



Molecular modelling of HIV-1 protease inhibition:

A drug design for HIV/AIDS

AR Motsilanyane



orcid.org/0000-0002-2555-1912

Thesis accepted in fulfilment of the requirements for

the degree *Doctor of Philosophy*

in Chemistry at the North-West University

PROMOTER: PROF Z Zondi

CO-PROMOTER: DR S Sosibo

GRADUATION: July 2024

The bottom half of the page features a blue and white wavy pattern, mirroring the style of the top section.

DECLARATION

I Andrew Rabontsi Motsilanyane, herewith declare that the theses entitled, “Molecular modelling of HIV-1 protease inhibition: A drug design for HIV/AIDS”, which I herewith submit to the North-West University (Mahikeng) as completion of the requirements set for the award of PhD degree, is my own work and has not already been submitted to any other university.

I understand and accept that the copies that are submitted for examination are the property of the University.

This thesis is submitted under NWU A rule 5.10.5.

Signature of candidate.....

University number

Signed at this.....day of 20.....

ABSTRACT

In this study, computer aided drug design (CADD) procedures were employed as a tool to discover potential drug candidates for HIV-1 protease. The use of these methods has proven to be useful over the years, with some recent drugs in the market being discovered through these computational tools. CADD saved time and minimized costs in discovering suitable HIV inhibitors. Using two approaches of finding a drug candidate, a wild type of HIV was used as a target to determine the binding affinity of these drugs. In the first approach, the ZINC database was screened using the Lipinski's rule of five. A pharmacophore was then constructed using the co-crystallised structure of 1rv7 with lopinavir and was used to screen all the compounds that complied with the Lipinski's rule of five. All compounds with RMSD less than 0.5 Å were identified and docked against 1rv7. Subsequently, in the second approach phytochemicals from five plants were studied, whereby a library of these compounds was prepared on the Molecular Operating Environment (MOE) software. A pharmacophore model was also constructed based on the features of the best three compounds obtained from the ZINC database. This pharmacophore was used to screen all the phytochemical prepared in the library. Compounds from the ZINC database were docked against 1rv7 and all the plant extracts were docked in 1rv7. The docked compounds were analysed based on the following interactions: Van der Waals, conventional hydrogen bonds, pi-Sigma, alkyl, pi-alkyl. The best three compounds were selected based on the SwissADME/T properties, to evaluate the adsorption, distribution, metabolism, excretion and toxicity characteristic profiling. This included the analysis of the following models, iLOGP, XLOGP, MLOGP, WLOGP and SILICOS-IT. Zinc_001456687980 and Zinc_000015276352 with a binding affinity of (-8.0 K_D) and (-6.2 K_D) respectively, were the two lead compounds from the ZINC database. Based on molecular dynamics simulation analysis, Epicatechin (-8 K_D), Juglanin (-7.6 K_D) and Catechin (-6.9 K_D) were the three lead phytochemical. The trajectories for all compounds were saved at every 100ps, with simulation time of 200ns. The root mean square fluctuation, the root mean square deviation and the protein-ligand contact calculation, and protein-ligand contacts were calculated from the MD trajectories. The IC₅₀ (nM) values for Zinc_000015276352, Zinc_001456687980, Juglanin, Epicatechin, Catechin and Dolutegravir were 2.046(nM), 1,573(nM), 1.954(nM), 0.979(nM), 1.042(nM) and 1,435 (nM) respectively.

ACKNOWLEDGEMENTS

- ❖ I would like to take this time to appreciate my lord savior Jesus Christ for giving me the grace and the strength to complete this research. If it was not of the mercies of GOD, this thesis could not have been completed. I give you all the glory.
- ❖ I would also like to thank my supervisors Professor Zimbili Zondi and Dr Sphelele Sosibo for offering me the opportunity to be their student throughout the course of this research work. I really appreciate all moments we had together.
- ❖ To the organic and computational chemistry team I appreciate all your contributions in this research.
- ❖ To my family: My wife Kitso and kids Bontsi, Reatlegile and Regolotswe, no words can express how much I love you and how I will forever cherish all the support you have shown me during my studies. I could have not have achieved this without your involvement, both emotionally and otherwise. Thank you so much.
- ❖ My parents: My late father, Mr. Samuel, and Mrs. Girly Motsilanyane thank you for laying a strong foundation for me to believe in education. I appreciate you so much.
- ❖ All my aunties, uncles, cousins and in-laws, God bless you for supporting me, directly or indirectly, through your prayers and motivations.
- ❖ My siblings Katlego and Neo; my nieces, Masego, Kamogelo, Esaleona and nephew Leatile, your support and love has sustained me to this point. You guys are the best siblings ever! May the Lord grant your heart desire.
- ❖ I take this time to thank the Department of Chemistry, North-West University, Mahikeng campus for offering me the resources to pursue my studies.
- ❖ Mr. Mike and other Laboratory colleagues, thank you for the kind of motivation you have gave me throughout my stay in lab 1074.
- ❖ To the Saviour's Embassy Church International, thank you so much for your prayers, love, and support. You are indeed the best spiritual home ever!
- ❖ Finally yet importantly, "Jesus is the way, the truth and the life." John 14:6. Psalm 127:1 "Unless the Lord builds the house, its builders labor in vain."

LIST OF PUBLIATIONS

1. Motsilanyane, A.R., Mkhize, Z. and Sosibo, S. Computational studies for selected medicinal plants against dolutegravir using ligand-based pharmacophore, molecular docking, admet predictions and molecular dynamics simulation. January 2022. Pakistan Jornal of Medical and Health Sciences, 16(1): 927-936.
2. Motsilanyane, A.R., Mkhize, Z. and Sosibo, S. Identification of zinc inhibitors for hiv-1 protease through structure-based pharmacophore, virtual screening, admet and molecular dynamics simulation, January 2022. Pakistan Jornal of Medical and Health Sciences, 16(1):937-943.

LIST OF ABBREVIATIONS

(HIV-1)	Human immunodeficiency virus-1
(HIV-1 PR)	HIV-1 protease
(AIDS)	Acquired Immuno-Deficiency Syndrome
(CADD)	Computer Aided Drug Design
(MDR)	Multi-Drug Resistant
(MD)	Molecular Dynamics
(IN)	Integrase
(RT)	Reverse Transcriptase
(PR)	Retroviral protease
(PDB)	Protein Data Bank
(Cp-THF)	Cyclopentyltetrahydrofuranyl
(NRTIs)	Nucleotide Reverse Transcriptase Inhibitors
(NNRTIs)	Non-Nucleoside Reverse-Transcriptase Inhibitors
(INSTIs)	Integrase strand transfer inhibitor
(NtARTIs)	Nucleotide analogue reverse-transcriptase inhibitors
(NRTTIs)	Non-nucleoside reverse-transcriptase translocator
(HAT)	Histone acetyl transferase

(BAF)	Barrier-to-autointegration factor
[HMG-I(Y)]	High-mobility group protein A1
SPTBN1	Spectrin β nonerythrocyte 1
FMLF	Formyl methionine-leucinephenylalanine
VLPs	Virus-like particles
QFT	Quantum field theory
QSAR	Quantity Structure Activity Relationship
DFT	Density functional theory

TABLE OF CONTENTS

DECLARATION	i
ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iii
LIST OF PUBLICATIONS	iv
LIST OF ABBREVIATIONS	v
LIST OF TABLES	xii
LIST OF FIGURES	xiii
CHAPTER 1: INTRODUCTION	1
1.1 INTRODUCTION	1
1.1.1 Historical background.....	1
1.1.2 Problem statement.....	2
1.1.3 Aim and objectives.....	6
1.1.4 Retroviral structure and genome organisation	6
1.1.5 Biochemistry of HIV-1 protease	7
1.2 STRUCTURAL INVESTIGATIONS OF RETROVIRAL PROTEASES AND APPROACHES TO DESIGNING HIV PROTEASE INHIBITORS	13
1.2.1 Crystallographic studies of retroviral proteases	13
1.2.2 Description of the HIV-1 PR molecule.....	14
1.2.3 Common structural features of the inhibitor complexes of HIV-1 PR.....	14
1.2.4 Subsites of the inhibitor-binding pockets.....	15
1.2.5 NMR studies	15
1.2.6 Computational studies.....	16
1.2.7 Database of three-dimensional structures of HIV-1 PR.....	16
1.3 CATALYSIS OF HIV PROTEASE	17
1.4 INHIBITORS OF HIV-1 PROTEASE	18
1.4.1 Natural product-inspired inhibitors of aspartic proteases	19
1.4.2 The first generation of clinical inhibitors.....	21
1.4.3 The second generation of protease inhibitors.....	23
1.4.4 Reverse transcriptase inhibitors	26
1.4.5 Integrase inhibitors.....	27
1.4.6 Protease inhibitors.....	27

1.4.7	Entry/fusion inhibitors	27
1.4.8	Multi-class combination drugs.....	28
	REFERENCE LIST	29
	CHAPTER 2: LITERATURE REVIEW.....	43
2.1	HIV DRUG TARGETS AND DISCOVERIES.....	43
2.1.1	Background on drug targets	43
2.1.2	The components of HIV-1 genome.....	43
2.1.3	Reverse transcriptase inhibitors	46
2.1.3.1	NRTIs class of medication	47
2.1.3.2	NNRTIs class of medication includes.....	49
2.4	INTEGRASE INHIBITORS.....	49
2.5	HIV-1 ANTIVIRAL THERAPEUTICS	58
2.5.1	CCR3 blockers	58
2.5.2	CXCR4 antagonist	58
2.5.3	CCR5 or CD195 antagonist	58
2.5.4	NCp7 inhibitors.....	59
2.5.5	Ribonucleases as antiviral therapeutic	59
2.6	RECENT DEVELOPMENT OF HIV-1 PROTEASE INHIBITORS	59
2.7	CADD STUDIES.....	63
2.7.1	Homology modelling	64
2.7.2	Pharmacophore visual screening.....	66
2.7.3	Molecular docking studies	66
2.7.4	Molecular Dynamics Simulation	72
2.7.5	QSAR.....	76
2.7.6	Pro-drugs.....	77
2.7.7	Novel co-crystal structure	78
2.7.8	Highly drug resistant HIV-1 protease mutant	79
	REFERENCE LIST	81
	CHAPTER 3: GENERAL EXPERIMENTAL PROCEDURES	99
3.1	MATERIALS AND METHODS.....	99
3.1.1	Selection and retrieval of Ligands	100
3.1.2	Target database	102

3.1.3	Protein preparation	102
3.1.4	Pharmacophore modelling	104
3.1.5	Docking analysis	105
3.1.6	Lead compound identification and SwissADME applications	107
3.2	MOLECULAR DOCKING	111
3.3	MOLECULAR DYNAMICS SIMULATION.....	112
3.3.1	Force field	112
3.3.2	Energy minimization.....	112
3.3.3	Integration Algorithm	113
3.3.4	Temperature coupling	113
3.3.5	Pressure coupling	113
3.3.6	Periodic Boundary Conditions (PBC).....	113
3.3.7	Constraint algorithms	113
3.4	SIMULATION QUALITY ANALYSIS	114
3.4.1	Protein-Ligand Interaction Analysis	114
	REFERENCE LIST	116
	CHAPTER 4: IDENTIFICATION OF ZINC INHIBITORS FOR HIV-1 PROTEASE THROUGH STRUCTURE-BASED PHARMACOPHORE, VIRTUAL SCREENING, ADMET AND MOLECULAR DYNAMICS SIMULATION	124
	ABSTRACT.....	124
4.1	INTRODUCTION	125
4.2	MATERIALS AND METHODS.....	127
4.3	RESULTS	130
4.3.1	Drug likeness analysis.....	130
4.3.2	Virtual screening of HIV-1 protease inhibitors using ZINC databank and AutoDock Vina applying the Lipinski's Rule of Five.....	131
4.3.3	Development of A pharmacophore model based on the co-crystal compound with protein by molecular operating environment. (<i>Structure based pharmacophore modelling</i>):	131
4.3.4	Ligand-Pharmacophore properties based on the RMSD.....	135
4.3.5	Molecular docking against target protein.....	135
4.3.6	SwissADME studies	135
4.4	MOLECULAR DYNAMICS SIMULATION.....	139

REFERENCE LIST	146
CHAPTER 5: COMPUTATIONAL STUDIES FOR SELECTED MEDICINAL PLANTS AGAINST DOLUTEGRAVIR USING LIGAND-BASED PHARMACOPHORE, MOLECULAR DOCKING, ADMET PREDICTIONS AND MOLECULAR DYNAMICS SIMULATION	151
ABSTRACT.....	151
5.1 INTRODUCTION	152
5.2 MATERIALS AND METHODS.....	155
5.3 RESULTS	158
5.3.1 Drug likeness analysis.....	158
5.3.2 Construction of A ligand-based pharmacophore model.....	159
5.3.3 Virtual screening of HIV-1 Protease inhibitors using PubChem Databank, AutoDock Vina and the Lipinski's Rule of Five	160
5.3.4 Molecular docking studies	163
5.3.5 SwissADME studies	165
5.3.6 Docking analysis for lead compounds	170
5.4 MOLECULAR DYNAMICS SIMULATION.....	174
REFERENCE LIST	182
CHAPTER 6: A DFT-BASED QSAR ANALYSIS OF EPICATECHIN, JUGLANIN, CATECHIN AND ZINC COMPOUNDS AGAINST DOLUTEGRAVIR AS POTENTIAL INHIBITORS FOR HIV-1 PROTEASE.....	190
ABSTRACT.....	190
6.1 INTRODUCTION	191
6.2 MATERIALS AND METHODS.....	192
6.2.1 Calculation of Density Functional Theory	192
6.2.2 QSAR Modelling	192
6.2.3 External validation QSAR model	192
6.3 RESULTS	193
6.3.1 DFT Calculation (Phytochemicals).....	193
6.3.2 Electronic behavior of the compounds (ZINC compounds)	193
6.3.3 QSAR Modelling	195
REFERENCE LIST	200
CHAPTER 7: CONCLUSION AND RECOMMENDATIONS	203

7.1	CONCLUSIONS.....	203
7.1.1	Identified Zinc Compound ZINC_001456687980.....	205
7.1.2	Identified Zinc Compound ZINC_000015276352.....	205
7.2	RECOMMENDATIONS.....	206
	REFERENCE LIST	208
	APPENDIX.....	209

LIST OF TABLES

Table 1. 1: General characteristics and current indications of licensed antiretroviral drugs	26
Table 2. 1: Proteins encoded by HIV-1 genome	45
Table 2. 2: NRTIs class of medicine.....	48
Table 2. 3: NNRTIs class of medicine	49
Table 2. 4: Integrase inhibitors.....	65
Table 2. 5: A set of integrase inhibitors and their small molecular derivatives	52
Table 2. 6: Summary of HIV and host targets	61
Table 4. 1: Interaction between the trio hot-spot residues of 1rv7 co-crystal structure and the corresponding residues of the 1rv7 receptor at the binding site (PDB:1rv7).....	132
Table 4. 2: SwissADME properties analysis for 10 best compounds	135
Table 4. 3: SwissADMET for two lead compounds	136
Table 4. 4: SwissADME/T properties for two lead compounds	137
Table 4. 5: The docked Van der Waals, Conventional Hydrogen Bonds, Pi-Sigma, Alkyl, Pi-Alkyl.....	139
Table 5. 1: Interaction between the trio hot-spot residues of three lead ZINC ligand's structure (from previous investigation): PDB ID: 1rv7	160
Table 5. 2: Phytochemical extracts that were screened.....	161
Table 5. 3: Lipophilicity of the 10 best phytochemicals.....	165
Table 5. 4: Swiss ADME properties	167
Table 5. 5: SwissADME/T properties for four Lead Compounds	169
Table 5. 6: The docked Van der Waals, Conventional Hydrogen Bonds, Pi-Sigma, Alkyl, Pi-Alkyl between 1rv7 protein and best three compounds	173
Table 6. 1: DFT calculations of the investigated compounds	194
Table 6. 2: The QSAR model	196
Table 6. 3: External validation parameters of the constructed QSAR model	196
Table 6. 4: pIC ₅₀ and IC ₅₀ of the lead compounds	197

LIST OF FIGURES

Figure 1. 1: A co-crystallized structure of 1rv7 obtained from Protein Data Bank	8
Figure 1. 2: Statine	20
Figure 1. 3: The first generation of clinical inhibitors	21
Figure 1. 4: The second generation of protease inhibitors	24
Figure 2. 1: An example of a computer-aided drug design (CADD) flowchart (Image by Hung-Jin Huang et al.) (304).....	63
Figure 2. 2: A typical flowchart representing an ensemble docking protocol for HIV-1 PR inhibitors	74
Figure 3. 1: A flowchart representation involved in the study.....	99
Figure 3. 2: A flowchart representing the processes preparing a macromolecule	106
Figure 4. 1: The Flowchart of adopted computer-aided drug design (CADD) methodology	129
Figure 4. 2: The flowchart highlighting the filtering of ZINC compounds	130
Figure 4. 3: An uncomplexed HIV-1 protease.....	131
Figure 4. 4: A Pharmacophore model of 1rv7 co-crystallized structure	133
Figure 4. 5: Structures of the best 10 compounds from molecular docking	134
Figure 4. 6: Binding interaction of ZINC compounds and 1rv7	138
Figure 4. 7: Ligand- Protein contacts.....	140
Figure 4. 8: Molecular dynamics Simulation diagrams at 200ns.....	141
Figure 4. 9: RMSF P/L for lead compounds.....	143
Figure 4. 10: Protein-Ligand contacts.....	144
Figure 4. 11: Protein-ligand contact histogram.....	145
Figure 5. 1: The Flowchart of adopted computer-aided drug design (CADD) methodology	156
Figure 5. 2: Flowchart of the results obtained from computer aided drug design (CADD)	159
Figure 5. 3: A Pharmacophore model of three ZINC Ligands structure.....	160
Figure 5. 4: The best 10 compounds from molecular docking	164
Figure 5. 5: Binding interaction of epicatechin, juglanin and dolutegravir	172
Figure 5. 6: Molecular dynamics Simulation diagrams at 200ns.....	177
Figure 5. 7: Residue wise Root Mean Square Fluctuation (RMSF).....	179
Figure 5. 8: Protein Secondary Structures element distribution by residue index throughout the protein structure	180
Figure 5. 9: Protein-ligand contact histogram.....	181

Figure 6. 1: HOMO and LUMO diagram of (1). Epicatechin (2). Juglanin (3). Catechin and (4). Dolutegravir	199
Figure 6. 2: HOMO and LUMO diagram of ZINC (1) Zinc_000015276352 (2) Zinc_001456687980 and (3) Dolutegravir	199

CHAPTER 1: INTRODUCTION

1.1 INTRODUCTION

1.1.1 Historical background

Human immunodeficiency virus-1 (HIV-1) stands to be a worldwide threat to the human population. This is after more than three decades of its discovery (Kohl *et al.*, 1988). This infectious virus is the main causative agent of acquired immunodeficiency syndrome (AIDS), which is a complex multi-factorial disease. According to the data of the AIDS epidemic in 2020, as reported by UNAIDS, there were 36.9 million people living with HIV, 1.8 million new HIV infections and 690 000 AIDS-related deaths in 2019 and this poses a global health concern (UNAIDS, 2020).

(HIV-1) is a member of the genus *Lentivirus* that is a subgroup of the family Retroviridae. This virus is typically transmitted *via* sexual intercourse, mother-to-child transmission (via both the birth and breastfeeding processes), the sharing of intravenous drug injection paraphernalia and through direct blood contact with an infected person (Baba *et al.*, 2000). The HIV-1 infection can destroy the immune system, and ultimately resulting in severe neurological disease, including behavioural abnormalities, HIV-1 associated dementia (HAD) and motor dysfunction.

HIV-1 protease (HIV-1 PR) is a dimeric enzyme, and it is from the family of aspartic proteases, meaning they are capable of breaking other proteins. To maximise the use of its genetic information, HIV, since it is small, does this by making a long polypeptide chain that contains proteins (Lv *et al.*, 2015). These protein precursors, Gag and Gag Pol, must be cleaved by protease at nine specific points to produce functional proteins. This leads to the Gag precursor that will eventually give rise to structural proteins and the Pol precursor that will yield enzymes that include reverse transcriptase, integrase, and protease. These enable an HIV protease to be an important protein in making more functional viruses, this allows it to be widely exploited as a drug target and that exhibits a broad substrate recognition (Leung *et al.*, 2000).

HIV-1 protease plays an indispensable role in the life cycle of the virus, and it is responsible for the cleavage of the polypeptides into functional proteins. However, HIV-1 protease inhibitors (PIs) can block the viral assembly and its maturation, thus impairing the activities of the HIV-1 protease by binding to the active site and locking it up. The mode of action lies primarily with the binding to the active site of the protease (Yedidi *et al.*, 2014), (Trylska *et al.*, 2007). Gupta *et al.* discovered that the inhibition of protease results in the release of the structurally disorganised viral particle that is non-infectious (Gupta *et al.*, 2008).

Presently there are more than two dozen of HIV-protease inhibitors. One of the existing challenge is that, most HIV-protease inhibitors are designed with a hydroxyl group in order to mimic the transition state of the substrate's scissile peptide bonds and this leads to the structural similarity of inhibitors and hence, yielding mutations in HIV-1 protease that is commonly associated with cross-resistance to the other inhibitors (Ghosh *et al.*, 2016).

Palmer *et al.* isolated the clinical multi-drug resistant (MDR) HIV-1 strain 769 from a patient failing protease inhibitor containing antiretroviral regimens and the protease of strain 767 (MDR 767) was seen to be resistant to all inhibitors tested. This information contributed to the identification and isolation of 36 mutation loci in the 99 amino acid residue HIV-1 protease. This led to various levels of resistance of protease inhibitor treatment (Palmer *et al.*, 1999).

1.1.2 Problem statement

In modern medicine, antibiotics are the cornerstone of global health care. There has been, however, a great rise in antimicrobial resistance (AMR). AMR is the ability of bacteria to resist the administered antibiotic (Kumar *et al.*, 2018). Antibiotics fight bacterial infections and carry the ability to prevent bacterial infections and treat illness. It has been over three decades since the human immunodeficiency virus (HIV), the causative agent for acquired immunodeficiency syndrome (AIDS), was identified. Since the beginning of the global pandemic of HIV/AIDS in the early 1980s, an estimated 78 million people have been infected with HIV and about 39 million people have died of AIDS-related causes, according to the Joint United Nations Program on

HIV/AIDS (UNAIDS). An estimated 37 million people worldwide are now living with HIV/AIDS (UNAIDS, 2019).

The HIV-1 protease enzyme is the enzyme in the human body crucial to the development of the HI-virus. HIV-1 protease contains two relatively small identical polypeptides. The active sites of these enzymes are two aspartic acid residues located at the centre of the protein. This enzyme takes long polypeptide chains as substrates and cut them via hydrolysis. Scientists have decided to make a drug that targets the enzyme and in turn targets the development of HIV. Further developments on drugs have been made that take advantage of the enzyme's easily accessible active site and act as competitive inhibitors. All competitive inhibitors consist of carbon rich groups and two oxygen molecules at the centre of the inhibitor that interact with the active site (Brik and Wong, 2003).

These inhibitors stay in the active site, preventing the enzyme from doing its normal function as well as further spreading of the disease. Furthermore, HIV-1 protease is responsible for the production of all viral enzymes and structural proteins necessary to produce mature, virulent virions (Imamichi, 2004), (Velazquez-Campoy *et al.*, 2003). During replication, HIV infects T-cells via membrane fusion. Viral RNA then enters the cell and is turned into DNA via RT (5). The DNA enters the nucleus of the cell and is incorporated into the host cell's DNA by Integrase (DNA by IN). HIV then exploits the natural transcription and translation mechanism of the host cell to provide the viral polyprotein. The polypeptide is then hydrolysed into mature proteins by HIV-1 protease (PR). The viral RNA and proteins then accumulate at the cell's surface and are released as infectious virions. Inhibition of PR represents intervention at a vital stage in the HIV life cycle (Yang *et al.*, 2012).

The life cycle of HIV-1 consists of the following stages: a) adsorption and fusion, b) reverse transcription, c) integration, d) viral gene expression, e) virus assembly and maturation and f) budding. When the infectious virion is fully matured, the following enzymes are seen present: the reverse transcriptase enzyme (RT), the integrase enzyme (IN) and the protease enzyme (PR). Two of each of these are the copies of the genomic RNA which are the RT and the IN and the other one is the functional protein which is the PR (Pommier *et al.*, 2000).

The HIV-1 virion enters the host cell that includes the CD4⁺ and/or the T-lymphocyte and during this process, RT will speed up the conversion of one copy of the genomic viral RNA into a double-stranded viral DNA (dsDNA), while the IN catalyses its integration into the host chromosome to form a proviral DNA (Costin *et al.*, 2007).

The creation of shorter strands of messenger RNA (mRNA) and the copies of the HIV-1 genomic material can be orchestrated through the host's cell cellular system, where these mRNA functions are important factors for making long chains of HIV-1 precursor proteins. The role of these precursors is to cut the HIV-1 into smaller active proteins, and this is how a new viral particle is created. This viral particle will bud from the host cell and it will infect other cells (Blanco *et al.*, 2003).

HIV-1 protease is prone to resistance development in large part due to the process in which it is synthesised *in-vivo*. Reverse transcriptase creates viral DNA from the viral RNA that is incorporated into the host cell by the virus during infection. This process, unlike eukaryotic DNA synthesis, is extremely error prone due to its lack of proofreading mechanisms, causing HIV-1 to develop many polymorphic mutations (Kozal *et al.*, 1996). The large error rate of this process increases the probability of genetic mutants to be introduced into the system and the evolutionary pressures of PIs influence this. Consequently, the development of resistance is exacerbated. HIV-1 protease mutations can be classified into two groups: primary and secondary mutations. Primary mutations are changes in residues directly involved in substrate binding and manifest themselves in the active site of the enzyme. The residues considered involved in the active site are residues 25–32, 47–53 and 80–84. Changes in any other residues are classified as secondary mutations (Gulnik *et al.*, 1995).

Secondary mutations are located away from the active site and are usually compensatory mutations to relegate the detrimental effects of primary mutations on binding to the protease's natural substrate. One of the major strategies utilised to combat protease drug resistance is the design of PLs (Protease Inhibitors) by promoting hydrogen bonding interactions with the backbone atoms in the HIV-1 protease active site (Colman, 2009). An overlay of the X-ray structures of inhibitor-

bound wild-type HIV-1 protease and various mutant proteases shows minimal distortion of active site backbone conformations. As can be seen, the overall structural changes are minimal, even around the flexible flap region (Laco *et al.*, 1997).

The difficulty in combating resistance in HIV-1 protease is the tolerance for many mutations that lower affinity for PIs but maintain affinity for its natural substrate. Various cleavage sites in the natural substrates are vast and dissimilar. It has been speculated that the shape and volume of the substrate, with respect to the binding site, are of greater importance to binding affinity than the specific interactions with protease residues (King *et al.*, 2004). The observed loss of affinity for many PIs in the presence of the double mutant V82F/I84V can be rationalised by this hypothesis. The replacement of the isopropyl moiety of valine with a phenyl ring in phenylalanine and the loss of a methyl group from isoleucine to valine alter the shape of the active site. This results in a loss of Van der Waals interactions and a gain in steric hindrances, leading to poor inhibitor binding (Ohtaka *et al.*, 2003).

Most approved protease inhibitors, in general, show less than optimum pharmacokinetic properties. This is because PIs are substrates for the isoenzyme CYP3A4, a subunit of the cytochrome P450 hepatic enzyme system, which is responsible for metabolic degradation of PIs. As a result, many approved PIs show low plasma through levels and a short plasma half-life. Interestingly, many PIs act as both inhibitors and inducers of the isoenzymes. As it turns out, ritonavir (RTV) is a potent inhibitor of CYP3A4 and administration of a small dose of RTV with PIs significantly improves the PI's pharmacokinetic parameters (Wensing *et al.*, 2010). This study therefore sought to use virtual screening, pharmacophore modelling, molecular docking, ADMET analysis, and molecular dynamics simulations to thoroughly investigate the binding of suitable inhibitors to HIV-1 protease to its suitable inhibitors (Panda *et al.*, 2022).

1.1.3 Aim and objectives

The aim of the study is to investigate the use of CADD in finding new HIV-1 protease inhibitors.

Objectives of this study are to:

- Screen suitable compounds from the ZINC database and create a structure-based pharmacophore from the co-crystallized structure of a multidrug-resistant HIV-1 protease PDB ID: 1RV7
- Create a ligand-based pharmacophore to screen phytochemicals (Using PUB-CHEM database) from five medicinal plants.
- Investigate the interaction of 1rv7 with discovered lead ZINC ligands and phytochemicals.
- Employ molecular dynamic simulation to predict the simulation trajectories that includes root mean square deviation (RMSD), root mean square fluctuation (RMSF), protein-ligand contacts for C- α atom and analyse their binding status in static conditions.
- Calculate drug-like descriptors, such as adsorption, distribution, metabolism, excretion, and toxicity (ADMET).
- Application of QSAR and DFT on discovered inhibitors.

Compare the binding affinity, ADMET analysis and molecular dynamic simulation of the discovered inhibitors, against Dolutegravir (a first-line drug for HIV in South African healthcare systems).

1.1.4 Retroviral structure and genome organisation

Retroviruses belong to the family Retroviridae, a group of RNA viruses classified into seven genera. Their RNA genome is linear, single stranded and has an approximate 7-12 kilobases (depending on the retrovirus species) with a positive polarity (Jones *et al.*, 1994). The genome consists of two identical copies and is condensed by association with the nucleocapsid protein (NC). The two copies of genomic RNA are found within the Capsid alongside the enzymes. The integrase (IN), reverse transcriptase (RT) and protease (PR) are the three functional virion proteins

found in capsid and are needed for replication in the host cell. The Retrovirus is enveloped and contains outer membrane glycoproteins known as GP120 and GP41, responsible for the recognition and binding of the host cell receptors and the engulfment inside the host cell (Galletto *et al.*, 2004).

1.1.5 Biochemistry of HIV-1 protease

1.1.5.1 Structure and function of HIV-1 protease

HIV-1 protease is a retropepsin, an enzyme involved with peptide bond hydrolysis in retroviruses and a retroviral aspartyl protease. This enzyme is essential for the life-cycle and development of HI-virus, the retrovirus that causes AIDS (Jetzt *et al.*, 2000).

The first team of researchers to discover a crystal structure of HIV protease was Navia *et al.* from Merck laboratories. Later, a more accurate structure was reported by Kent *et al.* HIV PR consists of subunits of 99 amino acid aspartyl protease which functions as a homodimer with only one active site, namely C₂-symmetric in the free form. Each subunit is made up of nine β -stands and a single α -helix. Four anti-parallel β -strands form the highly stable dimer interface, which constitutes the active site (Bauer *et al.*, 2022). The core of the active site is hydrophobic and consists of two aspartic acid residues contributed by both subunits. Flexible anti-parallel β -sheets from both monomers form two flaps that cover the active site, thereby restricting access to it. The flaps have different orientations when they are closed and when they are semi-open, showing different conformations with a ligand in the active site and having a free enzyme state respectively (Wlodawer *et al.*, 1989), (Navia *et al.*, 1989). More than 140 structures of HIV-1 PR, its enzymes complexes with various inhibitors and mutants have been reported to date. Below (Figure 1.1) (Lee *et al.*, 1998).

