthase BPPS [144]. A detailed introduction to ONIOM is given in chapter 2, section 2.5. This chapter focuses on the results obtained from the ONIOM calculations with reasonable computation time and costs.

6.1 The ONIOM models with respect to B3LYP/6-31G:UFF level

Arguably the most compelling, straightforward proof is that the behind every physical structure is energy and, in many healings, systems knowing the dynamic vital energy is of great importance. Unless otherwise stated, the ONIOM models were calculated at B3LYP/3-21G:UFF level through out the study. The models provides insights into the total energy of the systems while maintaining reasonable computational cost and time. The calculated values of total energies of the 3 models are shown in table 10. The values obtained are -1503.26 a.u for model 2ZCS, -962.30 a.u for model 3D4S, -881.58 a.u for model 3WFT. Particularly, the model ZCS has the highest total energy among the others.

Additionally, the ONIOM calculation allows for the evaluation of electric field derivatives such as dipole moments [145]. The model 3D4S had the highest dipole moment of 2084.2 Debye in relative to the others, meaning that the electronic structure and charge distribution when a ligand interacted was more than the others. Previously, in chapter 3, section 3.3, figure 3.4, the ligand SDZ 18 and protein 3D4S had more direct interaction with the amino acid PHE 193 and 290 in relative to other proteins. We assume that this might be the reason why the dipole moment is higher.

The molecular wavefunction symmetry can be described with point groups [146]. All the models had point group C1, they belong to a group with no symmetry. The model 2ZCS and 3D4S had imaginary frequency while model 3WFT had no imaginary frequency. Attempts to locate the minima for model were successful, while for the model 2ZCS and 3D4S it was not successful, no stable molecular structures were obtained. From a standpoint of obtaining these imaginary frequency values, we do not have deeper insights as to which models have local or global minima.

Further atomistic insights were uncovered, the RMS gradient norm was observed to be smallest for model 3D4S with the value of 0.000001167 and largest for model 3W7T with the value of 0.00000958. In quantum chemistry, the root mean square (RMS) gradient norm is used to calculate how close the molecular geometry is to the minimum energy structure [43], [147]. The model 3W7T suggests a larger movement of crystal structure geometry away from minimum than the 3D4S [147].

Table 6 : The ONIOM models.

ONIOM models	**Higher layer amino acid residues**	**Total Energy (a.u.)**	**Dipole moment (Debye)**	**RMS Gradient norm**	**Point group**	**Imaginary Frequency**
2ZCS	PHE 26	-1503.26	489.8	0.00000884	C1	1
3D4S	PHE 290	-962.20	2084.2	0.000001167	C1	1
3W7T	ICS 479	-881.58	0.3695	0.00000958	C1	NAN

6.2 Total intrinsic Reaction Coordinate (IRC)

In medicinal chemistry, QM is used to predict reaction mechanism (flow of electrons) between drug molecules and proteins based on energy [135]. Integrating QM and MM into one package allows the partitioning of total energy of the system to be determined with great convenience and efficiency [89], [135]. Enzymes residues that directly participate in chemical reactions, are taken as part of QM subsystem [148].

What has been illustrated in this section is the IRC profiles that provide insights into the energy profile of the ligand binding to the protein. They are used to trace the path of chemical reaction from transition state (TS) to products and reactants and has become important in the study of potential energy surfaces of chemical reactions [149], [150]. As it was discussed by Fukui, the mass-weighted cartesian coordinates when used develops a path which becomes the IRC [149]. The first point on the IRC starts from the initial state of the geometry which is known as the saddle point and each IRC point starts from the optimized geometry of the previous point [149]. The curve is determined by plotting the potential energy surface connecting the reactants and products through the transition state [149]. The ligand entry path into buried binding sites can be investigated using the reaction states where the ligand is placed in different sites on the protein surface [150].

In this section, we analyze the steps of the energy in the protein environment with regard to the amino acids interacting with the ligand. According to the research work of Susanta Das et al, they sharply emphasized the protein-ligand docking of multiple reactive states along a reaction coordinate in enzymes using the intrinsic reaction coordinate to study the terpene synthase reactions [144]. Figure 6.1 shows the different stages along the reaction coordinate in which stationary points on potential energy surface of the protein-ligand system are connected by a steepest energy path [149], [151]. The interacting amino acids was PHE 26 as it is previously shown on the table above. A point to note here is the energy changes of reaction pathway between the amino acids and the ligand. A more careful examination showed the highest energy was -1478 a.u / Hartree.

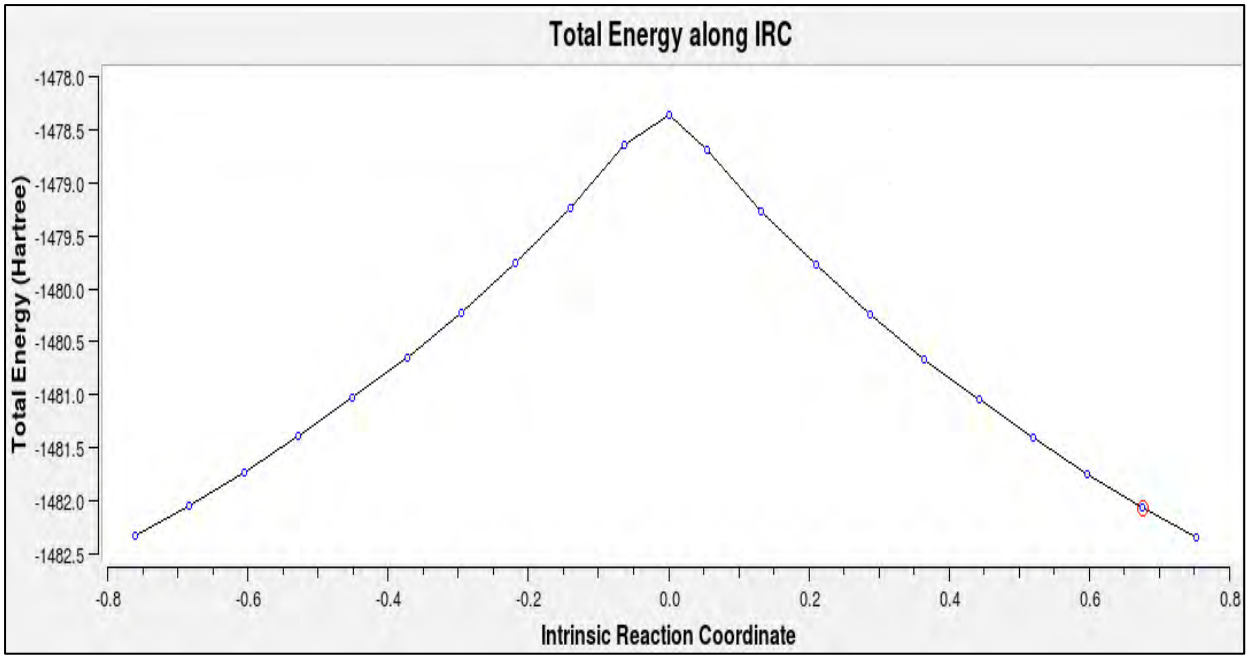


Figure 6.1 The intrinsic reaction coordinate (IRC) for SDZ 18 and 2ZCS.

6.3 RMS Gradient norm along intrinsic Reaction Coordinate (IRC)

Figure 6.2 shows a plot of the RMS (Root mean square) gradient norm along the IRC for the protein-ligand system. To shed further light on the ligand-protein interaction, we examined the RMS Gradient Norm along IRC. Highlighting the importance of gradient, it is used to measure the change in energy with respect to nuclear coordinates of the system. It is expressed in Hartree against Bohr which is simply the unit of length used to describe the atomic and molecular dimensions in quantum chemistry given in units of angstroms (Å). For this system, the IRC is constructed by transforming the steepest descent path in mass-weighted cartesian coordinates and the gradient equation is the internal coordinate equivalent of IRC of Fukui [149], [151]. We therefore believe that this might be the reasonable assumption because of the similarity of the curve in figure 6.1 and figure 6.2.

The ligand-bind energy reaction path used the step size of 0.010 hartree/Bohr and the highest value reached was 0.120 hartree/Bohr. A remark must be made that the overall monitoring of the RMS gradient norm along the IRC provides valuable information about the stability of protein-ligand binding in steps of 0.010 hartree/Bohr. The potential energy surface near the

reaction path is expanded as a Taylor series in the internal coordinates thus having many coordinates [149].

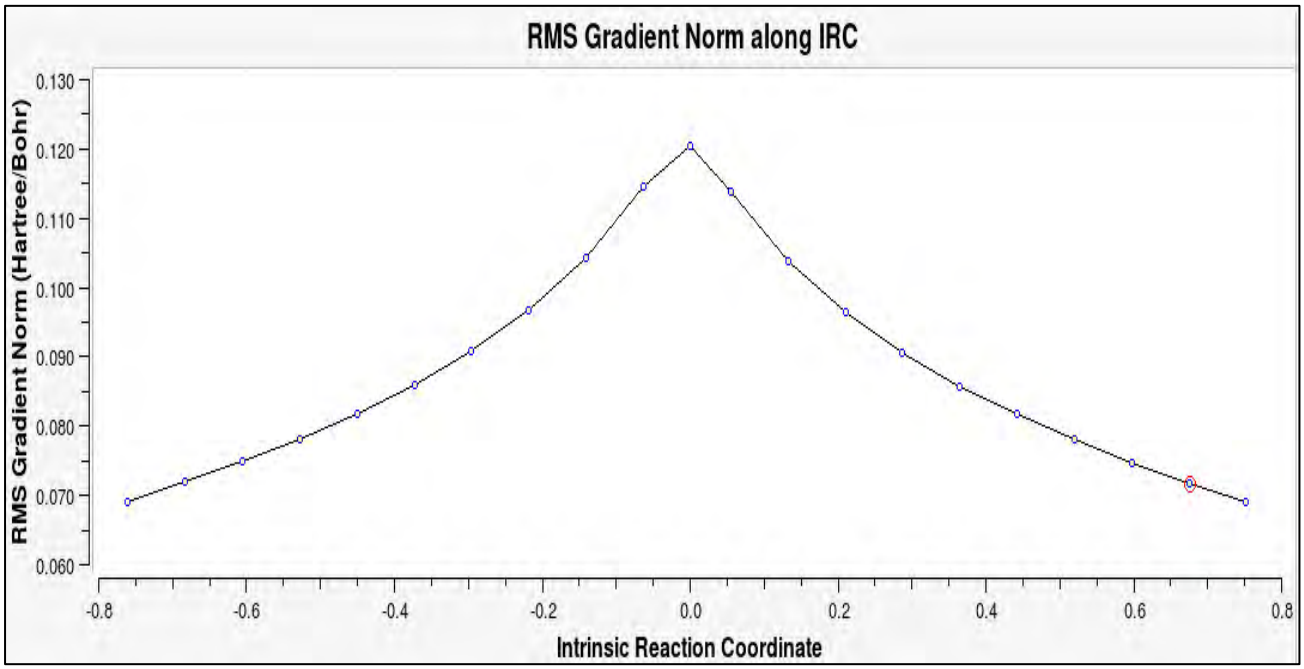


Figure 6.2 The RMS gradient norm (RMS) along IRC for SDZ18 and 2ZCS.

6.4 Summary of chapter

In this chapter, the hybrid QM/MM calculations were employed to obtain the total energy of our complexes and also a superior elucidation of the energetic changes with respect to the intrinsic reaction coordinate. The findings in table 6 show that the model 2ZCS had the highest total energy in relative to the others. More deeper insights into the complexes were provided by this theoretical method.

The region of the active side was treated as the high layer and most expensive computation to capture the chemical properties while the rest of the atoms were treated as the lower layer. The analysis of the IRC plot shows that during the breaking of the pi-pi stacking bond of the ligand SDZ 18 and amino acid the highest energy reached was -1478.55 a.u. We can justify that we have tried to understand the protein-ligand interactions at different stages using the cartesian coordinates. Indeed, we must admit that there are many insights that we can obtained using the ONIOM QM/MM, unfortunately, time and computation costs have not permitted for further examination of our models.

Chapter 7 : Conclusion and future work

This chapter provides a summary of the whole research project. On a scientific note, it is fitting to begin by referencing a passage from concluding discussion of the paper that motivated the work. There is some evidence that squalene monooxygenase inhibitors decrease LDL levels in dose-dependent manner in HepG2 cells, L6 myoblasts, rat, hamster and dogs similarly or greater than statins. Furthermore they decrease triglyceride, increase HDL and do not affect HMG-CoA reductase activity [2].

From a molecular modelling point of view, we have fully employed the theoretical studies to elucidate the impact of SM inhibitors on the cholesterol proteins with the aim of finding new compounds that can lower cholesterol. In this work, we have not only found the best binding ligands, SDZ 18 (-8.5 kcal mol^{-1}) and TNSA 84 (-8.0 kcal mol^{-1}) to the target proteins but more importantly demonstrated that the ligand stayed in the active side of proteins and furthermore uncovered the new theoretical insights. We believe that this is the first time that these ligands have been studied at the atomistic level.

Generally, the principal findings of this work were as follows:

1. Molecular docking was used to find the hit compounds using the scoring functions to calculate the negative sum of the energy of the ligand and protein, i.e., the most negative energy taken representing the better ligand-protein pose. The best binding ligands were SDZ 18 and TNSA as previously mentioned. Also noteworthy observations were the insights of the ligand-protein binding pocket which showed the hydrogen bond interaction with the ligand TNSA 84 and pi-pi stacking bond with SDZ 18. The molecular mechanics-generalized born surface area (MM-GBSA) was chosen as a unique approach that can be used to calculate the free energy in the binding pocket.
2. Molecular dynamics was also used to explore the stability and conformational flexibility of complex for 100 ns simulation. Perhaps the most convincing evidence for stability of ligand in the active site of the protein is determined by this method in modern science. Overall, the RMSD values that were obtained indicated that there were no significant fluctuations in the structure of the complex or changes resulting from the ligand binding in the active

side. The RMSF values calculated the average fluctuations in the binding pocket, and it was found that the RMSF plots showed minor residual fluctuations present in most of our compounds. In conclusion, .

3. Quantum mechanics calculations with DFT and ONIOM calculations were done to obtain further insight into our compounds and complex systems. In DFT, we have investigated the reactivity behavior of our molecules which is closely linked to their frontier molecular orbitals and also other characteristics were analyzed. The reactivity site of the molecules were determined with the electrostatic potential.

In conclusion, the theoretical calculations provided strong evidence that the antimycotic ligand, Terbinafine's derivative SDZ 97-469 (18) and the Trisnorsqualene alcohol (TNSA 84) bind to the target cholesterol proteins and the two ligands are maintained in the binding pocket during the 100 ns simulation. Thanks to the DFT technique, which was quite successful in the prediction of molecular properties and uncovering the chemical reactivity such as frontier molecular orbitals and the electrostatic potentials of the six ligands with high docking above -7.0 kcal mol^{-1} were SDZ 9, SDZ 13, SDZ 18, TNSA 81, TNSA 84, TNSA 85. The investigations done in this project has shed light on the idea that a universal medicine which is most powerful to perfect an imperfect body with very high cholesterol may result from the compounds. Great and wonderful are these compounds.

Future work

Finally, looking at the future, we herein propose an alternative method that in the future these compounds will be extracted from nature by the material scientist's skill and experiments done on them. The *in silico* results will therefore complement the future experimental work. The *in-vivo*, *in-vitro* and clinical trials should be performed as these ligands are therapeutic potential for lowering high cholesterol. They also show biological activity as it was investigated with the aid of quantitative structure-activity relationship (QSAR) and the results are shown in the appendix. In similar vein to previous researches, our findings suggest that more future analysis of the other properties of the SM inhibitors can in principle be carried out commencing with synthesis.

“The beginning of this divine science is the fear of the Lord, and its end is charity and love toward our neighbor. The all satisfying golden crop is properly devoted to the rearing and endowing of temples and hospices.’’

Jean d’Espagnet, *Hermetic Arcanum* (1628).

And so, in words of **John Bunyan** 1661,

“I penned it down, until at last it came to be, for the length and breadth, the bigness which you see.”

Mofeli Benedict Leoma

Grahamstown, South Africa

Rhodes university

23rd July 2023

References

[1] A. A. Sagatova, “Strategies to Better Target Fungal Squalene Monooxygenase,” 2021.

[2] A. Belter and M. Skupinska, “Squalene monooxygenase – a target for hypercholesterolemic,” no. July 2016, 2011, doi: 10.1515/BC.2011.195.

[3] S. Zaman *et al.*, “Cardiovascular disease and COVID-19: Australian and New Zealand consensus statement,” *Med. J. Aust.*, vol. 213, no. 4, pp. 182–187, 2020, doi: 10.5694/mja2.50714.

[4] G. Aggarwal *et al.*, “Association of Cardiovascular Disease With Coronavirus Disease 2019 (COVID-19) Severity: A Meta-Analysis,” *Curr. Probl. Cardiol.*, vol. 45, no. 8, p. 100617, 2020, doi: 10.1016/j.cpcardiol.2020.100617.

[5] A. Paramasivam, J. V. Priyadharsini, S. Raghunandhakumar, and P. Elumalai, “A novel COVID-19 and its effects on cardiovascular disease,” *Hypertens. Res.*, vol. 43, no. 7, pp. 729–730, 2020, doi: 10.1038/s41440-020-0461-x.

[6] M. Bansal, “Cardiovascular disease and COVID-19,” *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 14, no. 3, pp. 247–250, 2020, doi: 10.1016/j.dsx.2020.03.013.

[7] J. G. Robinson, G. Bakris, J. Torner, N. J. Stone, and R. Wallace, “Is it time for a cardiovascular primary prevention trial in the elderly?,” *Stroke*, vol. 38, no. 2, pp. 441–450, 2007, doi: 10.1161/01.STR.0000254602.58896.d2.

[8] A. Endo, “A historical perspective on the discovery of statins,” *Proc. Japan Acad. Ser. B Phys. Biol. Sci.*, vol. 86, no. 5, pp. 484–493, 2010, doi: 10.2183/pjab.86.484.

[9] N. K. Chua, H. W. Coates, and A. J. Brown, “Progress in Lipid Research Squalene monooxygenase : a journey to the heart of cholesterol synthesis,” *Prog. Lipid Res.*, vol. 79, no. March, p. 101033, 2020, doi: 10.1016/j.plipres.2020.101033.

[10] H. Li, X. H. Yu, X. Ou, X. P. Ouyang, and C. K. Tang, “Hepatic cholesterol transport and its role in non-alcoholic fatty liver disease and atherosclerosis,” *Prog. Lipid Res.*, vol. 83, no. May, p. 101109, 2021, doi: 10.1016/j.plipres.2021.101109.

[11] S. J. V. Vicente, G. R. Sampaio, C. K. B. Ferrari, and E. A. F. S. Torres, “Oxidation of Cholesterol in Foods and Its Importance for Human Health,”

Food Rev. Int., vol. 28, no. 1, pp. 47–70, 2012, doi: 10.1080/87559129.2011.594972.