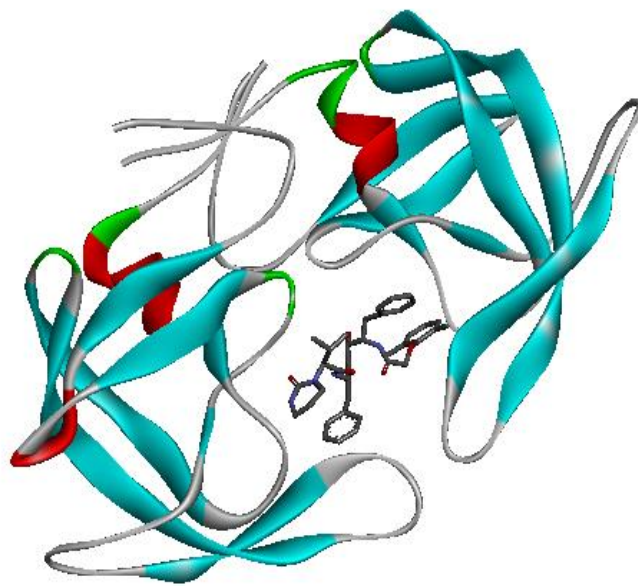


Figure 1. 1: A co-crystallized structure of 1rv7 obtained from Protein Data Bank

Reverse transcriptase consists three enzymatic activity, namely; RNase H, DNA-dependent DNA polymerase and RNA-dependent DNA polymerase. Env is the envelope protein, resides in the lipid layer and determines viral tropism. Gag is processed to the matrix and other core proteins that determine retroviral core. The PR cleavage site on the Gag and Gag-Pol precursor contains the unique amino acid sequences of Phe-Pro and Tyr-Pro (Velazquez-Campoy *et al.*, 2003).

1.1.5.2 The life cycle of HIV-1

There are seven distinct stages of the HIV life cycle, namely 1) binding/adsorption; 2) fusion; 3) reverse transcription; 4) integration; 5) replication/viral gene expression; 6) virus assembly and mutation and 7) budding. The retroviral lifecycle begins in the nucleus of an infected cell (King, 1994).

The free particle can infect new-targeted host cells such as CD4⁺ T-lymphocytes by binding to a cell surface receptor. The virus particle, with the help of the GP120 and GP41, attaches to the cell membrane of the host cell (Neamati *et al.*, 2000). The first interaction is with the viral envelope

glycoprotein GP120, which attaches to the target cell CCR5 or CXCR4 on the surface of the host's cell membrane. The association of GP120 and CCR5/CXR4 triggers a conformational shift and because of this conformational shift, GP41, that is folded inside the viral envelop, starts to open(Brik and Wong, 2003) .

Inside the cytoplasm, we have both copies of the mRNA, reverse transcriptase, protease, and integrase. Once these genetic materials are inside the host's cell, they only have the mRNA. To have that DNA, they will need to produce the DNA from the mRNA because they need to integrate their DNA section into the genome of the host cell (Liu, 2019). Inside the cell, the host the genome is has a DNA material and not the RNA, so they need to produce a genetic material that can be embedded into the host cell's genome and that will be possible if they convert their mRNA into DNA, meaning they need to produce DNA out of mRNA by a process called reverse transcription, the first stage of the retroviral life cycle. The RNA is therefore reverse-transcribed to the DNA by a viral encoded enzyme which is the reverse transcriptase (RT) (Arens *et al.*, 1993). The mRNA produces its first strand of DNA using the reverse transcriptase enzyme. The reverse transcriptase has two important sections for two important functionalities. One section is responsible for the polymerization and the second section is responsible for RNase degrading activity. Previously we had a hybrid of mRNA and DNA. They take this strand and copy another strand of the DNA and together they make up a dsDNA using reverse transcriptase enzyme. This produced dsDNA will be incorporated into the genome of the host. The reverse transcription generated DNA intermediate is called the pro-virus and it consists of 5' LTR that is the intervening viral genome and a 3' LTR. This pro-virus is carried in the nucleus, and it is integrated into the host cell's genome. The enzyme that is responsible for such an activity is called the integrase and in complex with a variety of other proteins, it forms what is called the pre-integration complex (PIC). A mature viral transcript is generated after transcription when the 5' end of the transcript is capped by 7-methylguanosine (Felber *et al.*, 1990).

Integration accounts for the virus's oncogenic activity since it is essentially "random" and could create a mutation within the gene. The virus thus persists with in the cell indefinitely through the aid of this integrase enzyme (Vaishnav *et al.*, 1991).

Integrase will cut sections from dsDNA that was prepared from the 3' ends. It cuts two nucleotides that have just been prepared from the 3' ends of the dsDNA, that is, they delete some sections so that they can produce a sticky substance that will cleave and hold on to the terminal. After the cleaving occurred, the DNA will be carried by the integrase, which also acts as a carrier, into the nucleus of the host's cell. After they enter the host's nucleus, they integrate their DNA into the host's genome and begin to rely on mechanisms for the replication, transcription and translation of the viral gene product (Malim *et al.*, 1990). The viral RNA attaches to the ribosome and through the process of translation, the ribosome reads the viral RNA and creates viral proteins, such as transcriptase, integrase and protease. The cell's own ribosome creates parts for new viruses. As this process continues, the ribosome also creates the capsid. However, the dsDNA carries information and for synthesis, more important proteins such as structural proteins and the enzyme reverse transcriptase, integrase and proteases (Inuzuka *et al.*, 2005).

Retroviruses contain open reading frames designated by the four main genes, Gag (encodes three main structural proteins: NC, CA, and MA) and the Pro that encodes a protease that is responsible for the cleavage of *gag-pol* precursors during virus maturation. The Pol is where the viral RNA genome is converted into a DNA intermediate, through the enzyme reverse transcriptase that is encoded by the Pol (Hu *et al.*, 2019). On the other hand, the integrase is an enzyme that allows the DNA intermediate to be merged or incorporated into the genome of the host cell. On the other hand, the precursor protein is translated and encoded during and after virus assembly (Nassal *et al.*, 1990).

The Gag, the Pol, Pro and Env proteins come together as they are synthesised to assemble progeny virions (Dicker *et al.*, 2021). These virions are assembled at the plasma membrane. The Gag recruits other env and other viral proteins. This is possible since the Gag displays the binding sites for these viral proteins. This central role of the Gag precursor protein, extends to function as a protein that is responsible for the packaging of viral RNA and it further binds the packaging (ψ) sequence near the 5' end of the RNA (Jamjoom *et al.*, 1977).

The mRNA will later migrate out of the cytoplasm and translated into protein products. Large protein products will be made through this process. This, however, will be a problem because the virus does not want the production of these large protein products and this leads to the third and final stage of the retroviral life cycle called proteolysis of the virus (Prabu-Jeyabalan *et al.*, 2000). The protease enzyme brings about the virus proteolysis. Protease will cleave to sections of the proteins to generate small sections from that protein, to make up a complete virus particle that includes all structural components and other important proteins. All these generated proteins will be taken inside the nucleus and well packaged into capsid, which contains RNA, transcriptase, integrase and protease (Mao *et al.*, 2012). The Gag precursor protein consists of matrix (MA), capsid (CA) and nucleocapsid (protein) NC and is targeted to the plasma membrane via hydrophobic post-translational modifications. One such important substance that attaches such proteins is myristic acid. At this point, the capsid does not have an envelope, however, this capsid gets the envelope once it is released from the cell through the membrane and passes through the endoplasmic reticulum. Inside this endoplasmic reticulum, there are different receptors and different molecules for glycosylation and different proteins are produced there. Glycosylated molecules like GP120 and GP41 are produced here (Mische *et al.*, 2005). The capsid will then, migrate through the endoplasmic reticulum and get its own envelope. On the envelope surface, the capsid has GP120 and GP41 and inside the capsid, there are different viral particles that have been prepared inside the nucleus. Thus, after budding from the cell, PR cleaves both Gag and Gag-Pro-Pol inactive precursors into active proteins that render the viral particle mature and infectious. This is how the viral life cycle can continue. This is now a matured virus that comes out of the host's cell and is ready to infect another healthy host (Göttlinger *et al.*, 1989).

1.1.5.3 Mechanism of the HIV protease

Suguna *et al.* described the most acceptable mechanism for aspartic protease, which is based on the crystal structure of *Rhizopys chinensis* aspartic protease complexed with a reduced peptide inhibitor. Based on the pH–rate profile of this enzyme, only one of the two active site aspartic acids are unprotonated in the active pH range (Hofmann *et al.*, 1984).

The aspartic protease mechanism suggests that, to generate an oxyanion tetrahedral intermediate, the Asp group that is closer to the nucleophilic water molecule is assigned the negative charge. (Windsor et al., 2018). This molecule is held together between the catalytic aspartates. This molecule is then activated by the negative aspartate side chain, which then in due process attacks the carbonyl group. This attack occurs in the substrate scissile bond to generate an oxyanion tetrahedral intermediate. The breakdown of the tetrahedral intermediate occurs through the protonation of the scissile amide (N atom) and its rearrangement, and the product becomes the hydrolysis product (Paiva et al., 2018). The use of the Lewis acid such as Zn^{2+} can be prevented by the general acid-base mechanism of the aspartic protease family. This is similar to the case in zinc metalloproteinase and the formation of covalent acyl intermediates (ANTONOV et al., 1981).

Since the full detailed mechanism of this enzyme is not fully understood, Jaskólski *et al.* worked on a new model that involves the study of the enzymatic mechanism for the HIV PR enzyme. This involves investigation of the crystal structure of a complex between a chemically synthesised HIV PR and an octapeptide inhibitor U-85548e (H-Val-Ser-Gln-Asn-Leu- ψ -[CH(OH)CH₂]-Val-Ile-Val-OH), that included other forms of the crystal structures of HIV PR complexes with various inhibitors (Jovanović et al., 2021). In their investigation, hydrolysis reaction was seen as a one-step process. This process was viewed as a process in which the nucleophilic water molecule and the acidic proton, attack the scissile peptide bond in a planned and organised manner. The major difference is between the simultaneous attack from the nucleophile and electrophile; this is when compared with previously proposed mechanisms. The possibility of covalent catalysis, in the chemical mechanism of HIV PR is discounted. This is as in the case with all the aspartic protease family (Jaskolski *et al.*, 1991).

Additionally, Hyland *et al.* studied the O incorporation into the transpeptidation product of the reaction of AcSQNYFLDG-NH₂ from AcSQNYPVV-NH₂ and FLDG-NH₂ carried out in H₂O enriched water and catalysed by HIV PR. This provided some evidence against the formation of a covalent intermediate among these molecules (Hyland *et al.*, 1991). Since the incorporation of O into the substrate AcSQNYPVV-NH₂ could be observed under the condition where the re-formation of substrate after hydrolysis was negligible, the involvement of an acyl enzyme

intermediate could be excluded. From the study of Hyland and colleagues, it was also found that the HIV PR-catalysed exchange of O from H₂O into the re-formed substrates occurs at a rate 0.01–0.012 times than that of the hydrolysis rate. Their results are in agreement with the formation of a kinetically competent enzyme-bound amide hydrate intermediate, the collapse of which is the rate-limiting chemical step in the reaction (Huff, 1991).

1.1.5.4 HIV protease: a target for AIDS therapy

Retroviral protease (PR) from the human immunodeficiency virus type 1 (HIV-1) was identified over two decades ago as a potential target for structure-based drug design and based on the previous works done by other researches, this, will advance our biochemical knowledge of HIV-I protease and develop novel therapeutics which targets the key viral enzyme (Huff, 1991).

1.2 STRUCTURAL INVESTIGATIONS OF RETROVIRAL PROTEASES AND APPROACHES TO DESIGNING HIV PROTEASE INHIBITORS

Ratner *et al.*, when comparing the genomic sequence of HIV-1 with other retroviruses, suggested that the HIV-1 genome encodes a protease. Moreover, Toh *et al.*, after observing a signature sequence (Asp-Thr-Gly), suggested that their protease might be related to the family of eukaryotic aspartic protease, which is exemplified by pepsin. These HIV-1 PR exhibit other characteristics of aspartic protease (Toh *et al.*, 1985). These characteristics include inhibition by pepstatin and inactivation by mutation of the putative active-site aspartates. The structure of HIV-1 PR constitutes a minor component of mature virions, and this is what initially hampered its structural studies. The use of recombinant and synthetic technology was implemented in producing the milligram quantities needed to further investigate its structure. HIV-1 PR was therefore cloned in a variety of vectors and prepared by total chemical synthesis (Debouck *et al.*, 1987).

1.2.1 Crystallographic studies of retroviral proteases

In 1989, a breakthrough that made a structure-assisted design of protease-targeted drugs against AIDS was determined. A eukaryotic aspartic protease as a template was used to model the structure

of HIV-1 PR. This model built, HIV-1 as a dimer of two identical aspartic protease-like domains. However, a more accurate model of HIV-1 PR was constructed using the three-dimensional (3D) structure of RSV PR and homology modelling techniques (Weber *et al.*, 1989).

1.2.2 Description of the HIV-1 PR molecule

The crystallographic studies provided the initial structures of retroviral proteases. The general topology of HIV-1 PR monomer is like the one of a single domain in pepsin-like protease. The main difference is that the dimer interface in the former is made up of four short strands, rather than the six long strands present in the pepsins (Rao *et al.*, 1991).

1.2.3 Common structural features of the inhibitor complexes of HIV-1 PR

These complexes are used in investigating the binding of substrate-based modified oligopeptide inhibitors and the binding of such inhibitors introduces substantial conformation change to the enzyme. All peptidomimetic inhibitors are bound to the enzymes active site (Bashore *et al.*, 2020). This is in an extended conformation. They are bound together with most of the inhibitors co-crystallised with HIV PR. This becomes important, so that when they are superimposed upon one another, they still have an aligned functional element. Between the Asp25/Asp250 carboxyls, is the hydroxyl group. This hydroxyl group is present in all the inhibitors excluding those containing peptide bond isosteres. This is at the non-scissile junction and occurs within the hydrogen-bonding distance (Wlodawer and Erickson, 1993).

On each aspartate, we have the bindings that occur to at least one carboxylate. One of the main features common to all complexes of HIV-1 PR is a hidden or buried water molecule that bridges the P2 and P100 CO group of the inhibitor and Ile50 and Ile150 NH groups of the flaps with the functionality substitution of this water leading to the design of urea- based inhibitors (Wlodawer *et al.*, 1989).

1.2.4 Subsites of the inhibitor-binding pockets

In the HIV-1 PR, there is subsites that accommodate side chains of the inhibitor, which can be identified as well. On each side of the non-scissile bond (S1-S3 and S10-S30), three subsites are very well defined. More subsites that are distant are not as clear. The most hydrophobic are the protease side chains comprising the pocket S1 and S10, with exception of the active side (Kong *et al.*, 1992).

There are two main components of the inhibitors that are involved in the polar contacts. These components contain the central group that is vital for such a process, and these are the main chain hydroxyl group of these inhibitors and the side chains of the active site aspartates. In some of the inhibitors, there is a standard amino acid. This amino acid is in the subsite S4 and does not form an actual binding pocket (Potempa *et al.*, 2018). However, only a small number of inhibitors display this property and only two has been reported. These have a moiety at P40, which are AG-4 and PS1. It is only the carbonyl most end of the P4 residue that is surrounded by the protease atom. The PS-1 has a proline at position PS1 and P40 which has one side of the glutamine fork directed into the S40 subunits. P5 or P50 residues of the inhibitor extends out of the binding groove of the enzyme and leads it to having little contacts with the protease (Burnouf *et al.*, 2004).

1.2.5 NMR studies

This a technique that has been successful for the past decade and used mainly for the determination of 3D structures of many macromolecules, including receptors and ligands/receptor complex (Tanoli *et al.*, 2015). It is a tool that has been used for rapidly determining the conformations of receptor-bound ligand. NMR has become an excellent tool for mapping enzymes-inhibitor interactions in a dynamic state when taken together with the data from crystallographic experiments. One of the initial methodologies on finding the complex of HIV-1 and its study was performed by Podlogar *et al.* They designed the protease inhibitors by using the NMR in relation to the computational methods. They worked on the novel macrocyclic inhibitor in which P1, P3 side chains were joined, and out of that, an excellent inhibitor of HIV- PR was derived. One of the

challenges that this technique overcame was when it was used beyond the limited solubility of the protein (Fesik, 1993), (Podlogar *et al.*, 1994).

1.2.6 Computational studies

The description of ligand-enzyme interactions or the prediction of ligand affinity uses a several quantum-chemical based computational techniques and the various computational techniques have been used in the binding of different inhibitors and the substrates to HIV PR (Kaushik *et al.*, 2018). These methods include the binding free energy, the molecular docking techniques and simple molecular mechanics calculation. These various methods are therefore aimed at predicting the inhibitor binding affinity. Over the years, a combination of *ab initio*, semi-empirical and empirical calculations have been used for the description of the catalytic mechanism of HIV-1 PR (Barbault and Maurel, 2015). This is in addition to the molecular dynamics calculations that have been used for studying enzymes dynamics and the influence of certain mutations on the structure. The prediction of the free energy of binding by computational methods has been possible through calculations that utilises molecular dynamics coupled with the free energy perturbation method, having acted as a reliable tool in the facilitation of such predictions (Sansom *et al.*, 1992).

The first docking studies of HIV-1 PR were performed by the Kuntz laboratory. One of their discoveries was the identification of haloperidol as an inhibitor of the enzyme. This interaction of HIV-1 PR with different inhibitors was a step in the right direction, as these crystal structures gave way to various docking techniques. Later, a different orientation of crystal structure determination revealed a different orientation for haloperidol. Docking methods and algorithms were tested using the known structural data and experimental characteristics by Monte Carlo docking (DesJarlais *et al.*, 1990, Rutenber *et al.*, 1993).

1.2.7 Database of three-dimensional structures of HIV-1 PR

The collection of HIV-1 PRs structure in a single database is beneficial. This is what was postulated in early 1993. This was done after observing the rapid progress in determining these HIV PRs (Counts *et al.*, 2015). There were two main reasons why this database was created.

Firstly, since there was a great emergence of various inhibitors, some reached approval by the FDA as anti-AIDS drugs, while others were only advancing to the clinical trials and others had no clinical relevance. The structures of the drugs that do not make it to the clinical phase are kept for future references and further scientific research. Secondly, new drug-design principles and synthetic strategies would be provided by this database that will supply a unique source of information about ligand-enzyme interactions in a well characterised system (Mulichak *et al.*, 1993).

The Protein Data Bank (PDB) is used to contain structural files that are deposited and stored in it. Sometimes these structures can directly be placed in the HIV PR database. A table of content of these complexes is formulated in a chronological order when these structures are deposited in the database and the branched links provide more detailed information about the inhibitors, their chemical formulas, 3D structures visualised by a Web-based browser applet and literature references (WINSLOW *et al.*, 1995)

The volume and surface calculations for the protein and the inhibitors are key information and should be given a lot of attention. These involve the construction of minimal and maximal binding volumes in the active site of HIV PR. The volume calculations are performed for all the stored structures from which the inhibitors and the water molecules will be removed and no further association could be discovered between the inhibitor and the binding cavity could be concluded on (Maseko *et al.*, 2019).

1.3 CATALYSIS OF HIV PROTEASE

HIV protease shares a common mechanism with all aspartic proteases. The enzymes of this protease lower the kinetic barrier to chemical reaction through catalysis. The mechanism of aspartic protease has been studied for decades and many unanswered questions have evolved. Aspartic proteases provide a 10-fold rate enhancement compared to the non-catalysed reaction (Jones, 2018). The family of aspartic protease facilitates amide hydrolysis by activating a water nucleophilic attack using acid-base catalysis (Eatemadi *et al.*, 2017). The reaction of aspartic protease begins by the formation of a Michael's complex. This is where a mono-protonated active

side binds to a water molecule and the peptide substrate. This aspartate deprotonates water leading to nucleophilic while the other aspartic acid protonates the scissile carbonyl oxygen to form the geminal-diol intermediate. The cationic gem-diol is formed by the proton abstracted from the water which is transferred to the scissile aide nitrogen and the scission of this bond leads to the formation of the products (Szecsi, 1992).

The least understood of the proton transfer and scission step comes at the second proton transfer step. In this step, the intermediate may be of similar energy to the transition state. This might constitute a single step and a proton transfer occurs between the nascent termini prior to product dissociation. Flap dynamics are important for substrate binding. This is the contribution made by crystallography, which has provided such substantial insight into this mechanism. This has made it possible to observe protease with both closed and open flaps. In an instance where the flaps are open the active site is solvated prior to the peptide substrate binding and a catalytically inactive variant of HIV protease can be prepared by installing the D25N mutation. This single atom change alters the active-site residue but still enable the protein to attain the proper fold. Many structures of inactivated HIV protease bound to the substrates have been solved to reveal the Michaelis complex and several structures of inefficiently cleaved substrates have trapped the gem-diol intermediates. A structure was solved with the product bound and the missing structure is the protonated gem-diol (Rodriguez et al., 1993);(Prabu-Jeyabalan et al., 2002).

1.4 INHIBITORS OF HIV-1 PROTEASE

PLs are the most potent of all the ARV drugs, since they improve the clinical outcome of HIV patients (85) (Kanters et al., 2017). Three ways that PLs improve the clinical outcome of HIV patients are firstly, they suppress HIV-1 PR by preventing the assembly of mature virion and this prevents viral replication and decrease viral load. Secondly, PIs inhibit PR cleavage of host cell proteins and this reduces PR-related cytotoxicity, apoptosis and necrosis in infected host cells. Thirdly, they have anti-apoptotic effects via a mechanism not directly related to HIV-1 PR suppression, thus, making them very essential to any HAART and their development of viral resistance may have dire public health consequences. Currently there are more than two dozen

antiretroviral drugs to treat HIV infection. This is fundamental because these inhibitors prevent substrate association. One such candidate is the Protease inhibitor, are the C3- substituted cyclopentyltetrahydrofuranyl (Cu-THF) compounds, which is also forms part of the flap-binding small molecule inhibitors (Rizza and Badley, 2008)

Ghosh *et al.* found that several Cp-THFs effectively inhibit a panel of dr-PRs in vitro investigation. This demonstrated their potential to combat dr-PRs. Another allosteric PI class being investigated is dimerization inhibitors. Bouras *et al.* first reported a dimerization inhibitor that interfered with the β -sheet dimerization interface of the HIV-1 PR monomer. These PLs inhibit HIV-1 PR by binding to the dimerization interface and preventing the formation of the active PR homodimer. All FDA-approved competitive inhibitors, including allosteric inhibitors that are currently in development and discovered through a structure-based design (Ghosh *et al.*, 2011).

1.4.1 Natural product-inspired inhibitors of aspartic proteases

Natural HIV protease inhibitors and traditional herbal medicine are known to repress HIV. These modes of treatment are gaining popularity in the treatment of diseases such as HIV (Newman and Cragg, 2020) .

Statins having inhibitory properties against protease

Actinomycetes is a natural product that was discovered in 1970, possessing the ability to secrete a natural product that can inhibit the digestive protease pepsin. Pepstatin is an inhibitor and a peptide that contains the noncanonical amino acid statine (Figure 1.2). An analogue of leucine, statine generate potent competitive inhibitors of aspartic protease when incorporated into peptide substrates at the P1 position. The hydroxyethylene group mimics the tetrahedral formed after the first step of catalysis of HIV protease (Marciniszyn Jr *et al.*, 1976).

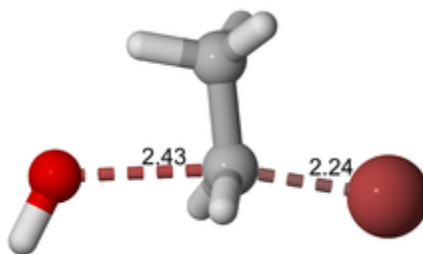


Figure 1. 2: Statine

COA-FS herbal medicine

This organisation, Centre of awareness-Food Supplement (COA-FS) based in Cape Coast, Ghana, identified the herbal medicine that contains six Africa plants, namely *Persea americana*, *Azadirachta indica*, *Carica papaya*, *Ocimum viridae*, *Spondias mombin* and *Vernonia amygdalina*. Nine of the selected phytochemical compounds (EGA,EPG,LNT,BIT,GA,IST,STG,PTA and NGN) were present in the COA-FS medicine (Kehinde *et al.*, 2019).

Ribosome-inactivating proteins (RIPs)

The mushroom RIPs marmorin from *Hypsizigus marmoreus*, hypsin from *H. marmors*, lyophyllin from *Lyophyllum shimeji* and velutin from *Flammulina velutipes* all possessed inhibitory traits for HIV-1 reverse transcriptase with an IC_{50} value of 30 μ M, 8 μ M, 7.9 μ M and 360 μ M respectively (Puri *et al.*, 2012).

Inhibitory compounds of other natural products

Some natural products including polyphenols, alkaloids, xanthenes polysaccharides, coumarins, terpenoids and proteins have been reported to have anti-HIV activity (Ng *et al.*, 1997).

1.4.2 The first generation of clinical inhibitors

First-generation inhibitors (Figure 1.3) were based on hydroxyethylene and hydroxyethylamine isosteres (Ghosh *et al.*, 2003).

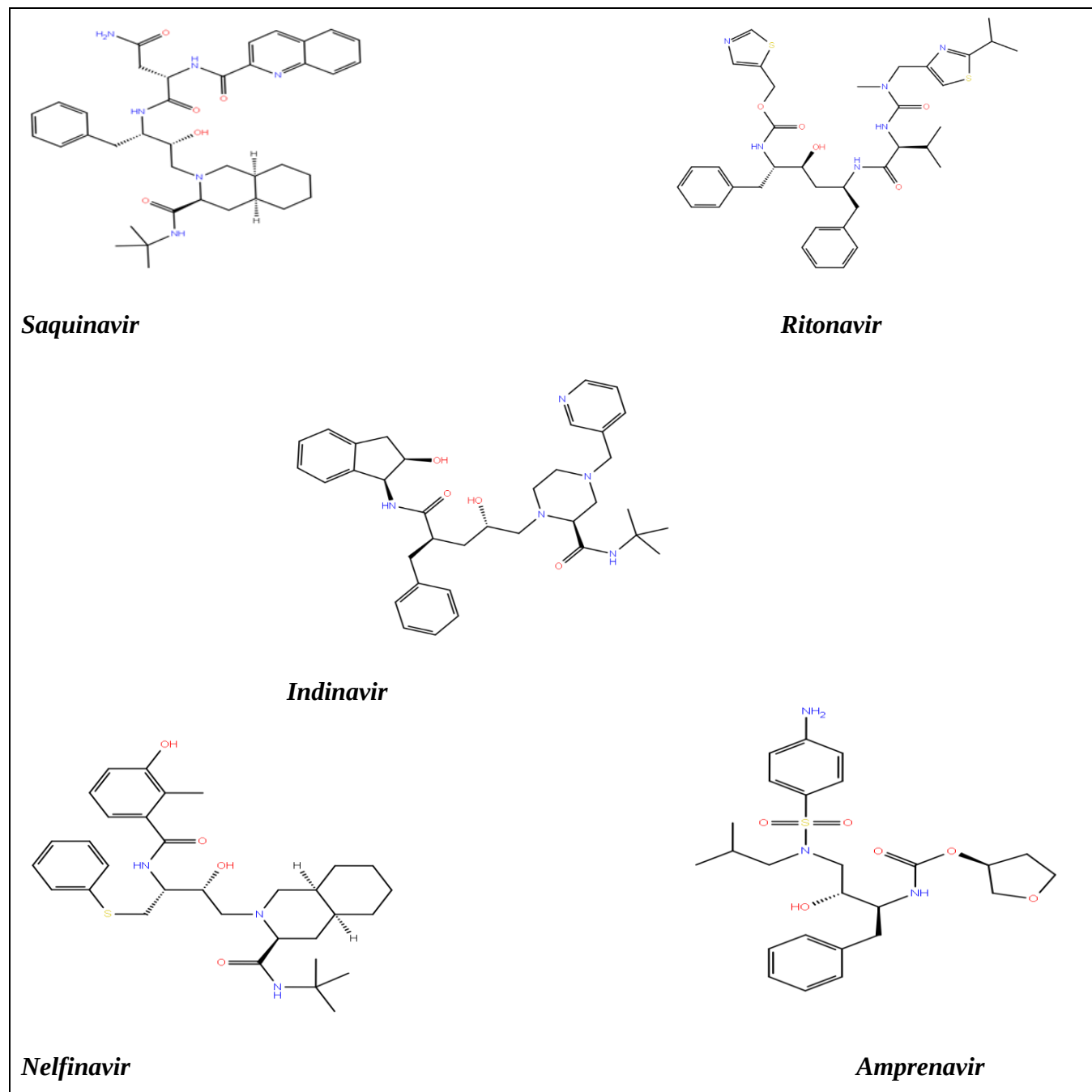


Figure 1. 3: The first generation of clinical inhibitors

Saquinavir

Saquinavir was discovered by Hoffmann-La Roche. In 1995 Saquinavir ($K_i = 0.12 \text{ nM}$) became the first protease inhibitor to gain regulatory approval for the treatment of AIDS. This antiretroviral drug blocks the HIV protease and is available as a generic medication. The design of this drug was based on the unusual property of HIV protease being able to accommodate a proline residue at the P1' position. Saquinavir has the hydroxyethylamine isostere replacement for the Phe-Pro cleavage site. Common side effects include diarrhoea, tiredness, vomiting and nausea, heart block, liver problem and QT prolongation. The structure has the R stereochemistry at the carbon bearing the central hydroxyl group (Gulick *et al.*, 1997).

Ritonavir

The FDA approved Ritonavir in 1996. Side effects include loss of appetite and numbness of the hands and feet. Ritonavir is a pharmacokinetic enhancer. It inhibits metabolic enzymes (cytochrome P-450) 3A, that degrade Lopinavir a protease inhibitor. Ritonavir ($K_i = 0.015 \text{ nM}$) therefore increases the plasma levels of Lopinavir (Kempf *et al.*, 1995).

Indinavir

Indinavir was discovered in 1996. It is a soluble white powder administered orally. It is combined with other retroviral drugs preventing the HIV-Protease to function normally. This antiretroviral drug ($K_i = 0.36 \text{ nM}$) is no longer recommended due to side effects (Dorsey *et al.*, 1994).

Nelfinavir

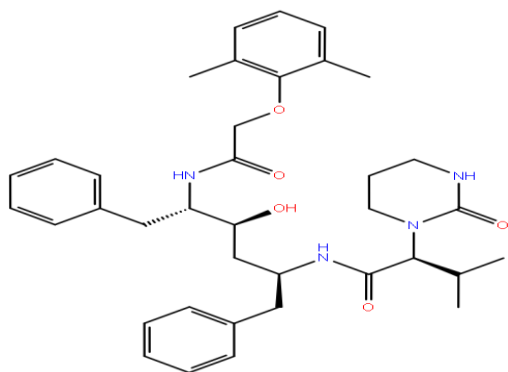
Nelfinavir was discovered in 1997. It is used in reducing viral replication. It has been used together with nucleoside reverse transcriptase inhibitors (NRTIs) for the treatment-naïve patients. ($K_i = 2 \text{ nM}$) (Kaldor *et al.*, 1997).

Amprenavir

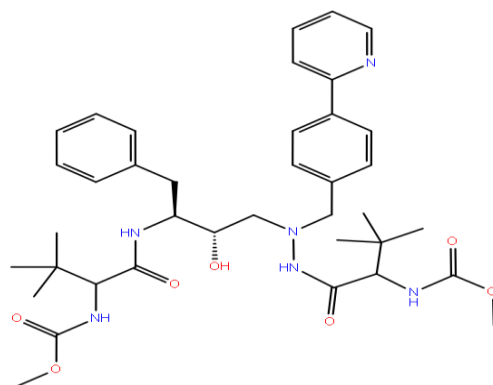
Discovered in 1999, is the last of the first generation of HIV-Protease, Amprenavir ($K_i = 0.6$ nM), approved in 1999, was designed with great hope to avoid severe side effects, bulking hydrophobic groups and amide. However, the first generation of HIV-Protease suffered from high susceptibility to drug resistance mutations. This short coming opened a doorway for the second generation of inhibitors (Kim *et al.*, 1995).

1.4.3 The second generation of protease inhibitors

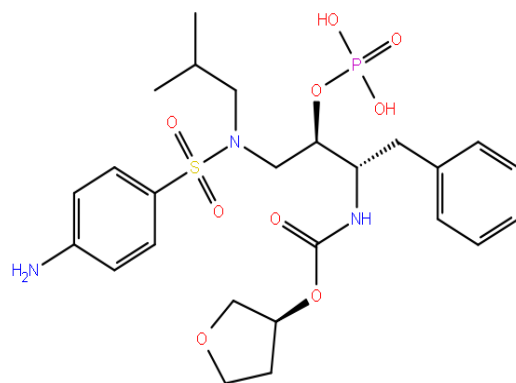
The second generation of protease inhibitors (Figure 1.4) came in as a result to improve the properties of the first-generation protease inhibitors, which led to low half-life, high metabolic clearance and largely poor oral bioavailability and required more frequent dosing (Ghosh *et al.*, 2011) , (Ghosh *et al.*, 2003).



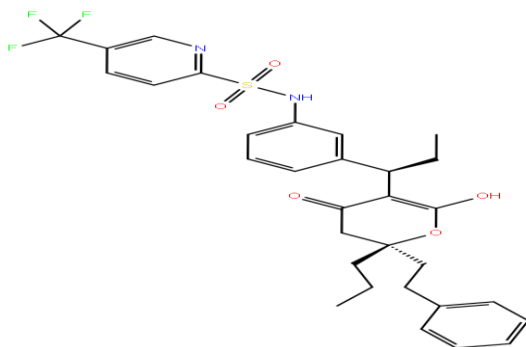
Lopinavir



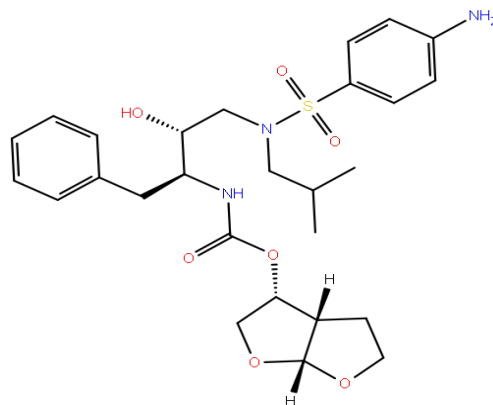
Atazanavir



Fosamprenavir



Tipranavir



Darunavir

Figure 1. 4: The second generation of protease inhibitors

Lopinavir

Lopinavir was discovered in 2000 and is classified as a protease inhibitor (Sham *et al.*, 1998).

Atazanavir

In 2003, Atazanavir was investigated to a combination suitable for oral dosing with a longer half-life than previous protease (Iwamoto *et al.*, 2008) .

Fosamprenavir

Fosamprenavir was discovered in (1999). The human body metabolises fosamprenavir to form amprenavir, which is the active ingredient (Caturegli, 2009).

Tipranavir

This antiretro-viral drug Tipranavir, should be taken in combination with other antiretroviral drugs and ritonavir (Flexner *et al.*, 2005). It was discovered in 2005.

Darunavir

Darunavir 2006, inhibits the cleavage of HIV gag and the gag-pol polyproteins (Ghosh *et al.*, 2006)

Table 1. 1: General characteristics and current indications of licensed antiretroviral drugs

Protease Inhibitor	Initial FDA Approval	Boosting	Daily Pill Burden	Optimal Formulation	Dose (mg) PI/ritonavir	Current Indication
Saquinavir HGC	1995	YES	2 X 3	Tabl. 500mg	1000/100BID	T.N AND T.E
Ritonavir	1996	—	—	Caps 100mg	—	Pharm. E
Indinavir	1996	NO	3 X 2	Caps 400mg	800 TID	T.E
		YES	2 X 3	Caps 400mg	400/100 BID	T.E
Nelfinavir	1997	NO	2 X 5	Tabl. 250mg	1250 BID	Prg. Only
		NO	2 X 2	Tabl. 700mg	1400 BID	T.N
Fosamprenavir (a)	1999	YES	2 X 3	Tabl. 700mg	1400/100 BID	T.N AND T.E
		YES	1 X 3	Tabl. 700mg	1400/100QD	T.N
Lopinavir (b) - ritonavir	2000	YES	2 X 2	Tabl. 200/50mg	400/100 BID	T.N AND T.E
		co-formulate	1 X 4	Tabl. 200/50mg	800/200QD	T.N
		YES co-formulated				
Atazanavir (c)	2003	NO	1 X 2	Caps 200mg	400 QD	T.N
		YES	1 X 2	Caps 300mg	300/100QD	T.N AND T.E
Tipranavir	2005	YES	2 X 4	Caps 250mg	500/200BID	T.E
Darunavir	2006	YES	2 X 2	Tabl. 600mg	600/100BID	T.E
		YES	1 X 3	Tabl. 400mg	800/100QD	T.N

a) Amprenavir was approved in 1999 and discontinued 2004, replaced with Fosamprenavir. The prodrug Fosamprenavir was approved in 2003. Unboosted or once daily fosamprenavir is not approved by the European Agency (EMA). b) Once daily Lopinavir is not approved by the EMA. c) Unboosted Atazanavir is not approved by the EMA. (Flamholz and Gisslén, 2008)

Note: T.N: Treatment of naïve patients; Therapy experience; Pharm. E: Pharmacokinetics; Prg. Only = Pregnant only; Tabl. = Tablet; Caps = Capsule; QD = Once a day; BID = Twice a day; TID = Three times a day

1.4.4 Reverse transcriptase inhibitors

Reverse transcriptase (RT) is the one that is responsible for the transformation of the single-stranded RNA viral genome into double stranded DNA. This enzyme is therefore regarded as a crucial target in the development and investigation of new HIV-1 protease inhibitors. Two types are briefly discussed below, namely, nucleotide reverse transcriptase inhibitors (NRTIs) and non-nucleoside Reverse-Transcriptase Inhibitors (NNRTIs) (Gavegnano and Schinazi, 2009).

NRTIs

NRTIs blocks and prevents HIV protease from making new copies of itself. It competes with the endogenous nucleotide at the polymerase active site after being phosphorylated in vivo, acting as chain terminators to prevent DNA extension (Perno *et al.*, 2006).

NNRTIs

When an HIV infects a host cell, it copies its own genetic code into the host cell's DNA, "programming" the cell to create new copies of the HIV. NNRTIs attaches itself to the viral enzyme, which is the reverse transcriptase and prevents the enzyme from converting RNA to DNA. NNRTIs bind in a flexible allosteric pocket located at approximately 10 Å from the DNA polymerizing processing site, changing the conformation of RT and restraining HIV-replication (Aquaro *et al.*, 2006).

1.4.5 Integrase inhibitors

These are a class of antiretroviral drug designed to block the activities of one of the viral enzymes', integrase. A very well established medical use was developed through the integrase inhibitors called integrase strand transfer inhibitor (INSTIs) (Summa *et al.*, 2008).

1.4.6 Protease inhibitors

Protease inhibitors are designed to block an HIV enzyme called protease by preventing newly formed immature HIV from becoming a mature virus that has the capacity to infect a new host cell (Yarchoan *et al.*, 1994).

1.4.7 Entry/fusion inhibitors

Enfuvirtide and maraviroc are the two approved entry inhibitors against HIV, to date. Enfuvirtide is also called Fuzeon, T-20 is derived from gp41 an HIV envelope protein, which inhibits hairpin formation that is critical for the fusion of viral envelope with host membrane (Perno *et al.*, 1998).

On the other hand, enfuvirtide inhibits HIV-1 entry into different target cells including macrophages, PBMCs and immature dendritic cells (Yi *et al.*, 2008). A further comprehensive study of these inhibitors in primary macrophages is further needed. In addition, maraviroc is a small molecule, which binds with CCR5 receptor reversibly and prevents the virus host interactions and it is so far the only CCR5 antagonist licensed for the treatment of HIV-infected patients. Aplaviroc, vicriviro and TAK-779 are no more considered for clinical development. This is due to serious side effects and a lack of clinical efficacy. New, cocktail of CCR5 antagonists are in the different stages of development. Notably, maraviroc has encountered its resistance and the mechanisms of resistance has been studied (Dorr *et al.*, 2005).

1.4.8 Multi-class combination drugs

Combination drugs combine multiple medications into one drug form. This type of regimen is usually used to treat people who have never taken HIV medications before (Desai *et al.*, 2012).

REFERENCE LIST

ANTONOV, V. K., GINODMAN, L. M., RUMSH, L. D., KAPITANNIKOV, Y. V., BARSHEVSKAYA, T. N., YAVASHEV, L. P., GUROVA, A. G. & VOLKOVA, L. I. 1981. Studies on the mechanisms of action of proteolytic enzymes using heavy oxygen exchange. *European journal of biochemistry*, 117, 195-200.

AQUARO, S., SVICHER, V., SCHOLS, D., POLLICITA, M., ANTINORI, A., BALZARINI, J. & PERNO, C. F. 2006. Mechanisms underlying activity of antiretroviral drugs in HIV-1-infected macrophages: new therapeutic strategies. *Journal of leukocyte biology*, 80, 1103-1110.

ARENS, M., JOSEPH, T., NAG, S., MILLER, J. P., POWDERLY, W. G. & RATNER, L. 1993. Alterations in spliced and unspliced HIV-1-specific RNA detection in peripheral blood mononuclear cells of individuals with varying CD4-positive lymphocyte counts. *AIDS research and human retroviruses*, 9, 1257-1263.

BABA, T. W., LISKA, V., HOFMANN-LEHMANN, R., VLASAK, J., XU, W., AYE HUNIE, S., CAVACINI, L. A., POSNER, M. R., KATINGER, H. & STIEGLER, G. 2000. Human neutralizing monoclonal antibodies of the IgG1 subtype protect against mucosal simian-human immunodeficiency virus infection. *Nature medicine*, 6, 200-206.

BARBAULT, F. & MAUREL, F. 2015. Simulation with quantum mechanics/molecular mechanics for drug discovery. *Expert Opinion on Drug Discovery*, 10, 1047-1057.

BASHORE, C., JAISHANKAR, P., SKELTON, N. J., FUHRMANN, J., HEARN, B. R., LIU, P. S., RENSLO, A. R. & DUEBER, E. C. 2020. Cyanopyrrolidine inhibitors of ubiquitin specific protease 7 mediate desulfhydration of the active-site cysteine. *ACS Chemical Biology*, 15, 1392-1400.

BLANCO, R., CARRASCO, L. & VENTOSO, I. 2003. Cell killing by HIV-1 protease. *Journal of Biological Chemistry*, 278, 1086-1093.

BRIK, A. & WONG, C.-H. 2003. HIV-1 protease: mechanism and drug discovery. *Organic & biomolecular chemistry*, 1, 5-14.

BURNOUF, D. Y., OLIERIC, V., WAGNER, J., FUJII, S., REINBOLT, J., FUCHS, R. P. & DUMAS, P. 2004. Structural and biochemical analysis of sliding clamp/ligand interactions suggest a competition between replicative and translesion DNA polymerases. *Journal of molecular biology*, 335, 1187-1197.

CATUREGLI, P. 2009. Drug Profile: Lexiva. *Discovery medicine*, 3, 20-21.

COLMAN, P. M. 2009. New antivirals and drug resistance. *Annual review of biochemistry*, 78, 95-118.

COSTIN, J. M., RAUSCH, J. M., GARRY, R. F. & WIMLEY, W. C. 2007. Viroporin potential of the lentivirus lytic peptide (LLP) domains of the HIV-1 gp41 protein. *Virology journal*, 4, 1-14.

COUNTS, C. J., HO, P. S., DONLIN, M. J., TAVIS, J. E. & CHEN, C. 2015. A functional interplay between human immunodeficiency virus type 1 protease residues 77 and 93 involved in differential regulation of precursor autoprocessing and mature protease activity. *PLoS One*, 10, e0123561.

DEBOUCK, C., GORNIK, J. G., STRICKLER, J. E., MEEK, T. D., METCALF, B. W. & ROSENBERG, M. 1987. Human immunodeficiency virus protease expressed in Escherichia coli exhibits autoprocessing and specific maturation of the gag precursor. *Proceedings of the National Academy of Sciences*, 84, 8903-8906.

DESAI, M., IYER, G. & DIKSHIT, R. 2012. Antiretroviral drugs: critical issues and recent advances. *Indian journal of pharmacology*, 44, 288.

DESJARLAIS, R., SEIBEL, G., KUNTZ, I., FURTH, P., ALVAREZ, J., DE MONTELLANO, P. O., DECAMP, D., BABE, L. & CRAIK, C. 1990. Structure-based design of nonpeptide inhibitors

specific for the human immunodeficiency virus 1 protease. *Proceedings of the National Academy of Sciences*, 87, 6644-6648.

DICKER, K., JÄRVELIN, A. I., GARCIA-MORENO, M. & CASTELLO, A. The importance of virion-incorporated cellular RNA-Binding Proteins in viral particle assembly and infectivity. *Seminars in cell & developmental biology*, 2021. Elsevier, 108-118.

DORR, P., WESTBY, M., DOBBS, S., GRIFFIN, P., IRVINE, B., MACARTNEY, M., MORI, J., RICKETT, G., SMITH-BURCHNELL, C. & NAPIER, C. 2005. Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. *Antimicrobial agents and chemotherapy*, 49, 4721-4732.

DORSEY, B. D., LEVIN, R. B., MCDANIEL, S. L., VACCA, J. P., GUARE, J. P., DARKE, P. L., ZUGAY, J. A., EMINI, E. A. & SCHLEIF, W. A. 1994. L-735,524: the design of a potent and orally bioavailable HIV protease inhibitor. *Journal of medicinal chemistry*, 37, 3443-3451.

EATEMADI, A., AIYELABEGAN, H. T., NEGAHDARI, B., MAZLOMI, M. A., DARAEI, H., DARAEI, N., EATEMADI, R. & SADRODDINY, E. 2017. Role of protease and protease inhibitors in cancer pathogenesis and treatment. *Biomedicine & Pharmacotherapy*, 86, 221-231.

FELBER, B. K., DRYSDALE, C. M. & PAVLAKIS, G. N. 1990. Feedback regulation of human immunodeficiency virus type 1 expression by the Rev protein. *Journal of virology*, 64, 3734-3741.

FESIK, S. W. 1993. NMR structure-based drug design. *Journal of biomolecular NMR*, 3, 261-269.

FLAMHOLC, L. & GISSLÉN, M. 2008. Once-daily fosamprenavir with ritonavir in the treatment of HIV infection in therapy-naïve patients. *Therapeutics and clinical risk management*, 4, 1281.

FLEXNER, C., BATE, G. & KIRKPATRICK, P. 2005. Tipranavir. *Nature reviews Drug discovery*, 4, 955-957.

GALETTO, R., MOUMEN, A., GIACOMONI, V., VÉRON, M., CHARNEAU, P. & NEGRONI, M. 2004. The structure of HIV-1 genomic RNA in the gp120 gene determines a recombination hot spot in vivo. *Journal of Biological Chemistry*, 279, 36625-36632.

GAVEGNANO, C. & SCHINAZI, R. F. 2009. Antiretroviral therapy in macrophages: implication for HIV eradication. *Antiviral Chemistry and Chemotherapy*, 20, 63-78.

GHOSH, A. K., ANDERSON, D. D. & MITSUYA, H. 2003. The FDA Approved HIV-1 Protease Inhibitors for Treatment of HIV/AIDS. *Burger's Medicinal Chemistry and Drug Discovery*, 1-74.

GHOSH, A. K., CHAPSAL, B. D., PARHAM, G. L., STEFFEY, M., AGNISWAMY, J., WANG, Y.-F., AMANO, M., WEBER, I. T. & MITSUYA, H. 2011. Design of HIV-1 Protease Inhibitors with C3-Substituted Hexahydrocyclopentafuranyl Urethanes as P2-Ligands: Synthesis, Biological Evaluation, and Protein–Ligand X-ray Crystal Structure. *Journal of medicinal chemistry*, 54, 5890-5901.

GHOSH, A. K., OSSWALD, H. L. & PRATO, G. 2016. Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS. *Journal of medicinal chemistry*, 59, 5172-5208.

GHOSH, A. K., RAMU SRIDHAR, P., KUMARAGURUBARAN, N., KOH, Y., WEBER, I. T. & MITSUYA, H. 2006. Bis-tetrahydrofuran: a privileged ligand for darunavir and a new generation of HIV protease inhibitors that combat drug resistance. *ChemMedChem: Chemistry Enabling Drug Discovery*, 1, 939-950.

GÖTTLINGER, H. G., SODROSKI, J. G. & HASELTINE, W. A. 1989. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proceedings of the National Academy of Sciences*, 86, 5781-5785.

GULICK, R. M., MELLORS, J. W., HAVLIR, D., ERON, J. J., GONZALEZ, C., MCMAHON, D., RICHMAN, D. D., VALENTINE, F. T., JONAS, L. & MEIBOHM, A. 1997. Treatment with

indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. *New England Journal of Medicine*, 337, 734-739.

GULNIK, S. V., SUVOROV, L. I., LIU, B., YU, B., ANDERSON, B., MITSUYA, H. & ERICKSON, J. W. 1995. Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochemistry*, 34, 9282-9287.

GUPTA, R., HILL, A., SAWYER, A. W. & PILLAY, D. 2008. Emergence of drug resistance in HIV type 1-infected patients after receipt of first-line highly active antiretroviral therapy: a systematic review of clinical trials. *Clinical infectious diseases*, 47, 712-722.

HOFMANN, T., HODGES, R. S. & JAMES, M. N. 1984. Effect of pH on the activities of penicillopepsin and Rhizopus pepsin and a proposal for the productive substrate binding mode in penicillopepsin. *Biochemistry*, 23, 635-643.

HU, J., PROTZER, U. & SIDDIQUI, A. 2019. Revisiting hepatitis B virus: challenges of curative therapies. *Journal of virology*, 93, e01032-19.

HUFF, J. R. 1991. HIV protease: a novel chemotherapeutic target for AIDS. *Journal of medicinal chemistry*, 34, 2305-2314.

HYLAND, L. J., TOMASZEK JR, T. A. & MEEK, T. D. 1991. Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. *Biochemistry*, 30, 8454-8463.

IMAMICHI, T. 2004. Action of anti-HIV drugs and resistance: reverse transcriptase inhibitors and protease inhibitors. *Current pharmaceutical design*, 10, 4039-4053.

INUZUKA, M., HAYAKAWA, M. & INGI, T. 2005. Serinc, an activity-regulated protein family, incorporates serine into membrane lipid synthesis. *Journal of Biological Chemistry*, 280, 35776-35783.

IWAMOTO, M., WENNING, L. A., MISTRY, G. C., PETRY, A. S., LIOU, S. Y., GHOSH, K., BREIDINGER, S., AZROLAN, N., GUTIERREZ, M. J. & BRIDSON, W. E. 2008. Atazanavir modestly increases plasma levels of raltegravir in healthy subjects. *Clinical infectious diseases*, 47, 137-140.

JAMJOOM, G., NASO, R. & ARLINGHAUS, R. 1977. Further characterization of intracellular precursor polyproteins of Rauscher leukemia virus. *Virology*, 78, 11-34.

JASKOLSKI, M., TOMASSELLI, A. G., SAWYER, T. K., STAPLES, D. G., HEINRIKSON, R. L., SCHNEIDER, J., KENT, S. B. & WLODAWER, A. 1991. Structure at 2.5-Å resolution of chemically synthesized Human Immunodeficiency Virus Type 1 protease complexed with a hydroxyethylene-based inhibitor. *Biochemistry*, 30, 1600-1609.

JETZT, A. E., YU, H., KLARMANN, G. J., RON, Y., PRESTON, B. D. & DOUGHERTY, J. P. 2000. High rate of recombination throughout the human immunodeficiency virus type 1 genome. *Journal of virology*, 74, 1234-1240.

JONES, J. S., ALLAN, R. W. & TEMIN, H. M. 1994. One retroviral RNA is sufficient for synthesis of viral DNA. *Journal of Virology*, 68, 207-216.

JONES, S. J. 2018. *Investigating Nonlinear Enzyme Kinetics as an Internal Control System for Nanoreactor Drug Release*. University of Leeds.

JOVANOVIĆ, M., TURKOVIĆ, N., IVKOVIĆ, B., VUJIĆ, Z., NIKOLIĆ, K. & GRUBIŠIĆ, S. 2021. 3D-QSAR, molecular docking and in silico ADMET studies of propiophenone derivatives with anti-HIV-1 protease activity. *Structural Chemistry*, 32, 2341-2353.

KALDOR, S. W., KALISH, V. J., DAVIES, J. F., SHETTY, B. V., FRITZ, J. E., APPELT, K., BURGESS, J. A., CAMPANALE, K. M., CHIRGADZE, N. Y. & CLAWSON, D. K. 1997. Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *Journal of medicinal chemistry*, 40, 3979-3985.

- KANTERS, S., PARK, J. J., CHAN, K., SOCIAS, M. E., FORD, N., FORREST, J. I., THORLUND, K., NACHEGA, J. B. & MILLS, E. J. 2017. Interventions to improve adherence to antiretroviral therapy: a systematic review and network meta-analysis. *The lancet HIV*, 4, e31-e40.
- KAUSHIK, A. C., KUMAR, A., BHARADWAJ, S., CHAUDHARY, R. & SAHI, S. 2018. *Bioinformatics techniques for drug discovery: applications for complex diseases*, Springer.
- KEHINDE, I., RAMHARACK, P., NLOOTO, M. & GORDON, M. 2019. The pharmacokinetic properties of HIV-1 protease inhibitors: A computational perspective on herbal phytochemicals. *Heliyon*, 5, e02565.
- KEMPF, D. J., MARSH, K. C., DENISSEN, J. F., MCDONALD, E., VASAVANONDA, S., FLENTGE, C. A., GREEN, B. E., FINO, L., PARK, C. H. & KONG, X.-P. 1995. ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans. *Proceedings of the National Academy of Sciences*, 92, 2484-2488.
- KIM, E., BAKER, C., DWYER, M., MURCKO, M. A., RAO, B., TUNG, R. & NAVIA, M. 1995. Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. *Journal of the American Chemical Society*, 117, 1181-1182.
- KING, N. M., PRABU-JEYABALAN, M., NALIVAICA, E. A., WIGERINCK, P., DE BÉTHUNE, M.-P. & SCHIFFER, C. A. 2004. Structural and thermodynamic basis for the binding of TMC114, a next-generation human immunodeficiency virus type 1 protease inhibitor. *Journal of virology*, 78, 12012-12021.
- KING, S. R. 1994. HIV: virology and mechanisms of disease. *Annals of emergency medicine*, 24, 443-449.
- KOHL, N. E., EMINI, E. A., SCHLEIF, W. A., DAVIS, L. J., HEIMBACH, J. C., DIXON, R., SCOLNICK, E. M. & SIGAL, I. S. 1988. Active human immunodeficiency virus protease is required for viral infectivity. *Proceedings of the National Academy of Sciences*, 85, 4686-4690.

- KONG, X.-P., ONRUST, R., O'DONNELL, M. & KURIYAN, J. 1992. Three-dimensional structure of the β subunit of E. coli DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell*, 69, 425-437.
- KOZAL, M. J., SHAH, N., SHEN, N., YANG, R., FUCINI, R., MERIGAN, T. C., RICHMAN, D. D., MORRIS, D., HUBBELL, E. & CHEE, M. 1996. Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. *Nature medicine*, 2, 753-759.
- KUMAR, M., CURTIS, A. & HOSKINS, C. 2018. Application of nanoparticle technologies in the combat against anti-microbial resistance. *Pharmaceutics*, 10, 11.
- LACO, G. S., SCHALK-HIHI, C., LUBKOWSKI, J., MORRIS, G., ZDANOV, A., OLSON, A., ELDER, J. H., WLODAWER, A. & GUSTCHINA, A. 1997. Crystal structures of the inactive D30N mutant of feline immunodeficiency virus protease complexed with a substrate and an inhibitor. *Biochemistry*, 36, 10696-10708.
- LEE, T., LACO, G. S., TORBETT, B. E., FOX, H. S., LERNER, D. L., ELDER, J. H. & WONG, C.-H. 1998. Analysis of the S3 and S3' subsite specificities of feline immunodeficiency virus (FIV) protease: development of a broad-based protease inhibitor efficacious against FIV, SIV, and HIV in vitro and ex vivo. *Proceedings of the National Academy of Sciences*, 95, 939-944.
- LEUNG, D., ABBENANTE, G. & FAIRLIE, D. P. 2000. Protease inhibitors: current status and future prospects. *Journal of medicinal chemistry*, 43, 305-341.
- LIU, M. A. 2019. A comparison of plasmid DNA and mRNA as vaccine technologies. *Vaccines*, 7, 37.
- LV, Z., CHU, Y. & WANG, Y. 2015. HIV protease inhibitors: a review of molecular selectivity and toxicity. *Hiv/aids (Auckland, NZ)*, 7, 95.

- MALIM, M. H., TILEY, L. S., MCCARN, D. F., RUSCHE, J. R., HAUBER, J. & CULLEN, B. R. 1990. HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence. *Cell*, 60, 675-683.
- MAO, Y., WANG, L., GU, C., HERSCHHORN, A., XIANG, S.-H., HAIM, H., YANG, X. & SODROSKI, J. 2012. Subunit organization of the membrane-bound HIV-1 envelope glycoprotein trimer. *Nature structural & molecular biology*, 19, 893-899.
- MARCINISZYN JR, J., HARTSUCK, J. A. & TANG, J. 1976. Mode of inhibition of acid proteases by pepstatin. *Journal of Biological Chemistry*, 251, 7088-7094.
- MASEKO, S., PADAYACHEE, E., MAPHUMULO, S., GOVENDER, T., SAYED, Y., MAGUIRE, G., LIN, J., NAICKER, T., BAIJNATH, S. & GERHARDUS, K. H. 2019. Kinetic and thermodynamic characterisation of HIV-protease inhibitors against E35D↑ G↑ S mutant in the South African HIV-1 subtype C protease. *Journal of enzyme inhibition and medicinal chemistry*, 34, 1451-1456.
- MISCHE, C. C., YUAN, W., STRACK, B., CRAIG, S., FARZAN, M. & SODROSKI, J. 2005. An alternative conformation of the gp41 heptad repeat 1 region coiled coil exists in the human immunodeficiency virus (HIV-1) envelope glycoprotein precursor. *Virology*, 338, 133-143.
- MULICHAK, A. M., HUI, J., TOMASSELLI, A., HEINRIKSON, R., CURRY, K., TOMICH, C.-S., THAISRIVONGS, S., SAWYER, T. & WATENPAUGH, K. 1993. The crystallographic structure of the protease from human immunodeficiency virus type 2 with two synthetic peptidic transition state analog inhibitors. *Journal of Biological Chemistry*, 268, 13103-13109.
- NASSAL, M., JUNKER-NIEPMANN, M. & SCHALLER, H. 1990. Translational inactivation of RNA function: discrimination against a subset of genomic transcripts during HBV nucleocapsid assembly. *Cell*, 63, 1357-1363.

- NAVIA, M. A., FITZGERALD, P. M., MCKEEVER, B. M., LEU, C.-T., HEIMBACH, J. C., HERBER, W. K., SIGAL, I. S., DARKE, P. L. & SPRINGER, J. P. 1989. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature*, 337, 615-620.
- NEAMATI, N., MARCHAND, C. & POMMIER, Y. 2000. HIV-I integrase inhibitors: Past, present, and future. *Advances in pharmacology*, 49, 147-165.
- NEWMAN, D. J. & CRAGG, G. M. 2020. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *Journal of Natural Products*, 83, 770-803.
- NG, T., HUANG, B., FONG, W. & YEUNG, H. 1997. Anti-human immunodeficiency virus (anti-HIV) natural products with special emphasis on HIV reverse transcriptase inhibitors. *Life sciences*, 61, 933-949.
- OHTAKA, H., SCHÖN, A. & FREIRE, E. 2003. Multidrug resistance to HIV-1 protease inhibition requires cooperative coupling between distal mutations. *Biochemistry*, 42, 13659-13666.
- PAIVA, P., SOUSA, S. F., RAMOS, M. J. & FERNANDES, P. A. 2018. Understanding the catalytic machinery and the reaction pathway of the malonyl-acetyl transferase domain of human fatty acid synthase. *ACS Catalysis*, 8, 4860-4872.
- PALMER, S., SHAFER, R. W. & MERIGAN, T. C. 1999. Highly drug-resistant HIV-1 clinical isolates are cross-resistant to many antiretroviral compounds in current clinical development. *AIDS (London, England)*, 13, 661.
- PANDA, M., PUROHIT, P. & MEHER, B. R. 2022. Structure-based virtual screening, ADMET profiling, and molecular dynamics simulation studies on HIV-1 protease for identification of active phytocompounds as potential anti-HIV agents. *Molecular Simulation*, 1-19.
- PERNO, C.-F., NEWCOMB, F. M., DAVIS, D. A., AQUARO, S., HUMPHREY, R. W., CALIÒ, R. & YARCHOAN, R. 1998. Relative potency of protease inhibitors in monocytes/macrophages

acutely and chronically infected with human immunodeficiency virus. *Journal of Infectious Diseases*, 178, 413-422.

PERNO, C. F., SVICHER, V., SCHOLS, D., POLLICITA, M., BALZARINI, J. & AQUARO, S. 2006. Therapeutic strategies towards HIV-1 infection in macrophages. *Antiviral research*, 71, 293-300.

PODLOGAR, B. L., FARR, R. A., FRIEDRICH, D., TARNUS, C., HUBER, E. W., CREGGE, R. J. & SCHIRLIN, D. 1994. Design, synthesis, and conformational analysis of a novel macrocyclic HIV-protease inhibitor. *Journal of medicinal chemistry*, 37, 3684-3692.

POMMIER, Y., MARCHAND, C. & NEAMATI, N. 2000. Retroviral integrase inhibitors year 2000: update and perspectives. *Antiviral research*, 47, 139-148.

POTEMPA, M., LEE, S.-K., YILMAZ, N. K., NALIVAICA, E. A., ROGERS, A., SPIELVOGEL, E., CARTER JR, C. W., SCHIFFER, C. A. & SWANSTROM, R. 2018. HIV-1 protease uses bi-specific S2/S2' subsites to optimize cleavage of two classes of target sites. *Journal of molecular biology*, 430, 5182-5195.