[12] N. M. F. S. A. Cerqueira *et al.*, "Cholesterol Biosynthesis: A Mechanistic Overview," *Biochemistry*, vol. 55, no. 39, pp. 5483–5506, Oct. 2016, doi: 10.1021/acs.biochem.6b00342.

[13] G. I. Harisa and F. K. Alanazi, "Low density lipoprotein bionanoparticles: From cholesterol transport to delivery of anti-cancer drugs," *Saudi Pharm. J.*, vol. 22, no. 6, pp. 504–515, 2014, doi: 10.1016/j.jsps.2013.12.015.

[14] L. Liscum, "Cholesterol biosynthesis," *Biochem. Lipids, Lipoproteins Membr.*, no. Ldl, pp. 399–421, 2008, doi: 10.1016/B978-044453219-0.50016-7.

[15] R. N. Horodinschi, A. M. A. Stanescu, O. G. Bratu, A. P. Stoian, D. G. Radavoi, and C. C. Diaconu, "Treatment with statins in elderly patients," *Med.*, vol. 55, no. 11, pp. 1–11, 2019, doi: 10.3390/medicina55110721.

[16] S. L. Di Stasi, T. D. MacLeod, J. D. Winters, and S. A. Binder-Macleod, "Effects of statins on skeletal muscle: A perspective for physical therapists," *Phys. Ther.*, vol. 90, no. 10, pp. 1530–1542, 2010, doi: 10.2522/ptj.20090251.

[17] M. Thurnher, G. Gruenbacher, and O. Nussbaumer, "Regulation of mevalonate metabolism in cancer and immune cells," *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids*, vol. 1831, no. 6, pp. 1009–1015, 2013, doi: 10.1016/j.bbalip.2013.03.003.

[18] L. P.H., L. J.O., L. A.M., and L. L.A., "Treatment of statin adverse effects with supplemental Coenzyme Q 10 and statin drug discontinuation," *BioFactors*, vol. 25, no. 1–4, pp. 147–152, 2005, [Online]. Available: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L44070842%0Ahttp://za2uf4ps7f.search.serialssolutions.com/?sid=EMBASE&issn=09516433&id=doi:&atitle=Treatment+of+statin+adverse+effects+with+supplemental+Coenzyme+Q+10+and+statin+drug

[19] A. Belter, M. Skupinska, M. Giel-Pietraszuk, T. Grabarkiewicz, L. Rychlewski, and J. Barciszewski, "Squalene monooxygenase - a target for hypercholesterolemic therapy," *Biol. Chem.*, vol. 392, no. December, p., 2011, doi: 10.1515/bc-2011-195.

[20] J. Hu, Z. Zhang, W. Shen, and S. Azhar, "Cellular cholesterol delivery , intracellular processing and utilization for biosynthesis of steroid hormones," no. Table 1, pp. 7–9, 2010.

[21] P. M. R. Cruz, H. Mo, W. J. McConathy, N. Sabnis, and A. G. Lacko, "The

role of cholesterol metabolism and cholesterol transport in carcinogenesis: A review of scientific findings, relevant to future cancer therapeutics," *Front. Pharmacol.*, vol. 4 SEP, no. September, pp. 1–7, 2013, doi: 10.3389/fphar.2013.00119.

[22] S. Gill, J. Stevenson, I. Kristiana, and A. J. Brown, "Cholesterol-dependent degradation of squalene monooxygenase, a control point in cholesterol synthesis beyond HMG-CoA reductase," *Cell Metab.*, vol. 13, no. 3, pp. 260–273, 2011, doi: 10.1016/j.cmet.2011.01.015.

[23] D. W. Russell, "Cholesterol biosynthesis and metabolism," *Cardiovasc. Drugs Ther.*, vol. 6, no. 2, pp. 103–110, 1992, doi: 10.1007/BF00054556.

[24] H. R. Waterham, "Defects of cholesterol biosynthesis," *FEBS Lett.*, vol. 580, no. 23, pp. 5442–5449, 2006, doi: 10.1016/j.febslet.2006.07.027.

[25] M. Spanova and G. Daum, "Squalene - biochemistry, molecular biology, process biotechnology, and applications," *Eur. J. Lipid Sci. Technol.*, vol. 113, no. 11, pp. 1299–1320, 2011, doi: 10.1002/ejlt.201100203.

[26] M. M. Schumacher, R. Elsabrouty, J. Seemann, Y. Jo, and R. A. DeBose-Boyd, "The prenyltransferase UBIAD1 is the target of geranylgeraniol in degradation of HMG CoA reductase," *Elife*, vol. 2015, no. 4, pp. 1–21, 2015, doi: 10.7554/eLife.05560.

[27] I. Buhaescu and H. Izzedine, "Mevalonate pathway: A review of clinical and therapeutical implications," *Clin. Biochem.*, vol. 40, no. 9–10, pp. 575–584, 2007, doi: 10.1016/j.clinbiochem.2007.03.016.

[28] U. Schönbeck and P. Libby, "Inflammation, Immunity, and HMG-CoA Reductase Inhibitors," *Circulation*, vol. 109, no. 21_suppl_1, 2004, doi: 10.1161/01.cir.0000129505.34151.23.

[29] M. Y. M. Van Der Wulp, H. J. Verkade, and A. K. Groen, "Regulation of cholesterol homeostasis," *Mol. Cell. Endocrinol.*, vol. 368, no. 1–2, pp. 1–16, 2013, doi: 10.1016/j.mce.2012.06.007.

[30] P. T. Kovanen, E. A. Nikkila, and T. A. Miettinen, "Regulation of cholesterol synthesis and storage in fat cells," *J. Lipid Res.*, vol. 16, no. 3, pp. 211–223, 1975, doi: 10.1016/s0022-2275(20)36728-6.

[31] N. K. Chua, H. W. Coates, and A. J. Brown, "Squalene monooxygenase: a journey to the heart of cholesterol synthesis," *Prog. Lipid Res.*, vol. 79, no. April, p. 101033, 2020, doi: 10.1016/j.plipres.2020.101033.

[32] J. Ye and R. A. DeBose-Boyd, "Regulation of cholesterol and fatty acid synthesis," *Cold Spring Harb. Perspect. Biol.*, vol. 3, no. 7, pp. 1–13, 2011,

doi: 10.1101/cshperspect.a004754.

[33] H. Yoshioka, H. W. Coates, N. K. Chua, Y. Hashimoto, A. J. Brown, and K. Ohgane, “A key mammalian cholesterol synthesis enzyme, squalene monooxygenase, is allosterically stabilized by its substrate,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 13, pp. 7150–7158, 2020, doi: 10.1073/pnas.1915923117.

[34] A. Chugh, A. Ray, and J. B. Gupta, “Squalene epoxidase as hypocholesterolemic drug target revisited,” vol. 42, pp. 37–50, 2003.

[35] N. Gupta and T. D. Porter, “Inhibition of human squalene monoozygenase by selenium compounds,” *J. Biochem. Mol. Toxicol.*, vol. 16, no. 1, pp. 18–23, 2002, doi: 10.1002/jbt.10014.

[36] S. Sarvagalla, S. B. Syed, and M. S. Coumar, *An Overview of Computational Methods, Tools, Servers, and Databases for Drug Repurposing*. Elsevier Inc., 2019. doi: 10.1016/B978-0-12-816125-8.00025-0.

[37] B. N. Diallo, T. Swart, H. C. Hoppe, and Ö. T. Bishop, “Potential repurposing of four FDA approved compounds with antiplasmodial activity identified through proteome scale computational drug discovery and in vitro assay GROningen MAchine for Chemical Simulations,” *Sci. Rep.*, pp. 1–15, 2021, doi: 10.1038/s41598-020-80722-2.

[38] A. Matondo, W. Dendera, B. K. Isamura, H. V. S. Mambo, and ..., “In silico drug repurposing of anticancer drug 5-FU and analogues against SARS-CoV-2 main protease: molecular docking, pharmacokinetics and chemical reactivity,” *... Appl. ...*, 2022, [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9391940/%0Ahttps://chemrxiv.org/engage/chemrxiv/article-details/62628aa5d048edcdf94d1171

[39] W. K. Wootters, “Quantum entanglement as a quantifiable resource,” *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 356, no. 1743, pp. 1717–1731, 1998, doi: 10.1098/rsta.1998.0244.

[40] U. Manual, “QikProp 4.4”.

[41] F. Ahmad *et al.*, “Dioxinodehydroeckol: A Potential Neuroprotective Marine Compound Identified by In Silico Screening for the Treatment and Management of Multiple Brain Disorders,” *Mol. Biotechnol.*, no. 0123456789, 2022, doi: 10.1007/s12033-022-00629-3.

[42] K. N. Houk and P. Liu, “Using Computational Chemistry to Understand & Discover Chemical Reactions,” *Daedalus*, vol. 143, no. 4, pp. 49–66, 2014, doi: 10.1162/DAED_a_00305.

[43] I. Hypercube, *Hyperchem computational chemistry*. 419 Phillip street, Waterloo, Ontario N2L 3X2 Canada, 1996.

[44] T. Van Mourik, M. Bühl, and M. P. Gaigeot, "Density functional theory across chemistry, physics and biology," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 372, no. 2011, 2014, doi: 10.1098/rsta.2012.0488.

[45] W. J. Hehre, *A Guide to Molecular Mechanics and Quantum Chemical Calculations*.

[46] A. Scherrer, F. Agostini, D. Sebastiani, E. K. U. Gross, and R. Vuilleumier, "On the mass of atoms in molecules: Beyond the born-oppenheimer approximation," *Phys. Rev. X*, vol. 7, no. 3, pp. 1–23, 2017, doi: 10.1103/PhysRevX.7.031035.

[47] T. I. Adelusi *et al.*, "Molecular modeling in drug discovery," *Informatics Med. Unlocked*, vol. 29, no. February, p. 100880, 2022, doi: 10.1016/j.imu.2022.100880.

[48] T. M. Dhameliya, P. R. Nagar, and N. D. Gajjar, "Systematic virtual screening in search of SARS CoV-2 inhibitors against spike glycoprotein: pharmacophore screening, molecular docking, ADMET analysis and MD simulations," *Mol. Divers.*, vol. 26, no. 5, pp. 2775–2792, 2022, doi: 10.1007/s11030-022-10394-9.

[49] S. S. Jadav, "Structure-based discovery of small molecule APC-Asef interaction inhibitors : In silico approaches and molecular dynamics simulations," 2020.

[50] X. Du *et al.*, "Insights into Protein – Ligand Interactions : Mechanisms , Models , and Methods," no. i, pp. 1–34, doi: 10.3390/ijms17020144.

[51] G. Verma *et al.*, "Pharmacophore modeling, 3D-QSAR, docking and ADME prediction of quinazoline based EGFR inhibitors," *Arab. J. Chem.*, vol. 12, no. 8, pp. 4815–4839, 2019, doi: 10.1016/j.arabjc.2016.09.019.

[52] X. Meng, H. Zhang, M. Mezei, and M. Cui, "Molecular Docking : A Powerful Approach for Structure-Based Drug Discovery," pp. 146–157, 2011.

[53] D. Gazgalis, M. Zaka, B. H. Abbasi, D. E. Logothetis, M. Mezei, and M. Cui, "Protein Binding Pocket Optimization for Virtual High-Throughput Screening (vHTS) Drug Discovery," 2020, doi: 10.1021/acsomega.0c00522.

[54] H. M. Li, Y. H. Xie, C. Q. Liu, and S. Q. Liu, "Physicochemical bases for protein folding, dynamics, and protein-ligand binding," *Sci. China Life Sci.*,

vol. 57, no. 3, pp. 287–302, 2014, doi: 10.1007/s11427-014-4617-2.

[55] A. A. T. Naqvi, T. Mohammad, G. M. Hasan, and M. I. Hassan, "Advancements in Docking and Molecular Dynamics Simulations Towards Ligand-receptor Interactions and Structure-function Relationships," *Curr. Top. Med. Chem.*, vol. 18, no. 20, pp. 1755–1768, 2019, doi: 10.2174/1568026618666181025114157.

[56] O. M. H. Salo-ahen *et al.*, "and Pharmaceutical Development," pp. 1–60, 2021.

[57] R. D. Taylor, P. J. Jewsbury, and J. W. Essex, "A review of protein-small molecule docking methods," pp. 151–166, 2002.

[58] S. Huang, S. Z. Grinter, and X. Zou, "Scoring functions and their evaluation methods for protein – ligand docking : recent advances and future directions," pp. 12899–12908, 2010, doi: 10.1039/c0cp00151a.

[59] P. Śledź and A. Caflisch, "Protein structure-based drug design: from docking to molecular dynamics," *Curr. Opin. Struct. Biol.*, vol. 48, pp. 93–102, 2018, doi: 10.1016/j.sbi.2017.10.010.

[60] L. Zheng, A. A. Alhossary, C. K. Kwoh, and Y. Mu, "Molecular dynamics and simulation," *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.*, vol. 1–3, no. V, pp. 550–566, 2018, doi: 10.1016/B978-0-12-809633-8.20284-7.

[61] M. A. González, "Force fields and molecular dynamics simulations," no. January, 2017, doi: 10.1051/sfn/201112009.

[62] S. Makeneni, D. F. Thieker, and R. J. Woods, "Applying Pose Clustering and MD Simulations to Eliminate False Positives in Molecular Docking," *J. Chem. Inf. Model.*, vol. 58, no. 3, pp. 605–614, 2018, doi: 10.1021/acs.jcim.7b00588.

[63] C. Mulakala and V. N. Viswanadhan, "Could MM-GBSA be accurate enough for calculation of absolute protein/ligand binding free energies?," *J. Mol. Graph. Model.*, vol. 46, pp. 41–51, 2013, doi: 10.1016/j.jmgm.2013.09.005.

[64] E. Wang *et al.*, "End-Point Binding Free Energy Calculation with MM / PBSA and MM / GBSA : Strategies and Applications in Drug Design," 2019, doi: 10.1021/acs.chemrev.9b00055.

[65] H. Sahakyan, "Improving virtual screening results with MM / GBSA and MM / PBSA rescoring," *J. Comput. Aided. Mol. Des.*, vol. 35, no. 6, pp. 731–736, 2021, doi: 10.1007/s10822-021-00389-3.

[66] A. M. Tolah, L. M. Altayeb, T. A. Alandijany, V. D. Dwivedi, S. A. El-kafrawy, and E. I. Azhar, "Computational and In Vitro Experimental Investigations Reveal Anti-Viral Activity of Licorice and Glycyrrhizin against Severe Acute Respiratory Syndrome Coronavirus 2," no. November, 2021.

[67] J. M. Hayes and G. Archontis, "MM-GB (PB) SA Calculations of Protein-Ligand Binding Free Energies," 2011.

[68] P. W. Kenny, "The nature of ligand efficiency," *J. Cheminform.*, vol. 11, no. 1, pp. 1–18, 2019, doi: 10.1186/s13321-019-0330-2.

[69] S. Genheden and U. Ryde, "The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities," *Expert Opin. Drug Discov.*, vol. 10, no. 5, pp. 449–461, 2015, doi: 10.1517/17460441.2015.1032936.

[70] L. Baldini, A. Casnati, F. Sansone, and R. Ungaro, "Calixarene-based multivalent ligands," 2007, doi: 10.1039/b603082n.

[71] V. Limongelli, "Ligand binding free energy and kinetics calculation in 2020," no. October 2019, pp. 1–32, 2020, doi: 10.1002/wcms.1455.

[72] M. Schauperl *et al.*, "Binding Pose Flip Explained via Enthalpic and Entropic Contributions," 2017, doi: 10.1021/acs.jcim.6b00483.

[73] P. A. Greenidge, C. Kramer, J. C. Mozziconacci, and R. M. Wolf, "MM/GBSA binding energy prediction on the PDBbind data set: Successes, failures, and directions for further improvement," *J. Chem. Inf. Model.*, vol. 53, no. 1, pp. 201–209, 2013, doi: 10.1021/ci300425v.

[74] R. Kalirajan, A. Pandiselvi, B. Gowramma, and P. Balachandran, "Design, ADMET Screening, MM-GBSA Binding Free Energy of Some Novel Isoxazole Substituted 9-Anilinoacridines as HER2 Inhibitors Targeting Breast Cancer," pp. 118–128, 2019, doi: 10.2174/2589977511666190912154817.

[75] T. Tuccinardi and T. Tuccinardi, "Expert Opinion on Drug Discovery What is the current value of MM / PBSA and MM / GBSA methods in drug discovery ?," *Expert Opin. Drug Discov.*, vol. 16, no. 11, pp. 1233–1238, 2021, doi: 10.1080/17460441.2021.1942836.

[76] L. Xu, H. Sun, Y. Li, J. Wang, and T. Hou, "Assessing the Performance of MM/PBSA and MM/GBSA Methods. 3. The Impact of Force Fields and Ligand Charge Models," vol. 8421, no. 1, 2013.

[77] T. three Initiates, *The kybalion*. 1908.

[78] J. D. D. Tarika, X. D. D. Dexlin, S. Madhankumar, D. D. Jayanthi, and T.

J. Beaula, “Tuning the Computational Evaluation of Spectroscopic, ELF, LOL, NCI analysis and Molecular Docking of Novel Anti COVID-19 Molecule 4-Dimethylamino Pyridinium 3, 5-Dichlorosalicylate,” *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 259, p. 119907, 2021, doi: 10.1016/j.saa.2021.119907.

[79] J. Chen, F. fei Min, L. yun Liu, and C. fu Liu, “Mechanism research on surface hydration of kaolinite, insights from DFT and MD simulations,” *Appl. Surf. Sci.*, vol. 476, no. October 2018, pp. 6–15, 2019, doi: 10.1016/j.apsusc.2019.01.081.

[80] S. Saen-oon, O. Aruksakunwong, and K. Wittayanarakul, “Insight into analysis of interactions of saquinavir with HIV-1 protease in comparison between the wild-type and G48V and G48V / L90M mutants based on QM and QM / MM calculations,” vol. 26, pp. 720–727, 2007, doi: 10.1016/j.jmgm.2007.04.009.

[81] S. B. Dixit, S. Y. Ponomarev, and D. L. Beveridge, “Root Mean Square Deviation Probability Analysis of Molecular Dynamics Trajectories on DNA,” pp. 1084–1093, 2006.

[82] U. Manual, “Desmond 4.4”.

[84] Eugen Merzbacher, *Quantum Mechanics*. 1970.

[85] W. F. Polik, “Undergraduate chemistry educaton: Report of an ACS Presidential Symposium,” *J. Chem. Educ.*, vol. 83, no. 1, pp. 17–18, 2006, doi: 10.1021/ed083p17.

[88] J. Hutter and A. Curioni, “Car – Parrinello Molecular Dynamics on Massively Parallel Computers”, doi: 10.1002/cphc.200500059.