PRABU-JEYABALAN, M., NALIVAICA, E. & SCHIFFER, C. A. 2000. How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. *Journal of molecular biology*, 301, 1207-1220.

PRABU-JEYABALAN, M., NALIVAICA, E. & SCHIFFER, C. A. 2002. Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. *Structure*, 10, 369-381.

PURI, M., KAUR, I., PERUGINI, M. A. & GUPTA, R. C. 2012. Ribosome-inactivating proteins: current status and biomedical applications. *Drug discovery today*, 17, 774-783.

RAO, J. M., ERICKSON, J. W. & WLODAWER, A. 1991. Structural and evolutionary relationships between retroviral and eukaryotic aspartic proteinases. *Biochemistry*, 30, 4663-4671.

RIZZA, S. A. & BADLEY, A. D. 2008. HIV protease inhibitors impact on apoptosis. *Medicinal Chemistry*, 4, 75-79.

RODRIGUEZ, E. J., ANGELES, T. S. & MEEK, T. D. 1993. Use of nitrogen-15 kinetic isotope effects to elucidate details of the chemical mechanism of human immunodeficiency virus 1 protease. *Biochemistry*, 32, 12380-12385.

RUTENBER, E., FAUMAN, E. B., KEENAN, R. J., FONG, S., FURTH, P. S., DE MONTELLANO, P. O., MENG, E., KUNTZ, I. D., DECAMP, D. L. & SALTO, R. 1993. Structure of a non-peptide inhibitor complexed with HIV-1 protease. Developing a cycle of structure-based drug design. *Journal of Biological Chemistry*, 268, 15343-15346.

SANSOM, C. E., WU, J. & WEBER, I. T. 1992. Molecular mechanics analysis of inhibitor binding to HIV-1 protease. *Protein Engineering, Design and Selection*, 5, 659-667.

SHAM, H. L., KEMPF, D. J., MOLLA, A., MARSH, K. C., KUMAR, G. N., CHEN, C.-M., KATI, W., STEWART, K., LAL, R. & HSU, A. 1998. ABT-378, a highly potent inhibitor of the human immunodeficiency virus protease. *Antimicrobial agents and chemotherapy*, 42, 3218-3224.

SUMMA, V., PETROCCHI, A., BONELLI, F., CRESCENZI, B., DONGHI, M., FERRARA, M., FIORE, F., GARDELLI, C., GONZALEZ PAZ, O. & HAZUDA, D. J. 2008. Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. *Journal of medicinal chemistry*, 51, 5843-5855.

SZECSI, P. B. 1992. The aspartic proteases. *Scandinavian Journal of Clinical and Laboratory Investigation*, 52, 5-22.

TANOLI, S. A., TANOLI, N. U., BONDANCIA, T. M., USMANI, S., UL-HAQ, Z., FERNANDES, J. B., THOMASI, S. S. & FERREIRA, A. G. 2015. Human serum albumin-specific recognition of the natural herbal extract of *Stryphnodendron polyphyllum* through STD NMR, hyphenations and docking simulation studies. *RSC Advances*, 5, 23431-23442.

TOH, H., ONO, M. & MIYATA, T. 1985. Retroviral gag and DNA endonuclease coding sequences in IgE-binding factor gene. *Nature*, 318, 388-389.

TRYLSKA, J., TOZZINI, V., CHIA-EN, A. C. & MCCAMMON, J. A. 2007. HIV-1 protease substrate binding and product release pathways explored with coarse-grained molecular dynamics. *Biophysical journal*, 92, 4179-4187.

UNAIDS 2019. *IV and AIDS in South Africa* [Online]. Available: <https://www.unaids.org/en/resources/fact-sheet>.

UNAIDS. 2020. *HIV and AIDS in South Africa* [Online]. Available: <https://www.unaids.org/en/resources/fact-sheet>.

VAISHNAV, Y., VAISHNAV, M. & WONG-STAAAL, F. 1991. Identification and characterization of a nuclear factor that specifically binds to the Rev response element (RRE) of human immunodeficiency virus type 1 (HIV-1). *The New biologist*, 3, 142-150.

VELAZQUEZ-CAMPOY, A., MUZAMMIL, S., OHTAKA, H., SCHON, A., VEGA, S. & FREIRE, E. 2003. Structural and thermodynamic basis of resistance to HIV-1 protease inhibition: implications for inhibitor design. *Current Drug Targets-Infectious Disorders*, 3, 311-328.

WEBER, I. T., MILLER, M., JASKOLSKI, M., LEIS, J., SKALKKA, A. M. & WLODAWER, A. 1989. Molecular modeling of the HIV-1 protease and its substrate binding site. *Science*, 243, 928-931.

WENSING, A. M., VAN MAARSEVEEN, N. M. & NIJHUIS, M. 2010. Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. *Antiviral research*, 85, 59-74.

WINSLOW, D. L., STACK, S., KING, R., SCARNATI, H., BINCSIK, A. & OTTO, M. J. 1995. Limited sequence diversity of the HIV type 1 protease gene from clinical isolates and in vitro susceptibility to HIV protease inhibitors. *AIDS research and human retroviruses*, 11, 107-113.

WINDSOR, I. W., GOLD, B. & RAINES, R. T. 2018. An $n \rightarrow \pi^*$ interaction in the bound substrate of aspartic proteases replicates the oxyanion hole. *ACS catalysis*, 9, 1464-1471.

WLODAWER, A. & ERICKSON, J. W. 1993. Structure-based inhibitors of HIV-1 protease. *Annual review of biochemistry*, 62, 543-585.

WLODAWER, A., MILLER, M., JASKOLSKI, M., SATHYANARAYANA, B. K., BALDWIN, E., WEBER, I. T., SELK, L. M., CLAWSON, L., SCHNEIDER, J. & KENT, S. 1989. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science*, 245, 616-621.

YANG, H., NKEZE, J. & ZHAO, R. Y. 2012. Effects of HIV-1 protease on cellular functions and their potential applications in antiretroviral therapy. *Cell & bioscience*, 2, 1-8.

YARCHOAN, R., LIETZAU, J. A., NGUYEN, B.-Y., BRAWLEY, O. W., PLUDA, J. M., SAVILLE, M. W., WYVILL, K. M., STEINBERG, S. M., AGBARIA, R. & MITSUYA, H. 1994. A randomized pilot study of alternating or simultaneous zidovudine and didanosine therapy in patients with symptomatic human immunodeficiency virus infection. *Journal of Infectious Diseases*, 169, 9-17.

YEDIDI, R. S., PROTEASA, G., MARTIN, P. D., LIU, Z., VICKREY, J. F., KOVARI, I. A. & KOVARI, L. C. 2014. A multi-drug resistant HIV-1 protease is resistant to the dimerization inhibitory activity of tlf-paff. *Journal of Molecular Graphics and Modelling*, 53, 105-111.

YI, Y., LOFTIN, L., WANG, L., RATCLIFFE, S. J., ISAACMAN-BECK, J. & COLLMAN, R. G. 2008. Entry coreceptor use and fusion inhibitor T20 sensitivity of dual-tropic R5X4 HIV-1 in primary macrophage infection. *Journal of acquired immune deficiency syndromes (1999)*, 47, 285.

CHAPTER 2: LITERITURE REVIEW

2.1 HIV DRUG TARGETS AND DISCOVERIES

2.1.1 Background on drug targets

Recent investigations and useful clues for further design of HIV-1 inhibitors are discussed on this review. This review gives out the findings and the recent discoveries made towards seeking a novel next- generation treatment for HIV-1 protease. Currently there are anti-HIV chemotherapeutics through various strategies by targeting the host-virus targets (Prasad et al., 2017). These strategies include targets into the CD4 receptors; target into the viral envelope glycoprotein gp120, targeting the CXCR4 or the CCR5 co-receptors, targeting the glycoprotein gp41; targeting the substrate-binding site and non- substrate binding site which is the allosteric site and targeting the integrase, reverse transcriptase and the protease and the cellular transcriptional factors (Shi et al., 2016). This further includes inhibitors targeting the entrance channel of HIV-1 such as the gag, nef, TPST2, Protein disulfide isomerase and natural seminal amyloids to mention just a few (Puhl *et al.*, 2019).

2.1.2 The components of HIV-1 genome

To understand the virus genome mechanisms, the protein functionality must be understood. This will assist in proposing relevant virus targets and further contribute to discovering relevant virus targets. Structurally its genome has a less than 10kb that has nine different proteins coded in it. These proteins are involved in splicing processes that produce more than 30 distinct mRNAs. These are prerequisites for structural and enzymatic proteins. Inside this genome, the gag, pol and env are viral polyproteins; the rev and tat are essential regulatory elements and the vir, vpu, vpr and nef are accessory regulatory proteins (Takahashi et al., 2015). Found within the virion particles are vpr, vif and nef together with a viral protease. HIV protease plays a vital role in cleaving the precursor Gag polyprotein to produce structural proteins (Herbein *et al.*, 2010). Pol codes for viral enzymes reverse transcriptase (RT), RNase H, integrase (IN) and HIV protease (Coffin *et al.*, 1997).

The enzyme that is responsible to transcribe DNA from RNA template is the reverse transcriptase. In addition, the integrase is essential for binding or integrating the double stranded viral DNA into the host genome. Gag will code for the precursor gag polyprotein which is processed by necessary protease during maturation to matrix protein (MA), p17, capsid protein (CA), p24, P6 protein, spacer peptide (SP1), (p2), nucleocapsid protein (NC), p7 and spacer peptide 2(SP2), p1 (King, 1994).

The “envelope” codes for gp160, is cleaved by a furin (the host protease), within the endoplasmic reticulum of the host cell. The glycoprotein surfaces, gp 120, attach to the host’s cell and CD4 is produced after post translational processing. This glycoprotein attaches to the CD4 receptors that are present on the lymphocytes. The gp41 or TM embeds on the viral envelope to enable the virus to fuse with the target cells (Mushahwar, 2006).

A matrix p17 surrounds the capsid and the nucleocapsid associates with the genomic RNA (one molecule per hexamer) and it protects the RNA from digestion by. This matrix ensures the biological integrity of the virion particle. The outer coat of the HIV envelope is made up of two fatty molecules layers. These layers contain spikes like protein used to penetrate its way into the host cell (Health, 1998).

The viral elements such as reverse transcriptase, integrase and protease have been intensively investigated, for more than 30 years as therapeutic targets against HIV/AIDS. Lysine tRNA is the primer of the magnesium dependent reverse transcriptase. A matured virus is formed when the capsid that is surrounded by the host cell envelop buds from the host cell, taking some of the host-cell membrane with it (Li and De Clercq, 2016).

HIV trans-activator, tat, regulates the reverse transcriptase of the viral genome RNA (Hayash, 2016). This ensures the regulation of the release of virions from infected cells securing efficient synthesis of viral mRNAs. It binds to the bulged genomic RNA stem-loop secondary structure near the 5’ LTR region forming the trans-activator response element (TAR) and is expressed as 72-amino acid one-exon tat as well as the 86-101 amino-acid two-exon tat (Singleton, 2006). Regulator of expression of virion proteins, rev protein binds to the viral genome via an arginine-rich RNA-binding motif that also acts as a NLS (nuclear localisation signals) which is required for

its transportation into the nucleus from the cytosol during viral replication. It is important for the synthesis of major viral proteins which plays a vital role in viral replication, and it also function to recognise a complex stem-loop structure of the mRNA located in the intron separating coding exon of tat and rev, which is known as the rev response elements (RRE) (Baba, 2004; Feinberg and Greene, 1992).

Vpr is a virion-associated, nucleocytoplasmic shuttling regulatory protein. The role of this protein is to replicate the HIV-1. (Liu *et al.*, 2020).

Table 2. 1: Proteins encoded by HIV-1 genome

Class /Composite	Name	Abbreviation	Primary Protein	Processed Protein Product
Viral particle	Group-specific antigen	<i>Gag</i>	Gag polyprotein	MA, CA, SP1, NC, SP2, P6
Structural protein	Envelope	<i>Env, Pol</i>	Gp 160, Pol polyprotein	gp120, gp41, RT, RNase H, IN, PR
Essential regulatory elements	trans-activator regulator of expression of virion proteins	<i>Tat, Rev</i>	Tat, Rev	
Accessory regulatory protein	lentivirus protein R protein U negative factor	<i>Vpr, Vpu, Nef</i>	Vpr, Vpu, Nef,	
Vif		<i>Vif</i>	Vpu	

Vpu functions in degrading the CD4. (Granoff & Webster, 1999).

Nef negative factor. It is an N-terminal myristoylated membrane-associated phosphoprotein. Involved in multiple functions during the viral replication.(Lata et al., 2018).

Vif required by HIV-1 to synthesise infected viruses in lymphocytes, macrophages and certain human cell lines (De Clercq, 2005).

2.1.3 Reverse transcriptase inhibitors

There are three forms of reverse transcriptase inhibitors (RTIs), namely, nucleoside analogue reverse-transcriptase inhibitors (NARTIs or NRTIs), Nucleotide analogue reverse-transcriptase inhibitors (NtARTIs or NtRTIs) and Non-nucleoside reverse-transcriptase translocator (NNRTIs) (1, 600).and tetrahydroimidazo [4,5] [1,4]benzodiazepine-2(1H-one – thione (TIBO) are the first two classes of compounds that were identified as NNRTIs. The unexpected findings of TIBO compounds paved way for the definition of this NNRTIs class and its mode of action, which has been useful in research since its discovery in the late 1980s, showing its great capability to inhibit reverse transcriptase (Basavapathruni & Anderson, 2007).

The HEPT were originally believed to be NRTIs. This is after they were described before the TIBO compounds. However, it was later discovered that these HEPT compounds share the common mechanism of action with the TIBO compounds. TIBO and HEPT compounds were first to be identified as highly specific and potent HIV-1 reverse transcriptase inhibitors (de Béthune, 2010).

The analogues of the naturally occurring deoxynucleotides compete with this deoxynucleotides that is needed to synthesise the viral DNA. However, the (NRTIs) and (NtRTIs) have the same antiviral effect; lack a 3'-hydroxyl group on the deoxyribose moiety (Prajapati *et al.*, 2009). This a vital key factor for these inhibitors as its incorporation with the deoxynucleotide cannot form the next 5'-3' phosphodiester bond needed to extend the DNA chain. The phosphodiester bonds occur when two of the hydroxyl groups in phosphoric acid react with the hydroxyl group on another molecules to form two ester bonds and because this process is halted, and it leads to another process known as chain termination. This makes all NRTIs and Nerites to be classified as competitive substrate inhibitors. These inhibitors however compete for not only viral processes but also for host DNA synthesis. It acts as a chain terminator for both having an antiviral effect and a drug toxicity or side effects (Hetherington *et al.*, 2001).

NNRTIs (Table 2.3) on the other hand, uses a different mode of action. They block the reverse transcriptase by binding directly to the enzyme. NNRTIs inhibit, rather than incorporate with the viral DNA, the movement of protein domains of reverse transcriptase needed to carry out DNA

synthesis (Mallery et al., 2018). This makes NNRTIs to be classified as non-competitive inhibitors of reverse transcriptase, for they reduce the activity of the enzyme and binds equally well to the enzyme. There are already about 31 anti-HIV drugs that have been developed with about six main types or classes, however there are challenges that involves the clinical management of HIV infection which includes the emergence of multidrug- resistant strains, a high requirement of patient's compliance, harsh adverse effects of the antiviral drug and significant expenses of the treatment. All these challenges emphasise a need to further pursue novel anti-HIV therapeutics (Fischl *et al.*, 1987).

2.1.3.1 NRTIs class of medication

Nucleoside reverse transcriptase inhibitors (Table 2.2) were the first kind of inhibitors to be approved. This kind of agents are used for the treatment of HIV infection by interfering with the transcription of double stranded viral RNA into DNA. It is referred to as the “backbone” of a first-line HIV treatment combination. It includes the following drugs (Havens *et al.*, 2013):

Table 2. 2: NRTIs class of medicine

Name of Inhibitor	Type of Action	Reference
Abacavir	It inhibits the reverse transcriptase enzyme competitively and is phosphorylated to active metabolites that compete for incorporation into DNA.	(Satake <i>et al.</i> , 1994)
Zidovudine	Zidovudine has an antiviral activity that causes a decline in serum HIV-1 antigen.	(Gandhi <i>et al.</i> , 2015)
Tenofovir	It also proved to be highly effective in patients that have never had an antiretroviral therapy.	(Manfredi & Calza, 2008)

Tenofovir alafenamide	Tenofovir diphosphate is incorporated into viral DNA instead of the natural substrate deoxyanosine 5-triphosphate resulting in DNA chain termination and impairment of HIV replication.	(Nelson & Schiavone, 2004)
Tenofovir disoproxil	The ECS50 (50% effective concentration) value of tenofovir against HIV-1 virus ranged between 0.04 μ M (combination of tenofovir disoproxil with other drugs).	(Giudici <i>et al.</i> , 2000)
Lamivudine	The pharmacokinetics and the antiretroviral activity of lamivudine in combination with zidovudine was studied in pregnant women infected with HIV-1 and their neonates. After administering these drugs from 38 weeks of pregnancy to one week after delivery, both treatment regimens were well tolerated in women and new-borns.	(Günthard <i>et al.</i> , 2016)
Emtricitabine	A nucleoside reverse transcriptase inhibitor, is phosphorylated by cellular enzymes to emtricitabine 5'-triphosphate which, in turn, inhibits the activity of HIV-1 (HIV) reverse transcriptase by competing with the endogenous substrate	(Cohen <i>et al.</i> , 2014).

2.1.3.2 NNRTIs class of medication includes

Table 2. 3: NNRTIs class of medicine

Name of Inhibitor	Type of Action	Reference
Doravirine	Inhibits HIV-1 replication by non-competitively inhibiting HIV-1 reverse transcriptase	(Das <i>et al.</i> , 2007).
Nevirapine	Slows down damage to the immune system and prevent the occurrence of AIDS-defining illnesses	(Stellbrink, 2007).
Rilpivirine	Decreasing the amount of HIV in the blood	(FDA.Administration, 2017).
Efavirenz	Approved specifically for treatment of HIV infections in patients who failed therapy with zidovudine.	(Moelling & Broecker, 2015).
Etravirine	High genetic barrier to the emergence of viral resistance	(Moelling and Broecker, 2015).

2.4 INTEGRASE INHIBITORS

HIV IN is a relatively new therapeutic target compared to inhibitors that are much more prominent in the market, such as HIV RT and HIV PR enzymes (Choi et al., 2018). Raltegravir (RAL) was the first IN inhibitor to be discovered. HIV IN has unique features because it has the least resemblance to any human protein which gives it a unique feature makes it an attractive anti-HIV target for reducing off-target side effects (Barrie *et al.*, 1996).

Viral replication is carried out through the process of integration of HIV genetic material with the host cell. This is essential in the life cycle of the virus (D'Urbano et al., 2018). After the virus has entered the host cell, its RNA is reverse transcribed into double-stranded cDNA by an enzyme called reverse transcriptase. Two consecutive steps account for the viral genomic integration inside the host cell: the first stage is the viral cDNA 3'-processing (3'-P) in the cytoplasm followed by the second stage, namely strand transfer (ST) in the host cell. Following reverse transcriptase, the viral cDNA is now primed for integration with the host DNA. The integration process is carried out by the viral integrase enzyme by forming a complex with the viral cDNA in cytoplasm along with cofactors to complete the integration process in the host nucleus (Ridky *et al.*, 1996).

In the first stage, two nucleotides (GT dinucleotides) are trimmed from a CAGT tetranucleotide sequence to expose CA-3'-OH at both 3'-ends of the viral cDNA (Choi et al., 2018). This process is facilitated by integration in the host cell. A pre-integration complex (PIC) is then formed by the trimmed viral cDNA that is bound by the IN, together with the viral and host cofactors. This complex is translocated to the nucleus to integrate with the host cell. This process of importation of the PIC into the host is very complex (76). The viral protein includes the matrix p17, capsid protein, the viral RT, the nucleocapsid protein p7/p9, IN and accessory protein Vpr (Yeang, 2015). It also comprises a number of cellular proteins such as IN interactor 1 (INI1), p300 histone acetyl transferase (HAT), Lens Epithelium-Derived Growth factor (LEDGF/p75), barrier-to-autointegration factor (BAF) and a high-mobility group protein A1 HMG-I(Y) among others (Judes et al., 2015). The exposed CA-3'-OH of the viral DNA will act as a nucleophile to attack the phosphodiester bonds on the complementary strands of the host DNA through a transesterification reaction. This takes place during the ST step (Pearl and Taylor, 1987).

The two divalent metal ions facilitate this reaction, which helps to stabilise the IN-DNA complex. ST takes place concomitantly for both extremities of viral DNA. The points that are not coupled in the integration intermediate are the 5'-end of the viral DNA and the 3'-end of the host DNA. These are at the insertion points (Liedtke et al., 2014). At each junction between the inserted viral DNA and the host DNA, a five-base, single stranded gap is left and at the 5'-ends of the viral DNA, an unpaired two- base fragment. Completion of integration requires cellular enzymes to remove

the unpaired nucleotides at the 5'-ends of the viral DNA and repair the single-strand gaps between viral and host DNA (Stellbrink, 2007).

Table 2.4: Integrase inhibitors

Name of Inhibitor	Type of Action	Reference
Raltegravir (RAL)	Although it became a successful anti-HIV drug in (2007), it was later rendered ineffective against the virus, this is because of the identification of multiple mutations in IN CCD	(Chouchana <i>et al.</i> , 2020) (Deeks, 2018)
Elvitegravir (EVG)	It is a potent, boosted, once-daily, HIV integrase inhibitor with antiviral activity against wild type and drug-resistant strains of HIV.	(Abbas and Herbein, 2014)
Bictegravir (BIC)	Exhibits potent and selective in vitro antiretroviral activity in both T-cell lines and primary human T lymphocytes	(Yi <i>et al.</i> , 1998)
Dolutegravir (DTG) (used as a control in this study)	It is administered once daily without a pharmacological enhancer and has relatively low cross-resistance with RAL	(Iordanskiy <i>et al.</i> , 2013)

Table 2. 5: A set of integrase inhibitors and their small molecular derivatives. Failed INSTIs is marked with (*). This, however, has become instrumental for the clarity and behaviour regarding the interactions between the HIV IN and its inhibitors that targets the IN active site for many other proposed HIV IN.

Name of inhibitor	Inhibitor activity	Developments of some approved inhibitors and (*) = Failed inhibitors
1. Raltegravir (RAL)	A potent and well-tolerated Anti-HIV drug	
2. Elvitegravir (EVG)	Antiviral activity against wild-type and drug-resistant strains of HIV	
3. Dolutegravir(DTG)	It is superior to the earlier generation INIs	
4. 1-(5-chloroindol-3-yl)-3-chloroindol-3-yl)-3-hydroxyl-3- 2H- tetrazol-5-yl Propanone (5-CITEP)	Provided the prototypical scaffold (β -diketo acid (Ghorban et al.)) for many INIs developed later. This inhibitor binds centrally in the active site between the two aspartates, D64 and D116 and the glutamate E152, and makes a close contact with the protein to compete with viral DNA for binding to IN (158).	
5. L- 731,988	Exhibit potent activity against HIV-1 IN <i>in vitro</i> The selectivity for ST (IC_{50} 80 nM for ST and IC_{50} = 6 μ M for 3'- P) (159)	(*) Identified by Merck 5 and 6 failed in suppressing virus production at the concentration of 50 μ M
6. 8-hydroxy-[1,6] - naphthyridine -7- carboxamides (L-870,810)	Exhibits potent inhibition against IN ST activity (IC_{50} = 8 nM) and robust antiviral activity with an EC_{95} value of 15 nM in a cell based assay. Has improved bioavailability (i.e. oral bioavailability and half-life)	However, the liver and kidney toxicity that appeared after long-term use in dogs halted the drug candidate for further clinical development.
7. (L-870,812)	Data not available	
8. (Z)-1-[5-(4-fluorobenzyl) furan-2-yl] -3-hydroxy-3-(1H-1,2,4-triazol-3-yl)propanone (S-1360)		(*) Failed to deliver efficacy in human

Name of inhibitor	Inhibitor activity	Developments of some approved inhibitors and (*) = Failed inhibitors
9. 4-hydroxy-2-oxo- 1, 2-dihydro-1,5-naphthyridine. (INSTI GSK364735)	Inhibits HIV replication with robust potency. Showed EC50 = 1.2 nM on peripheral blood mononuclear cell (PBMCs) and MT-4 cell EC50 = 5 nM.	Clinical development was terminated early due to hepatotoxicity in long-term safety studies in monkeys.
10. GS-9160	Data not available	Possessed poor pharmacokinetics profile and unfavourable bioavailability
11. MK-2048		Designation for a molecule in its pre- clinical drug discovery
12. Cabotegravir S/GSK 1265744		Undergoing phase clinical trial
13. Bictegravir (BIC)	As compared to RAL, EVG and DTG; BIC shows improved resistance.	In 2016, bictegravir was in a phase III trials as part of a single tablet regimens.
14. 1-hydroxy-2-oxo-1,2-dihydro-1,8- naphthyridine-3- carboxamide	Potent against variants that are resistant to HIV RT and PR	
15. 1,8-dihydroxy-2 – xo1,2-dihydroquinoline-3 carboxamide	Potent against variants that are resistant to HIV RT and PR.	
16. 4-[(5-bromo-4- {[2,4-dioxo-3-(2-oxo-2-phenylethyl)-1,3-thiazolidin -5-ylidene]methyl}-2-ethoxyphenoxy)methyl] benzoic acid (D77)		(*) Failed to inhibit LEDGF/p75-IN interactions in Alpha Screen assays

17. 1,4-bis(5-(naphthalene-1-yl)thiophen-2-yl)naphthalene	Interact with IN-LEDGF p75	
18. BI-1001	A series of 2-(quinolin-3-yl)acetic acid derivatives inhibit IN-LEDGF/p75 binding and HIV-1 replication in a cell culture.	
19. LZ-14	2,3-dihydroxybenzamide scaffold led to the identification of the selective IN-LEDGF/p75	
20. 3-hydroxypicolinamide	Forms several H-bonds and hydrophobic interactions with important allosteric residues such as Glu 170, Hist 171, Ala 128, Trp 132.	
21. CX14442	When compared with BI-1001, it has a larger tert-butyl group in lieu of the methyl group of the ester functionality, therefore demonstrates better interactions with the highly hydrophobic-binding pocket of the CCD dimer interface and resulted in strong affinity.	
22. N-aryl-naphthylamines (chemical library)	Many compounds were identified with its general structure and it inhibits LEDGF/p75 and IN interaction. Their activities (IC_{50}) against IN-LEDGF interaction ranged from 70.07 and 2.5 μ M	
23. 4-(4-hydroxy-1 benzoic acid	Has higher LEDGF/p75-IN inhibition with an IC_{50} value of 2.5 μ M	
24. BI 224436	An allosteric inhibitor that targets a noncatalytic site and disrupts chromatin and IN from interacting with LEDGF/p75. It also inhibits 3'-P step.	

25. LEDGIN1	At 100 μ M inhibited LEDGF/p75-IN interaction by 36%, whereas at 9 μ M BI-A Inhibited 50% of IN 3'-P activity.	
26. LEDGIN6 (CX0516)	Highly selective for disrupting IN-LEDGF binding (IC_{50} = 1.37 μ M)	
27. BI-A	At 9 μ M, it inhibited 50% of IN 3'-P activity. This leads to many other compounds that Inhibits HIV replication with EC_{50} value ranging from 10 to 100 nM.	
28. LEDGIN7 (CX05045)	Exhibits a micromolar IC_{50} value Has a tert-butylether motif at the α -position of a carboxylic acid and Thiophene ring. It has structural similarities with other LEDGINs.	
29. MUT-A	Has a tert-butylether motif at the α -position of a carboxylic acid and Thiophene ring. It has structural similarities with other LEDGINs.	
30. DW-D-5	Shows promising antiviral properties with EC_{50} value of 2.0 μ M in the inhibition of HIV-1 replication. It also inhibited the HIV-IN ST activity with an IC_{50} value of 0.85 μ M when combined with the FDA approved HIV drugs. Co-treatment of 1 μ M with 5 nM RAL exhibited 51.59%. RAL alone had an inhibition of 35.34%.	
31. KF115		When BI-1001 (IC_{50}) as a structural prototype KF115 was developed by altering the quinolone moiety into biaryl moiety that contains pyridine ring.

32. KF116	Has an EC ₅₀ value of 0.024 µM	Used in resistance of Pyridine-based inhibitor.
33. 8-hydroxyquinoline	It inhibit the required protein-protein interaction. Their potency is sevenfold greater than that of compounds when they are structurally modified.	Molecular modelling studies revealed that this compound is bound to IN at IN-LEDGF/p75 interface.
34. 2-(3,4-dihydroxyphenyl)-7-((4- fluorobenzyl)oxy)-3,5-dihydroxy-4H-chromen-4-one (a) And 2-(3,4-dihydroxyphenyl)-3, 5-dihydroxy-7--4H-chromen-4-one (b)	Has the highest inhibitory activity with IC ₅₀ value of 4 µM against ST activity within this class. Substituting with hydrophilic morpholino group at C7 position improved activity against LEDGF/p75. Interaction with IC ₅₀ values that range from 0.97 to 1.7 µM.	
35. Lovastatin	IC ₅₀ = 1.97 µM against the protein-protein Interactions	
36. 2-hydroxyisoquinoline-1,3(2H, 4H)-diones (HIDs)	Has high genetic barrier and favourable cross-resistance profile with RAL. It works by inhibit both ST and 3'-P activities to analogue synthesized and found with IC ₅₀ that ranged from 10 to 260 nM. Exhibits against RNase H and IN.	Because of its high cytotoxicity it was deemed that further optimisation of the scaffold is necessary before translation as a Clinical candidate.
37. (4-[1-[(4-fluorophenyl)methyl]-7-[N-(3-chloroprop-1-yl) piperazin-1-yl]-4-(1H)-quinolin-3-yl]-2-hydroxy-4-oxo-2-buten-1-yl)butanoic acid	Inhibited the ST step with an IC ₅₀ value below 50 nM and displayed RNase H activity with IC ₅₀ of 5.7 µM. Their ST activities were in low micromolar to nanomolar range.	

38. (4-[1-(2,4-difluorophenyl) methyl-4(1H)-quinolinon-3-yl] 2-hydroxy-4-oxo-2-butenic acid)	Highest ST activity $IC_{50} = 10$ nM and RNase H activity $IC_{50} = 36\mu M$ Analogues reported as IN and RT RNase H inhibitor.	
39. 6-(1-benzyl-1H-pyrrol-2-yl) -2,4-dioxo-5-hexenoic acid	Most potent INI with IC_{50} value of 26 nM and displayed RNase H activity with IC_{50} value of $2.5 \mu M$	
40. (ethyl 6-(4-(4-fluoroben -zoyl) -1-phenyl-1H-pyrrol-2-yl) -2-hydroxy-4 -oxohexa-2,5-dienoate)	The most active compound against IN and RNase H. IN $IC_{50} = 1.8 \mu M$; RNase H $IC_{50} = 1.2 \mu M$	
41. CHI-1043	Developed from altering the quinolone.	
42. CHIBA-3002	Was found to inhibit HIV-1 IN LEDGF/p75 interaction at micromolar concentration.	
43. CHIBA-3003	Was found to inhibit HIV-1 IN LEDGF/p75 interaction at micromolar concentration.	
454. CHIBA-3053	Was found to inhibit HIV-1 IN LEDGF/p75 interaction at micromolar concentration.	
45. LEDGF/p75	The most potent statin with an IC_{50} of 1.97 To $1.7 \mu M$.	
46. (2S)-2-(6-chloro-2-methyl-4-phenylquinolin-3-yl) pentanoic acid (CX04328)	Have high genetic barrier and favourable cross-resistance profiles with RAL.	

47. GS-A	Showed potent antiviral activities with EC ₅₀ values ranging from 10 to 287 Nm.	
48. GS-B	Showed potent antiviral activities with EC ₅₀ values ranging from 10 to 287 Nm.	
49. GS-C	Showed potent antiviral activities with EC ₅₀ values ranging from 10 to 287 Nm.	
50. GSK1264	Inhibited both LEDGINs and 3'-P promoting IN polymerization.	

2.5 HIV-1 ANTIVIRAL THERAPEUTICS

2.5.1 CCR3 blockers

CCR3 is a chemokine receptor that is mainly involved in the entry of M-tropic HIV-1 strain into the microglia of the central nervous system. It is also known that this antigen is present on the surface of lymphocytes and monocytes in semen (Carter and Ehrlich, 2008).

2.5.2 CXCR4 antagonist

To enter a host cell or a target cell, HIV-1 requires CD4+ and co-receptors. CXCR4 is therefore a co-receptor that is involved in the entrance of T-cell tropic (T-tropic) HIV-1 isolates (Jiang *et al.*, 2019), (Peterle *et al.*, 2018; Sancineto, 2015).

2.5.3 CCR5 or CD195 antagonist

On the other hand, CCR5 proteins are receptors that are involved in the entrance of macrophage-tropic (M-tropic) HIV-1 isolates. The CCR5 antagonists are inhibiting a protein called CCR5/CCR5/CD195. It is a transmembrane chemokine receptor involved in envelope binding of HIV. Drugs like maraviroc targets CCR5 and inhibit this chemokine receptor causing a minimal cross-resistance with other medication. This poses as an advantage (Ilinskaya & Mahmud, 2014).

2.5.4 NCp7 inhibitors

Nitrosobenzamide (NOBA) was the first inhibitor of NCp7 to be reported during the late 1990s. These compounds are also referred to as covalent NCp7 or irreversible inhibitors and/or zinc ejectors (Yarchoan et al., 1994). They were specifically designed and developed to recognise NCp7 among cellular proteins containing ZFs. The development of Nitrosobenzamide paved a way for the development of other small molecules that have the properties to inhibit NCp7: these included azadicarbonamide (ADA), dithiane, thiocarbamates (TICAs), S-acyl 2-mercatobenzamides (SAMTs), dithiobis(benzamides)(DIBAs), benzisothiazolone (BITAs), diselenobis (benzamides) (DSeBAs), transition metal complexes, pyridinioalkanoyl thioesters (PATEs and thioether prodrugs (Piot *et al.*, 2015).

2.5.5 Ribonucleases as antiviral therapeutic

Looking into the properties of RNases, they share a conserved set of disulfide bonds that confer a great stability. They possess cytotoxic ribonucleolytic activity and they do not require delivery agents. They also possess the endogenous ability to gain entry to the cytosol. These features make them an attractive protein to serve as the basis of chemotherapeutics (Maartens *et al.*, 2014).

2.6 RECENT DEVELOPMENT OF HIV-1 PROTEASE INHIBITORS

The development of protease targets was essential in therapeutic inhibition since this enzyme plays a critical role in processing the gag and gag-pol gene products (Shamsi et al., 2016). It is the gag and the gag-pol gene products that must be processed into essential viral proteins required for the assembly of a new mature virus. After the discovery of such vital information about the behaviour of the protease and the gag and gag-pol gene association, efforts to develop an HIV-protease inhibitor began. In the mid-1990s, a combination of HIV-1 protease inhibitor drugs and those with reverse transcriptase marked the beginning of a highly active antiretroviral therapy (Ifeanyi et al., 2017). This combination was approved because it became evident that combination chemotherapy was significantly more effective than dosing the drug sequentially. Structural proteins are produced by the HIV protease that catalyses the hydrolysis of gag and gag-pol polyproteins at various

cleavage sites. These structural proteins include the viral envelope glycoprotein, the RT, IN and PR enzymes for the new virion particles (Kramer-Hämmerle *et al.*, 2005).

From this advent of HAART, many other different treatment regimens have become known and new therapies with other targets have been developed. These regimens aim to be convenient, potent and well tolerable and typically reduce HIV blood concentration to undetectable levels within a few weeks of treatment. ART regimens induce a robust and sustainable increase of CD4 T-cell counts (Huang *et al.*, 2010; Hung-JinHuanga, 2010).

There are, however, significant setbacks to current treatment, despite massive advancements in HIV/AIDS therapies, namely drug toxicity, the evolution of different patterns of systematic complications, drug-drug interactions and a lifelong drug administration with unknown long-term side effects (Holmberg *et al.*, 2004). There is a vast increase in neurocognitive disorders, possibly due to the poor central nervous system (CNS) penetration of current anti-HIV therapies since it is a major sanctuary of HIV-1 infection. One of the major problems encountered, was the emergence of drug resistance. This drug resistance challenge renders the current therapies ineffective within months in some cases (Cavasotto & Phatak, 2009; Westbrook *et al.*, 2003).

Despite all challenges emerging from drug resistance, there still stand a need to appreciate the immense progress that has been made by scientists in understanding the mechanisms and the significant properties that takes place on the target cell. Table 2.5 gives a summative overview of this association.

Table 2. 6: Summary of HIV and host targets

Target	Location	Mechanism	Significant properties
TPST2	Host	Entry	Highlights sulfation
SLC35B2	Host	Entry	Highlights sulfation
Gag, capsid	Virus	Entry	Cleaves to two structural protein, matrix & capsid
Protein disulfide isomerase	Host	Entry	Catalysed reduction of disulfides in gp120
Galectin-9	Host	Entry	Traps PDI on the surface of the T-cells making them more susceptible to infections.
CD4-mimic	N/A	Entry	Targets the Phe43.
Nef	Virus	Entry	Key role on the chemocytokine network
c-SRC protein Tyrosine kinase	Host	Early stages of virus entry	Promote viral infection
Natural seminal amyloids	Host	Entry inhibitor	Increase infectivity by > 400 000
BRAF-histone deacetylase complex	Host	Transcription	Promotes Tat-mediated LTR activation
Protein phosphatase Host phosphatase 1	Host	Transcription	Sensitive to the reduction of intracellular iron
LEDGF/p75	Host	Transcription	Involved in HIV integration

Tat	Virus	Transcription	Required for robust transcription of the integrated viral genome
mTOR	Host	Transcription	Controls HIV latency
SIRT-AMPK pathway	Host	Transcription	Substrate for the deacetylase activity of sirtuin
RNAse H	Virus	Transcription	Responsible for initiating the second strand of viral DNA synthesis
Ku70	Host	Transcription	Required for early and late stages of viral lifecycle
Cytidine deaminases	Host	Transcription	Important in viral replication
Y-box-binding protein 1	Host	Transcription	Supports HIV replication
B23/ Nucleophosmin	Host	Transcription	Play a role in nuclear import
ALCAM	Host	Spread of HIV	Highlights sulfation
JAK pathway	Host	Cell-cell viral transmission	Involved in HIV persistence
Mucosa-associated lymphoid tissue translocation 1	Host	Suppression of reactivation	Reactivates and induces apoptosis of infected cells
Toll-like receptor	Host	Replication	Recognition receptors
Dendritic cell Immunoreceptor	Host	Infection/	Involved in infections via transmission to CD4 ⁺

2.7 CADD STUDIES

The use of computer aided drug design is an essential tool to finding new inhibitors for emergence of drug-resistant mutations. CADD has been increasingly used to discover and design potential drugs. CADD comprises protein structures, computer technology and mathematical modelling that is used to predict the energies of small-molecules. The binding patterns can also be predicted through these computer techniques. Figure 2.1 below shows the general CADD flowchart used in finding potential drug candidate.

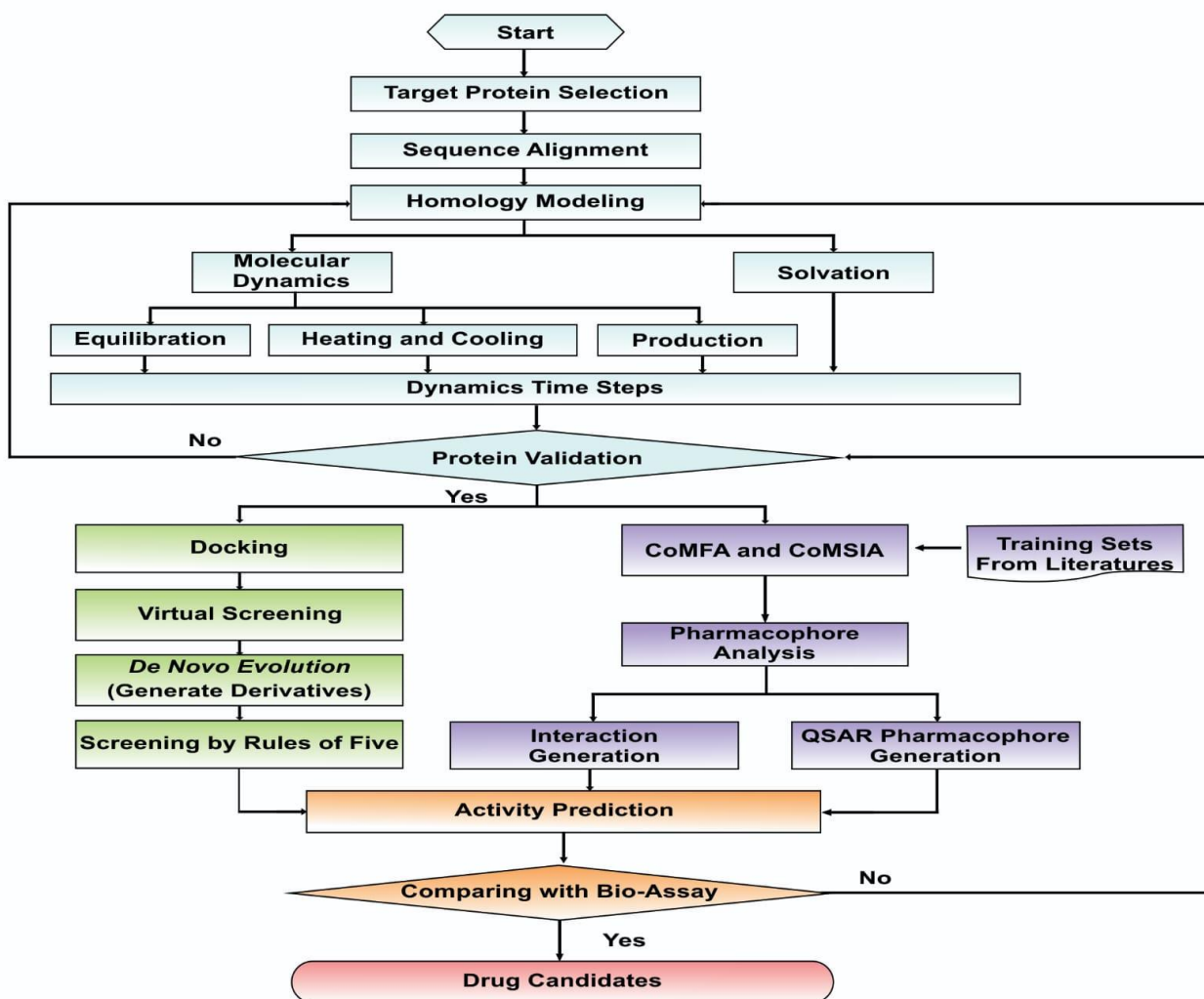


Figure 2. 1: An example of a computer-aided drug design (CADD) flowchart (Image by Hung-Jin Huang *et al.*) (304)

To understand the structural requirements of HIV inhibitors, a continuous study is needed to be carried out. Numerous mechanisms have evolved which include a computer aided drug design (CADD) that provides a clearer interaction between the ligand and the acceptors (Hassan Baig et al., 2016). The pharmacophores and the 3D- QSAR contour maps provide rich information for rotational designs. There is also a growing availability on the co-crystal structures and mechanism providing a promising lead to the discovery of more protease inhibitors. This new lead will also give way to a new proposed binding site (Das and Chakraborty, 2021).

2.7.1 Homology modelling

Homology modelling is considered the most accurate of the computational structure prediction methods. It is one of the computational structure prediction methods that is used to determine a protein 3D structure from its amino acid sequence (Studer et al., 2021). This straightforward, easy to apply modelling consists of multiple steps. The functionality of the model depends on the quality of the generated protein 3D structure and maximising the quality of homology modelling is crucial. With the worldwide protein data bank containing more than 144, 000 experimentally determined protein three-dimensional (3D) structures, there has been a great achievement with the introduction of experimental methods that include X-ray crystallography and NMR spectroscopy. When using the X-ray crystallography for analysis, the molecules should be crystallised and for NMR, proteins molecules should be small (Schauperl & Denny, 2022).

Therefore, this model gives us the platform to fill in the gap between protein sequences available and protein 3D structures determined experimentally. Homology modelling aims at building 3D structure of proteins from their sequences by using templates with an accuracy which is similar to the experimental methods, since there are great advances in the integration between homology modelling with other computer-aided drug design methods and post-modelling applications (Muhammed & Aki-Yalcin, 2019).

In one of such studies on homology modelling with other CADD, Sharma *et al.* worked on the homology model-guided 3D-QSAR of HIV-1 Integrase Inhibitors (Sharma et al., 2012). They reported the exploration of binding modes of potent HIV-1 integrase inhibitors, namely MK-0518

(raltegravir) and GS-9137 (elvitegravir). Their studies also included chalcone and related amide IN inhibitors. These are synthesised and developed 3D-QSAR modes for integrase inhibition. Homology models of DNA-bound HIV-1 IN were constructed. This is based on the X-ray crystal structure of the foamy virus IN-DNA complex (PDB ID: 3L2T). This structure was used in docking. Their binding modes of elvitegravir and raltegravir in homology models were in accordance with those in the virus structure that displayed foam and this revealed the interactions, which are important for inhibitor-IN binding. The following inhibitors were further studied by Sharma *et al.*: elvitegravir, raltegravir and their respective analogues. Other HIV-1 IN inhibitors that are related to this study were also investigated, for example the synthesised 3-keto salicylic acid integrase inhibitor series. These studies were conducted to gain more insight into these structural requirements (Sharma *et al.*, 2012; Shah *et al.*, 2014).

Firstly, 103 compounds, docking-based alignment III and the atom fit alignments I and II were studied and the data that was obtained assisted to develop a 3D-QSAR. A model 1 of 3D-QSAR for atom fit alignment I was developed, a model 2 for atom fit alignment II and lastly, a model 3 for docking-based alignment III. Each of these indexes comprised a comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA) 3D-QSARs. (Maignan *et al.*, 1998).

In their observation, model 3 related to the docking-based alignment performed better than model 1 and 2. This was analysed concerning the statistical significance and their robustness. Model 3 represented the structure-based while model 1 and 2 represented the ligand-based models (Krishnan *et al.*, 2010).

Model 3, which is the docking-based, used an expanded data set. This data set was the one of 296 compounds that built a more comprehensive 3D-QSAR. On the other hand, model 4 displayed a good q^2 value of 0.75 for CoMSIA 3D-QSARs and 0.70 for CoMFA (Narkhede & Degani, 2007). On an external validation test set, the one of 59 compounds displayed an acceptable and good predictive performance and its predictive r^2 values up to 0.71. Relevant insights were provided in these studies, which included the discovery of new predictive tools for structure-based design. This

was achieved by successfully investigating the HIV-integrase-DNA homology model together with the comprehensive 3D-QSAR model. This information contributed to further optimisation of integrase inhibitors (Fikkert *et al.*, 2004; Shimura *et al.*, 2008).

2.7.2 Pharmacophore visual screening

Pharmacophore modelling is another method that is frequently used in drug discovery (Bhadoriya *et al.*, 2015). In this method, the steric and electronic feature a query molecule possesses. This is essential for the receptor-ligand interaction. Vadivelan and colleagues worked on the potential anti-HIV lead compounds that could be generated by analogue based studies using 3D-QSAR and pharmacophore models. They used a training set of 36 molecules to develop the best model. The MFA model was generated based on the feature query of the biologically active conformation of the most active compound (pIC₅₀ of 8.57M) (Vadivelan *et al.*, 2011). They also composed a best pharmacophore model with three characteristics: one hydrogen bond acceptor, one hydrophobic aliphatic and one aromatic ring. A knowledge-based screening was used to evaluate their predictive ability and validate both models. A total of 10,000 molecules were generated based on the knowledge of the binding interaction of ligands to the RT. Also, an investigation of the common features necessary for the molecule biological activity and a cross validation was done with both models. The results suggested that these techniques yield almost the same results. Spallarossa *et al.* performed docking studies on RT structures using Glide, in an attempt to produce reliable true positives and negatives (Cichero *et al.*, 2009). This was after the screening had produced some false-positives and a few false-negatives. The combined approach developed in their study showed a possible way to critically assess the identification and optimisation of lead compounds through a better understanding of protein and ligand features (Chen *et al.*, 2000).

2.7.3 Molecular docking studies

Molecular docking is a method that analyses the conformation of molecules into the binding site of a macromolecular target. This method is employed as a fast and inexpensive technique in academic and industrial platforms. Several software systems have been developed over the past

years, among which are software systems such as FlexX, Molegro Virtual docker, GOLD, AutoDock Vina, DockThor and AutoDock (Qayum et al., 2019)

When using these software systems, the first step in a docking calculation is to obtain the target structure. This structure normally consists of a large biological molecule which is either the protein RNA or DNA which can be readily obtained from the protein data bank Berman et al., 2000; Berman et al., 2000). In cases where the experimental 3D structure of the target is unavailable, computational prediction methods such as comparative and ab initio modelling can be used to obtain this 3D structure of the protein (Hetényi & Van der Spoel, 2006).

The binding site is usually known, however, when it is missing, two approaches are employed, namely to algorithmically predict the binding sites or what is called a “blind docking” simulation is carried out (Iorga et al., 2006). This is done when the binding site location on which to focus the docking calculations is unknown. One of these software systems used to detect the binding sites is MolDock, which uses an integrated cavity detection algorithm to identify potential binding sites. Other software systems, such as DoGSiteScorer functions as an algorithm that determines possible pockets and their drug ability scores, describing the potential of the binding site to interact with a small drug-like molecule. Another software system that uses small molecular probes to identify surface regions in the receptor that are prone to interact with small molecules is the Fragment Hotspot Maps, which also functions as generating distant restraints known to greatly increase protein-small molecular docking accuracy. The predicted interaction sites can be provided as the centre of the sampling space (Volkamer *et al.*, 2012).

One of the main fundamental strategies in docking calculations is to employ a grid representation that includes pre-calculating potential energies for interaction within the binding site and speeding up the docking runs. It consists of the discretization of the binding site, and it is at each grid point, where the interactions of the Lennard-Jones and electrostatic potentials are calculated (Brooijmans & Kuntz, 2003) (324).

In other to obtain ligand structures, small molecule database PubChem and ZINC were used to obtain plant phytochemicals and for screening respectively. These two methods are vital in

retrieving many compounds for virtual screening. However, if structures are not readily available, 3D atomic coordinates of these compounds from the 2D structures can be used, by means of several software systems such as ChemSketch, ChemDraw, Avogadro and or Concord. For small molecule ligands a stereochemically defined geometry with relevant protonation state is needed, since conformations will be explored by the docking software in the context of the target's binding site (Liu *et al.*, 2018) (325).

De Clercq *et al.* (2009) reported that the most explored RTIs are the NNRTIs, with a large number of chemical and structurally diverse compounds identified as genuine inhibitors that suppress HIV-1 replication, which was based on the investigation by Sluis-Cremer *et al.* (2004), who had previously reported that the NNRTI binding pockets consist of hydrophobic residues with significant aromatic character (Y181, Y188, F227, W229, Y232 and Y318) and hydrophilic residues (K101, K103, S105, D192, E224 and H235 and E138) (De Clercq, 2009), (Sluis-Cremer *et al.*, 2004).

For a successful docking, it is critical to evaluate free torsions, protonation and charge assignments. Hsiou *et al.* (1996) reported that in the open state of the binding pocket, when the structure is co-crystallised with NNRTIs, the solvent accessible entrance is formed by L100, K101, K103, V179, Y181 and E138. This is due to the significant torsional shifts of the Y181 and Y188 residues that accommodate the ligand. The binding pocket is blocked in the absence of the ligand, since the side chains of Y181 and Y188 are situated at the hydrophobic core, representing a closed state of the pocket. This inherent flexibility of the binding pocket provides a challenge to molecular docking (Hsiou *et al.*, 1998).

There are also different software systems that can evaluate the pKa of the amino acids residue, such as PropKs and H++. This is important since the protonation state of the target's amino acid residue can be critical to the ligand interactions and consequently to the binding affinity predictions. Docking program searches the algorithm that analyses and generates ligand pose at a target's binding site, considering the roto-translational (Carbone *et al.*, 2022). Charges are usually assigned through algorithms that distribute the net charge of a molecule among its constituents'

atoms as partial atom-centered charges. One important aspect is the degree of freedom of the ligand. Ligand protonation affects the net charge of the molecule and the partial charge of individual atoms (Trott & Olson, 2010). Each docking program employs a different charge assignment protocol where the AutoDock used the Gasteiger-Marsili atomic charges whereas the AutoDock Vina does not require the assignment of the atomic charge, since the terms that compose its scoring function are charge-independent. The MolDock program automatically prepares the charge and the protonation state assigned and the DockThor automatically generates the topology files, which are the atom type and partial charge for the protein, ligand and cofactors according to the MMFF94S force field (de Magalhães *et al.*, 2014).

There has also been several algorithms developed to use information from protein and ligand pharmacophores. An extended software FLEXX called FLEXX-PHARM uses algorithms that try to match the distance between each of the ligands and protein pharmacophoric points. It also applies pharmacophoric features as constraints into a docking calculation (Sarkar and Sen, 2020).

Docking programs also have an important aspect of the algorithm, with higher computational cost, which is the deterministic search (Torres *et al.*, 2019). The orientation and the conformation of the ligand in each iteration is determined by the previous state and the new state has equal or lower value than the previous one. The cost is solely based on the computational cost of each individual docking run and in terms of accuracy. This is in cases where the overall size of the ligand contains many rotatable bonds, and this impacts most docking algorithms in a negative way. Each rotatable bond inherently increases the ligand's degree of freedom, and this increases the number of conformation (Wang *et al.*, 2019).

In some algorithms, different search strategies can be combined, and MD simulations used to analyse the time-resolved trajectory of the ligand-bound system. This system can further identify or pinpoint the best docking solution. There has been a generation of thousands of ligands orientation and in such kind of development additional scoring functions may be used to rank their conformation. This is based on the binding energy, qualitative numerical measure to approximate interaction energies and free energy (de Magalhães *et al.*, 2014).

The scoring functions can be grouped into three major types namely: force field, empirical and knowledge based. The force field-based functions consist of the sum of the energy terms, where the potential energy will account for the bonded and non-bonded terms. The bonded terms are the bond length, angle and dihedrals, whereas the non-bonded are the Van der Waals and electrostatic terms. DrugScore of FlexX and PMF are knowledge-based functions that are based on frequency of atom pairs and their interactions can be observed experimentally and determined through the 3D structures of ligands-target complexes (Sarkar and Sen, 2020)..

Elokey and Doerksen recommended Glide as one of the programs that uses soft docking receptors by scaling the Van der Waals radii (Elokely & Doerksen, 2013). Bahare *et al.* used two different programs to evaluate the accuracy of their docking procedure that consists of the docking of the NNRTI TNK 651 that was extracted from an X-ray crystallographic RT structure (PDB code: 1RT2). These programs were Glide and FlexX. Glide employed a hybrid approach that combined one or more docking algorithms in the generation of the ligand poses. This work was reported by Moitessier *et al.* On the other hand, Moitessier reported that the FlexX program was based on incremental construction, where the ligand was built dynamically in the active site and frequently counting on libraries of favoured conformations. Their findings were: the RMSD value between the docking prediction and the experimental conformation of TNK 651 were 0.370 Å and 1.254 Å, with Glide and FlexX, respectively (Sarkar and Sen, 2020).

The recent work by Fraczek and colleagues suggested that different docking methods could provide good binding mode predictions. However, results should not rely only on the docking scores when trying to rank active compounds with potencies. This is for 2RKI, all programs showed good predictions of ligand orientation in the binding site when compared to the reference pose. Their outcomes were as follows: Glide SP (97.3%), IFD (97.3%), AutoDock Vina (94.6%), FlexX (65.8%), Molegro Virtual Docker (61.3%) and Glide XP (59.5%). And the predictions for 3DLG were, IFD (88.3%), Glide SP (82.9%), AutoDock Vina (75.7%), Molegro Virtual Docker (71.2%) and Glide XP (67.6%), FlexX (36%) (Fraczek *et al.*, 2013).

These results were obtained after they had assessed the ability of docking programs and their scoring functions to predict the relative biological activity of triazole NNRTIs (Ren *et al.*, 2008).

Sherman *et al.* worked on a total of 110 types of NNRTI that were submitted to different docking protocols and that included the 1,2,4-triazole. These protocols were done using the Molegro Virtual Docker, the FlexX and the Glide XP and SP for the softened van der Waals potential, with 76 otherazole type submitted. To search for ligand flexibility the AutoDock Vina was used and for the receptor flexibility, the Induced Fit Docking was employed. Their IFD methods combined an iterative procedure to obtain initial poses allowing flexibility into rigid receptors (Sherman *et al.*, 2006). This was followed by a technique for modelling receptor conformational changes, present in the refinement module of prime program that explores flexibility as highlighted in the investigation by Jacobson *et al.* This method allowed efficient small backbone movement. This however would be inappropriate to more severe conformational changes due to an increase in complexity and computational cost (Jacobson *et al.*, 2004).

The RT structures that were 2RKI in complex used together with a triazole and 3DLG, both in complex with a benzophenone, were prepared and used in the docking procedures. They assumed the triazole binding mode, as the reference pose to be (36%) with a Molegro Virtual Docker of (71%). From these results, they learned that none of the scoring functions reached a perfect ranking of the compounds. This ranking is based in accordance with their activities. Finally, in their work, Glide XP achieved the highest correlation, namely Spearman's ρ of 0.7, which corresponds to around 0.75 probability of identifying the most active compound from two compounds (Kirschberg *et al.*, 2008).

2.7.4 Molecular Dynamics Simulation

RMSD analysis

In a docking study of HIV-1 reverse transcriptase with phytochemicals by Abhik S. *et al.*, natural products were used as important sources of drug discovery (Seal *et al.*, 2011). These natural products contained natural compounds such as Geranin, Gallotannin, Tiliroside, Curcumin, Kaempferol-3-o-glucoside and Trachelogenin, which were also found to be very effective according to their binding and ligand efficiency score. Abhik retrieved a three-dimensional structure of target HIV reverse transcriptase (PDBID: 1REV) from a protein data bank at 2.6 Å RMSD resolution. RMSD has often been used to measure the quality of reproduction of a known (i.e. crystallographic) binding posed by a computational method, such as docking. RMSD values representation mainly for analysing stability of protein and predicting conformational changes of protein. Phytochemicals with anti-HIV activity were also obtained from the Dr. Duke database that can search against PubChem and the ChempSpider database for the 2D structure. With the aid of the 2D structure collected from Dr. Duke, an open babel was used to convert the 2D into 3D structures and minimised using Hyperchem's MM+ force field. They used the Molegro Virtual Docker to detect the active sites and then continued to perform docking by the Moldock function. Moldock is an implementation of evolutionary algorithms (EAs) focussed on the molecular docking simulations and docking was performed with all the potential active sites that were on HIV reverse transcriptase. Molecules were prepared and bonds, charges, bond order, explicit hydrogens and flexible torsions were assigned. To reduce the overall computing time, they used the ignored distant atoms options far away from the binding site (Seal *et al.*, 2011). RMSD value is used to validate the docking protocol.

This means that a crystallographic complex protein with ligand in it should be considered and the docking of the same complex performed. To do this they referred and checked the RMSD values. Their docking was validated since the docking protocol was able to produce similar docking poses of a ligand with respect to the biological configuration of the same ligand in the similar crystal structure of complex protein. The lower the value of RMSD the higher the accuracy of docking.

Literature also affirms to that fact RMSD values less than 1.5 Å are always worth considering (Vyas *et al.*, 2008). Abhik and colleagues did a pose clustering by the tabu based clustering method. By using this clustering technique, each solution is added to a 'tabu list'. During the docking simulation, the poses are compared to the ligands in this 'tabu list'. They did this to observe if the pose being docked was closer to one of the ligands in the list than specified by the RMSD threshold. The energy penalty was added to the scoring function. They set the energy penalty to 100, the RMSD threshold was 2.0 and the RMSD calculation by atom ID (fast) were set. Poses were later sorted by rerank score, after the docking simulation was over. The reranking score function is computationally more expensive than the scoring functions used at determining the best pose. This comparison is made among several poses originating from the same ligand. The ligand efficiency is mostly defined as the ratio of the free energy of binding over the number of heavy atoms in a molecule (Ghorban *et al.*, 2004).

They calculated the binding affinities using Molegro data modelling and MVD was used to find Ligand Efficiency 1 (LE 1). They further derived the coefficients for binding affinity using multiple linear regression. This is one model that is usually used to collaborate using a data set of more than 200 structurally diverse complexes from the PDBbind database, with known binding affinities (expressed in kJ/mol), which resulted in the Pearson correlation coefficient of 0.60 when doing 10-fold cross validation. However, it is worth noting that this particular model was trained and employed only on strongly interacting ligands in their optimal conformation that is known from the PDB complexes (Ghorban *et al.*, 2004). Since their binding affinity measure was trained using known binding modes only, it might sometimes assign too strong binding affinities to weakly or non-binding molecules and therefore they recommended ranking the results of a virtual screening run using the rerank score. Using the MolDock software, they further improved these scoring functions with a new hydrogen bonding term and a new charge scheme. Based on their docking criterion they further concluded that MolDock showed better overall performance in docking simulation when compared with other software (Razzaghi-Asl *et al.*, 2012).