[89] H. M. Senn and W. Thiel, “Theoretical Chemistry QM / MM Methods for Biomolecular Systems Angewandte,” pp. 1198–1229, 2009, doi: 10.1002/anie.200802019.

[90] L. P. Wang and C. Song, “Car-Parrinello Monitor for More Robust Born-Oppenheimer Molecular Dynamics,” *J. Chem. Theory Comput.*, vol. 15, no. 8, pp. 4454–4467, 2019, doi: 10.1021/acs.jctc.9b00439.

[91] S. F. Sousa *et al.*, “Application of quantum mechanics / molecular mechanics methods in the study of enzymatic reaction mechanisms,” vol. 7, no. April, 2017, doi: 10.1002/wcms.1281.

[92] F. Remacle and R. D. Levine, “An electronic time scale in chemistry,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 18, pp. 6793–6798, 2006, doi: 10.1073/pnas.0601855103.

[93] R. P. M. E. V Subramanian and P. K. Chattaraj, “Variation of electrophilicity during molecular vibrations,” pp. 257–266, 2005, doi: 10.1007/s00214-005-0634-3.

[94] N. Argaman, G. Makov, and N. Argaman, “Density functional theory : An introduction Density functional theory : An introduction,” vol. 69, no. 2000, pp. 68–79, 2020, doi: 10.1119/1.19375.

[95] I. B. Obot, D. D. Macdonald, and Z. M. Gasem, “Density functional theory (DFT) as a powerful tool for designing new organic corrosion inhibitors . Part 1 : An overview,” *Corros. Sci.*, vol. 99, pp. 1–30, 2015, doi: 10.1016/j.corsci.2015.01.037.

[96] H. L. Woodcock, M. Hodošček, and B. R. Brooks, “Exploring SCC-DFTB paths for mapping QM/MM reaction mechanisms,” *J. Phys. Chem. A*, vol. 111, no. 26, pp. 5720–5728, 2007, doi: 10.1021/jp0714217.

[97] G. Gahungu and J. Zhang, “Molecular geometry , electronic structure and optical properties study of meridianal tris (8-hydroxyquinolinato) gallium (III) with ab initio and DFT methods,” vol. 755, pp. 19–30, 2005, doi: 10.1016/j.theochem.2005.06.017.

[98] A. Bendjeddou, T. Abbaz, S. Maache, R. Rehamnia, A. K. Gouasmia, and D. Villemin, “Quantum chemical descriptors of some paminophenyl tetrathiafulvalenes through density functional theory (DFT),” *Rasayan J. Chem.*, vol. 9, no. 1, pp. 18–26, 2016.

[99] I. Pangestika *et al.*, “Inhibitory effects of tangeretin and trans-ethyl caffeate on the HMG-CoA reductase activity: Potential agents for reducing cholesterol levels,” *Saudi J. Biol. Sci.*, vol. 27, no. 8, pp. 1947–1960, 2020, doi: 10.1016/j.sjbs.2020.06.010.

[100] A. P1, “Vibrational and DFT study of 5- (3-pyridyl-methylidene),” vol. 48, pp. 289–296, 2008, doi: 10.1016/j.vibspec.2008.01.012.

[101] H. M. Berman *et al.*, “The Protein Data Bank,” vol. 28, no. 1, pp. 235–242, 2000.

[102] L. MacroModel, Schrödinger, “Schrödinger Release 2023-2,” 2023.

[103] J. S. Scott and M. J. Waring, “Practical application of ligand efficiency metrics in lead optimisation,” *Bioorganic Med. Chem.*, vol. 26, no. 11, pp. 3006–3015, 2018, doi: 10.1016/j.bmc.2018.04.004.

[104] S. Schultes, C. De Graaf, E. E. J. Haaksma, I. J. P. De Esch, R. Leurs, and O. Kra, “Fragment based drug discovery Ligand efficiency as a guide in fragment hit selection and optimization,” pp. 157–162, 2010, doi:

10.1016/j.ddtec.2010.11.003.

[105] Planey, “Lipophilicity Indices for Drug Development,” *J. Appl. Biopharm. Pharmacokinet.*, no. January, pp. 6–12, 2013, doi: 10.14205/2309-4435.2013.01.01.6.

[106] U. Catacondensed and N. Computer, “References and notes (1),” pp. 31–36, 1988.

[107] A. Sinyani, K. Idowu, L. Shunmugam, H. Mathambo, R. Khan, and F. Group, “A molecular dynamics perspective into estrogen receptor inhibition by selective flavonoids as alternative therapeutic options,” *J. Biomol. Struct. Dyn.*, vol. 0, no. 0, pp. 1–13, 2022, doi: 10.1080/07391102.2022.2062786.

[108] and D. E. S. Kevin J. Bowers, Edmond Chow, Huafeng Xu, Ron O. Dror, Michael P. Eastwood, Brent A. Gregersen, John L. Klepeis, Istvan Kolossvary, Mark A. Moraes, Federico D. Sacerdoti, John K. Salmon, Yibing Shan, “calable Algorithms for Molecular Dynamics Simulations on Commodity Clusters.,” 2006.

[109] P. V. R. Rao, K. Srishailam, B. V. Reddy, and G. R. Rao, “Theoretical (DFT) and experimental (FT-IR & FT Raman) approach to investigate the molecular geometry and vibrational properties of 2 , 5- and 2 , 6-dihydroxytoluenes,” *J. Mol. Struct.*, vol. 1240, p. 130617, 2021, doi: 10.1016/j.molstruc.2021.130617.

[110] K. Roy Dennington, Todd A. Keith, and John M. Millam, Semichem Inc., Shawnee Mission, “Gaussview, Version 6.1.” 2016.

[111] C. Tuma, A. D. Boese, and N. C. Handy, “Predicting the binding energies of H-bonded complexes: A comparative DFT study,” *Phys. Chem. Chem. Phys.*, vol. 1, no. 17, pp. 3939–3947, 1999, doi: 10.1039/a904357h.

[112] L. Rao, H. Ke, G. Fu, X. Xu, and Y. Yan, “Performance of several density functional theory methods on describing hydrogen-bond interactions,” *J. Chem. Theory Comput.*, vol. 5, no. 1, pp. 86–96, 2009, doi: 10.1021/ct800237n.

[113] K. E. Riley, B. T. Op’t Holt, and K. M. Merz, “Critical assessment of the performance of density functional methods for several atomic and molecular properties,” *J. Chem. Theory Comput.*, vol. 3, no. 2, pp. 407–433, 2007, doi: 10.1021/ct600185a.

[114] F. F. Contreras-torres and V. A. Basiuk, “Imidazo [1 , 2- a] pyrazine-3 , 6-diones Derived from r -Amino Acids : A Theoretical Mechanistic Study of Their Formation via Pyrolysis and Silica-Catalyzed Process,” pp. 7431–

7440, 2006.

[115] M. I. Ban, “L . L . Stach6 , M . I . B6n , Global searching of reaction paths Main results,” vol. 11, pp. 405–421, 1992.

[116] R. W. W. Hooft, C. Sander, and G. Vriend, “Objectively judging the quality of a protein structure from a Ramachandran plot,” vol. 13, no. 4, pp. 425–430, 1997.

[117] P. Kumar and A. Arya, “Ramachandran plot - A simplified approach,” *Pathfind. Res. Train. Found. India*, no. January, p. 6, 2018, [Online]. Available: https://www.researchgate.net/publication/330158666_Ramachandran_plot-_A_simplified_approach

[118] R. A. Laskowski, N. Furnham, and J. M. Thornton, “The Ramachandran plot and protein structure validation,” *Biomol. Forms Funct. A Celebr. 50 Years Ramachandran Map*, pp. 62–75, 2012, doi: 10.1142/9789814449144_0005.

[119] W. Zhang *et al.*, “Exploring the Potential Mechanism of Guchang Zhixie Wan for Treating Ulcerative Colitis by Comprehensive Network Pharmacological Approaches and Molecular Docking Validation as Well as Cell Experiments,” *Chem. Biodivers.*, vol. 18, no. 1, 2021, doi: 10.1002/cbdv.202000810.

[120] A. K. Kahlon, S. Roy, and A. Sharma, “Molecular docking studies to map the binding site of squalene synthase inhibitors on dehydrosqualene synthase of staphylococcus aureus,” *J. Biomol. Struct. Dyn.*, vol. 28, no. 2, pp. 201–210, 2010, doi: 10.1080/07391102.2010.10507353.

[121] V. Fey, D. Jambulingam, H. Sara, S. Heron, C. Sipeky, and J. Schleutker, “Biocpr–a tool for correlation plots,” *Data*, vol. 6, no. 9, 2021, doi: 10.3390/data6090097.

[122] H. Zhang, L. Zhang, C. Gao, R. Yu, and C. Kang, “Pharmacophore screening, molecular docking, ADMET prediction and MD simulations for identification of ALK and MEK potential dual inhibitors,” *J. Mol. Struct.*, vol. 1245, p. 131066, 2021, doi: 10.1016/j.molstruc.2021.131066.

[123] C. H. Reynolds, S. D. Bembenek, and B. A. Tounge, “The role of molecular size in ligand efficiency,” *Bioorganic Med. Chem. Lett.*, vol. 17, no. 15, pp. 4258–4261, 2007, doi: 10.1016/j.bmcl.2007.05.038.

[124] D. Chemistry, “Impact of Lipophilic Efficiency on Compound Quality,” pp. 19–21, 2012.

[125] C. H. Reynolds, B. A. Tounge, and S. D. Bembenek, “Ligand binding efficiency: Trends, physical basis, and implications,” *J. Med. Chem.*, vol. 51, no. 8, pp. 2432–2438, 2008, doi: 10.1021/jm701255b.

[126] I. D. Kuntz, K. Chen, K. A. Sharp, and P. A. Kollman, “The maximal affinity of ligands,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 18, pp. 9997–10002, 1999, doi: 10.1073/pnas.96.18.9997.

[127] R. Then, “Dihydropteroate synthase,” *xPharm Compr. Pharmacol. Ref.*, pp. 1–7, 2007, doi: 10.1016/B978-008055232-3.60524-0.

[128] L. Shunmugam and M. E. S. Soliman, “Targeting HCV polymerase: A structural and dynamic perspective into the mechanism of selective covalent inhibition,” *RSC Adv.*, vol. 8, no. 73, pp. 42210–42222, 2018, doi: 10.1039/c8ra07346e.

[129] R. D. Jawarkar *et al.*, “ORIGINAL ARTICLE QSAR based virtual screening derived identification of a novel hit as a SARS CoV-229E 3CL pro Inhibitor : GA-MLR QSAR modeling supported by molecular Docking , molecular dynamics simulation and MMGBSA calculation approaches,” *Arab. J. Chem.*, vol. 15, no. 1, p. 103499, 2022, doi: 10.1016/j.arabjc.2021.103499.

[130] L. H. S. Santos, R. S. Ferreira, and E. R. Caffarena, “Chapter 2,” vol. 2053, 2019.

[131] A. Matondo *et al.*, “In silico Drug Repurposing of Anticancer Drug 5-FU and Analogues Against SARS-CoV-2 Main Protease : Molecular Docking , Molecular Dynamics Simulation , Pharmacokinetics and Chemical Reactivity Studies,” no. August, pp. 59–77, 2022.

[132] S. Details, P. Information, and L. Information, “Protein-Ligand RMSD,” pp. 1–10, 2023.

[133] B. K. Isamura and P. L. A. Popelier, “Toward a simple yet efficient cost function for the optimization of Gaussian process regression model hyperparameters,” no. August 2020, 2023, doi: 10.1063/5.0151033.

[134] K. M. Merz, “Using quantum mechanical approaches to study biological systems,” *Acc. Chem. Res.*, vol. 47, no. 9, pp. 2804–2811, 2014, doi: 10.1021/ar5001023.

[135] N. Ye, Z. Yang, and Y. Liu, “Applications of density functional theory in COVID-19 drug modeling,” *Drug Discov. Today*, vol. 27, no. 5, pp. 1411–1419, 2022, doi: 10.1016/j.drudis.2021.12.017.

[136] W. Kohn, A. D. Becke, and R. G. Parr, “Density Functional Theory of

Electronic Structure," vol. 0, no. 96, pp. 12974–12980, 1996.

[137] L. D. Nghiem, "Pharmaceutical Retention Mechanisms by Nanofiltration Membranes," vol. 39, no. 19, pp. 7698–7705, 2005.

[138] M. Hagar, H. A. Ahmed, G. Aljohani, and O. A. Alhaddad, "Investigation of some antiviral N-heterocycles as COVID 19 drug: Molecular docking and DFT calculations," *Int. J. Mol. Sci.*, vol. 21, no. 11, 2020, doi: 10.3390/ijms21113922.

[139] R. W. P. Wagemans, J. H. Van Lenthe, P. E. De Jongh, A. J. Van Dillen, and K. P. De Jong, "Hydrogen Storage in Magnesium Clusters : Quantum Chemical Study," vol. 104, no. 16, pp. 16675–16680, 2005.

[140] V. N. Hegde *et al.*, "Structural , computational and in silico studies of 4-bromo-3-flurobenzonitrile as anti-Alzheimer and anti-Parkinson agents," *J. Biomol. Struct. Dyn.*, vol. 0, no. 0, pp. 1–25, 2023, doi: 10.1080/07391102.2023.2226755.

[141] S. Kenouche and C. Sandoval-ya, "The antioxidant capacity of myricetin . A molecular electrostatic potential analysis based on DFT calculations ☆," vol. 801, no. May, 2022, doi: 10.1016/j.cplett.2022.139708.

[142] M. J. Frisch, "Geometry Optimization with QM / MM , ONIOM , and Other Combined Methods . I . Microiterations," vol. 9627775, 2003.

[143] V. Tripathy and K. Raghavachari, "EE-ONIOM-CT Method to Efficiently Account for the Missing Interactions in ONIOM : Energies and Analytic Gradients," 2023, doi: 10.1021/acs.jctc.3c00497.

[144] S. Das *et al.*, "EnzyDock: Protein–Ligand Docking of Multiple Reactive States along a Reaction Coordinate in Enzymes," *J. Chem. Theory Comput.*, vol. 15, pp. 5116–5134, 2019, doi: 10.1021/acs.jctc.9b00366.

[145] K. S. Byun and K. Morokuma, "A new ONIOM implementation in Gaussian98 . Part I . The calculation of energies , gradients , vibrational frequencies and electric field derivatives ☪," 1999.

[146] C. Cao, J. Hu, W. Zhang, X. Xu, and D. Chen, "Toward a Larger Molecular Simulation on the Quantum Computer: Up to 28-Qubit System Accelerated by Point Group Symmetry," pp. 1–12, 2022.

[147] T. R. Cundari, J. U. N. Deng, and W. Fu, "PM3 (tm) Parameterization Using Genetic Algorithms," vol. 3, 2000.

[148] C. M. Clemente, L. Capece, and M. A. Mart, "Best Practices on QM / MM Simulations of Biological Systems," 2022, doi: 10.1021/acs.jcim.2c01522.

[149] C. Gonzalez and H. B. Schlegel, “Reaction Path Following In Mass-Weighted Internal Coordinates c ’/ y,” no. 1, pp. 5523–5527, 1990.

[150] B. R. Brooks *et al.*, “CHARMM : The Biomolecular Simulation Program,” 2009, doi: 10.1002/jcc.

[151] L. Deng, T. Ziegler, L. Fan, L. Deng, T. Ziegler, and L. Fan, “A combined density functional and intrinsic reaction coordinate study on the ground state energy surface of H2CO A combined density functional and intrinsic reaction coordinate study on the ground state energy surface of H 2 CO,” vol. 3823, no. 1993, 1999, doi: 10.1063/1.466129.