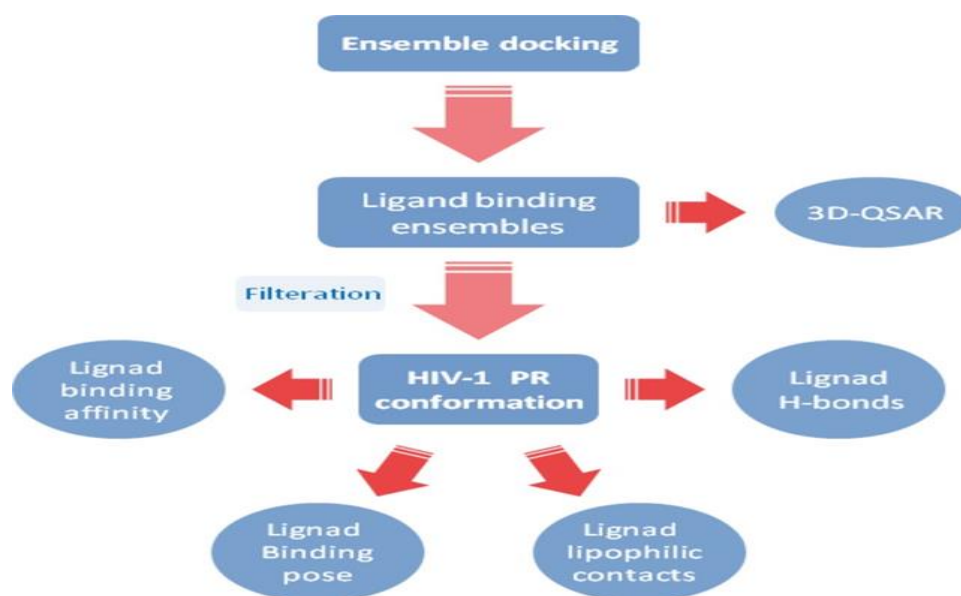


Figure 2. 2: A typical flowchart representing an ensemble docking protocol for HIV-1 PR inhibitors

Using the same docking protocol done by Abhik as shown above, Razzaghi-Asl *et al.* (Nima Razzaghi-Asla, 2015) worked on the biomolecular conformation on docking simulation performing molecular docking via advanced AutoDock4.2 software. Where after, enabling the docking of Amprenavir into the active site of various conformations of HIV-1 PR, it predicted different interaction modes/energies with all required holo/apo PDB structures retrieved from the Brookhaven protein databank and a performance of a docking method was checked via its ability in reproducing a binding mode for co-crystallographic (cognate) ligand. The RMSD of the backbone carbon atoms ($C\alpha$) in the selected PDB structure ranged from 0.22-0.85 and 0.24-0.93 Å in chains A and B of HIV-1 PR. They were interested in finding the factors that might be determinant in inducing conformation of HIV-1 PR/Amprenavir complex (PDB ID: 1HPV) where a binding system comprising of the three major constituents namely the ligand, enzyme and their interaction were taken into consideration and the RMSD (Å) of the backbone carbon atoms ($C\alpha$) in the chain A of individual HIV-1 PR structure with regard to the HIV PR apo structure (PDB ID: 3IXO) were reported the RMSD from reference structure (kcal/mol) ranged from 0.38-0.87 (Seal *et al.*, 2011) (345).

Abhik on the other hand reported that, the binding pose of each ligand at their highly bound cavity, was studied through selected photochemicals with potential Anti-HIV compound 4,5,6,7-tetrahydroimidazo[4,5,1-*jk*][1,4]benzodiazepin-2(1H)-thione (TIBO) and the re-rank score and the Moldock score were found to be good compared to (TIBO). In the same study, by using a similar structural analogue of the ligands in all the cavities and the phytochemicals, a higher binding score than a structural analogue was shown (DeGoey *et al.*, 2011).

RMSF analysis

The lower the RMSD values reflects Minimum Å and the lower RMSD is acceptable. Therefore, the ligand binding poses, energy and interaction fully depends on residual fluctuation (RMSF). After protein ligand docking, the docked conformers undergo identification of RMSF analysis. The RMSF plot will depict the ratio of fluctuation in residue level (Li *et al.*, 2008). Hence, both dynamic plots are important to carry out the prediction of structural stability on protein. In one of the studies that showed the fluctuations in both the chains of Ab-bound protease, an investigation conducted by Padariya and colleagues was done, which was the inhibition of the activity of HIV-1 protease through antibody binding and mutations probed by molecular dynamics simulations (Padariya & Kalathiya, 2017). Based on the computed values of the WT-free chains, RIT-bound, Ab-bound and mutated protease (G40E and G40R), they observed that the overall fluctuations in both chains of Ab-bound protease were quenched including the functionally important regions such as, Elbow (with residue 36-46), catalytic site (residues 25-27), cantilever (residue 59-75), dimer interface (residues 4 to 10 and 90 to 99) and the flaps (residue 43-58). They also observed that in mutants and RIT-bound protease, the overall fluctuation is reduced, except the elbow region where change in fluctuation is not that significant. Padariya also observed in their molecular dynamics' simulation study of COM5 ligand bound HIV-1 protease, that chain A of protease is more dominant in terms of hydrogen-bonding interactions than chain B. This is also reflected in their RMSF plot. In their results, F11.2.32 binds to the elbow region of chain A of HIV-1 protease (Tarasova *et al.*, 2015).

A fluctuation was observed in the elbow region of one of the chains. This was a region of chain A of Ab-bound protease. However, when compared with other mutants, G40 and G40E these fluctuations of the protease were reduced drastically. When they looked at all these other cases, it is only in the active site where the fluctuations were reduced. This did not include the chain in B of G40E. Many other factors were observed, but the one that caught their attention was the change in salt-bridges network. This included the hydrogen bonding that was calculated along with the RMSF values. These network changes occurred in the regions of Ab/RIT bound and mutant proteases. An increased hydrogen-bonding interaction displayed what they named, rigidification. What led to this was the correlation they observed when RMSF decreases, as this rigidity occurred between the mutants and the Ab-bound proteases (Li *et al.*, 2008).

2.7.5 QSAR

Structural activity relationship is described in terms of physicochemicals, topology, hydrophobicity, electronic properties, steric effects, hydrogen bond donors and hydrogen acceptors. In QSAR studies, molecules with predicted toxic or poor pharmacokinetics properties are eliminated, through filtering virtually large compound libraries. These molecules are eliminated earlier on to narrow the libraries to drug-like or lead-like compounds. There are some QSAR studies that reported on RT. In one recent study by Tarasova *et al.*, the use of data from publicly and commercially available database was addressed to produce an accurate and predictive QSAR model using RT as the case study (Tarasova *et al.*, 2015). They chose two databases, namely Thomson Reuters' Integrity and ChEMBL, to collect all RTIs against both WT and mutants RT, where several methods for the creation of modelling sets from the chosen database were proposed and their accuracy investigated. After choosing the GUSAR program to build the QSAR model, the integrity data yielded high-performance QSAR on all RT forms. The ChEMBL on the other hand provided a more consistent and higher quality QSAR model than the other methods employed in the same database.

The mix-and-match QSAR did not allow other methods to work within the databases which was caused by the lack of unified and standardised descriptors between the Integrity and ChEMBL and this revealed to be a problem to data aggregation.

A paper by Li *et al.* that was published in 2008 showed the combination of 3D-QSAR with MD simulations. According to Dudek *et al.*, the 3D-QSAR methodology consists of obtaining compound descriptors from an experimentally determined ligand and aligning conformers of the chosen dataset in space. In their research, a series of diaryltriazine analogues was extracted and separated in two data sets, namely the 8A-G and 9A-R. This category is a NNRTIs kind of analogue that has been revealed from literature, based on its structure and activity (IC_{50}). To select compounds with low, moderate, and high activity the steric and electrostatic interactions were calculated (Li *et al.*, 2008). These calculations were made by means of the molecular field analysis (MFA) method, leading to the evaluation of five physicochemicals properties that relate to steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptors parameter. This evaluation was done by the comparative molecular similarity analysis method. From each data set, the two most active molecules, 8E with pIC_{50} of 1.21 M and 9H with pIC_{50} of 2.04 M, were further analysed by 2.0 ns MD simulation, which showed more hydrophobic contacts and hydrogen bonds for 9H. This kind of property might make it more active and stable than 8E. The superposition of these compounds showed similar binding modes with RT. This indicated that a conventional 3D-QSAR model of these two types of RTI could be constructed (Dudek *et al.*, 2006).

2.7.6 Pro-drugs

Currently, combination therapy (cART) is one the most effective treatment for HIV-1 (PIs) are an important component of some regimens of cART. Known for sub-optimal ADME properties, PLs have resulted in a poor oral bioavailability, leading to high drug doses. PLs are usually combined with pharmacokinetic enhancers and/or special formulations. This enables them to effectively deliver PLs, leading to high pill burden and sometimes reducing patient compliance (Clas *et al.*, 2014). Over the years, there have been drawbacks due to pharmacokinetic profile and resistance in addition to other factors that challenge antiviral agents (Midde *et al.*, 2016). These two factors,

the pharmacokinetic profile and drug resistance, make them excellent targets for research, especially in viral pandemics such as HIV and hepatitis C methods and probably in the current fight against Covid-19. The prodrug mechanisms are employed in the type of drug synthesised into ester prodrugs, macromolecules prodrugs, targeted delivery prodrugs, non-nucleoside drugs, phosphate prodrugs, amino acids prodrugs, acyl migration prodrugs and other conjugates. This is according to the work done by Subbaiah *et al.*, who investigated design strategies in the prodrugs of HIV-1 protease inhibitors to improve the pharmaceutical properties (Bowman *et al.*, 2005).

Their results showed that, when utilising an ester prodrug approach, a novel isopropyl ester prodrug is a potent HIV integrase inhibitor. VivaGel, which is a dendrimer macromolecular prodrug, showed to be very effective and it is now undergoing clinical trials. It was, however, concluded that the most promising prodrugs are the nucleoside analogues and non-nucleoside analogues prodrugs. These are attempts to improve the ADME properties of existing drugs via prodrug and to reduce side effects, resistance and the ability of drugs to easily move within the body (Subbaiah *et al.*, 2017).

2.7.7 Co-crystal structure

LEDGF, SFRS1 and transcriptional co-activator p75 are the proteins that can be employed as cellular binding partners. These proteins have important properties that include the disposition of the *Lentivirus*, which is a property of the lens epithelium derived growth factor (LEDGF). Functioning within the components of transcription, it mainly integrates these protein partners. A crystal structure was presented by Hare and his colleagues when they were studying the binding factors of these partners. They worked on the N-terminal crystal structures with CCD of HIV-2 IN in complex, which is the catalytic core domain. These were all together formed in a complex with the integrase binding domain of the lens epithelium derived growth factor. More information was extrapolated from this complex of IN-LEDGF interface (Jain *et al.*, 2007).

The effects on this complex were observed by the acidic residue and the positive residues, which were observed on the NTD and the IBD respectively, characterising the mutations together with the lentiviral Ins and these are crucial in stimulating the concerted lentiviral DNA. This integration

was observed when the LEDGF association was tested *in vitro* revealing the important role, which NTD-IBD contacts play during the early steps of HIV-1 replication. A new mutant was derived from this new structure, namely the HIV-1 IN complex. This HIV-1 IN functioned only when presented with a complementary LEDF mutant. Such findings provided a structural basis for the high affinity lentiviral IN–LEDGF interaction. It also gave some further insight in the development of LEDGF-based targeting technologies for gene therapy (Maertens *et al.*, 2003).

2.7.8 Highly drug resistant HIV-1 protease mutant

The development of HIV-1 protease inhibitors has shown their emergence by the initial selection of amino acid changes. This occurs in or near the substrate-binding cleft of the viral protease, namely at codon 30, 50, 82 or 84. It is now clear that these resistance mutations are mainly selected at those positions. Nalam *et al.* Kolli *et al.* and Prabu-Jeyabalan *et al.* demonstrated that the PIs protrude beyond the substrate-binding envelope and are in direct contact with the inhibitor, resulting in an overall enlargement of the catalytic region and such mutations are called primary or major resistance mutations (Nalam *et al.*, 2010; Kolli *et al.*, 2009) (292; Prabu-Jeyabalan *et al.*, 2006). The compensatory or “minor” mutations emerge at codon 20, 36 and 71. Nijhuis *et al.* revealed that these codons do not by themselves have a substantial effect on the drug resistance, but they improve resistance and in time, they replicate viruses containing major mutations. From the works done by Mammano *et al.*, the amino acid changes were observed in the viral protease as well as in the substrate (Mammano *et al.*, 1998).

Furthermore, Kim *et al.*, Amiel *et al.*, Winters *et al.* and Kozisek and colleagues revealed that the Gag substrate changes are mainly observed in the NC/p1 and p1/p6 CS and are believed to increase the affinity of the drug-resistant viral protease. These have an altered substrate-binding pocket for its substrate and occasionally during PI therapy, amino acid insertions that range from one to six amino acids are selected (Kim *et al.*, 2001). This selection was found at different sites of the viral protease sequence. The presence of these positively correlates with the protease resistance mutations conferring reduced susceptibility to contemporary PIs atazanavir, lopinavir,

amprenavir and tipranavir. These are the findings from a study done by Kozísek and colleagues (Amiel *et al.*, 2011; Winters *et al.*, 2005).

This leads to the development of multi drug and highly resistant variants that are often employing the combination of such mechanisms that act synergistically to evade inhibition. An inactive (D25N) variant of a multidrug resistant clinical isolate deemed MDR 769 was discovered. It carried in total 10 mutations, namely I54V, I62V, L10I, M36V, M46L, L63P, A71V, L90M, V82A and I84V and was studied by X-ray crystallography and a variant of MDR769 with V82T displayed 11- 2600-fold resistance to the above-mentioned Pls. The findings were compared to the HIV-1 protease (wild type) and analysed of a 1.3Å resolution crystal structure. Three major differences were described by the wild type and the MDR 769 (Kožíšek *et al.*, 2008; Wang *et al.*, 2011). Firstly, the MDR 767 flaps displayed a ‘wide open’ form because of structural arrangements. Secondly, a distortion of the S1/S1’ was observed. This was a result of L90 that reshapes the drug interactions of the MDR769 active site. This is a common occurrence by a domino effect of perturbations in the 80’s loop and that is what causes the distortion. Finally, there was a further expansion of the active site by an approximately 3.0Å. This occurrence was through the dual reduction I atom volume from the following mutations V82A and 184V (Martin *et al.*, 2005; Yedidi *et al.*, 2011).

REFERENCE LIST

ABBAS, W. & HERBEIN, G. 2014. Plasma membrane signaling in HIV-1 infection. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1838, 1132-1142.

AMARO, R. E., BAUDRY, J., CHODERA, J., DEMIR, Ö., MCCAMMON, J. A., MIAO, Y. & SMITH, J. C. 2018. Ensemble docking in drug discovery. *Biophysical journal*, 114, 2271-2278.

AMIEL, C., CHARPENTIER, C., DESIRE, N., BONNARD, P., LEBRETTE, M. G., WEISS, L., PIALOUX, G. & SCHNEIDER, V. 2011. Long-term follow-up of 11 protease inhibitor (PI)-naïve and PI-treated HIV-infected patients harbouring virus with insertions in the HIV-1 protease gene. *HIV medicine*, 12, 138-144.

BABA, M. 2004. Inhibitors of HIV-1 gene expression and transcription. *Current topics in medicinal chemistry*, 4, 871-882.

BHADORIYA, K. S., SHARMA, M. C. & JAIN, S. V. 2015. 2, 4-Dihydropyrano [2, 3-c] pyrazole: Discovery of new lead as through pharmacophore modelling, atom-based 3D-QSAR, virtual screening and docking strategies for improved anti-HIV-1 chemotherapy. *Journal of Taibah University for Science*, 9, 521-530.

BARRIE, K. A., PEREZ, E. E., LAMERS, S. L., FARMERIE, W. G., DUNN, B. M., SLEASMAN, J. W. & GOODENOW, M. M. 1996. Natural variation in HIV-1 protease, Gag p7 and p6, and protease cleavage sites within gag/pol polyproteins: amino acid substitutions in the absence of protease inhibitors in mothers and children infected by human immunodeficiency virus type 1. *Virology*, 219, 407-416.

BASAVAPATHRUNI, A. & ANDERSON, K. S. 2007. Reverse transcription of the HIV-1 pandemic. *The FASEB Journal*, 21, 3795-3808.

BERMAN, H. M., WESTBROOK, J., FENG, Z., GILLILAND, G., BHAT, T. N., WEISSIG, H., SHINDYALOV, I. N. & BOURNE, P. E. 2000. The protein data bank. *Nucleic acids research*, 28, 235-242.

BOHACEK, R. S. & MCMARTIN, C. 1997. Modern computational chemistry and drug discovery: structure-generating programs. *Current opinion in chemical biology*, 1, 157-161.

BOWMAN, M. J., BYRNE, S. & CHMIELEWSKI, J. 2005. Switching between allosteric and dimerization inhibition of HIV-1 protease. *Chemistry & biology*, 12, 439-444.

BROOIJMANS, N. & KUNTZ, I. D. 2003. Molecular recognition and docking algorithms. *Annual review of biophysics and biomolecular structure*, 32, 335-373.

CARBONE, J., GHIDINI, A., ROMANO, A., GENTILUCCI, L. & MUSIANI, F. 2022. PacDOCK: A Web Server for Positional Distance-Based and Interaction-Based Analysis of Docking Results. *Molecules*, 27, 6884.

CARTER, C. A. & EHRLICH, L. S. 2008. Cell biology of HIV-1 infection of macrophages. *Annu. Rev. Microbiol.*, 62, 425-443.

CHAHAR, H. S., BAO, X. & CASOLA, A. 2015. Exosomes and their role in the life cycle and pathogenesis of RNA viruses. *Viruses*, 7, 3204-3225.

CAVASOTTO, C. N. & PHATAK, S. S. 2009. Homology modelling in drug discovery: current trends and applications. *Drug discovery today*, 14, 676-683.

CHEN, J. C.-H., KRUCINSKI, J., MIERCKE, L. J., FINER-MOORE, J. S., TANG, A. H., LEAVITT, A. D. & STROUD, R. M. 2000. Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proceedings of the National Academy of Sciences*, 97, 8233-8238.

- CHEREPANOV, P., AMBROSIO, A. L., RAHMAN, S., ELLENBERGER, T. & ENGELMAN, A. 2005. Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proceedings of the National Academy of Sciences*, 102, 17308-17313.
- CHOI, E., MALLAREDDY, J. R., LU, D. & KOLLURU, S. 2018. Recent advances in the discovery of small-molecule inhibitors of HIV-1 integrase. *Future Science OA*, 4, FSO338.
- CHOUCHANA, L., PARIENTE, A., PANNIER, E., TSATSARIS, V. & TRELUYER, J.-M. 2020. Dolutegravir and neural tube defects: a new insight. *The Lancet Infectious Diseases*, 20, 405-406.
- CICHERO, E., CESARINI, S., SPALLAROSSA, A., MOSTI, L. & FOSSA, P. 2009. Computational studies of the binding mode and 3D-QSAR analyses of symmetric formimidoester disulfides: a new class of non-nucleoside HIV-1 reverse transcriptase inhibitor. *Journal of molecular modeling*, 15, 357-367.
- CLAS, S.-D., SANCHEZ, R. I. & NOFSINGER, R. 2014. Chemistry-enabled drug delivery (prodrugs): recent progress and challenges. *Drug discovery today*, 19, 79-87.
- COFFIN, J. M., HUGHES, S. H. & VARMUS, H. E. 1997. *Retroviruses*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 1997.
- COHEN, C., WOHL, D., ARRIBAS, J. R., HENRY, K., VAN LUNZEN, J., BLOCH, M., TOWNER, W., WILKINS, E., EBRAHIMI, R. & PORTER, D. 2014. Week 48 results from a randomized clinical trial of rilpivirine/emtricitabine/tenofovir disoproxil fumarate vs. efavirenz/emtricitabine/tenofovir disoproxil fumarate in treatment-naïve HIV-1-infected adults. *Aids*, 28, 989-997.
- D'URBANO, V., DE CRIGNIS, E. & RE, M. C. 2018. Host restriction factors and Human immunodeficiency Virus (HIV-1): a dynamic interplay involving all phases of the viral life cycle. *Current HIV research*, 16, 184-207.

DAS, K., SARAFIANOS, S. G., CLARK JR, A. D., BOYER, P. L., HUGHES, S. H. & ARNOLD, E. 2007. Crystal structures of clinically relevant Lys103Asn/Tyr181Cys double mutant HIV-1 reverse transcriptase in complexes with ATP and non-nucleoside inhibitor HBY 097. *Journal of molecular biology*, 365, 77-89.

DE BÉTHUNE, M.-P. 2010. Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery, development, and use in the treatment of HIV-1 infection: a review of the last 20 years (1989–2009). *Antiviral research*, 85, 75-90.

DE CLERCQ, E. 2005. Antiviral drug discovery and development: where chemistry meets with biomedicine. *Antiviral research*, 67, 56-75.

DE CLERCQ, E. 2009. Another ten stories in antiviral drug discovery (part C):“old” and “new” antivirals, strategies, and perspectives. *Medicinal research reviews*, 29, 611-645.

DE MAGALHÃES, C. S., ALMEIDA, D. M., BARBOSA, H. J. C. & DARDENNE, L. E. 2014. A dynamic niching genetic algorithm strategy for docking highly flexible ligands. *Information Sciences*, 289, 206-224.

DE MAGALHÃES, C. S., BARBOSA, H. J. & DARDENNE, L. E. Selection-insertion schemes in genetic algorithms for the flexible ligand docking problem. *Genetic and Evolutionary Computation Conference*, 2004. Springer, 368-379.

DEEKS, E. D. 2018. Bictegravir/emtricitabine/tenofovir alafenamide: a review in HIV-1 infection. *Drugs*, 78, 1817-1828.

DEGOEY, D. A., GRAMPOVNIK, D. J., CHEN, H.-J., FLOSI, W. J., KLEIN, L. L., DEKHTYAR, T., STOLL, V., MAMO, M., MOLLA, A. & KEMPF, D. J. 2011. P1-substituted symmetry-based human immunodeficiency virus protease inhibitors with potent antiviral activity against drug-resistant viruses. *Journal of medicinal chemistry*, 54, 7094-7104.

DUDEK, A. Z., ARODZ, T. & GÁLVEZ, J. 2006. Computational methods in developing quantitative structure-activity relationships (QSAR): a review. *Combinatorial chemistry & high throughput screening*, 9, 213-228.

ELOKELY, K. M. & DOERKSEN, R. J. 2013. Docking challenge: protein sampling and molecular docking performance. *Journal of chemical information and modelling*, 53, 1934-1945.

FDA.ADMINISTRATION 2017. "FDA approval of Isentress (raltegravir)". Archived from the original

FEINBERG, M. B. & GREENE, W. C. 1992. Molecular insights into human immunodeficiency virus type 1 pathogenesis. *Current opinion in immunology*, 4, 466-474.

FIKKERT, V., HOMBROUCK, A., VAN REMOORTEL, B., DE MAEYER, M., PANNECOUQUE, C., DE CLERCQ, E., DEBYSER, Z. & WITVROUW, M. 2004. Multiple mutations in human immunodeficiency virus-1 integrase confer resistance to the clinical trial drug S-1360. *Aids*, 18, 2019-2028.

FISCHL, M. A., RICHMAN, D. D., GRIECO, M. H., GOTTLIEB, M. S., VOLBERDING, P. A., LASKIN, O. L., LEEDOM, J. M., GROOPMAN, J. E., MILDVAN, D. & SCHOOLEY, R. T. 1987. The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex. *New England Journal of Medicine*, 317, 185-191.

FRACZEK, T., SIWEK, A. & PANETH, P. 2013. Assessing molecular docking tools for relative biological activity prediction: a case study of triazole HIV-1 NNRTIs. *Journal of chemical information and modelling*, 53, 3326-3342.

GANDHI, M., TEAM, I. S., GLIDDEN, D. V., TEAM, I. S., LIU, A., TEAM, I. S., ANDERSON, P. L., TEAM, I. S., HORNG, H. & TEAM, I. S. 2015. Strong correlation between concentrations of tenofovir (TFV) emtricitabine (FTC) in hair and TFV diphosphate and FTC triphosphate in dried blood spots in the iPrEx open label extension: implications for pre-exposure prophylaxis adherence monitoring. *The Journal of infectious diseases*, 212, 1402-1406.

GHORBAN, D. L., MADADKAR, S. A., SHAFIEI, A. & MAHMOUDIAN, M. 2004. Flexible ligand docking studies of matrix metalloproteinase inhibitors using Lamarckian genetic algorithm.

GIUDICI, B., VAZ, B., BOSSOLASCO, S., CASARI, S., BRAMBILLA, A. M., LÜKE, W., LAZZARIN, A., WEBER, T. & CINQUE, P. 2000. Highly active antiretroviral therapy and progressive multifocal leukoencephalopathy: effects on cerebrospinal fluid markers of JC virus replication and immune response. *Clinical infectious diseases*, 30, 95-99.

GRANOFF, A. & WEBSTER, R. G. 1999. *Encyclopaedia of virology*, Elsevier.

GÜNTHARD, H. F., SAAG, M. S., BENSON, C. A., DEL RIO, C., ERON, J. J., GALLANT, J. E., HOY, J. F., MUGAVERO, M. J., SAX, P. E. & THOMPSON, M. A. 2016. Antiretroviral drugs for treatment and prevention of HIV infection in adults: 2016 recommendations of the International Antiviral Society–USA panel. *Jama*, 316, 191-210.

HASSAN BAIG, M., AHMAD, K., ROY, S., MOHAMMAD ASHRAF, J., ADIL, M., HARIS SIDDIQUI, M., KHAN, S., AMJAD KAMAL, M., PROVAZNÍK, I. & CHOI, I. 2016. Computer aided drug design: success and limitations. *Current pharmaceutical design*, 22, 572-581.

HAYASH, T. 2016. Preventive effect of ascorbic acid against biological function of human immunodeficiency virus trans-activator of transcription. *Journal of Intercultural Ethnopharmacology*, 5, 205.

HAVENS, P. L., KISER, J. J., STEPHENSEN, C. B., HAZRA, R., FLYNN, P. M., WILSON, C. M., RUTLEDGE, B., BETHEL, J., PAN, C. G. & WOODHOUSE, L. R. 2013. Association of higher plasma vitamin D binding protein and lower free calcitriol levels with tenofovir disoproxil fumarate use and plasma and intracellular tenofovir pharmacokinetics: cause of a functional vitamin D deficiency? *Antimicrobial agents and chemotherapy*, 57, 5619-5628.

HEALTH, N. I. O. 1998. "Human Immunodeficiency Viruses: Molecular Virology, pathogenesis, diagnosis and treatment". 13, 75–87.

HERBEIN, G., GRAS, G., KHAN, K. A. & ABBAS, W. 2010. Macrophage signaling in HIV-1 infection. *Retrovirology*, 7, 1-13.

HETÉNYI, C. & VAN DER SPOEL, D. 2006. Blind docking of drug-sized compounds to proteins with up to a thousand residues. *FEBS letters*, 580, 1447-1450.

HETHERINGTON, S., MCGUIRK, S., POWELL, G., CUTRELL, A., NADERER, O., SPREEN, B., LAFON, S., PEARCE, G. & STEEL, H. 2001. Hypersensitivity reactions during therapy with the nucleoside reverse transcriptase inhibitor abacavir. *Clinical therapeutics*, 23, 1603-1614.

HINDLE, S. A., RAREY, M., BUNING, C. & LENGAUER, T. 2002. Flexible docking under pharmacophore type constraints. *Journal of computer-aided molecular design*, 16, 129-149.

HOLMBERG, S. D., PALELLA JR, F. J., LICHTENSTEIN, K. A. & HAVLIR, D. V. 2004. The case for earlier treatment of HIV infection. *Clinical infectious diseases*, 39, 1699-1704.

HSIOU, Y., DAS, K., DING, J., CLARK JR, A. D., KLEIM, J.-P., RÖSNER, M., WINKLER, I., RIESS, G., HUGHES, S. H. & ARNOLD, E. 1998. Structures of Tyr188Leu mutant and wild type HIV-1 reverse transcriptase complexed with the non-nucleoside inhibitor HBY 097: inhibitor flexibility is a useful design feature for reducing drug resistance. *Journal of molecular biology*, 284, 313-323.

HUANG, H.-J., YU, H. W., CHEN, C.-Y., HSU, C.-H., CHEN, H.-Y., LEE, K.-J., TSAI, F.-J. & CHEN, C. Y.-C. 2010. Current developments of computer-aided drug design. *Journal of the Taiwan Institute of Chemical Engineers*, 41, 623-635.

HUNG-JINHUANGA 2010. Current developments of computer-aided drug design. *Journal of the Tai. Inst. Chem. Eng.*, 41, 623-635.

IFEANYI, O. E., AMILO GRACE, I., UZOMA, O. G. & NNATUANYA ISAAC, N. 2017. Human Immunodeficiency Virus infection and cardiovascular. *Int. J. Curr. Res. Med. Sci*, 3, 9-37.

ILINSKAYA, O. & MAHMUD, R. S. 2014. Ribonucleases as antiviral agents. *Molecular Biology*, 48, 615-623.

IORDANSKIY, S., SANTOS, S. & BUKRINSKY, M. 2013. Nature, nurture and HIV: the effect of producer cell on viral physiology. *Virology*, 443, 208-213.

IORGA, B., HERLEM, D., BARRÉ, E. & GUILLOU, C. 2006. Acetylcholine nicotinic receptors: finding the putative binding site of allosteric modulators using the “blind docking” approach. *Journal of molecular modeling*, 12, 366-372.

JACOBSON, M. P., PINCUS, D. L., RAPP, C. S., DAY, T. J., HONIG, B., SHAW, D. E. & FRIESNER, R. A. 2004. A hierarchical approach to all-atom protein loop prediction. *Proteins: Structure, Function, and Bioinformatics*, 55, 351-367.

JAIN, R., DUVVURI, S., KANSARA, V., MANDAVA, N. K. & MITRA, A. K. 2007. Intestinal absorption of novel-dipeptide prodrugs of saquinavir in rats. *International journal of pharmaceutics*, 336, 233-240.

JIANG, N., FANG, Y., FANG, Y., WANG, S.-Y. & JI, S.-J. 2019. Iodine-promoted one pot reaction of pyridin-2-amine with arylmethyl ketone and selenosulfonate: synthesis of 3-(alkylselanyl)-2-arylimidazo [1, 2-a] pyridine under transition-metal free conditions. *Organic Chemistry Frontiers*, 6, 654-659.

JOO, K., LEE, J. & LEE, J. 2011. Methods for accurate homology modelling by global optimization. *Homology Modelling*. Springer.

JUDES, G., RIFAÏ, K., NGOLLO, M., DAURES, M., BIGNON, Y.-J., PENAULT-LLORCA, F. & BERNARD-GALLON, D. 2015. A bivalent role of TIP60 histone acetyl transferase in human cancer. *Epigenomics*, 7, 1351-1363.

KIM, E.-Y., WINTERS, M. A., KAGAN, R. M. & MERIGAN, T. C. 2001. Functional correlates of insertion mutations in the protease gene of human immunodeficiency virus type 1 isolates from patients. *Journal of virology*, 75, 11227-11233.

KING, S. R. 1994. HIV: virology and mechanisms of disease. *Annals of emergency medicine*, 24, 443-449.

KIRSCHBERG, T. A., BALAKRISHNAN, M., HUANG, W., HLUHANICH, R., KUTTY, N., LICLICAN, A. C., MCCOLL, D. J., SQUIRES, N. H. & LANSDON, E. B. 2008. Triazole derivatives as non-nucleoside inhibitors of HIV-1 reverse transcriptase—Structure–activity relationships and crystallographic analysis. *Bioorganic & medicinal chemistry letters*, 18, 1131-1134.

KOLLI, M., STAWISKI, E., CHAPPEY, C. & SCHIFFER, C. A. 2009. Human immunodeficiency virus type 1 protease-correlated cleavage site mutations enhance inhibitor resistance. *Journal of virology*, 83, 11027-11042.