Appendix

Code that was used to generate the post docking matrices and the plots.

```
import csv
import pandas as pd
import csv

data = [
 ["ligands", "smiles", "6GQF", "3COX", "3D4S", "2ZCS", "3W7F", "1M17", "pIC50"],
 ["SDZ13", "CC(C)(C)C#C/C=C/CN(C)Cc1ccccc1", -6.5, -6.4, -5.2, -6.2, -5.2, -4.2, 5.412289],
 ["SDZ14", "CC(C)(C)C#C/C=C/CN(C)Cc1ccc(cc1)OC", -5.1, -5.0, -3.8, -5.0, -4.7, -3.2, 4.319664],
 ["SDZ15", "CC(C)(C)C#C/C=C/CN(C)Cc1cccc2sc(c3c12)cccc3", -7.9, -6.7, -5.8, -5.8, 0.0, -4.9, 6.431798],
 ["SDZ16", "CC(C)(C)C#CC/C=C/CNCc1cccc(c12)cccc2", -6.4, -5.4, -4.0, -5.3, -6.5, -4.4, 5.665546],
 ["SDZ17", "CC(C)(C)C#CC/C=C/COCc1cccc(c12)cccc2", -6.8, -5.6, -5.8, -5.3, -6.4, -4.4, 6.585027],
 ["SDZ18", "CC(C)(C)C#CC/C=C/C[C@@H](C)Cc1cccc(c12)cccc2", -7.5, -6.5, -6.0, -6.8, -8.4, -6.3, 6.494850],
 ["SDZ9", "CC(C)(C)C#C/C=C/CN(C)Cc(ccc1)c(c12)scc2Cl", -7.7, -6.9, -6.1, -5.3, -7.3, -5.0, 4.301030],
 ["TNSA73", "CC(C)=CCC/C(C)=C/CC/C(C)=C/CC\\C=C(C)\\CC\\C=C(C)\\C=C(C)\\C=C/C(=O)O", -6.4, -6.7, -4.8, 0.0, -6.7, -
4.5, 5.397940],
 ["TNSA74", "OCCC/C(C)=C/CC/C(C)=C/CC\\C=C(C)\\CCC=C(C)C", -7.1, -5.6, -5.7, 0.0, -6.8, -4.4, 4.000000],
 ["TNSA75", "CC(C)=CCC\\C=C(C)\\CC\\C=C(C)\\CCCO", -3.2, -3.0, -3.0, -4.2, -3.8, -2.7, 3.397940],
 ["TNSA76", "CC(C)=CCC/C(C)=C/CC/C(C)=C/CC\\C=C(C)\\CC\\C=C(C)\\C=C/C(=O)O", -7.0, -6.8, -5.3, 0.0, -6.2, -5.6,
4.397940],
 ["TNSA77", "CC(C)=CCC/C(C)=C/CC/C(C)=C/CC\\C=C(C)\\CC\\C=C(C)\\C=C/C(=O)O", -7.2, -5.6, -5.6, 0.0, -6.7, -4.6,
5.397940],
 ["TNSA78", "CC(C)=CCC/C(C)=C/CC/C(C)=C/CC\\C=C(C)\\CC\\C=C(C)\\C=C/C(=O)O", -7.2, -5.6, -5.1, -5.4, -6.7, -4.6,
5.397940],
 ["TNSA79", "CC(C)=CCC/C(C)=C/CC/C(C)=C/CC\\C=C(C)\\CC\\C=C(C)\\C=C/C(=O)O", -6.8, -5.5, -5.3, 0.0, -5.2, -3.7,
4.522879],
 ["TNSA80", "CCC#CNCCC/C(C)=C/CC/C(C)=C/CC\\C=C(C)\\CC\\C=C(C)\\C=C/C(=O)O", -4.7, -7.2, -4.8, -4.9, 0.0, -2.8,
5.698970],
 ["TNSA81", "FC(F)=CCC/C(C)=C/CC/C(C)=C/CC\\C=C(C)\\CCC=C(C)C", -8.1, -5.9, -5.2, -5.4, -6.6, -4.3, 5.267606],
 ["TNSA82", "FC(F)=CC/C(C)=C/CC/C(C)=C/CC\\C=C(C)\\CC\\C=C(C)\\C=C(C)\\C=C/C(=O)O", -7.2, -6.1, -5.6, -5.6, -5.6, -
5.7, 5.346787],
 ["TNSA83", "FC(F)=CC/C(C)=C/CC/C(C)=C/CC\\C=C(C)\\CC\\C=C(C)\\C=C(C)\\C=C/C(=O)O", -7.2, -6.1, -5.6, -5.6, -5.6, -
5.7, 5.000000],
```

```
 ["TNSA84", "CC(C)=CCCC(\\C=O)=C\\CC/C(C)=C/CC\\C=C(C)\\CC\\C=C(C)\\C=C/C(=O)O", -5.2, -8.4, -5.7, 0.0, -6.1, -4.1,
4.154902],
 ["TNSA85", "CC(C)=CCC/C(CN)=C/CC/C(C)=

]

filename = "data.csv"

with open(filename, "w", newline="") as csvfile:

 writer = csv.writer(csvfile)

 writer.writerows(data)

print("CSV file created successfully!")

df = pd.read_csv('data.csv')

ndf = df

ndf.head(9)

ndf = df.drop(["Ligands"], axis=1)

ndf = ndf.astype(float)



  ##########################################
  #                                        #
  # correlations using numpy and searbon   #
  #                                        #
  ##########################################

# nnumpy
import numpy as np

n = 23
```

```
gqf = ndf['6GQF'].tolist()

#p = ndf['3COX'].tolist()

d4s = ndf['3D4S'].tolist()

zcs = ndf['2ZCS'].tolist()

w7f = ndf['3W7F'].tolist()

m17 = ndf['1M17'].tolist()


Xresults = np.array([gqf,x,d4s,zcs,w7f,m17])

X = np.array(Xresults).reshape(n,6)

X_train = X[: n // 2]

X_test = X[n // 2 :]

print("Corr(X)")

print(np.round(np.corrcoef(X.T), 2))



#seaborn
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

correlation_matrix = ndf.corr()

# Plot the correlation heatmap using Seaborn
sns.heatmap(correlation_matrix, annot=True, cmap='viridis')
```

```
plt.title('Correlation Heatmap')

plt.show()


  ##########################################
  #                                        #
  # post docking matric calculation        #
  #                                        #
  ##########################################


from rdkit import Chem
from rdkit.Chem import AllChem
from rdkit.Chem import QED

def matric_index(ligands_name,smiles_list,docking_list,pic50_list):
 LE2 = []
 HA_list = []
 names = []
 SEI_list = []
 BEI_list = []
 LLE_list = []
 LELP_list = []
 lig_eff = []
 molecules =[]
 #smiles = df2["smiles"].tolist()
 #docking = df2["docking"].tolist()

 for p,i,j,k in zip(ligands_name,smiles_list,docking_list,pic50_list):

 name = p

 molecules.append(Chem.MolFromSmiles(i))

 molecule = Chem.MolFromSmiles(i)

 adme = QED.properties(molecule)
```

```
per = QED.default(molecule)

heavy_atoms = molecule.GetNumHeavyAtoms()

HA_list.append(heavy_atoms)

#ligand efficincy
# lig eff = deltG/N (equation)
ligind_eff = round(j/heavy_atoms,2)

lig_eff.append(ligind_eff)

LE_2 = round(k/heavy_atoms,2)

LE2.append(LE_2)

#ligand lipophilic efficiency
#LELP = logp/ligand eff (equation)
if ligind_eff ==0:

 LELP = None

 print(f'{p}: ligand effency = {ligind_eff}')

else:

 if adme[1]!=0:

 LELP = round(adme[1]/ligind_eff,2)

 elif adme[0] ==0:

 LELP = None

 else:

 pass
```

```
LELP_list.append(LELP)

#LLE = PIC50 - LOGP
LLE = round(k-adme[1],2)

LLE_list.append(LLE)

#BEI =-(deltG/MW) equantion
BEI = round(-(j/(adme[0])),2)

BEI_list.append(BEI)

# SEI = -(delG/(PSA/100))
if adme[4]==0:

 SEI = None

else:

 SEI =round( -(j/(adme[4]/100)),2)

SEI_list.append(SEI)

names.append(p)

data = pd.DataFrame({'ligands' : names,
 'HA': HA_list,
 'ligand Efficiancy' : lig_eff,
 'LE(pIc50/HA)': LE2,
 'LLEP' : LELP_list,
 'LLE' : LLE_list,
 'BEI' : BEI_list,
 'SEI' :SEI_list },
 columns=['ligands','HA','ligand Efficiancy','LE(pIc50/HA)', 'LLEP','LLE', 'BEI', 'SEI'])

return data
```

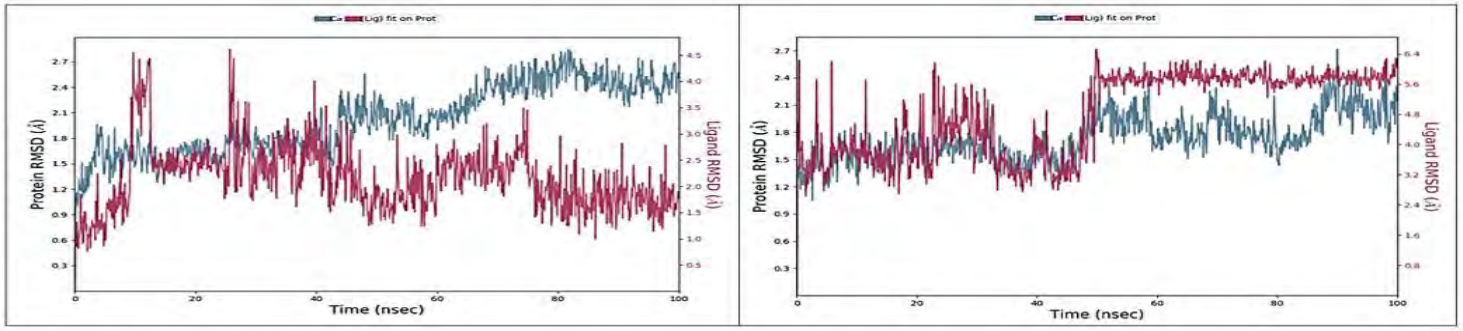
```
''''Calculating post docking matrices for each ligand against the proteins 2ZCS, 6GQF, 3D4S, 3COX, 1M17, 3W7F'''

M17 = matric_index(df2['ligands'],df2['smiles'],df2['1M17'],df2['pIC50'])

M17 = M17.fillna(0)

M17

cox = matric_index(df2['ligands'],df2['smiles'],df2['3COX'],df2['pIC50'])

cox = cox.fillna(0)

GQF = matric_index(df2['ligands'],df2['smiles'],df2['6GQF'],df2['pIC50'])

GQF = GQF.fillna(0)

ZCS = matric_index(df2['ligands'],df2['smiles'],df2['2ZCS'],df2['pIC50'])

ZCS = ZCS.fillna(0)

D4= matric_index(df2['ligands'],df2['smiles'],df2['3D4S'],df2['pIC50'])

D4 = D4.fillna(0)

W7 = matric_index(df2['ligands'],df2['smiles'],df2['3W7F'],df2['pIC50'])

W7 = W7.fillna(0)

W7


  ##########################################
  #                                        #
  # Plots for analysis                     #
  #                                        #
  ##########################################
```

```
import matplotlib.pyplot as plt
import numpy as np

X = M17["HA"]

#line plot
x1= np.array([8,9,12,19,23,30,35])

y1 =np.array([0,-0.29,-0.35,-0.20,-0.25,-0.19,-0.22])

y = np.sin(x1) + np.random.normal(0, 0.2, len(x1))

Y1 = M17["ligand Efficiancy"]
Y2 = cox["ligand Efficiancy"]
Y3 = GQF["ligand Efficiancy"]
Y4 = ZCS["ligand Efficiancy"]
Y5 = D4["ligand Efficiancy"]
Y6 = W7["ligand Efficiancy"]

spline = np.linspace(x1.min(), x1.max(), 1000)

smooth_curve = np.interp(spline, x1, y1)

plt.figure()

plt.scatter(X, Y6, color='purple', label='3W7F')

# Plot the line graph
plt.plot(spline, smooth_curve, linewidth=2)


plt.xlabel('HA')

plt.ylabel('LE')

plt.title('Multiple Scatter Plots')
```

```
plt.xlim(0, 50)

plt.ylim(-0.6, 1)

plt.legend()

plt.show()

X = M17["HA"]

Y6 = W7["LE(pIc50/HA)"]

plt.figure()

plt.scatter(X, Y6, color='purple', label='3W7F')

plt.xlabel('HA')

plt.ylabel('LE(pIc50/HA)')

plt.title('Multiple Scatter Plots')

plt.xlim(0, 50)

plt.ylim(-0, 1)

plt.show()

##

X = M17["HA"]

Y1 = M17["LLEP"]
Y2 = cox["LLEP"]
Y3 = GQF["LLEP"]
Y4 = ZCS["LLEP"]
Y5 = D4["LLEP"]
```

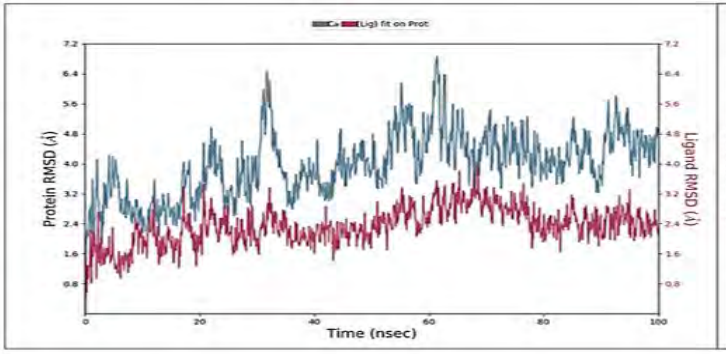
```
Y6 = W7["LLEP"]

plt.figure()

plt.scatter(X, Y1, color='red', label='1M17')
plt.scatter(X, Y2, color='blue', label='3COX')
plt.scatter(X, Y3, color='green', label='6GQF')
plt.scatter(X, Y4, color='orange', label='2ZCS')
plt.scatter(X, Y5, color='yellow', label='3D4S')
plt.scatter(X, Y6, color='purple', label='3W7F')

plt.xlabel('HA')

plt.ylabel('LLEP')

plt.title('Multiple Scatter Plots')

plt.xlim(0, 50)

plt.ylim(-60, 20)

plt.legend()

plt.show()
```

Molecular dynamics calculations

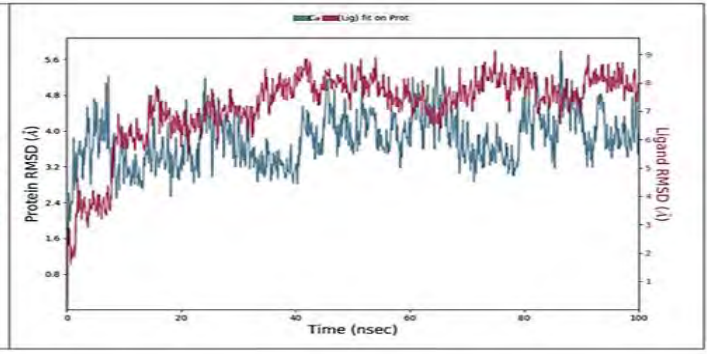


(a) Protein 2ZCS and ligand SDZ13.

(b) Protein 2ZCS and ligand SDZ15.



(e) Protein 2ZCS and ligand SDZ15.



(f) Protein 3D4S and ligand TNSA84.

Figure A 1 The RMSD Plots for 100 ns simulation for ligands and proteins.

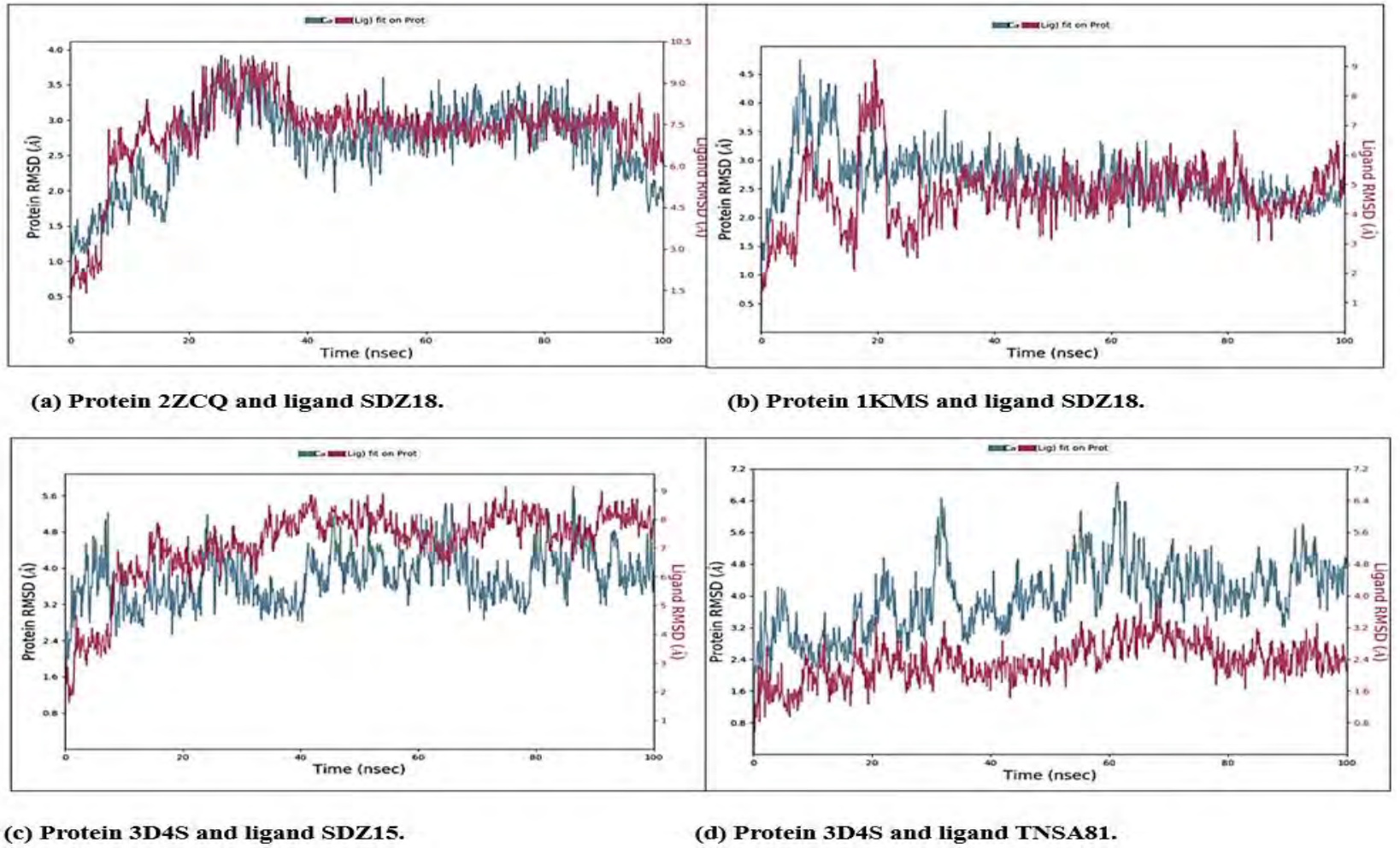
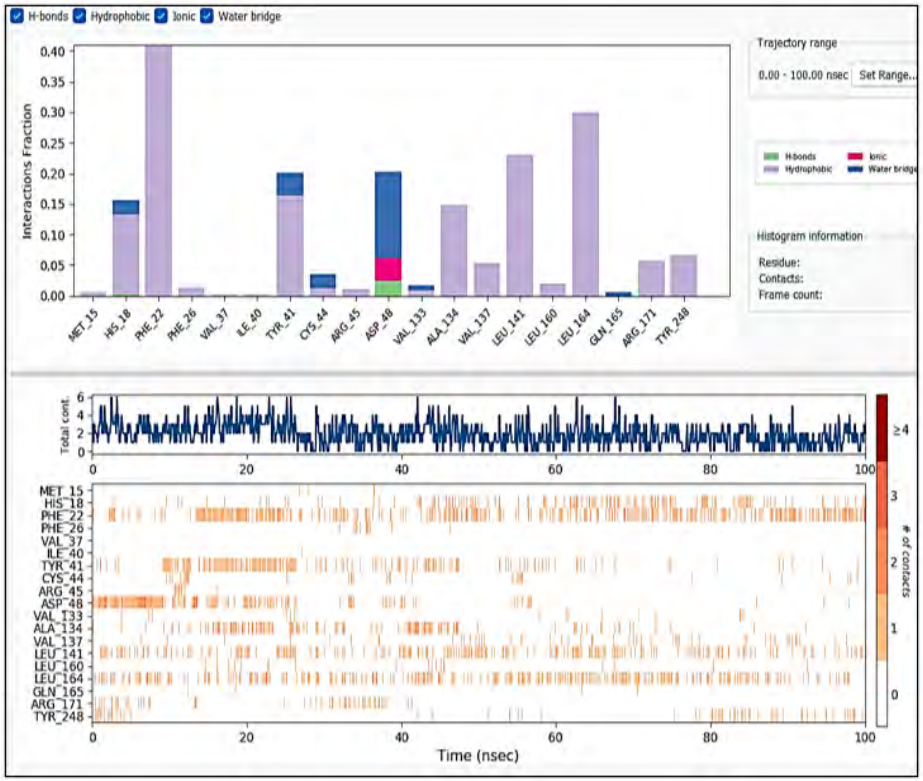
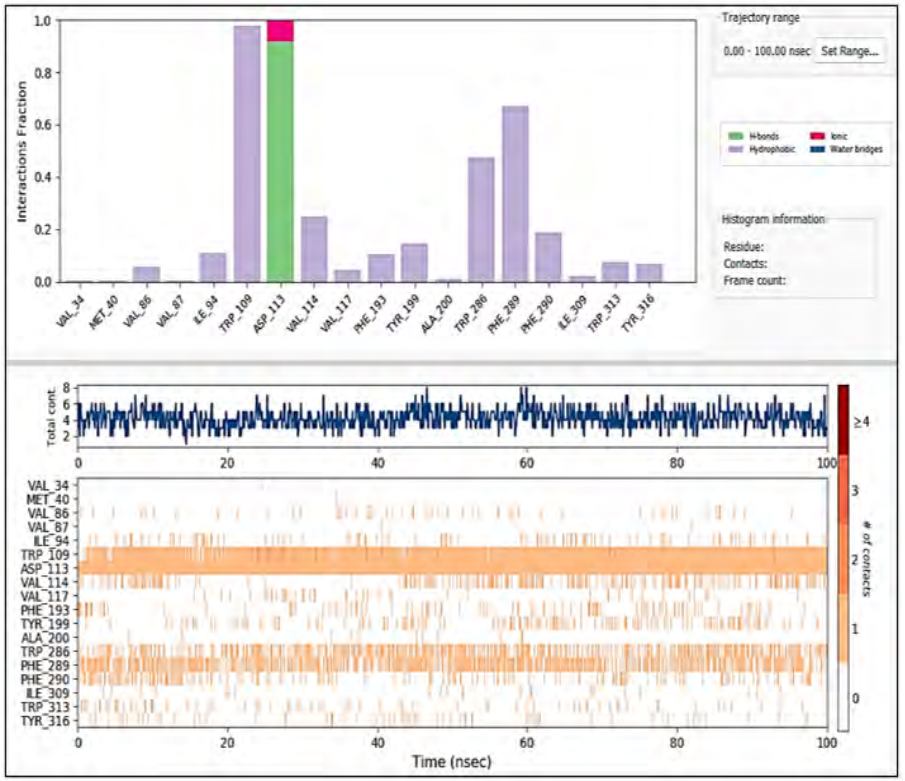


Figure A 2 The RMSD Plots for 100 ns simulation for ligands and proteins.

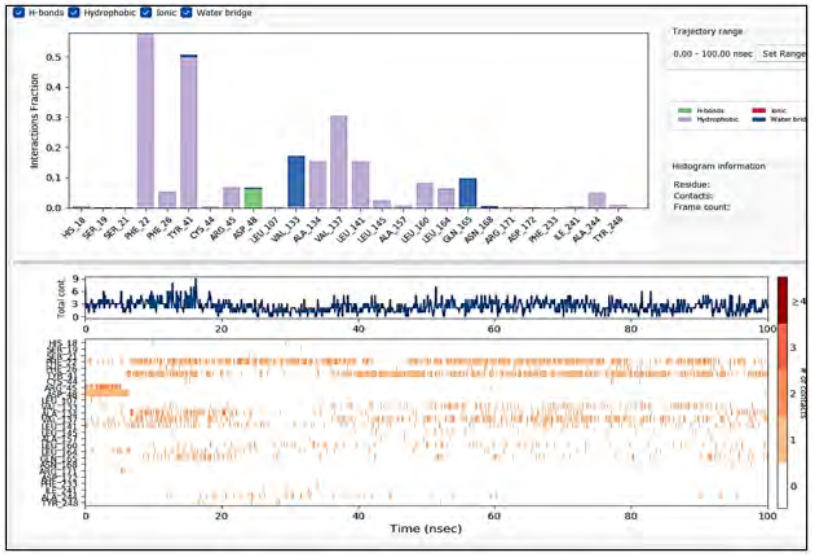


(a) Protein 2ZCS and ligand SDZ15.

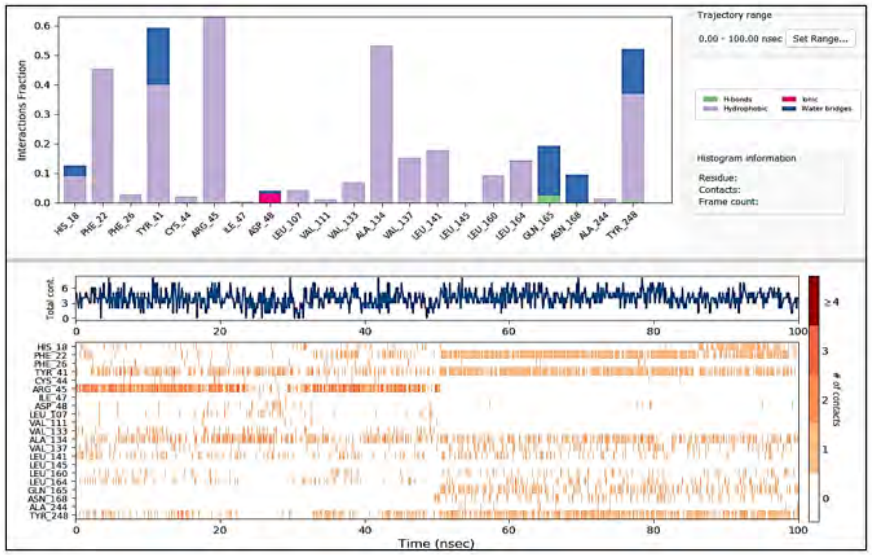


(b) Protein 2ZCS and ligand SDZ15.

Figure A 3 The Protein-ligand contacts and contacts timeline diagram.

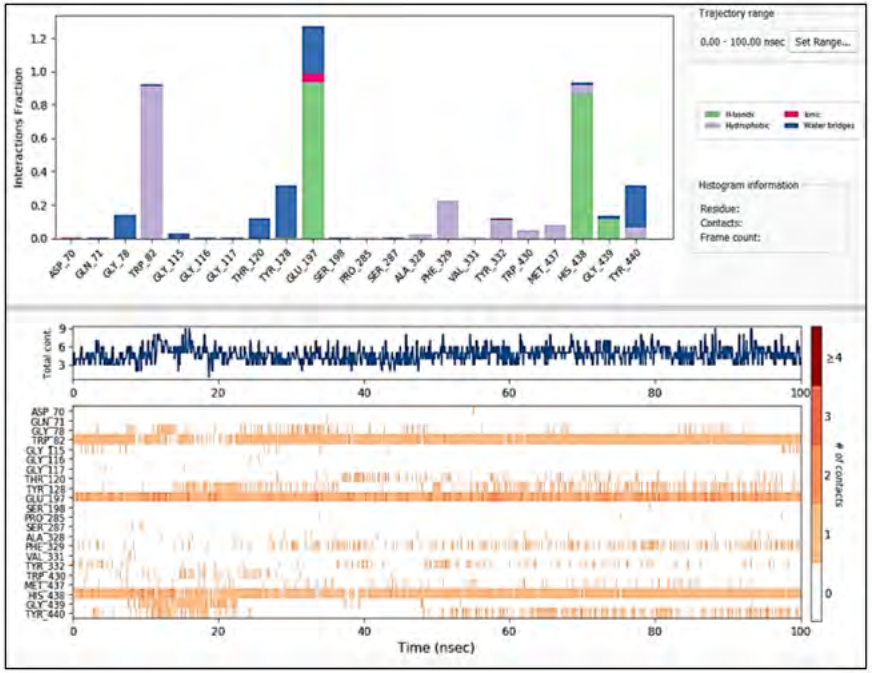


(a) Protein 1KMS and ligand SDZ18.

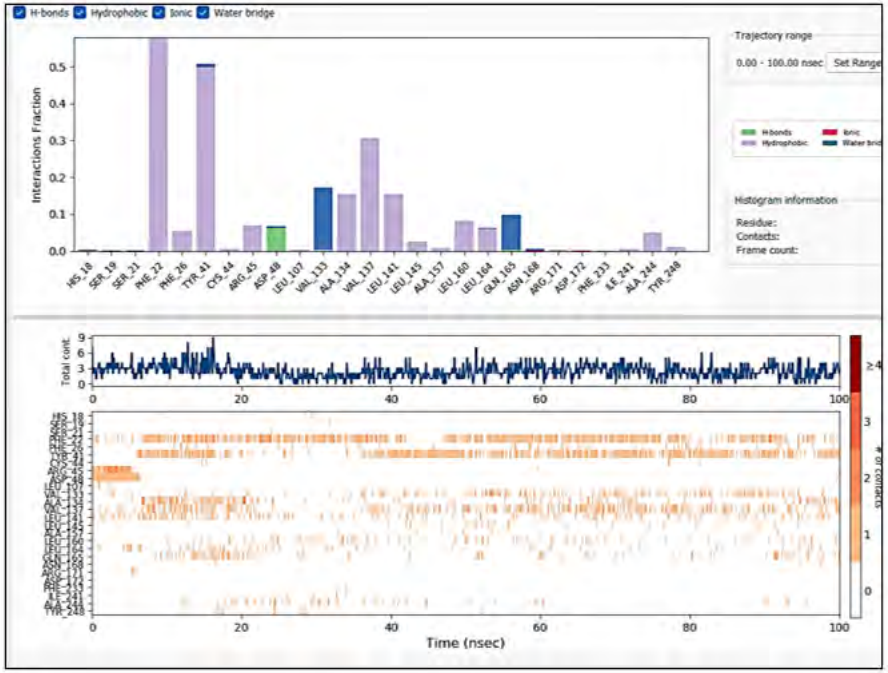


(b) Protein 3D4S and ligand TNSA84.

Figure A 4 The Protein-ligand contacts and contacts timeline diagram.

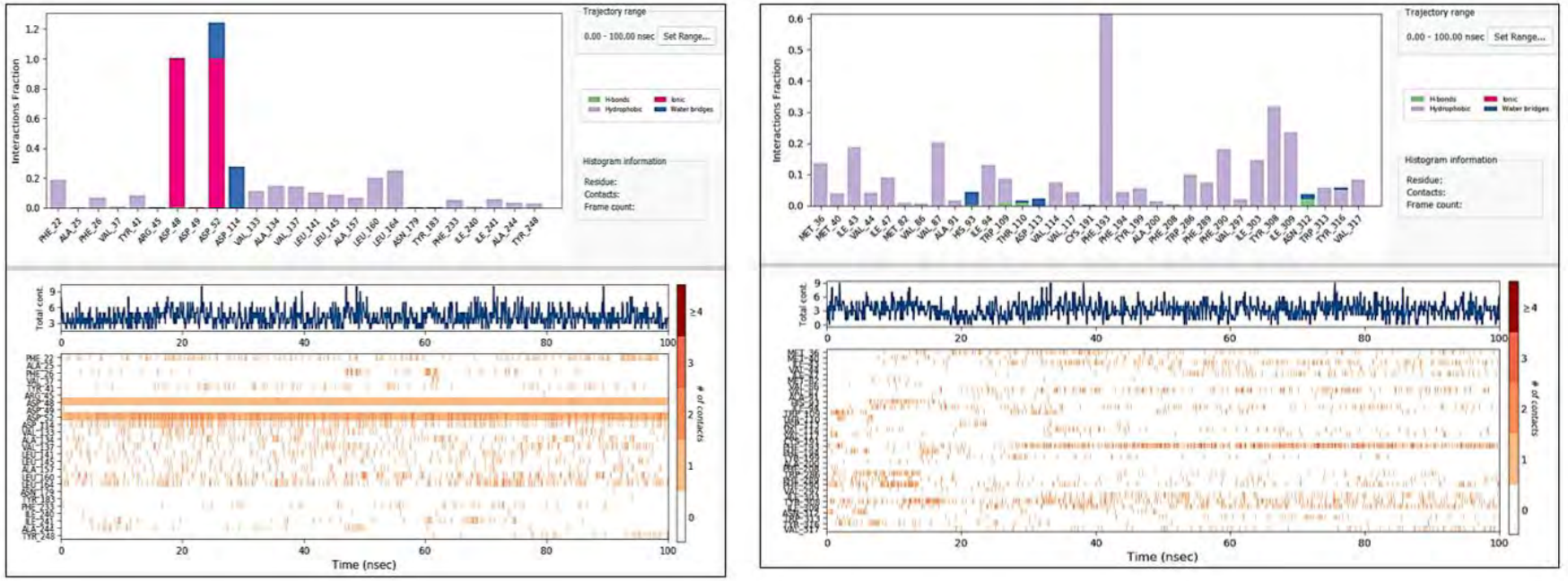


(a) Protein 2ZCQ and ligand SDZ18.



(b) Protein 1KMS and ligand SDZ18.

Figure A 5 The Protein-ligand contacts and contacts timeline diagram.



(a) Protein 2ZCQ and ligand TNSA74.

(b) Protein 1KMS and ligand SDZ18.

Figure A 6 The Protein-ligand contacts and contacts timeline diagram.

Pharmacophore modelling, 3D-QSAR and ADME prediction

A library of **106 compounds** (obtained from squalene monooxygenase – a target for hypercholesterolemic therapy paper by Agnieszka Belter et al) was reduced to **23 compounds** for the further study using the pharmacophore modelling, 3D-QSAR and AMDET techniques and results are shown in this section.

1. Pharmacophore modelling results

(a) Antimycotic squalene monooxygenase inhibitors: Naftifine and its derivatives

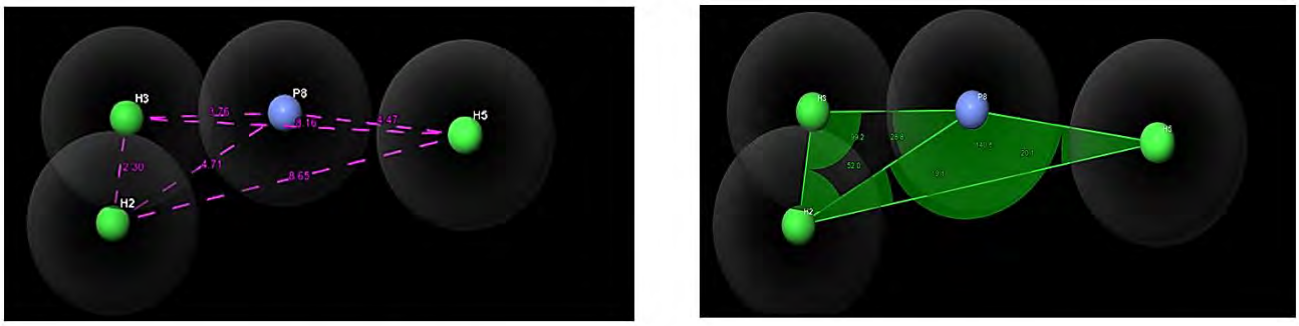


Figure A 7 The pharmacophores with intersite distances and angles. All distances are in A. Pink spheres with arrow, hydrogen bond acceptor (A), green sphere, hydrophobe (H), blue, Donor (D).

Table A 1: Pharmacophore distances.

Entry	Distance (Å)
H2 to H3	2.30
H3 to P8	3.26
H2 to P8	4.71
H2 to H5	8.65
P8 to H5	4.47
H3 to H5	8.15

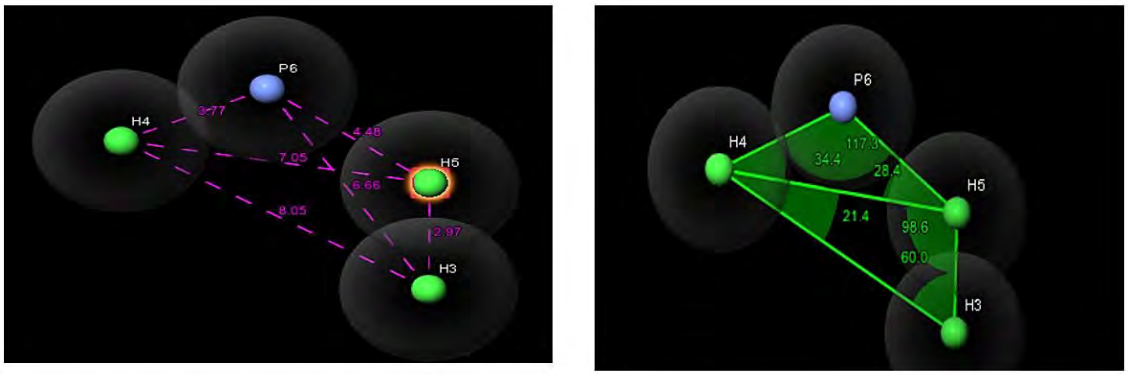
Table A 2 : Pharmacophore angles

Angle	Site 1	Site 2	Site 3	Angle value
H2P9H3	H2	P9	H3	28.8
H2P9H5	H2	P9	H5	20.1
H2H3P8	H2	H3	P8	99.2
H3H2P9	H3	H2	P9	52.0
P8H2H5	P8	H2	H5	19.1
H2P8H5	H2	P8	H5	140.5

Table A 3 : The results of hypothesis shown below as vector, volume, site scores, survival scores.

Pharmacophore	Survival Score	Site Score	Vector Score	Volume Score	PhaseHypo Score
HHHP_1	3,889	0,816	0	0,582	1,143
HHHP_2	3,895	0,787	0	0,606	1,142
HHHP_3	3,874	0,766	0	0,597	1,141
HHHP_4	3,837	0,733	0	0,559	1,138
HHHP_5	3,824	0,752	0	0,573	1,136
HHHP_6	3,784	0,710	0	0,601	1,135

(b) Antimycotic squalene monooxygenase inhibitors. Terbinafine and its derivatives



(a) The pharmacophoric distances between the points

(b) The pharmacophoric angles between the intersites.

Figure A 8 The pharmacophores with intersite distances and angles. All distances are in A. Pink spheres with arrow, hydrogen bond acceptor (A), green sphere, hydrophobe (H), blue, Donor (D).

Table A 4 : Pharmacophores distances.

Entry	Distance (Å)
H4 to H3	8.05
H3 to H5	2.97
H4 to H5	7.05
P6 to H5	4.48
P6 to H4	3.77
P6 to H5	4.48

Table A 5 : Pharmacophore angles.