KORB, O., OLSSON, T. S., BOWDEN, S. J., HALL, R. J., VERDONK, M. L., LIEBESCHUETZ, J. W. & COLE, J. C. 2012. Potential and limitations of ensemble docking. *Journal of chemical information and modelling*, 52, 1262-1274.

KOŽÍŠEK, M., ŠAŠKOVÁ, K. R. G., ŘEZÁČOVÁ, P. N., BRYNDA, J. I., VAN MAARSEVEEN, N. M., DE JONG, D., BOUCHER, C. A., KAGAN, R. M., NIJHUIS, M. & KONVALINKA, J. 2008. Ninety-nine is not enough: molecular characterization of inhibitor-resistant human immunodeficiency virus type 1-protease mutants with insertions in the flap region. *Journal of virology*, 82, 5869-5878.

KRAMER-HÄMMERLE, S., ROTHENAIGNER, I., WOLFF, H., BELL, J. E. & BRACK-WERNER, R. 2005. Cells of the central nervous system as targets and reservoirs of the human immunodeficiency virus. *Virus research*, 111, 194-213.

KRISHNAN, L., LI, X., NARAHARISETTY, H. L., HARE, S., CHEREPANOV, P. & ENGELMAN, A. 2010. Structure-based modelling of the functional HIV-1 intasome and its inhibition. *Proceedings of the National Academy of Sciences*, 107, 15910-15915.

LATA, S., MISHRA, R. & BANERJEA, A. C. 2018. Proteasomal degradation machinery: favorite target of HIV-1 proteins. *Frontiers in microbiology*, 9, 2738.

LI, G. & DE CLERCQ, E. 2016. HIV genome-wide protein associations: a review of 30 years of research. *Microbiology and Molecular Biology Reviews*, 80, 679-731.

LI, Z., HAN, J. & CHEN, H. F. 2008. Revealing interaction mode between HIV-1 reverse transcriptase and diaryltriazine analogy inhibitor. *Chemical biology & drug design*, 72, 350-359.

LIEDTKE, M. D., TOMLIN, C. R., LOCKHART, S. M., MILLER, M. M. & RATHBUN, R. C. 2014. Long-term efficacy and safety of raltegravir in the management of HIV infection. *Infection and drug resistance*, 7, 73.

LIU, C.-H., GRANDI, N., PALANIVELU, L., TRAMONTANO, E. & LIN, L.-T. 2020. Contribution of Human Retroviruses to Disease Development—A Focus on the HIV–and HERV–Cancer Relationships and Treatment Strategies. *Viruses*, 12, 852.

LIU, Z., LIU, Y., ZENG, G., SHAO, B., CHEN, M., LI, Z., JIANG, Y., LIU, Y., ZHANG, Y. & ZHONG, H. 2018. Application of molecular docking for the degradation of organic pollutants in the environmental remediation: A review. *Chemosphere*, 203, 139-150.

LUCA SANCINETO; ALICE MARIOTTI; LUANA BAGNOLI; FRANCESCA MARINI ;JENNY DESANTIS ;NUNZIO IRACI; CLAUDIO SANTI; CHRISTOPHE PANNECOUQUE; ORIANA TABARRINI Design and Synthesis of DiselenoBisBenzamides (DISEBAs) as Nucleocapsid Protein 7 (NCp7) Inhibitors with anti-HIV Activity. *J. Med. Chem*, 58, 9601–9614.

MAARTENS, G., CELUM, C. & LEWIN, S. R. 2014. HIV infection: epidemiology, pathogenesis, treatment, and prevention. *The Lancet*, 384, 258-271.

MAERTENS, G., CHEREPANOV, P., PLUYMERS, W., BUSSCHOTS, K., DE CLERCQ, E., DEBYSER, Z. & ENGELBORGH, Y. 2003. LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *Journal of Biological Chemistry*, 278, 33528-33539.

MAIGNAN, S., GUILLOTEAU, J.-P., ZHOU-LIU, Q., CLÉMENT-MELLA, C. & MIKOL, V. 1998. Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases. *Journal of molecular biology*, 282, 359-368.

MALLERY, D. L., MÁRQUEZ, C. L., MCEWAN, W. A., DICKSON, C. F., JACQUES, D. A., ANANDAPADAMANABAN, M., BICHEL, K., TOWERS, G. J., SAIARDI, A. & BÖCKING, T. 2018. IP6 is an HIV pocket factor that prevents capsid collapse and promotes DNA synthesis. *Elife*, 7.

MAMMANO, F., PETIT, C. & CLAVEL, F. 1998. Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients. *Journal of virology*, 72, 7632-7637.

MANFREDI, R. & CALZA, L. 2008. HIV infection and the pancreas: risk factors and potential management guidelines. *International journal of STD & AIDS*, 19, 99-105.

MARTIN, P., VICKREY, J. F., PROTEASA, G., JIMENEZ, Y. L., WAWRZAK, Z., WINTERS, M. A., MERIGAN, T. C. & KOVARI, L. C. 2005. “Wide-open” 1.3 Å structure of a multidrug-resistant HIV-1 protease as a drug target. *Structure*, 13, 1887-1895.

MIDDE, N. M., PATTERS, B. J., RAO, P., CORY, T. J. & KUMAR, S. 2016. Investigational protease inhibitors as antiretroviral therapies. *Expert opinion on investigational drugs*, 25, 1189-1200.

MOELLING, K. & BROECKER, F. 2015. The reverse transcriptase–RNase H: from viruses to antiviral defence. *Annals of the New York Academy of Sciences*, 1341, 126-135.

MORRIS, G. M., GOODSELL, D. S., HALLIDAY, R. S., HUEY, R., HART, W. E., BELEW, R. K. & OLSON, A. J. 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of computational chemistry*, 19, 1639-1662.

MUHAMMED, M. T. & AKI-YALCIN, E. 2019. Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chemical biology & drug design*, 93, 12-20.

MUSHAHWAR, I. K. 2006. Human immunodeficiency viruses: molecular virology, pathogenesis, diagnosis and treatment. *Perspectives in Medical Virology*, 13, 75-87.

NALAM, M. N., ALI, A., ALTMAN, M. D., REDDY, G. K. K., CHELLAPPAN, S., KAIRYS, V., ÖZEN, A. E. L., CAO, H., GILSON, M. K. & TIDOR, B. 2010. Evaluating the substrate-envelope hypothesis: structural analysis of novel HIV-1 protease inhibitors designed to be robust against drug resistance. *Journal of virology*, 84, 5368-5378.

NARKHEDE, S. S. & DEGANI, M. S. 2007. Pharmacophore Refinement and 3D-QSAR Studies of Histamine H3 Antagonists. *QSAR & Combinatorial Science*, 26, 744-753.

NELSON, M. & SCHIAVONE, M. 2004. Emtricitabine (FTC) for the treatment of HIV infection. *International journal of clinical practice*, 58, 504-510.

NIMA RAZZAGHI-ASLA 2015. Effect of Biomolecular Conformation on Docking Simulation: A Case Study on a Potent HIV-1 Protease Inhibitor. *Iranian Journal of Pharmaceutical Research* 14, 785-802.

PADARIYA, M. & KALATHIYA, U. 2017. Comparative molecular dynamics study of dimeric and monomeric forms of HIV-1 protease in ligand bound and unbound state. *Gen. Physiol. Biophys*, 36, 141-154.

PEARL, L. H. & TAYLOR, W. R. 1987. A structural model for the retroviral proteases. *Nature*, 329, 351-354.

PETERLE, M. M., SCHEIDE, M. R., SILVA, L. T., SABA, S., RAFIQUE, J. & BRAGA, A. L. 2018. Copper-catalyzed three-component reaction of oxadiazoles, elemental Se/S and aryl iodides: synthesis of chalcogenyl (Se/S)-oxadiazoles. *ChemistrySelect*, 3, 13191-13196.

PIOT, P., KARIM, S. S. A., HECHT, R., LEGIDO-QUIGLEY, H., BUSE, K., STOVER, J., RESCH, S., RYCKMAN, T., MØGEDAL, S. & DYBUL, M. 2015. Defeating AIDS—advancing global health. *The Lancet*, 386, 171-218.

PRABU-JEYABALAN, M., NALIVAICA, E. A., ROMANO, K. & SCHIFFER, C. A. 2006. Mechanism of substrate recognition by drug-resistant human immunodeficiency virus type 1 protease variants revealed by a novel structural intermediate. *Journal of virology*, 80, 3607-3616.

PRAJAPATI, D. G., RAMAJAYAM, R., YADAV, M. R. & GIRIDHAR, R. 2009. The search for potent, small molecule NNRTIs: A review. *Bioorganic & medicinal chemistry*, 17, 5744-5762.

PRASAD, M., RANJAN, K., BRAR, B., SHAH, I., LALMBE, U., MANIMEGALAI, J., VASHISHT, B., GAURY, M., KUMAR, P. & KHURANA, S. K. 2017. Virus-host interactions: new insights and advances in drug development against viral pathogens. *Current Drug Metabolism*, 18, 942-970.

PUHL, A. C., DEMO, A. G., MAKAROV, V. A. & EKINS, S. 2019. New targets for HIV drug discovery. *Drug discovery today*, 24, 1139-1147.

QAYUM, M., NISAR, M., RAUF, A., KHAN, I., KALEEM, W. A., RAZA, M., KARIM, N., SALEEM, M. A., BAWAZEER, S. & UYSAL, S. 2019. In-vitro and in-silico anticancer potential of taxoids from *Taxus wallichiana* Zucc. *Biologia Futura*, 70, 295-300.

RAREY, M., KRAMER, B., LENGAUER, T. & KLEBE, G. 1996. A fast flexible docking method using an incremental construction algorithm. *Journal of molecular biology*, 261, 470-489.

- RAZZAGHI-ASL, N., EBADI, A., EDRAKI, N., MEHDIPOUR, A., SHAHABIPOUR, S. & MIRI, R. 2012. Response surface methodology in docking study of small molecule BACE-1 inhibitors. *Journal of molecular modelling*, 18, 4567-4576.
- REN, J., CHAMBERLAIN, P. P., STAMP, A., SHORT, S. A., WEAVER, K. L., ROMINES, K. R., HAZEN, R., FREEMAN, A., FERRIS, R. G. & ANDREWS, C. W. 2008. Structural basis for the improved drug resistance profile of new generation benzophenone non-nucleoside HIV-1 reverse transcriptase inhibitors. *Journal of medicinal chemistry*, 51, 5000-5008.
- RIDKY, T. W., CAMERON, C. E., CAMERON, J., LEIS, J., COPELAND, T., WLODAWER, A., WEBER, I. T. & HARRISON, R. W. 1996. Human Immunodeficiency Virus, Type 1 Protease Substrate Specificity Is Limited by Interactions between Substrate Amino Acids Bound in Adjacent Enzyme Subsites (*). *Journal of Biological Chemistry*, 271, 4709-4717.
- SANCINETO, L. 2015. Design and Synthesis of DiselenoBisBenzamides (DSeBAs) as Nucleocapsid Protein 7 (NCp7) Inhibitors with anti-HIV Activity. *J. Med. Chem*, 58, 9601–9614.
- SATAKE, A., ITOH, K., SHIMMOTO, M., SAIDO, T. C., SAKURABA, H. & SUZUKI, Y. 1994. Distribution of lysosomal protective protein in human tissues. *Biochemical and biophysical research communications*, 205, 38-43.
- SCHAUPERL, M. & DENNY, R. A. 2022. AI-Based protein structure prediction in drug discovery: Impacts and challenges. *Journal of Chemical Information and Modeling*, 62, 3142-3156.
- SEAL, A., AYKKAL, R., BABU, R. O. & GHOSH, M. 2011. Docking study of HIV-1 reverse transcriptase with phytochemicals. *Bioinformation*, 5, 430.
- SHAH, K., GUPTA, S., MISHRA, H., SHARMA, P. K. & JAYASWAL, A. 2014. Examining structural analogy of elvitegravir as potential inhibitors of HIV-1 integrase. *Archives of virology*, 159, 2069-2080.

SHAMSI, T. N., PARVEEN, R. & FATIMA, S. 2016. Characterization, biomedical and agricultural applications of protease inhibitors: A review. *International journal of biological macromolecules*, 91, 1120-1133.

SHARMA, H., CHENG, X. & BUOLAMWINI, J. K. 2012. Homology model-guided 3D-QSAR studies of HIV-1 integrase inhibitors. *Journal of chemical information and modelling*, 52, 515-544.

SHERMAN, W., DAY, T., JACOBSON, M. P., FRIESNER, R. A. & FARID, R. 2006. Novel procedure for modelling ligand/receptor induced fit effects. *Journal of medicinal chemistry*, 49, 534-553.

SHI, S., NGUYEN, P. K., CABRAL, H. J., DIEZ-BARROSO, R., DERRY, P. J., KANAHARA, S. M. & KUMAR, V. A. 2016. Development of peptide inhibitors of HIV transmission. *Bioactive materials*, 1, 109-121.

SHIMURA, K., KODAMA, E., SAKAGAMI, Y., MATSUZAKI, Y., WATANABE, W., YAMATAKA, K., WATANABE, Y., OHATA, Y., DOI, S. & SATO, M. 2008. Broad antiretroviral activity and resistance profile of the novel human immunodeficiency virus integrase inhibitor elvitegravir (JTK-303/GS-9137). *Journal of virology*, 82, 764-774.

SINGLETON, P. S., D, 2006. "Hiv". *Dictionary of microbiology & molecular biology* 3.

SLUIS-CREMER, N., TEMIZ, N. A. & BAHAR, I. 2004. Conformational changes in HIV-1 reverse transcriptase induced by nonnucleoside reverse transcriptase inhibitor binding. *Current HIV research*, 2, 323-332.

STELLBRINK, H.-J. 2007. Antiviral drugs in the treatment of AIDS: what is in the pipeline? *European journal of medical research*, 12, 483-495.

STUDER, G., TAURIELLO, G., BIENERT, S., BIASINI, M., JOHNER, N. & SCHWEDE, T. 2021. ProMod3—A versatile homology modelling toolbox. *PLoS computational biology*, 17, e1008667.

SUBBAIAH, M. A., MEANWELL, N. A. & KADOW, J. F. 2017. Design strategies in the prodrugs of HIV-1 protease inhibitors to improve the pharmaceutical properties. *European journal of medicinal chemistry*, 139, 865-883.

TAKAHASHI, M., BURNETT, J. C. & ROSSI, J. J. 2015. Aptamer–siRNA chimeras for HIV. *Gene Therapy for HIV and Chronic Infections*, 211-234.

TARASOVA, O. A., URUSOVA, A. F., FILIMONOV, D. A., NICKLAUS, M. C., ZAKHAROV, A. V. & POROIKOV, V. V. 2015. QSAR modelling using large-scale databases: case study for HIV-1 reverse transcriptase inhibitors. *Journal of chemical information and modelling*, 55, 1388-1399.

TORRES, P. H., SODERO, A. C., JOFILY, P. & SILVA-JR, F. P. 2019. Key topics in molecular docking for drug design. *International journal of molecular sciences*, 20, 4574.

TROTT, O. & OLSON, A. J. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31, 455-461.

VADIVELAN, S., DEEKSHA, T., ARUN, S., MACHIRAJU, P. K., GUNDLA, R., SINHA, B. N. & JAGARLAPUDI, S. A. 2011. Virtual screening studies on HIV-1 reverse transcriptase inhibitors to design potent leads. *European journal of medicinal chemistry*, 46, 851-859.

VERDONK, M. L., COLE, J. C., HARTSHORN, M. J., MURRAY, C. W. & TAYLOR, R. D. 2003. Improved protein–ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics*, 52, 609-623.

- VOLKAMER, A., KUHN, D., GROMBACHER, T., RIPPMANN, F. & RAREY, M. 2012. Combining global and local measures for structure-based druggability predictions. *Journal of chemical information and modelling*, 52, 360-372.
- VYAS, V., JAIN, A., JAIN, A. & GUPTA, A. 2008. Virtual screening: a fast tool for drug design. *Scientia Pharmaceutica*, 76, 333-360.
- WANG, Y., LIU, Z., BRUNZELLE, J. S., KOVARI, I. A., DEWDNEY, T. G., REITER, S. J. & KOVARI, L. C. 2011. The higher barrier of darunavir and tipranavir resistance for HIV-1 protease. *Biochemical and biophysical research communications*, 412, 737-742.
- WANG, X., FENG, L. & ZHAO, H. 2019. Fast image encryption algorithm based on parallel computing system. *Information Sciences*, 486, 340-358.
- WEISS, R., TEICH, N., VARMUS, H. & COFFIN, J. 1985. RNA tumor viruses.
- WESTBROOK, J., FENG, Z., CHEN, L., YANG, H. & BERMAN, H. M. 2003. The protein data bank and structural genomics. *Nucleic acids research*, 31, 489-491.
- WINTERS, M. A., KAGAN, R. M., HESELTINE, P. N. & MERIGAN, T. C. 2005. New two-amino acid insertion near codon 70 of the HIV type 1 protease gene. *AIDS Research & Human Retroviruses*, 21, 311-313.
- YARCHOAN, R., LIETZAU, J. A., NGUYEN, B.-Y., BRAWLEY, O. W., PLUDA, J. M., SAVILLE, M. W., WYVILL, K. M., STEINBERG, S. M., AGBARIA, R. & MITSUYA, H. 1994. A randomized pilot study of alternating or simultaneous zidovudine and didanosine therapy in patients with symptomatic human immunodeficiency virus infection. *Journal of Infectious Diseases*, 169, 9-17.
- YEANG, T. S. 2015. *CELL-PENETRATING ANTIBODIES FOR TARGETING HIV-1 P24 CAPSID PROTEIN*. Universiti Sains Malaysia.

YEDIDI, R. S., PROTEASA, G., MARTINEZ, J. L., VICKREY, J. F., MARTIN, P. D., WAWRZAK, Z., LIU, Z., KOVARI, I. A. & KOVARI, L. C. 2011. Contribution of the 80s loop of HIV-1 protease to the multidrug-resistance mechanism: crystallographic study of MDR769 HIV-1 protease variants. *Acta Crystallographica Section D: Biological Crystallography*, 67, 524-532.

YI, Y., RANA, S., TURNER, J. D., GADDIS, N. & COLLMAN, R. G. 1998. CXCR-4 is expressed by primary macrophages and supports CCR5-independent infection by dual-tropic but not T-tropic isolates of human immunodeficiency virus type 1. *Journal of Virology*, 72, 772-777.

ZAID, H., RAIYN, J., OSMAN, M., FALAH, M., SROUJI, S. & RAYAN, A. 2016. In silico modeling techniques for predicting the tertiary structure of human H4 receptor. *Frontiers in Bioscience-Landmark*, 21, 597-619.

CHAPTER 3: GENERAL EXPERIMENTAL PROCEDURES

3.1 MATERIALS AND METHODS

Computer aided drug design is an important tool used to reduce costs of drug research and to save time in finding relevant drug for a particular disease. This application is widely used in pharmaceutical industry. Below is Figure 3.1 that shows the approach that was employed in this project.

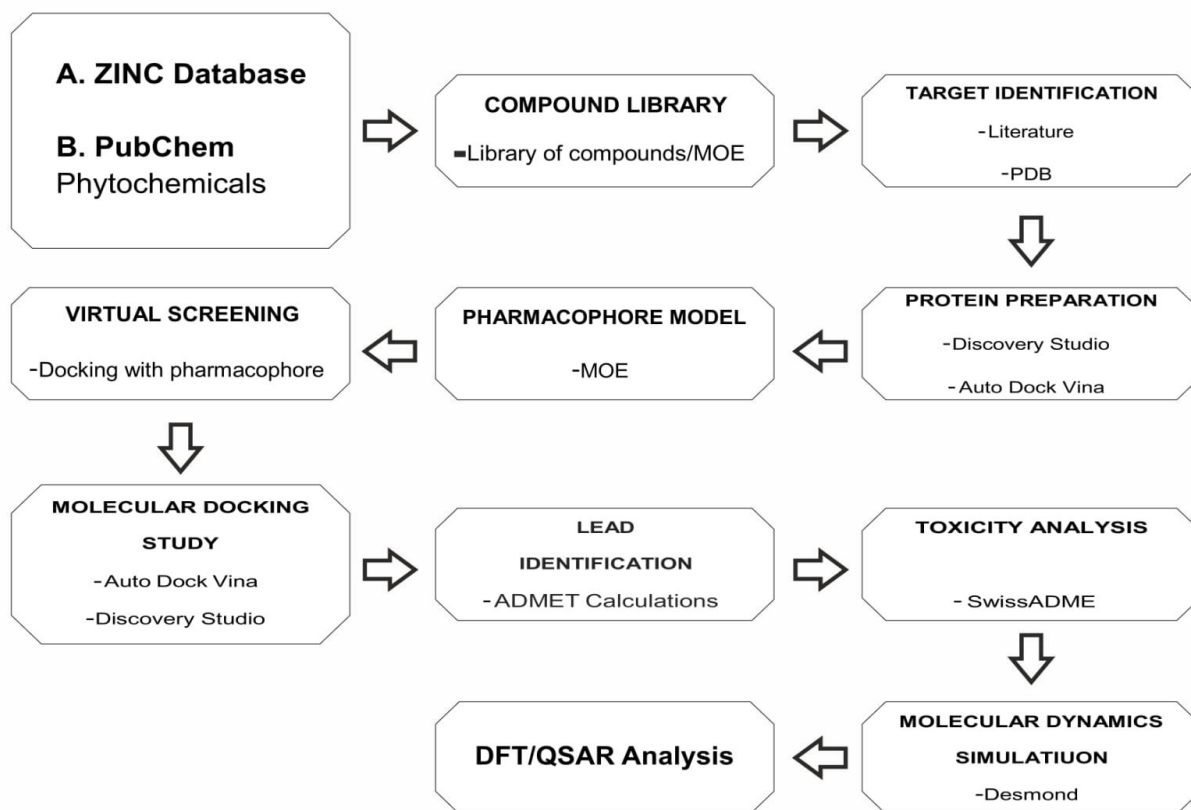


Figure 3. 1: A flowchart representation involved in the study

3.1.1 Selection and retrieval of Ligands

Ligand and compound database (ZINC database)

ZINC is a free and accessible database that enables the use of stored compounds for virtual screening (Irwin and Shoichet, 2005). This server plays an important role in searching for relevant ligands that are suitable for drug discovery and pharmacophore modelling, for a specific target. ZINC database assisted in finding new compounds that are in line with existing inhibitors for a particular enzyme (Shoichet, 2004), (Zhong *et al.*, 2012). These compounds were optimized and arranged accordingly for preparation, synthesis and further computational use. Both biological and chemoinformatics are the backbone for this database as they pull together the needed information to create a query language for researchers and commercial availability (Doytchinova *et al.*, 2018).

In the quest of finding a possible drug candidate, the Lipinski's rule of five was used as a filter for virtual screening of ZINC database (www.zinc.docking.org) (Zhang *et al.*, 2021). The assessment of the molecular parameters of the compounds is accomplished by molinspiration online server (<http://www.molinspiration.com/>). The properties of the compounds according to the Lipinski's rule of five are MW < 500 Da; CLogP < 5; H-bond donor < 5; H-bond acceptor <10 (Lipinski *et al.*, 1997).

This database can further be interrogated based on the ligands that have active biological properties and can bind a particular protein. Such compounds are deposited in the database and useful in finding an inhibitor for HIV-1 protease. The Lipinski's rule of five assisted in finding such compounds that have a better 3D representation for pharmacophore screening and better docking. Docking was also improved by using the latest version of Omega (OpenEye Scientific Software (<http://eyesopen.com>)) (CRISAN and BORA).

The ligands retrieved from this database were useful in generating a hypothesis based on the newly discovered ligands for a particular protein. These ligands were used in designing a drug for HIV-1 protease.

Ligand and compound database (PubChem database)

In all the Phytochemicals that were selected for this investigation, (89) of them were retrieved from the PubChem database. Just like the ZINC database, PubChem is a public repository for ligands and chemical substances and their biological activities that are deposited into the database. Over 219 million chemical substances, 89 million compounds and 230 million bioactivity outcomes are available in the database and about 94 million new structures were deposited in 2017 only. It contains a broad range of biochemical entities that includes, compounds (<https://www.ncbi.nlm.nih.gov/pccompound>), amino acids, lipids and 3-D structure of small molecules (which were used in this investigation) (<https://www.ncbi.nlm.nih.gov/pcsubstance>). These plant extracts were each collected using the phytochemical name and their structures saved for virtual screening (Kharisma *et al.*, 2021).

PubChem (<https://pubchem.ncbi.nlm.nih.gov>) provided the links for the articles (<https://www.pubmed.gov>) that mention some of these phytochemicals and these provided a strong background of the nature of some of these plant extracts. The following information about the phytochemicals was retrieved from PubChem: an information of already approved FDA antiretroviral drugs (Sahin, 2021); the interactions with protein target (for phytochemicals that had already being investigated) (Akhtar *et al.*, 2019), (Panwar and Singh, 2018); investigational drug and drug indications and many more. This information about these phytochemical was key in deriving a hypothesis pertaining to the use and application of plant extracts in the fight against HIV (Shiri *et al.*, 2018).

The phytochemicals retrieved from PubChem were used to determine the interaction of the protein, providing the measured binding affinities of the lead phytochemicals. Using the BindingDB (<https://www.bindingdb.org>), the binding properties of the phytochemicals that had been investigated were provided, and this revealed much needed information of the binding nature of these plant extracts (Copoiu *et al.*, 2020).

In order to have a cause of study on these PubChem compounds and the extent of the disease in question, the Guide to Pharmacology (<http://www.guidetopharmacology.org>), pertaining to these

small molecules, was considered. This guide was created under, International Union of Basic and Clinical Pharmacology (IUPHAR) and it provided in depth pathophysiology of important drug targets to the HIV-1 protease (Ju *et al.*, 2020).

ChEBi

Some of the small molecules in this study were substantiated in the ChEBI database (<http://www.ebi.ac.uk/chebi/>). This database focusses on the classification of compounds based on their biological activity and structural features. libChEBI provides a simple interface, to the contents of the ChEBI database and the memory is accessible through Python, Java and MATLAB APIs (Sharma *et al.*, 2020).

3.1.2 Target database

PDB

The target used in this study, is a 1rv7 protein, which was retrieved from the protein data bank (PDB) (<http://www.rcsb.org/pdb/>) (Burley *et al.*, 2019). It is a strain of HIV-1 protease that was isolated from a patient who was failing extensive antiretroviral therapy based on persistent viremia. Using X-ray diffraction, 1rv7 structure was saved and deposited into the PDB as a crystal structure of a multidrug-resistant human immunodeficiency virus type 1 protease in order to conduct future scientific studies. 1rv7 crystal structure has a resolution of 2.70 Å and an R-value free of 0.350. PDB provided the structure summary, the 3-D view and the sequence overview of the macromolecular under study, to name but a few. The sequence length for this protein is 99, with chains A and B (Bradley C Logsdon 2004); (Kneller *et al.*, 2020),

3.1.3 Protein preparation

Biovia discovery studio

The macromolecule that was retrieved from PDB was prepared using Biovia discovery studio visualizer. The components of this studio includes a hardware with the following requirements: a

processor component with a minimum hardware Intel-compatible x86_64 architecture and RAM with a minimum hardware 4 GB per core and 16 GB for a single core processor (additional memory will increase performance), with a disk space of 1GB. The operation was performed on a Windows 10, 64-bit computer (Valentine et al., 2021) (Sillapachaiyaporn et al., 2021).

Since the structure that was retrieved from the PDB database, was a co-crystallized structure of a multi-drug resistant strain of HIV-1 protease and a ligand, (Lopinavir), it was necessary to have the strain without any ligand interaction in order to carry out the docking analysis of the novel compounds for this study (Bradley C Logsdon 2004).

The protein was prepared using the following steps (Sharma et al., 2021) (Egbuna et al., 2021) (Najm et al., 2021) (Zuhrufa and Julianto, 2021):

Step 1. Delete Water molecule, since it will increase the affinity between the ligand and the protein.

Step 2. Choose active site and see all the active sites.

Step 3. Choose Ligand Group and analyze all small molecules Step 4. Select ligand 1 and select Receptor Ligand Interaction tab.

Step 4. Define and Edit Binding site and click from current selection: a red colour ball covers the ligand, SBD Site Sphere.

Step 5. To expand the sphere, click SBD Site Sphere and Click Expand. Step 6. Select SBD Site Sphere and Right Click and select Attribute and note X, Y, Z Coordinates for configuration file.

Step 7. Select Ligand group and go to Edit and delete Ligand Groups. Step 8. Delete Chain B with Edit option.

Step 9. Add Polar Hydrogen by going to Chemistry Tab and select Hydrogen and click Add polar And Save as PDB Format.

3.1.4 Pharmacophore modelling

MOE

The Molecular Operating Environment used a general-purpose comprehensive molecular modelling platform that has a clean and intuitive interface. The MOE software has eleven different force fields. The modelling of pharmacophore was performed on this software. Vital features responsible for the biological activity of the co-crystallized structure of 1rv7 (for structure-based pharmacophore modelling) (Appiah-Kubi *et al.*, 2021) and of the three ZINC compounds obtained through screening (for ligand-based pharmacophore modelling) were constructed on this platform (Katiyar *et al.*, 2021).

The MOE carries many built-in functions for modelling both small molecules and macromolecules. The advantage of this software is that it has a SVL scripting, is cross-platform (Linux, Mac OS, and Windows), and can serve as a front-end for a variety of other programs, such as, GAMESS, ChemDraw, AMBER, Corina, ADF, Gaussian, GOLD, Mogul, MOPAC, Omega and NAMD (Elghoneimy *et al.*, 2021). A pharmacophore was created by applying the following steps: Step 1: Selection and preparation of the training set; Step 2: Flexible alignment. Step 3: Building the pharmacophore query- (El-Hassab *et al.*, 2020) (Elseginy, 2021) (Quazi *et al.*, 2021) and Step 4: Pharmacophore search of the test set. The training compounds were selected using the MOE builder that was saved as a molecular database file. The database viewer enables the display of database files. Using the active ligands that binds to the target, a pharmacophore is constructed and in conditions where the target is not available, meaning the already existing ligand-protein interaction is absent, that is referred to as a ligand-based pharmacophore modelling approach (Muhammed and Esin, 2021).

In this study, the ligands were obtained from the ZINC database and were used to create a ligand-based pharmacophore, which function is to screen the phytochemicals that were selected for molecular modelling. In addition, in conditions where the structure of the target in the co-crystallized structure, in this study 1rv7 co-crystallized structure is known and available, a pharmacophore model was built, based on the structure properties of the target. This is referred to

as structure-based pharmacophore (Sanders *et al.*, 2012). The three structures from the ligand and from the target-inhibitor structure were then constructed.

3.1.5 Docking analysis

AutoDock Vina

AutoDock Vina was downloaded and installed on Windows XP by double clicking the download MSI file. The software was run by opening the Command Prompt and Vina that was installed in the default location, type "*Program Files\The Scripps Research Institute\Vina\vina.exe*" *-help*. The MGL tools software was also downloaded. Receptor = 1rv7.pdbqt; exhaustiveness = 8; num_modes arg = 9; maximum number of binding modes to generate energy_range arg =3; maximum energy difference between the best binding mode and the worst one displayed (kcal/mol); center_x = 19.1732; center_y = 41.6246; center_z = 0.0403; size_x = 37.819451558; size_y = 37.9149091911; size_z = 58.311979351 (Perryman *et al.*, 2014).