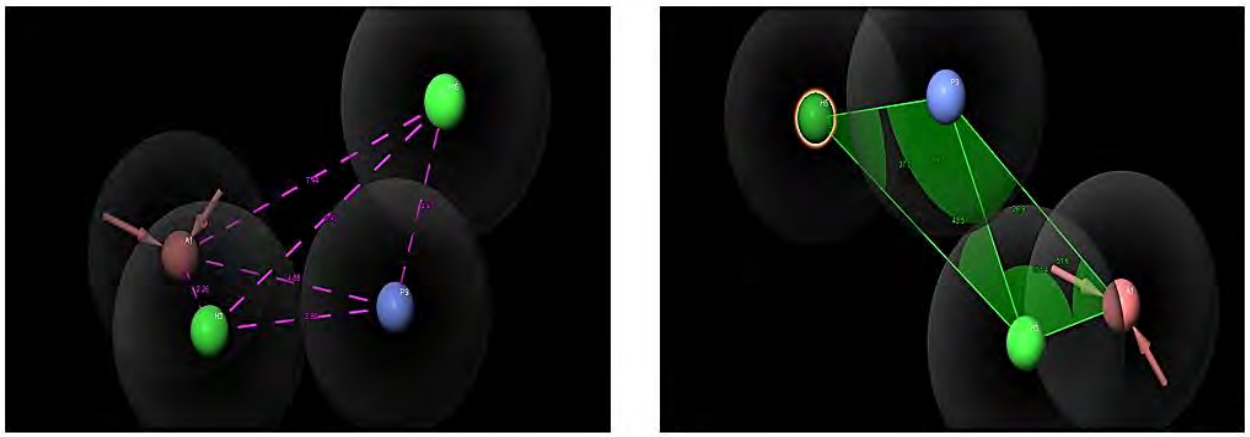
Angle	Site 1	Site 2	Site 3	Angle value
H4P6H5	H4	P6	H5	117.3
P6H4H5	P6	H4	H5	34.4
P6H5H4	P6	H5	H4	28.8
H4H5H3	H4	H5	H3	98.6
H4H3H5	H4	H3	H5	60.0
H5H4H3	H2	P8	H5	21.4

The results of hypothesis are shown below as vector, volume, site scores, survival scores and the best hypothesis was HHP_1. The hypothesis was selected for compounds and taken to the 3D-QSAR.

Table A 6 : Pharmacophore results.

Pharmacophore	Survival Score	Site Score	Vector Score	Volume Score	PhaseHypoScore
HHHP_1	3,782	0,86	0	0,673	0,952
HHHP_3	3,71	0,821	0	0,68	0,951
HHHP_4	3,707	0,836	0	0,663	0,951
HHHP_5	3,675	0,821	0	0,647	0,949
HHHP_6	3,657	0,769	0	0,669	0,948
HHHP_7	3,625	0,754	0	0,674	0,946
HHHP_8	3,598	0,749	0	0,598	0,944
HHHP_9	3,592	0,783	0	0,634	0,944
HHHP_10	3,549	0,736	0	0,625	0,941

(c) Antimycotic squalene monooxygenase inhibitors. Terbinafine's derivatives



(a) The pharmacophoric distances between the points **(b) The pharmacophoric angles between the intersites.**

Figure A 9 The pharmacophores with intersite distances and angles. All distances are in A. Pink spheres with arrow, hydrogen bond acceptor (A), green sphere, hydrophobe (H), blue, Donor (D).

Table A 7 : Pharmacophoric distances.

Entry	Distance (Å)
A1 to H3	2.26
A1 to P9	4.88
H3 to P9	3.98
H3 to H6	5.40
P9 to H6	4.47

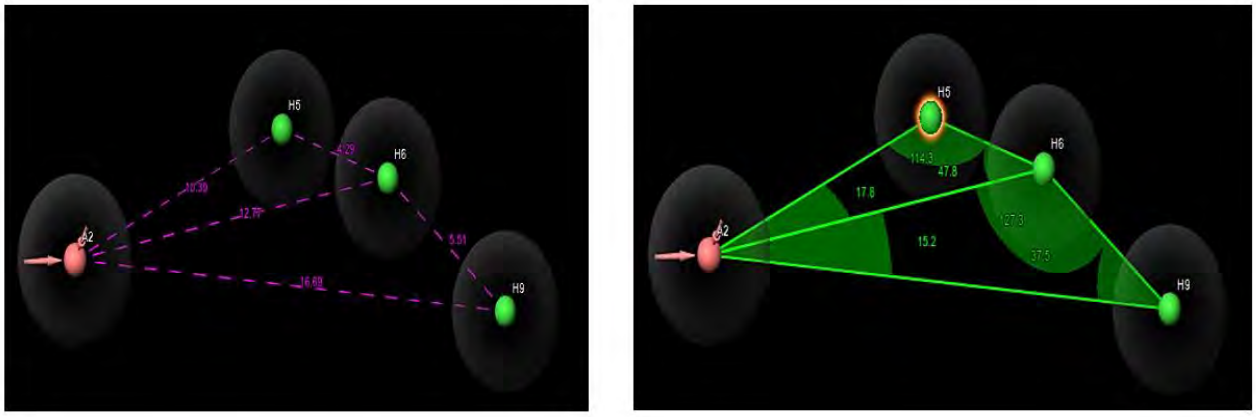
Table A 8 : Pharmacophoric angles.

Angle	Site 1	Site 2	Site 3	Angle
H6P9H3	H6	P9	H3	99.5
H3P9A1	H3	P9	A1	26.9
H6H3P9	H6	H3	P9	43.5
P9H6H3	P9	H6	H3	37.0
P9H3A1	P9	H3	A1	101.4
H3A1P9	H3	A1	P9	51.6

Table A 9 : Pharmacophore results.

Pharmacophore	Survival Score	Site Score	Vector Score	Volume Score	PhaseHypoScore
AHHHP_1	6,063	0,792	0,921	0,464	0,852
AHHHP_2	6,028	0,729	0,903	0,407	0,849
AHHHP_3	6,026	0,718	0,868	0,388	0,849
AHHHP_4	6,008	0,760	0,863	0,358	0,847
AHHHP_5	6,006	0,750	0,968	0,418	0,846
AHHHP_6	5,995	0,716	0,951	0,461	0,845
AHHHP_7	5,985	0,702	0,884	0,382	0,844
AHHHP_8	5,960	0,662	0,833	0,370	0,843
AHHHP_10	5,973	0,749	0,980	0,413	0,842
AHHHP_9	5,603	0,662	0,969	0,442	0,832
AHHP_1	5,556	0,749	0,967	0,445	0,856
AHHP_6	5,599	0,839	0,977	0,567	0,887
AHHP_2	5,546	0,826	0,905	0,576	0,798
AHHP_8	5,547	0,806	0,986	0,685	0,777
AHHP_7	5,587	0,821	0,966	0,697	0,786
AHHP_3	5,524	0,803	0,899	0,676	0,798
AHHP_10	5,560	0,804	0,867	0,666	0,798
AHHP_5	5,544	0,829	0,868	0,677	0,888
AHHP_9	5,560	0,779	0,900	0,777	0,780
AHHP_4	5,520	0,794	0,87	0,787	0,888

(d) Antimycotic squalene monooxygenase inhibitors. Terbinafine's derivatives



(a) The pharmacophoric distances between the points **(b) The pharmacophoric angles between the intersites.**

Figure A 10 The pharmacophores with intersite distances and angles. All distances are in A. Pink spheres with arrow, hydrogen bond acceptor (A), green sphere, hydrophobe (H), blue, Donor (D).

Table A 10 : Pharmacophoric distances.

Entry	Distance (Å)
A2 to H5	10.39
A2 to H6	12.77
A2 to H9	16.69
H5 to H6	4.29
H6 to H9	5.51

Table A 11: Pharmacophoric angles.

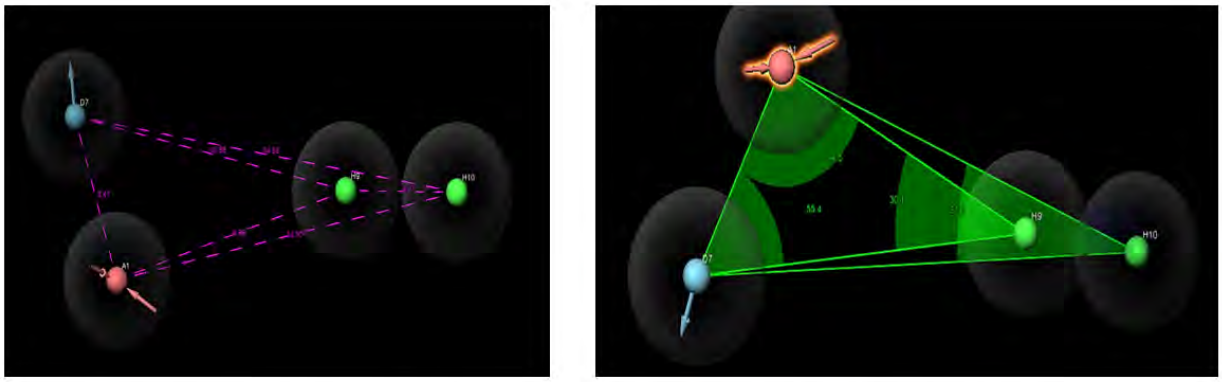
Angle	Site 1	Site 2	Site 3	Angle
A2H5H6	A2	H5	H6	114.3
H5A2H9	H5	A2	H9	17.8
H6A2H9	H6	A2	H9	15.2
H6H9A2	H6	H9	A2	37.5
H5H6A2	H5	H6	A2	47.8
A2H6H9	A2	H6	H9	127.3

The results of hypothesis are shown below as vector, volume, site scores, survival scores. The best hypothesis AHHH_1 was selected for compounds and take to the 3D-QSAR.

Table A 12 : Pharmacophore results.

Pharmacophore	Survival Score	Site Score	Vector Score	Volume Score	PhaseHypoScore
AHHH_1	3,994	0,574	0,921	0,464	0,852
AHHH_2	3,954	0,479	0,903	0,407	0,849
AHHH_3	3,953	0,622	0,868	0,388	0,849
AHHH_4	3,919	0,634	0,863	0,358	0,847
AHHH_5	3,907	0,468	0,968	0,418	0,846
DHHH_1	3,879	0,568	0,951	0,461	0,845
AHHH_6	3,863	0,645	0,884	0,382	0,844
AHHH_7	3,851	0,577	0,833	0,37	0,843
AHHH_8	3,843	0,615	0,9	0,413	0,842

(e) Antimycotic squalene monooxygenase inhibitors



(a) The pharmacophoric distances between the points. **(b) The pharmacophoric angles between the intersites.**

Figure A 11 The pharmacophores with intersite distances and angles. All distances are in A. Pink spheres with arrow, hydrogen bond acceptor (A), green sphere, hydrophobe (H), blue, Donor (D).

.

Table A 13 : Pharmacophoric distances.

Entry	Distance (Å)
A1 to D7	5.47
A1 to H9	8.99
A1 to H10	12.90
H9 to H10	4.11
D7 to H10	14.93

Table A 14 : Pharmacophoric angles.

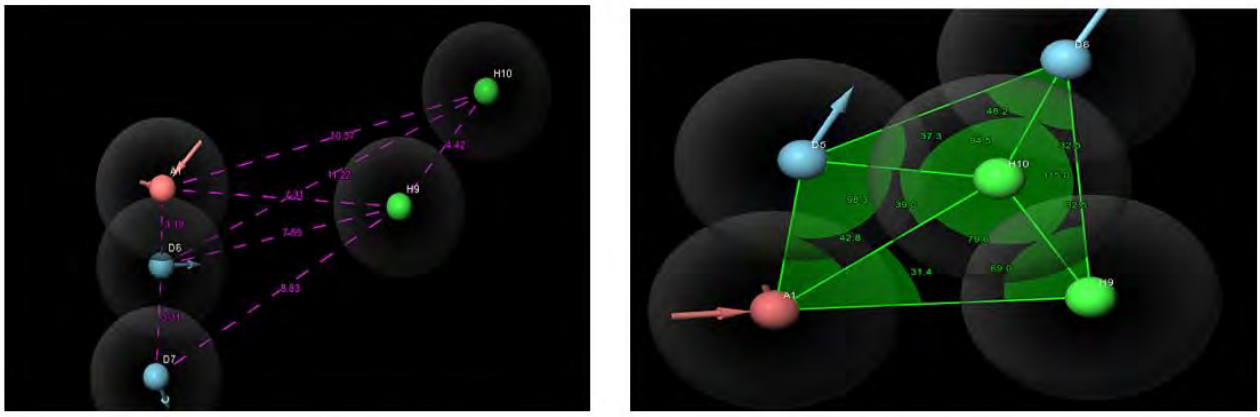
Angle	Site 1	Site 2	Site 3	Angle
A1D7H9	A1	D7	H9	55.4
D7A1H9	D7	A1	H9	94.5
D7H9A1	D7	H9	A1	30.1
A1H10D7	A1	H10	D7	21.1

Table A 15 : Pharmacophore results.

Pharmacophore	Survival Score	Site Score	Vector Score	Volume Score	PhaseHypoScore
AAAH_3(12)	4,336	0,752	0,867	0,485	1,038
ADDHH_2(10)	4,397	0,336	0,767	0,472	1,014
AADHH_3(11)	4,378	0,301	0,832	0,485	1,013
AAADH_3(11)	4,252	0,524	0,827	0,485	1,013
AAAH_4(12)	4,31	0,662	0,888	0,485	1,012
AAADH_1(11)	4,329	0,581	0,823	0,485	1,012
ADHH_1(10)	4,339	0,507	0,859	0,485	0,992
AADDH_1(10)	4,187	0,415	0,866	0,485	0,962
ADHH_2(9)	4,308	0,539	0,864	0,485	0,946
AAHH_3(9)	4,351	0,640	0,909	0,485	0,943

AAHH_4(10)	4,378	0,474	0,907	0,485	0,943
AAHH_1(9)	4,489	0,577	0,932	0,485	0,940
AAADH_2(10)	4,315	0,525	0,883	0,485	0,935
AAAHH_1(8)	4,753	0,499	0,850	1,000	0,919
AADHH_2(8)	4,546	0,467	0,842	1,000	0,917
ADDHH_1(8)	4,438	0,360	0,842	1,000	0,909
AAHH_2(10)	4,485	0,547	0,842	1,000	0,896
AAHH_2(9)	4,462	0,738	0,842	0,541	0,867
AAAH_1(9)	4,415	0,725	0,842	0,502	0,866
AADHH_1(8)	4,566	0,439	0,842	0,457	0,844

(f) Antimycotic squalene monooxygenase inhibitors from garlic extract



(a) The pharmacophoric distances between the points. **(b) The pharmacophoric angles between the intersites.**

Figure A 12 The pharmacophores with intersite distances and angles. All distances are in A. Pink spheres with arrow, hydrogen bond acceptor (A), green sphere, hydrophobe (H), blue, Donor (D).

Table A 16 : Pharmacophoric distances.

Entry	Distance (Å)
D6 to D7	3.31
A1 to H10	10.37
A1 to H9	7.33
H9 to H10	4.42
D7 to H10	8.83
D6 to H10	11.22
D6 to H9	7.99

Table A 17 : Pharmacophoric angles.

Angle	Site 1	Site 2	Site 3	Angle
A1H9H10	A1	H9	H10	79.0
H10A1H9	H10	A1	H9	31.4
H9H10D6	H9	H10	D6	115
H10H9D6	H10	H9	D6	32.0
A1D5H10	A1	D5	H10	95.0
D5A1H10	D5	A1	H10	42.8
D5D6H10	D5	D6	H10	48.2
A1H10H9	A1	H10	H9	79.6

Table A 18: Pharmacophore results.

Pharmacophore	Survival Score	Site Score	Vector Score	Volume Score	PhaseHypoScore
DDDHH_2	5,274	1,000	1,000	1,000	0,826
ADDHH_1	5,038	1,000	1,000	1,000	0,812
ADDHH_2	5,024	1,000	1,000	1,000	0,811
ADDHH_3	5,015	1,000	1,000	1,000	0,811
ADDHH_4	5,004	1,000	1,000	1,000	0,810
ADDHH_5	4,994	1,000	1,000	1,000	0,807
ADDHH_6	4,946	1,000	1,000	1,000	0,805
AADHH_1	4,915	1,000	1,000	1,000	0,805
AADHH_2	4,91	1,000	1,000	1,000	0,797
DDHH_1	4,779	1,000	1,000	1,000	0,796
DDHH_2	4,765	1,000	1,000	1,000	0,796
DDHH_3	4,76	1,000	1,000	1,000	0,795
DDHH_4	4,753	1,000	1,000	1,000	0,795
DDHH_5	4,751	1,000	1,000	1,000	0,795
DDHH_6	4,749	1,000	1,000	1,000	0,795
DDHH_7	4,747	1,000	1,000	1,000	0,795
DDHH_8	4,745	1,000	1,000	1,000	0,795
DDHH_9	4,744	1,000	1,000	1,000	0,795
DDHH_10	4,744	1,000	1,000	1,000	0,795
DDHH_11	5,274	1,000	1,000	1,000	0,826

2. 3D-QSAR results

(a) Antimycotic squalene monooxygenase inhibitors: Naftifine and its derivatives

Table A 19 : 3D-QSAR.

# Factors	SD	R^2	R^2 Scramble	R^2 CV	F	P	Stability	RMSE	Q^2	Pearson-r
1	1.5136	0.6546	0.4193	-0.8697	26.5	0.00147	-0.942	1.78	0	0

# Factors	H-bond donor	Hydrophobic/non-polar	Positive ionic	Electron- withdrawing	other
1	0.037097	0.90812	0.046318	0.005201	0

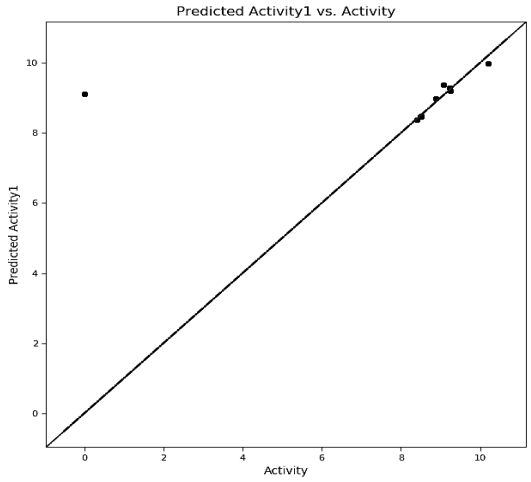


Figure A 13 Scatter plot of the observed versus the predicted activity for the training set.

(b) Antimycotic squalene monooxygenase inhibitors: Terbinafine and its derivatives

Table A 20: 3D-QSAR results.

# Factors	SD	R^2	R^2 Scramble	R^2 CV	F	P	Stability	RMSE	Q^2	Pearson-r
1	0.5524	0.9775	0.9220	-0.1392	130.5	0.00144	-0.275	0.31	0	0

# Factors	H-bond donor	Hydrophobic/non-polar	Positive ionic	Electron- withdrawing	other
1	0.024959	0.923757	0.046013	0.002445	0

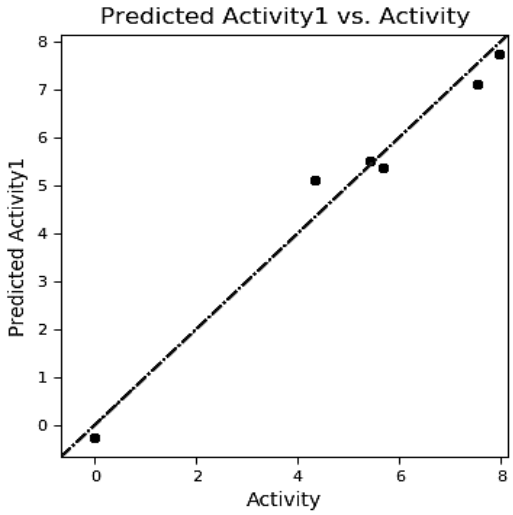


Figure A 14 Scatter plot of the observed versus the predicted activity for the training set.

(c) Antimycotic squalene monooxygenase inhibitors. Terbinafine's derivatives

Table A 21 : 3D-QSAR results.

#Factors	SD	R^2	R^2 Scramble	R^2 CV	F	P	Stability	RMSE	Q^2	Pearson-r
1	0.8240	0.6366	0.4515	-0.1559	61.3	3.34e-09	0.409	0.88	0.0630	0.4504

#Factors	H-bond donor	Hydrophobic/non-polar	Positive ionic	Electron- withdrawing	other
1	0.033567	0.792476	0.033115	0.001551	0

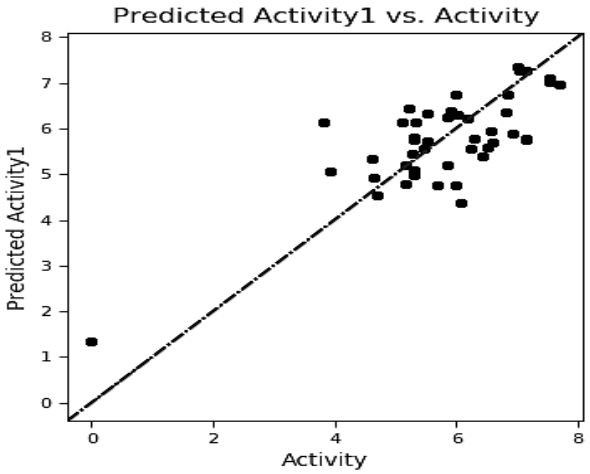


Figure A 15 Scatter plot of the observed versus the predicted activity for the training set.

(d) Antimycotic squalene monooxygenase inhibitors

Table A 22 : 3D-QSAR results

# Factors	SD	R^2	R^2 Scramble	R^2 CV	F	P	Stability	RMSE	Q^2	Pearson-r
1	0.3116	0.9718	0.9358	-0.0587	206.8	7.08e-06	-0.000951	0.64	0.2091	1.00000

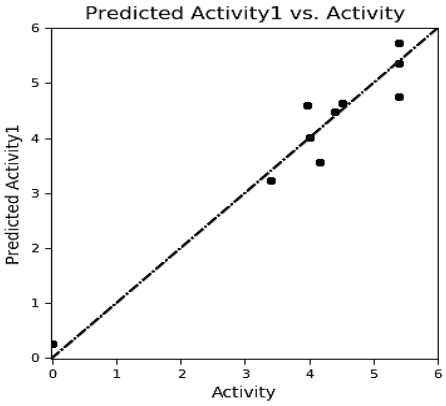


Figure A 16 Scatter plot of the observed versus the predicted activity for the training set.

(e) Antimycotic squalene monooxygenase inhibitors

Table A 23 : 3D-QSAR results.

# Factors	SD	R^2	R^2Scramble	R^2 CV	F	P	Stability	RMSE	Q^2	Pearson-r
1	0.3344	0.7322	0.2524	0.3203	172.2	1.1e-19	0.797	9.04	0.00000	0.00000

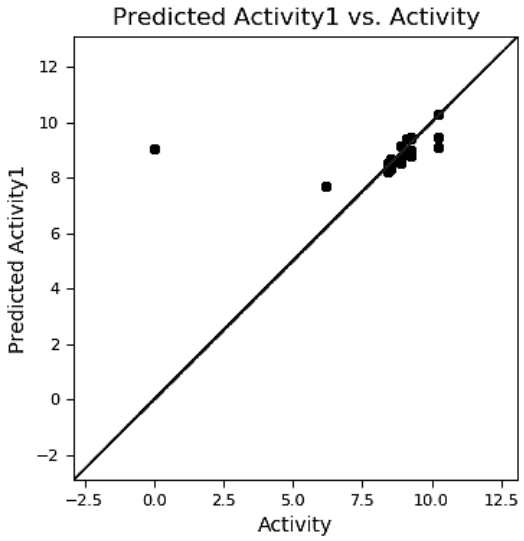


Figure A 17 Scatter plot of the observed versus the predicted activity for the training set.

(f) Antimycotic squalene monooxygenase inhibitors

Table A 24 : 3D-QSAR results.

# Factors	SD	R^2	R^2 Scramble	R^2 CV	F	P	Stability	RMSE	Q^2	Pearson-r
1	2.3433	0.3233	0.6113	-0.5592	3.8	0.0863	-0.0856	1.37	-0.0623	0.2651

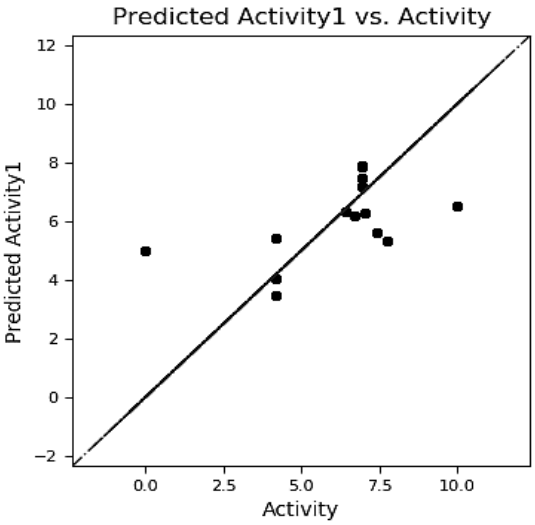
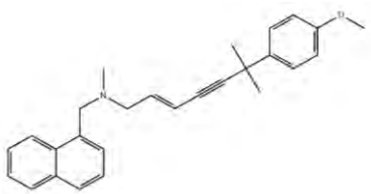
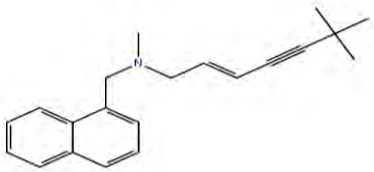


Figure A 18 Scatter plot of the observed versus the predicted activity for the training set

3. ADME results

(a) Table A 25 (i) : Antimycotic squalene monooxygenase inhibitors: Naftifine and its derivatives

Structure	mol MW	dHB	aHB	QPo/w	QPS	QPCaco	QBB	QMDCK	QKhsa	%Human oral absorption	Rule of Five
SDZ 87-469 (8) 	251,454	0	2	3,922	-3,662	2470,586	0,407	1454,789	0,777	100	0
SDZ 87-469 (9) 	223,401	0	2	3,445	-3,184	2470,586	0,322	1454,789	0,57	100	0
SDZ 87-469 (10)	401,718	0	2	6,237	-7,561	2470,586	0,819	1454,789	2,103	100	1

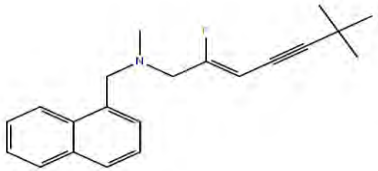
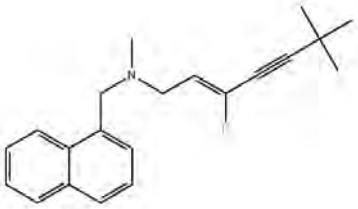
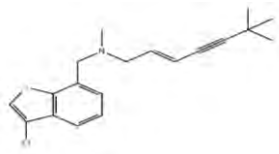
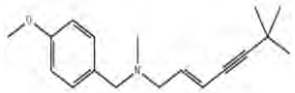
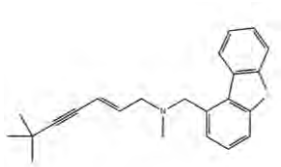
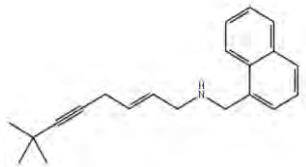
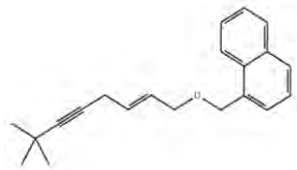
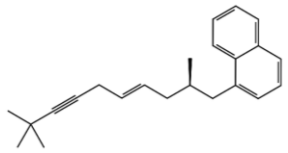
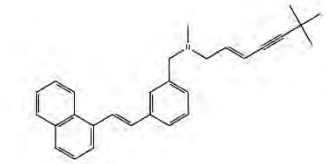
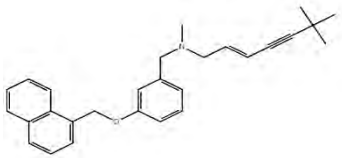
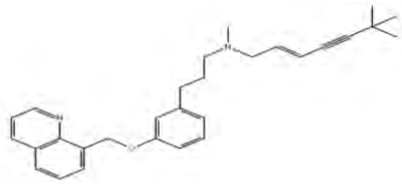
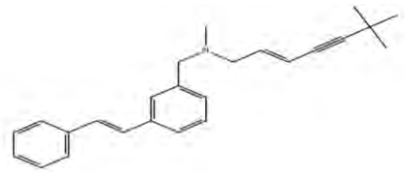
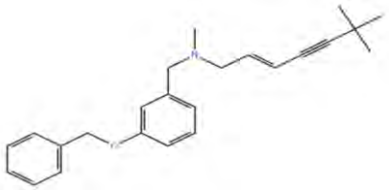
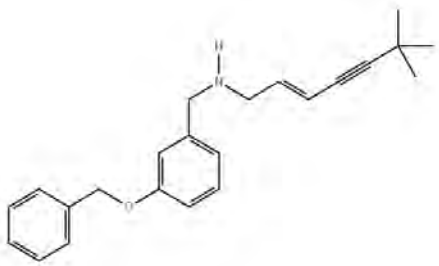
											
SDZ 87-469 (11) 	305,546	0	2	4,836	-5,203	2470,586	0,57	1454,789	1,269	100	0

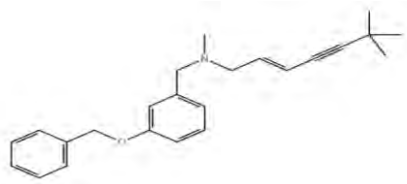
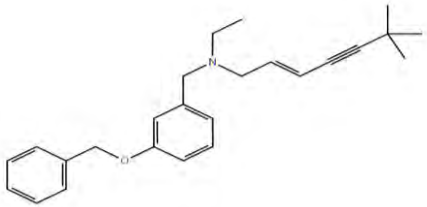
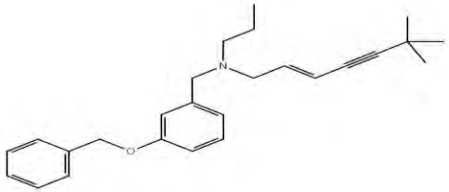
Table A 26 (ii) : Antimycotic squalene monooxygenase inhibitors. Terbinafine and its derivatives.

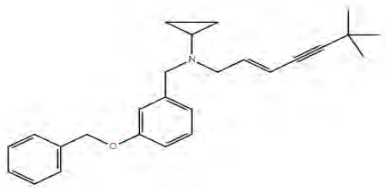
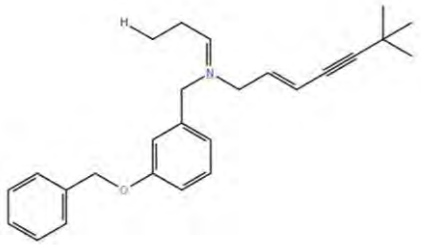
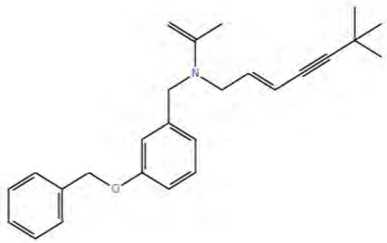
Structure	Mol MW	DHB	AHB	QPPw	QPo/w	QPlogS	QPPCaco	QBB	MDCK	QKhsa	%Human oral absorption	Rule of five
SDZ 87-469 (12) 	331.902	0	2	1,611	5,459	-5,93	2470,586	0,681	1455	1,38	100	1
SDZ 87-469 (13) 	241.375	0	3,7	2,983	4,436	-4,07	2318,083	0,39	1358	0,58	100	0
SDZ 87-469 (14) 	271.402	0	2	1,403	4,676	-4,14	2327,871	0,453	1364	0,81	100	0
SDZ 87-469 (15)	347.517	0	2,5	2,043	5,09	-5,23	2470,586	0,34	2049	1,19	100	1

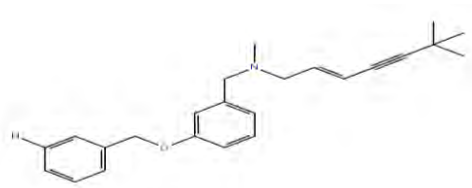
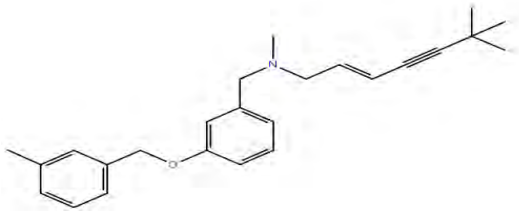
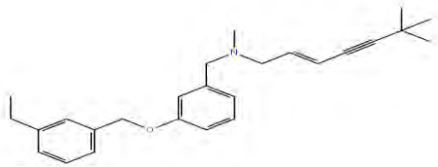
												
SDZ 87-469 (16) 	291.435	0	0	-1,82	10,726	-11,6	9906,038	1,59	5899	1,88	1	100	1
SDZ 87-469 (17) 	292.420	0	1,7	1,214	5,144	-8,45	9906,038	0,734	5899	1,2	1	100	1
SDZ 87-469 (18) 	304.474	1	1,5	2,593	5,52	-5,64	2061,907	0,393	1197	1,23	3	100	1

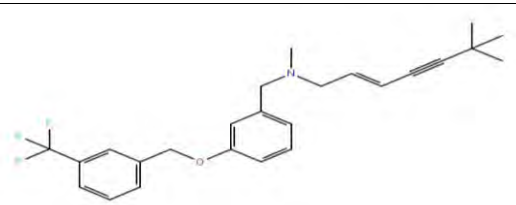
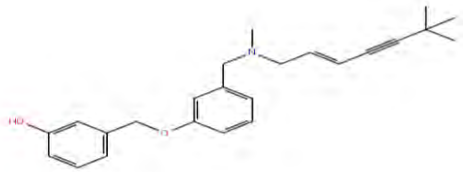
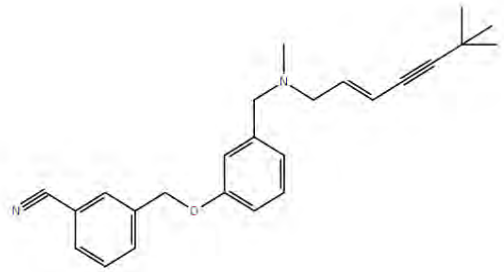
SDZ 87-469 (19) 	455,781	0	4,95	3,489	6,658	-7,393	2470,6	-0,355	2770,644	1,617	1	100	1
SDZ 87-469 (20) 	467,849	0	4,45	3,162	7,821	-7,29	2462,2	0,117	1449,463	1,826	1	100	1
SDZ 87-469 (21) 	467,849	0	4,45	3,162	7,821	-7,29	2462,2	0,117	1449,463	1,826	1	100	1

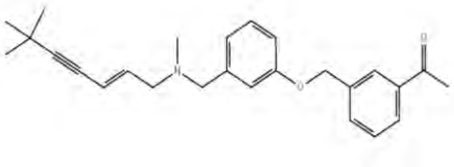
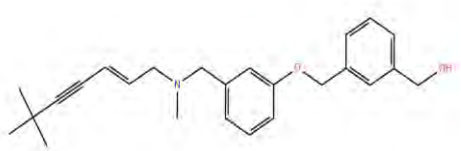
SDZ 87-469 (22) 	451,85	0	3,7	2,61	6,568	-7,591	2470,6	0,26	1454,789	2,026	1	100	1
SDZ 87-469 (23) 	466,864	2	4,7	6,196	6,86	-5,691	470,02	0,321	267,732	1,689	1	100	1
SDZ 87-469 (24) 	505,821	0	3,7	2,869	8,392	-9,778	2470,6	0,256	4833,874	2,327	1	100	2

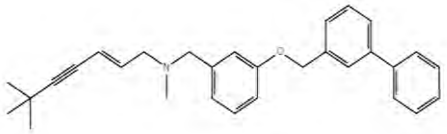
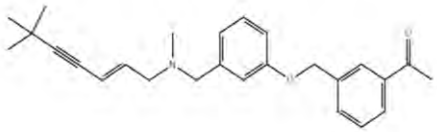
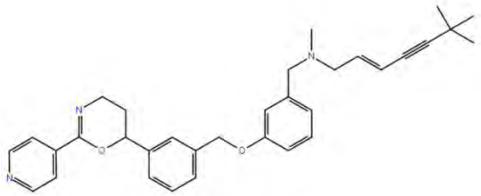
SDZ 87-469 (25) 	505,821	0	3,7	2,511	6,573	-6,928	2470,6	0,012	3620,44	1,906	1	100	2
SDZ 87-469 (26) 	467,849	0	5,4	3,953	5,155	-5,661	2470,6	-0,61	1454,789	1,454	3	100	1
SDZ 87-469 (27) 	451,85	0	3,7	2,4	5,586	-5,938	2470,6	0,085	1454,789	1,779	3	100	1

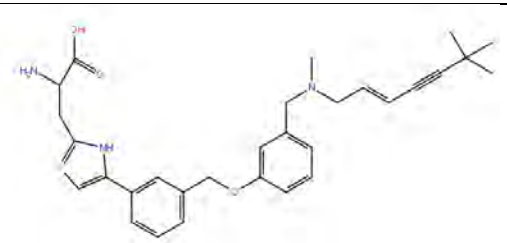
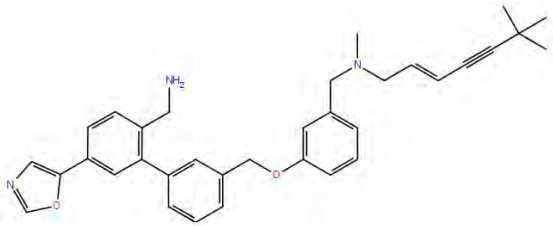
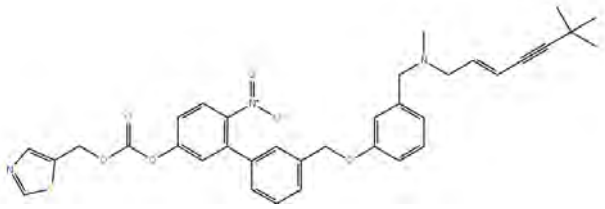
SDZ 87-469 (28) 	467,849	1	5,4	5,806	7,333	-7,84	847,29	-0,525	457,551	1,752	1	100	1
SDZ 87-469 (29) 	451,85	0	3,7	2,383	5,507	-5,805	2470,6	0,071	1454,789	1,796	3	100	1
SDZ 87-469 (30) 	455,814	0	3,7	2,543	6,475	-7,121	2470,6	0,127	2227,588	1,877	1	100	1

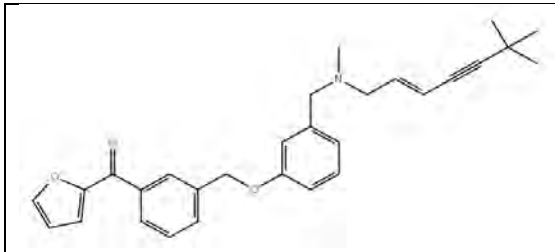
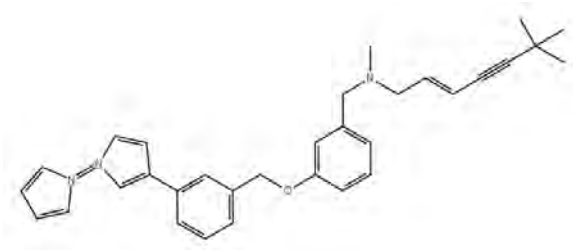
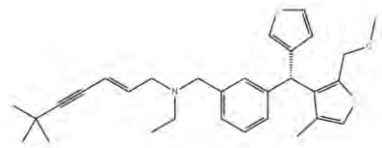
SDZ 87-469 (31) 	535,848	1	3,7	4,071	8,968	-8,978	1142,9	-0,077	2471,951	2,298	1	100	2
SDZ 87-469 (32) 	482,864	3	6,4	9,587	5,252	-4,047	76,221	-0,575	37,477	1,182	1	78,422	1
SDZ 87-469 (33) 	482,864	3	6,4	9,866	5,912	-6,115	77,157	-0,753	37,975	1,436	1	82,384	1
SDZ 87-469 (34)	482,864	3	6,4	9,862	5,79	-5,915	71,577	-0,781	35,015	1,396	1	81,084	1

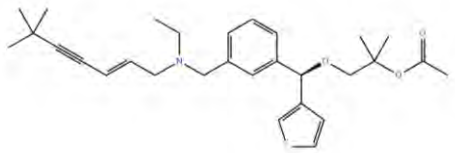
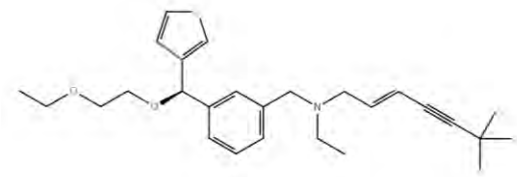
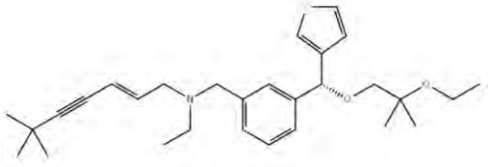
													
SDZ 87-469 (35) 	469,883	0,8	4,2	3,828	5,43	-5,594	2470,6	-0,22	2062,062	1,612	3	100	1
SDZ 87-469 (36) 	453,822	1	5,4	5,7	6,58	-5,933	1006,9	-0,301	551,415	1,454	1	100	1

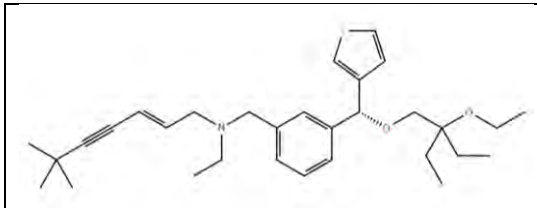
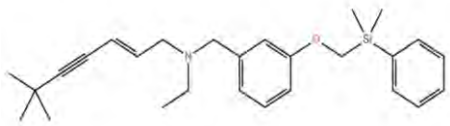
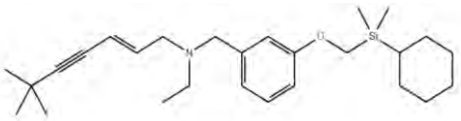
SDZ 87-469 (37) 	483,91	0	4,2	3,068	7,358	-8,679	2470,6	0,122	2072,935	2,138	1	100	1
SDZ 87-469 (38) 	467,849	0	5,4	4,346	6,995	-8,761	2470,6	-0,283	1454,789	1,897	1	100	1
SDZ 87-469 (39)	437,823	0	3,7	2,723	7,094	-8,478	2470,6	0,353	1454,789	2,072	1	100	1

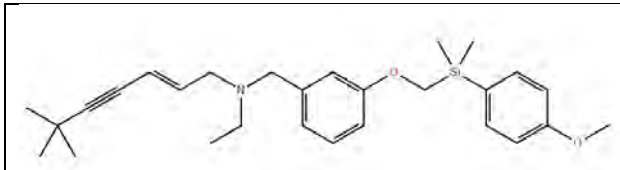
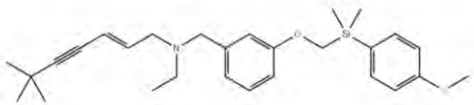
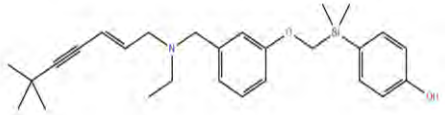
													
SDZ 87-469 (40) 	437,823	0	3,7	2,475	5,936	-6,528	2470,6	0,147	1454,789	1,812	1	100	1
SDZ 87-469 (41) 	483,835	0	4,95	2,933	6,12	-6,164	2470,6	-0,45	2762,444	1,537	1	100	1
SDZ 87-469 (42)	427,728	0	5,9	4,241	5,3	-5,446	2470,6	-0,943	2776,48	0,982	3	100	1

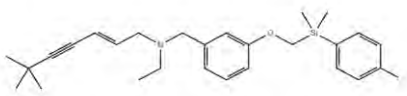
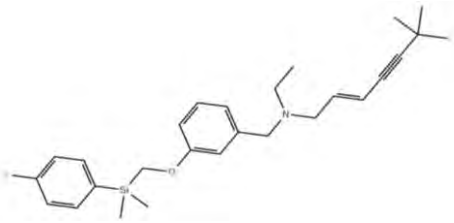
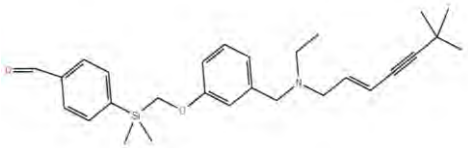
													
SDZ 87-469 (43) 	471,781	1	6,65	6,508	6,549	-5,803	1236,2	-0,216	1272,58	1,209	1	100	1
SDZ 87-469 (44) 	452,781	1	5,7	6,653	6,356	-6,421	439,08	0,602	384,822	1,56	1	100	1
SDZ 87-469 (45)	449,759	0	5,4	4,813	7,18	-7,537	2299,1	0,261	1345,936	1,63	1	100	1

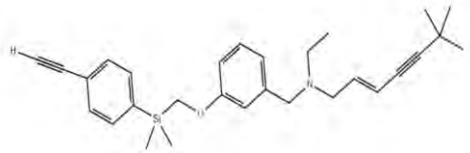
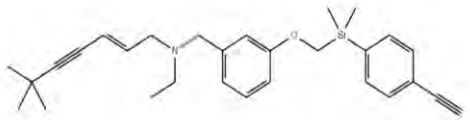
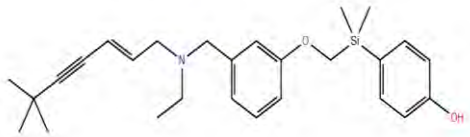
													
SDZ 87-469 (46) 	447,786	0	3,7	2,983	6,25	-7,382	2470,6	0,203	1454,789	1,976	1	100	1
NB-598 (47) 	395,668	1	5,4	5,613	5,531	-4,835	1576,2	-0,028	895,013	0,98	3	100	1
NB-598 (48)	394,683	2	4,7	6,376	5,034	-3,132	313,31	0,292	172,71	1,058	3	88,138	1

													
FR19478 (49) 	381,641	1	5,4	6,057	5,451	-5,586	992,38	-0,212	542,799	1,045	3	100	1
FR19478 (50) 	433,64	0	3,7	3,075	7,596	-8,097	2297,7	0,527	5292,775	1,717	1	100	1
FR19478 (51)	393,695	0	3,7	2,771	6,291	-7,287	2470,6	0,21	1454,789	1,69	1	100	1

													
FR19478 (52) 	379,668	0	3,7	2,941	6,794	-6,877	2302,4	0,263	1348,067	1,544	1	100	1
FR19478 (53) 	365,641	0	3,7	2,823	5,499	-6,117	2470,6	0,07	1454,789	1,346	3	100	1
FR19478 (54)	393,695	0	3,7	2,511	7,169	-6,797	2457,9	0,176	1446,706	1,617	1	100	1

													
FR19478 (55) 	393,695	0	3,7	2,558	6,323	-7,18	2470,6	0,216	1454,789	1,644	1	100	1
FR19478 (56) 	379,668	0	3,7	2,749	6,184	-7,107	2470,6	0,191	1454,789	1,564	1	100	1
FR19478 (57)	365,641	0	3,7	2,823	5,499	-6,117	2470,6	0,07	1454,789	1,346	3	100	1

													
FR19478 (58) 	351,615	1	3,2	3,861	6,319	-6,535	2130,6	0,223	1239,677	1,414	1	100	1
FR19478 (59) 	361,61	0	3,7	2,609	5,688	-4,283	2325,6	0,274	1362,746	1,051	3	100	1
FR19478 (60)	363,669	0	2	1,29	7,561	-7,935	2288,9	0,259	1339,501	1,973	1	100	1

													
FR19478 (61) 	420,721	1	5,2	5,976	5,953	-5,292	462,06	0,556	262,837	1,455	3	96,535	1
FR19478 (62) 	419,733	0	3,7	2,741	5,117	-5,474	2470,6	0,001	1454,789	1,508	3	100	1
FR19478 (63)	417,76	0	2	1,427	8,47	-9,012	2283,5	0,257	1336,087	2,377	1	100	1

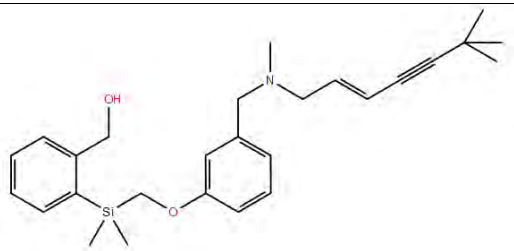
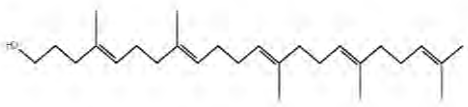
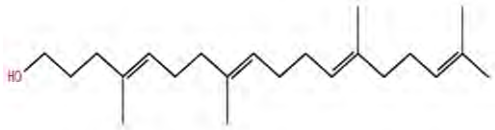
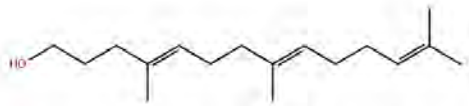
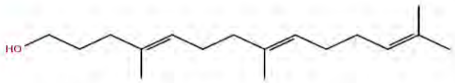
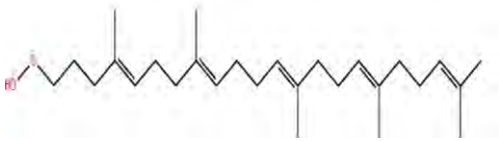
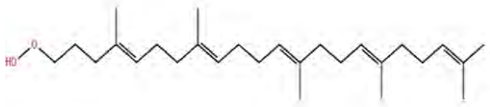
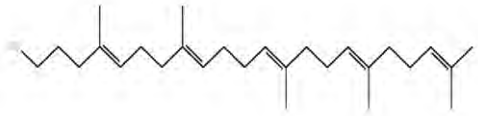
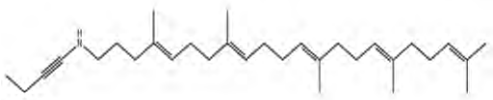
												

Table 27 A: Antimycotic squalene monooxygenase inhibitors (Trisnor-squalene alcohol).

Structure	mol MW	DHB	atHB	QPPw	QPPo/w	QPlogS	QCaco	QBB	MDCK	Qkhsa	%Human oral absorption	Rule of five
TNSA (73) 	454,778	1	1,7	1,242	9,198	-8,316	4261,908	-0,98	2370,71	2,227	100	1
TNSA (74) 	326,605	1	1,7	1,418	7,258	-7,857	3298,257	-1,06	1797,03	1,506	100	1
TNSA (75) 	256,471	1	1,7	1,94	5,409	-5,72	3299,763	-0,75	1797,91	0,864	100	1
TNSA (76) 	382,712	1	1,7	1,069	8,653	-9,465	2867,636	-1,36	1544,83	2,013	100	1

TNSA (77) 	410,766	1	1,7	0,725	9,452	-10,23	3296,644	-1,42	1796,08	2,262	100	1
TNSA (78) 	412,738	1	3,4	2,635	8,315	-9,437	1500,509	-1,89	767,085	1,821	100	1
TNSA (79) 	412,799	0,8	0,5	-0,858	10,055	-15,589	9906,038	1,77	10000	2,541	100	1
TNSA (80) 	430,747	0	0	-2,249	14,566	-16,409	9906,038	2,028	10000	2,955	100	1

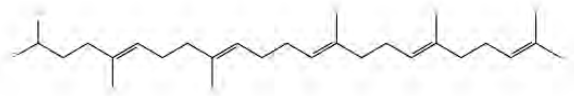
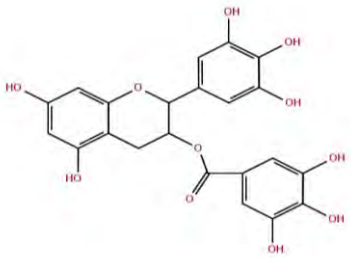
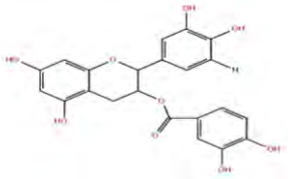
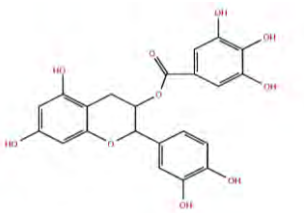
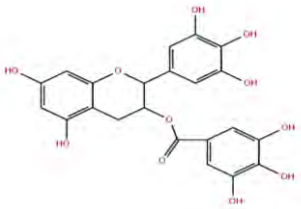
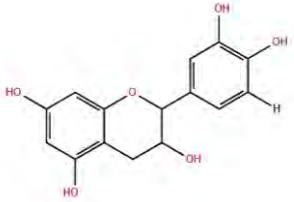
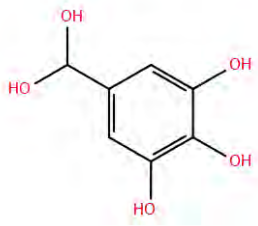
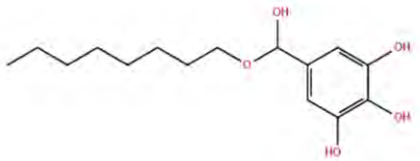
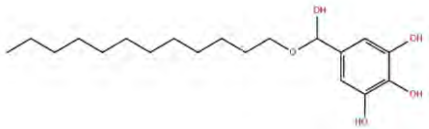
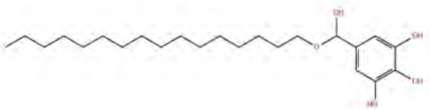
TNSA (81) 	438,819	1	1,7	0,658	10,212	-11,304	3345,177	-1,5	1824,67	2,553	100	1

Table 27 A: Anticholesterolemic squalene monooxygenase inhibitors (Epigallocatechin-3-O-gallate (EGCG))

Structure	mol MW	dHB	aHB	QPw	QPo/w	QS	QCaco	QBB	MDCK	Qkhsa	%Horal	Rule of five
EGCG (86) 	388,587	4	8,5	12,7	2,783	-2,717	628,3	-1,6	299,3	-0,206	93,32	0
ECG (87) 	548,845	4	8,5	14,7	5,477	-7,898	245,5	-2,09	108,4	1,169	75,877	2
ECG (88)	345,521	5	8,3	14,6	1,575	-2,599	39,25	-2,19	16,53	-0,366	64,691	0

												
ECG(89) 	345,521	5	8,3	14,6	1,726	-2,878	45,48	-2,17	19,39	-0,333	66,721	0
Garlic acid (90) 	345,521	5	8,3	14,5	1,669	-2,719	48,14	-2,12	20,62	-0,359	66,83	0
Garlic acid (91)	402,613	4	8,5	12,6	3,825	-5,483	245,6	-2,8	108,5	0,079	92,123	0

												
Garlic acid (92) 	346,506	4	8,5	12,6	2,308	-2,428	692,6	-1,59	332,6	-0,414	91,297	0
Garlic acid (93) 	290,399	4	8,5	13,8	1,041	-2,821	243,6	-1,99	107,5	-0,643	75,758	0
Garlic acid (94) 	178,185	5	8,5	16,5	-1,74	-0,918	77,65	-1,53	31,24	-0,882	50,58	0

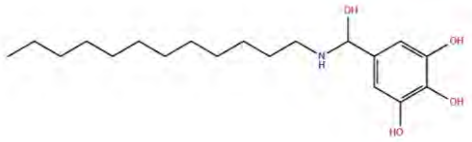
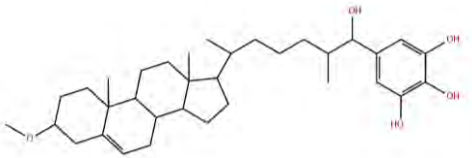
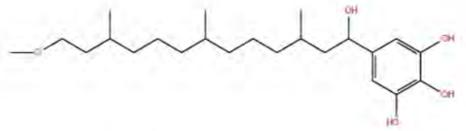
Garlic acid (95) 	302,367	5	10,2	18,8	-0,67	-2,385	75,28	-1,67	30,21	-0,647	56,587	0
Garlic acid (96) 	478,536	9	18,7	33	-2,71	-2,435	2,773	-3,93	0,852	-1,241	0	2
Garlic acid (97) 	476,563	8	17	29,8	-1,7	-2,844	6,215	-3,47	2,039	-0,991	18,264	1

Table 28 A: Anticholesterolemic squalene monooxygenase inhibitors from garlic extract.

Structure	mol MW	dHB	aHB	QPw	QPo/w	QCaco	QBB	QPPMDCK	QKhsa	%Human OralAbsorption	Rule of five
Selenium compound 106 	150,297	0	1	0,581	3,191	9906,038	0,369	10000	-0,064	100	0
Selenium compound 105 	182,357	0	0,5	-0,02	4,256	9906,038	0,594	10000	0,229	100	0
Allyl compounds 100	195,276	5	6,6	13,579	-0,959	52,422	-1,045	35,086	-0,977	52,107	0