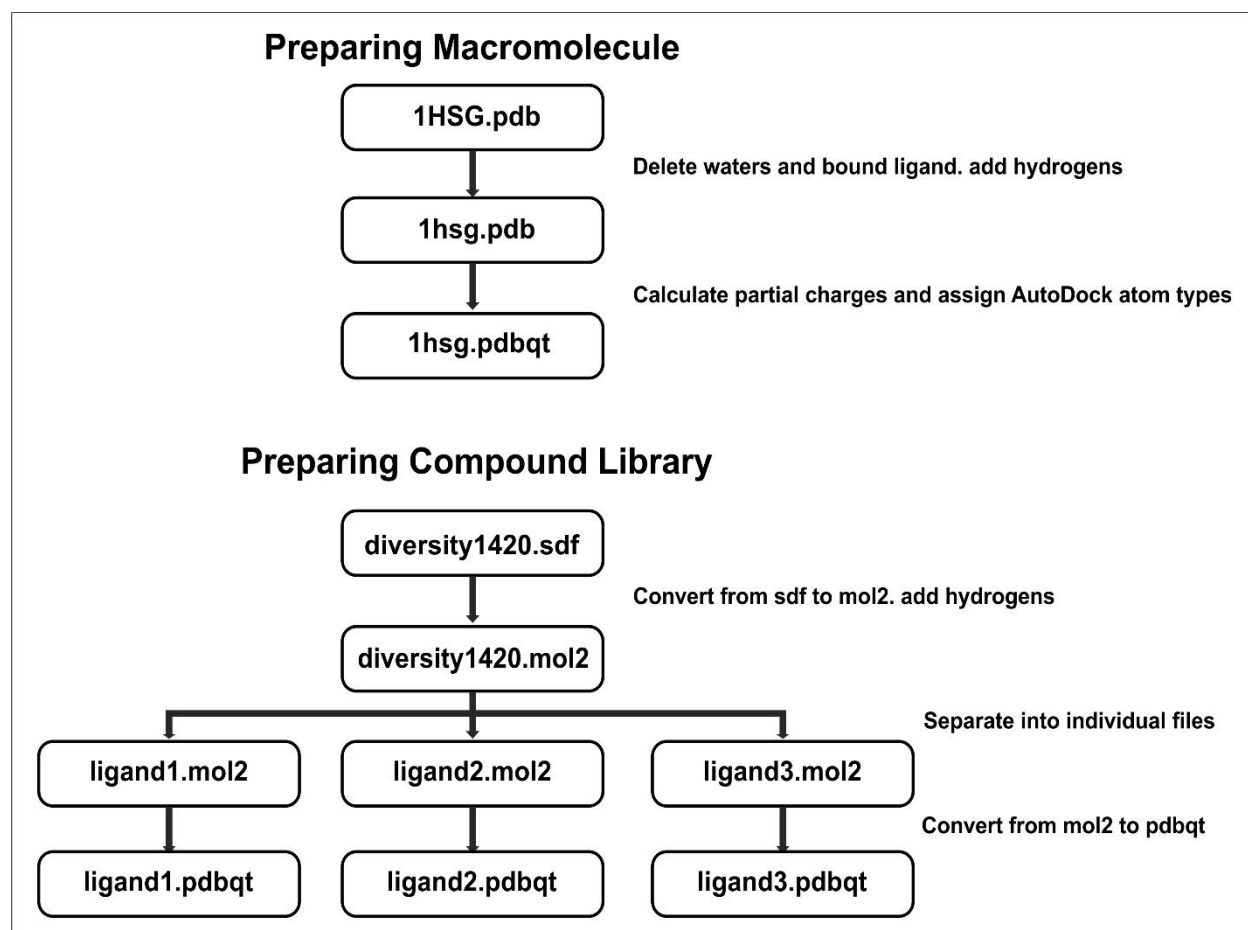


Figure 3. 2: A flowchart representing the processes preparing a macromolecule

As described above, the exhaustiveness parameter is the number of runs that the individual runs were executed in parallel. This exhaustiveness would limit the parallelism whereby each run was able to produce several results.

These results will be merged, clustered and sorted out automatically by the system so that they produce the results. The coordinates X, Y and Z and the sizes X, Y and Z are the search spaces where the docking was set to take place. AutoDock Vina was also used to prepare the protein, water molecules were deleted and the polar hydrogens were added together with the Kolhman's Charge (Rani *et al.*, 2020). The receptor was saved in a pdbqt format in the docking folder. Moreover, in situations where PyMol was used, the format of the ligand was exported from sdf to

pdb format. Using the command prompt, docking between ligand and protein was executed (Rauf *et al.*, 2015), (Butt *et al.*, 2020), .

Chimera

In this study, UCSF Chimera was one of the visualizing software's. It was used to analyse molecular structure and their related data. This software was downloaded following the link: "*C:\Program Files\Chimera\bin\chimera.exe*" and by clicking the new copy chimera icon. Structures were visualized by launching a session for the protein-ligand interaction features by using the preset button, which included visualization through ribbon, all atoms and hydrophobicity surface (Urooj *et al.*) (Noroozi-Aghideh *et al.*, 2020).

Biovia discovery studio visualizer

Molecular docking was performed on the BIOVIA Discovery Studio 2021. v21.1.0.20298. Software (Avram *et al.*, 2021). The protein was downloaded from the PDB and opened in BIOVIA. ZINC database inhibitors were used to dock ZINC compounds with 1rv7 and in addition, phytochemicals were also downloaded from PubChem database and docked with the protein. Many operations were performed in the discovery studio, for example the small molecule tap, receptor-ligand interaction. The receptor size was measured based on the size of the protein which is (center_x = 19.1732; center_y = 41.6246; center_z = 0.0403; size_x = 37.819451558; size_y = 37.9149091911; size_z = 58.311979351), in order to perform a blind docking. Just like in the AutoDock Vina, the ligands and the protein were prepared before docking. However, the ligand and the protein can be prepared automatically using LibDock options. The residues involved in binding were analysed and the 2D structures were obtained.

3.1.6 Lead compound identification and SwissADME applications

The in silico software tools that was used to predict the ADMET properties was the SwissADME <http://www.swissadme.ch>. This software assisted in determining the safety of the lead inhibitors, which are the candidate drugs, before clinical trial. Different parameters that include the passive

gastro-intestinal (HIA), the absorption and blood-brain barrier (BBB), Log S, MW and Log P value, shown by ($\log P_{o/w}$), which is the partition coefficient between n-octanol and water, were analysed. Log P is the classical descriptor for *Lipophilicity* and is determined by the intermolecular relationships between an organic substance and solvent (Mvondo *et al.*). Other physicochemical properties that were considered included the size, flexibility, polarity, saturation and ac philicity. The descriptors should follow the following criterion: a pink area that represents the optimal range, for each property., *Size*: MW between 150 and 500 g/mol; *flexibility*, no more than rotatable bonds; *Solubility*, log S not higher than 6; *Saturation*, Fraction of carbons in the sp³ hybridization not less than 0.25; *Lipophilicity*: XLOGP3 between - 0.7 and + 5.0 and *Polarity*: TPSA between 20 and 130 Å² (Parlak *et al.*, 2021) .

Lipophilicity is the main physico-chemical determinant that influences the permeability and bioavailability.

Log P calculation can be predicted using the following logarithm equations (Jacek Kujawski, 2012):

The properties of *n*- octanol that are used in the estimation processes of Log P, is described as being similar to those of lipid bilayer membrane. In addition, looking into how chemicals are distributed into n-octanol, simulates the dependency of lipophilicity on hydrogen bonding ability of the analyte with the solvent used. This technique is based on the following equation

(T. Fujita, 1964):

$$a) \pi = \log P_X - \log P_H, \quad (\text{Equation 1})$$

Where P_x is the partial coefficient between *n*- octanol and water.

And P_H represent a parent component.

Log P determination will be given by the following equation (A.J. Leo, 1993):

$$b) \log P_{\text{methyl group}} = \log P_{\text{toluene}} - \log P_{\text{benzene}} \quad (\text{Equation 2})$$

This determination is based on the determination factor for methyl group.

$$c) \log P = \sum a_n f_n + \sum b_m F_m \quad (\text{Equation 3})$$

Equation 3 represents Log P of molecules (R. Rekker, 1997), (El-Hassab *et al.*, 2020).

$$d) b_{ij} = a_i S_i a_j S_j T_{ij} R_{ij} + r_{ij}, \quad (\text{Equation 4})$$

All the features of hydrophobicity were put into account and they include: i) the calculated hydrophobic atomic constant for each atom (for macromolecule or small molecule); ii) the definition of secondary and tertiary features provided by hydrophobic interaction and iii) the hydrophobic and polar nature of the surrounding receptor (G.E. Kellogg, 1992).

These features can be calculated by the following equation, as shown above, on equation 4:

$$e) \sum \sum b_{ij} = \sum \sum a_i S_i a_j S_j T_{ij} R_{ij} + r_{ij}, \quad (\text{Equation 5})$$

$b_{ij} > 0$ identifies favourable interactions, while $b_{ij} < 0$ the unfavourable ones.

Equation 5 shows an equation that represents the whole interaction between two molecules. In this study, it will be used to predict the protein and ligand interaction. In some instances, equation 6 can represent the protein and DNA interactions (G.E. Kellogg, 1992).

$$f) A_a/A_s = C_a \sin(\Theta_a)/C_s \sin(\Theta_s) \quad (\text{Equation 6})$$

The drug must also have some specific environmental and behavioural traits once administered. These include the water solubility, drug-likeness and pharmacokinetics (H. Mo, 2010).

Water solubility

Water soluble molecules is one other important aspect in the development of drug activities (Mvondo *et al.*) (Atallah *et al.*, 2013). The formulation can influence absorption. SILICOS-IT

(<http://silicos-it.be.s3-website-eu-west-1.amazonaws.com/software/filter-it/1.0.2/filter-it.html>)

(Yadav and Mohite, 2020), a SwissADME solubility predictor was also employed in determining the solubility of a discovered drug (Savjani *et al.*, 2012). This is very essential since, there are drugs suitable for adult usage only. These drugs should be highly soluble in water and able to deliver enough active pharmaceutical dosage, using a small volume of that soluble. Another topological method that was used was to employ the solubility of the drug was to apply the ESOL model, whereas the Ali *et al.* was used to generate similarities between the two methods, which is the difference in the seminal general solubility equation. Both of these methods avoid the melting point parameter (Alam *et al.*, 2012).

Drug-likeness

The filtered compounds were also predicted in an attempt to see if the molecule has the capacity to become an oral drug with respect to bioavailability, through the structure of the known drugs, preferably those drugs that are in use and have minimum side effects. In this study, the drug likeness search began when the co-crystallized, complex structure of 1rv7 was used to create a pharmacophore in order to mimic the existing drug (Brenk *et al.*, 2008). The Abbot Bioavailability score was used in this case to predict the probability of a compound to have at least 10% oral bioavailability measurable Caco-2 permeability (Silva *et al.*, 2020).

Pharmacokinetics

The skin permeability coefficient (K_p) is a multiple linear regression model. This assessment was made by using the SwissADME application. This is based on the following predictions, the more the Log K_p , the less the skin permeant is the molecule (Zhao *et al.*, 2017) (Mishra and Dahima, 2019).

3.2 MOLECULAR DOCKING

Molecular docking is the more favourable position between a protein and a ligand. This comes as a result of finding treatment for diseases relating to plants, animals and human beings. Numerous softwares have been designed over the years in order to simplify and at the same time have more chemical parameters between the ligand and the protein. The following equation shows how the system in AutoDock 4.4, calculates the protein-ligand binding free energy ΔG (Guanglin, 2016):

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + (V_{bound}^{P-L} - V_{unbound}^{P-L}) + \Delta S_{conf}, \quad (\text{Equation 7})$$

where V : stands for the potential energy, P the protein, L represent the ligand and ΔS_{conf} is the entropic changes upon ligand binding.

The hydrogen-bond, electrostatics, van der Waals and desolvation energies are calculated using the following equation (Guanglin, 2016):

$$V = W_{vdW} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + W_{hbond} \sum_{i,j} E.T \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{elec} \sum_{ij} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} W_{sol} \sum_{ij} (S_i V_j + S_j V_i) \exp\left(-\frac{r_{ij}^2}{2\delta^2}\right) \quad (\text{Equation 8})$$

where other parameters includes the weights (W_x), X , representing *vdw*, *hbond*, *elec*, and *sol*.

3.3 MOLECULAR DYNAMICS SIMULATION

In molecular modelling, the interaction of particles at a specified period of time in an attempt to study their movements is defined as molecular dynamics. This particles can be differentiated by the ability to produce a trajectory, giving the insight into the dynamic behaviour of the system. (Guanglin, 2016)

3.3.1 Force field

The potential energy of an atomic system is modelled by the force field. Such force field that can be used to simulate biomolecular systems, such force fields includes CHARMM, AMBER, OPLS, GROMOS37, CHARMM and AMBER. The CHARMM force field uses the following potential energy V (Guanglin, 2016):

$$V = \sum_{bonds} k_b (b - b_o)^2 + \sum_{angel} k_\delta (\delta - \delta_o)^2 + \sum_{dihedrals} k_\phi [1 + \cos(n\phi - \delta)] \\ + \sum_{impropers} k_w (w - w_o)^2 + \sum_{urey-Bradley} k_{UB} (r_{1,3} - r_{1,3})^2 + \sum_{\phi, \psi} V_{cmap} + \sum_{nonbonded} \epsilon_{ij} [\\ (\frac{R_{min.if}}{r_{if}})^{12} - 2(\frac{R_{min.if}}{r_{if}})^6] + \sum_{nonbonded} \frac{q_i q_j}{4\pi\epsilon_o r_{if}} \quad (Equation 8)$$

whereas the AMBER force field applies the following potential energy V :

$$V = \sum_{bonds} k_b (b - b_o)^2 + \sum_{angel} k_\delta (\delta - \delta_o)^2 + \sum_{dihedrals} \frac{1}{2} V_n [1 + \cos(n\phi - \\ \delta)] + \sum_{impropers} k_w (w - w_o)^2 + \sum_{nonbonded} \epsilon_{ij} [(\frac{R_{min.if}}{r_{if}})^{12} - 2(\frac{R_{min.if}}{r_{if}})^6] \\ + \sum_{nonbonded} \frac{q_i q_j}{4\pi\epsilon_o r_{if}} \quad (Equation 9)$$

3.3.2 Energy minimization

Energy minimization is used to remove bad contacts of atoms, before submitting it to molecular dynamics simulation. Three methods namely, steepest descent, conjugate gradient and the L-BFGS are the methods that can be used for energy minimization.

3.3.3 Integration Algorithm

The Leap-frog algorithm is the commonly used algorithm. Moreover, updates the position r and v by (Guanglin, 2016):

$$v(t + \frac{1}{2} \Delta t) = v(t - \frac{1}{2} \Delta t) + \frac{\Delta t}{m} f(t), \quad (\text{Equation 10})$$

$$r(t + \Delta t) = r(t) + \Delta t v(t + \frac{1}{2} \Delta t) \quad (\text{Equation 11})$$

3.3.4 Temperature coupling

The temperature deviation (by the Berendsen algorithm) from the reference value T_0 is slowly corrected according to (Guanglin, 2016):

$$\frac{dT}{dt} = \frac{T_0 - T}{t} \quad (\text{Equation 12})$$

where T is coupling parameter.

3.3.5 Pressure coupling

A “pressure bath” can be used to control the pressure of the system (Guanglin, 2016).

3.3.6 Periodic Boundary Conditions (PBC)

Periodic boundary conditions (PBC) is used in molecular dynamic simulation to minimise the edge effects of a finite system (Guanglin, 2016).

3.3.7 Constraint algorithms

In many molecular dynamic codes, the SHAKE methods can be used to implement constraints. (Desmond 4.2 User Manual, 2015), (Saurabh *et al.*, 2020 ; Guanglin, 2016)

3.4 SIMULATION QUALITY ANALYSIS

The total energy, pressure, volume, temperature and the potential energy are displayed on the simulation quality analysis, which display a simulation summary. This analysis includes the slope of a linear fit to the values of the property as function of time, the standard deviation and the average value. Simulation of protein 1rv7 and the entire compound used in this study were performed on a 1 CPU, at temperature 300.0 K, and time 200ns with a protein having 198 residues, protein chain [A, B], Res. in chains [99,99], atoms 3110, heavy atoms 1510 and charge +8 (Desmond 4.2 User Manual, 2015).

3.4.1 Protein-Ligand Interaction Analysis

The Simulation Interaction Diagram Panel is opened using the following pathway:

Choose Application→ Desmond → Simulation Interaction Diagram.

This interaction diagram creates and graphically displays all the information needed about the behaviour, association and interaction of both the protein and the ligand. The information obtained from interactions is used to validate a ligand pose (Xu *et al.*, 2021).

Protein-Ligand RMSD

An atom set for the protein RMSD value plots, was obtained from the C- α . In this section, the RMSD should fluctuate around some thermal average of 1 – 3 Å. This indicates that the system has reached equilibrium. However, RMSD that is decreasing or increasing implies that the system has not reached equilibrium, and a further analysis might be recommended (Desmond 4.2 User Manual, 2015).

Protein RMSF

Each residue in the protein chain is determined by the P-RMSF, which is the time-average fluctuations of the square deviation. The deviation is prescribed for a set of residue atoms over the

entire simulation of 200ns. In the RMSF plots, the peaks indicates the areas of the protein that fluctuate the most (Achiri *et al.*, 2021) .

Ligand RMSF

The L-RMSF displays the root mean-square fluctuation for each atom in the ligand. RMSF can be represented in a given alignment, in the basis of the binding site. The tom index tap is used in the plots and is shown on the figure/s obtained. This display is facilitated by selecting the Atom number (Hollingsworth and Dror, 2018).

Protein Secondary Structure Elements

These plots display the percentage of time each residue contributes to the three secondary structure elements, namely strand (Blue), loop (white) and helix (maroon) (Desmond 4.2 User Manual, 2015).

Protein-Ligand contacts

The protein-ligand tab shows information pertaining to the protein-ligand contacts. The contacts were displayed by hydrogen bonds, which also shows the breakdown of different contacts. These hydrogen bonds are defined using the following criterion, the DH...A-X atom arrangement: a D-A distance less than 2.5 Å, a D-H-A angle greater than 120° an H-A-X angle greater than 90° (Desmond 4.2 User Manual, 2015).

REFERENCE LIST

A.J. LEO 1993. Calculating log P_{oct} from structures. *Chem. Rev*, 93, 1281-1306.

ACHIRI, R., FOUZIA, M., BENOMARI, F. Z., DJABOU, N., BOUFELDJA, T., MUSELLI, A. & DIB, M. E. A. 2021. Chemical composition/pharmacophore modelling-based, virtual screening, molecular docking and dynamic simulation studies for the discovery of novel superoxide dismutase (SODs) of bioactive molecules from aerial parts of *Inula Montana* as antioxydant's agents. *Journal of Biomolecular Structure and Dynamics*, 1-22.

AKHTAR, A., HUSSAIN, W. & RASOOL, N. 2019. Probing the pharmacological binding properties, and reactivity of selective phytochemicals as potential HIV-1 protease inhibitors. *Universitas Scientiarum*, 24, 441-464.

ALAM, M. A., ALI, R., AL-JENOABI, F. I. & AL-MOHIZEA, A. M. 2012. Solid dispersions: a strategy for poorly aqueous soluble drugs and technology updates. *Expert opinion on drug delivery*, 9, 1419-1440.

APPIAH-KUBI, P., IWUCHUKWU, E. A. & SOLIMAN, M. E. 2021. Structure-based identification of novel scaffolds as potential HIV-1 entry inhibitors involving CCR5. *Journal of Biomolecular Structure and Dynamics*, 1-12.

ATALLAH, P., WAGENER, K. B. & SCHULZ, M. D. 2013. ADMET: The future revealed. *Macromolecules*, 46, 4735-4741.

AVRAM, S., STAN, M. S., UDREA, A. M., BUIU, C., BOBOC, A. A. & MERNEA, M. 2021. 3D-ALMOND-QSAR Models to Predict the Antidepressant Effect of Some Natural Compounds. *Pharmaceutics*, 13, 1449.

BRADLEY C LOGSDON , J. F. V., PHILIP MARTIN, GHEORGHE PROTEASA, JAY I KOEPKE, STANLEY R TERLECKY, ZDZISLAW WAWRZAK, MARK A WINTERS, THOMAS C MERIGAN, LADISLAU C KOVARI, 2004. Crystal structures of a multidrug-

resistant human immunodeficiency virus type 1 protease reveal an expanded active-site cavity. 6, 3123-3132.

BRENK, R., SCHIPANI, A., JAMES, D., KRASOWSKI, A., GILBERT, I. H., FREARSON, J. & WYATT, P. G. 2008. Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. *ChemMedChem*, 3, 435.

BURLEY, S. K., BERMAN, H. M., BHIKADIYA, C., BI, C., CHEN, L., DI COSTANZO, L., CHRISTIE, C., DALENBERG, K., DUARTE, J. M. & DUTTA, S. 2019. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic acids research*, 47, D464-D474.

BUTT, S. S., BADSHAH, Y., SHABBIR, M. & RAFIQ, M. 2020. Molecular Docking Using Chimera and Autodock Vina Software for Nonbioinformaticians. *JMIR Bioinformatics and Biotechnology*, 1, e14232.

COPOIU, L., TORRES, P. H., ASCHER, D. B., BLUNDELL, T. L. & MALHOTRA, S. 2020. ProCarbDB: a database of carbohydrate-binding proteins. *Nucleic acids research*, 48, D368-D375.

CRISAN, L. & BORA, A. NEW TRENDS AND STRATEGIES IN THE CHEMISTRY OF ADVANCED MATERIALS WITH RELEVANCE IN BIOLOGICAL SYSTEMS, TECHNIQUE AND ENVIRONMENTAL PROTECTION 13th edition, 07-08 October, 2021 Timisoara, ROMANIA.

DESMOND 4.2 USER MANUAL 2015. Desmond User Manual Schrödinger, LLC.

DOYTCHINOVA, I., ATANASOVA, M., VALKOVA, I., STAVRAKOV, G., PHILIPOVA, I., ZHIVKOVA, Z., ZHELEVA-DIMITROVA, D., KONSTANTINOV, S. & DIMITROV, I. 2018. Novel hits for acetylcholinesterase inhibition derived by docking-based screening on ZINC database. *Journal of enzyme inhibition and medicinal chemistry*, 33, 768-776.

EGBUNA, C., PATRICK-IWUANYANWU, K. C., ONYEIKE, E. N., KHAN, J. & ALSHEHRI, B. 2021. FMS-like tyrosine kinase-3 (FLT3) inhibitors with better binding affinity and ADMET properties than sorafenib and gilteritinib against acute myeloid leukemia: in silico studies. *Journal of Biomolecular Structure and Dynamics*, 1-12.

EL-HASSAB, M. A. E.-M., EL-BASTAWISSY, E. E. & EL-MOSELHY, T. F. 2020. Identification of potential inhibitors for HCV NS5b of genotype 4a by combining dynamic simulation, protein–ligand interaction fingerprint, 3D pharmacophore, docking and 3D QSAR. *Journal of Biomolecular Structure and Dynamics*, 38, 4521-4535.

ELGHONEIMY, L. K., ISMAIL, M. I., BOECKLER, F. M., AZZAZY, H. M. & IBRAHIM, T. M. 2021. Facilitating SARS CoV-2 RNA-Dependent RNA polymerase (RdRp) drug discovery by the aid of HCV NS5B palm subdomain binders: In silico approaches and benchmarking. *Computers in biology and medicine*, 134, 104468.

ELSEGINY, S. A. 2021. Virtual screening and structure-based 3D pharmacophore approach to identify small-molecule inhibitors of SARS-CoV-2 Mpro. *Journal of Biomolecular Structure and Dynamics*, 1-17.

G.E. KELLOGG, G. S. J., D.J. ABRAHAM, 1992. New tools for modeling and understanding hydrophobicity and hydrophobic interactions. *Med. Chem. Res*, 1.

H. MO, K. M. B., D.A. COLBY, 2010. practical deuteriumfree NMR method for the rapid determination of 1-octanol/water partition coefficient of pharmaceutical agents. *Bioorg. Med. Chem. Lett.*, 20, 6712 - 6715.

HOLLINGSWORTH, S. A. & DROR, R. O. 2018. Molecular dynamics simulation for all. *Neuron*, 99, 1129-1143.

IRWIN, J. J. & SHOICHET, B. K. 2005. ZINC– a free database of commercially available compounds for virtual screening. *Journal of chemical information and modeling*, 45, 177-182.

JACEK KUJAWSKI, H. P., ANNA MYKA, BEATA DRABIŃSKA, MAREK K. BERNARD, 2012. The log P Parameter as a Molecular Descriptor in the Computer-aided Drug Design – an Overview. *Department of Organic Chemistry, Faculty of Pharmacy, Poznan University of Medical Sciences*, 18, 81-88.

JU, J., LI, X., KUMAR, S., JOCKUSCH, S., CHIEN, M., TAO, C., MOROZOVA, I., KALACHIKOV, S., KIRCHDOERFER, R. N. & RUSSO, J. J. 2020. Nucleotide analogues as inhibitors of SARS-CoV polymerase. *Pharmacology research & perspectives*, 8, e00674.

KATIIYAR, K., SRIVASTAVA, R. K. & NATH, R. 2021. Identification of novel anti-cryptosporidial inhibitors through a combined approach of pharmacophore modeling, virtual screening, and molecular docking. *Informatics in Medicine Unlocked*, 100583.

KHAERUNNISA, S., KURNIAWAN, H., AWALUDDIN, R., SUHARTATI, S. & SOETJIPTO, S. 2020. Potential inhibitor of COVID-19 main protease (Mpro) from several medicinal plant compounds by molecular docking study. *Preprints*, 2020, 2020030226.

KHARISMA, V. D., WIDYANANDA, M. H., ANSORI, A. N. M., NEGE, A., NAW, S. & NUGRAHA, A. 2021. Tea catechin as antiviral agent via apoptosis agonist and triple inhibitor mechanism against HIV-1 infection: A bioinformatics approach. *Journal of Pharmacy & Pharmacognosy Research*, 9, 435-445.

KNELLER, D. W., AGNISWAMY, J., HARRISON, R. W. & WEBER, I. T. 2020. Highly drug-resistant HIV-1 protease reveals decreased intra-subunit interactions due to clusters of mutations. *The FEBS journal*, 287, 3235-3254.

LIPINSKI, C. A., LOMBARDO, F., DOMINY, B. W. & FEENEY, P. J. 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, 23, 3-25.

MISHRA, S. & DAHIMA, R. 2019. In vitro ADME studies of TUG-891, a GPR-120 inhibitor using SWISS ADME predictor. *Journal of Drug Delivery and Therapeutics*, 9, 366-369.

MUHAMMED, M. T. & ESIN, A.-Y. 2021. Pharmacophore Modeling in Drug Discovery: Methodology and Current Status. *Journal of the Turkish Chemical Society Section A: Chemistry*, 8, 749-762.

MVONDO, J. G. M., MATONDO, A., MAWETE, D. T., BAMBI, S.-M. N., MBALA, B. M. & LOHOHOLA, P. O. In Silico ADME/T Properties of Quinine Derivatives using SwissADME and pkCSM Webservers.

NAJM, A. A. K., AZFARALARIFF, A., DYARI, H. R. E., OTHMAN, B. A., SHAHID, M., KHALILI, N., LAW, D., SYED ALWI, S. S. & FAZRY, S. 2021. Anti-breast cancer synthetic peptides derived from the Anabas testudineus skin mucus fractions. *Scientific Reports*, 11, 1-20.

NOROOZI-AGHIDEH, A., SAFAVI, A., GHODOUSI, E. S., RAJAEINEJAD, M. & TAGHAVI MOGHADDAM, P. 2020. Homology modeling and molecular docking of hABCC3/MRP3 with chemotherapeutic agents in acute leukemia. *Jundishapur Journal of Natural Pharmaceutical Products*, 15.

PANWAR, U. & SINGH, S. K. 2018. Structure-based virtual screening toward the discovery of novel inhibitors for impeding the protein-protein interaction between HIV-1 integrase and human lens epithelium-derived growth factor (LEDGF/p75). *Journal of Biomolecular Structure and Dynamics*, 36, 3199-3217.

PARLAK, C., ALVER, Ö., OUMA, C. N. M., RHYMAN, L. & RAMASAMI, P. 2021. Interaction between favipiravir and hydroxychloroquine and their combined drug assessment: in silico investigations. *Chemical Papers*, 1-8.

PERRYMAN, A. L., SANTIAGO, D. N., FORLI, S., SANTOS-MARTINS, D. & OLSON, A. J. 2014. Virtual screening with AutoDock Vina and the common pharmacophore engine of a low diversity library of fragments and hits against the three allosteric sites of HIV integrase: participation in the SAMPL4 protein–ligand binding challenge. *Journal of computer-aided molecular design*, 28, 429-441.

QUAZI, S., GAVAS, S., MALIK, J. A., SUMAN, K. S. & HAIDER, Z. 2021. In-silico pharmacophore and Molecular Docking based drug discovery against Marburg Virus's Viral Protein 35; A potent of MAVD. *bioRxiv*.

REKKER 1997. The Hydrophobic Fragmental Constant. *Elsevier, Amsterdam*.

RANI, N., KHARE, N. & JHA, A. K. 2020. Molecular Docking Study of Dihydrohelenalin against DNMT1 to Treat Oral Cancer.

RAUF, M. A., ZUBAIR, S. & AZHAR, A. 2015. Ligand docking and binding site analysis with pymol and autodock/vina. *International Journal of Basic and Applied Sciences*, 4, 168.

SAHIN, K. 2021. Investigation of novel indole-based HIV-1 protease inhibitors using virtual screening and text mining. *Journal of Biomolecular Structure and Dynamics*, 39, 3638-3648.

SANDERS, M. P., ROUMEN, L., VAN DER HORST, E., LANE, J. R., VISCHER, H. F., VAN OFFENBEEK, J., DE VRIES, H., VERHOEVEN, S., CHOW, K. Y. & VERKAAR, F. 2012. A prospective cross-screening study on G-protein-coupled receptors: lessons learned in virtual compound library design. *Journal of medicinal chemistry*, 55, 5311-5325.

SAURABH, S., SIVAKUMAR, P. M., PERUMAL, V., KHOSRAVI, A., SUGUMARAN, A. & PRABHAWATHI, V. 2020. Molecular Dynamics Simulations in Drug Discovery and Drug Delivery. *Integrative Nanomedicine for New Therapies*, 275-301.

SAVJANI, K. T., GAJJAR, A. K. & SAVJANI, J. K. 2012. Drug solubility: importance and enhancement techniques. *International Scholarly Research Notices*, 2012.

SHARMA, S., SHARMA, A. & GUPTA, U. 2021. Molecular Docking Studies on the Anti-Fungal Activity of Allium Sativum (Garlic) Against Mucormycosis (Black Fungus) by BIOVIA Discovery Studio Visualizer 21.1. 0.0.

SHARMA, S., SRIVASTAVA, S., SHRIVASTAVA, A., MALIK, R., ALMALKI, F., SAIFULLAH, K., ALAM, M. M., SHAQUIZZAMAN, M., ALI, S. & AKHTER, M. 2020. Mining of potential dipeptidyl peptidase-IV inhibitors as anti-diabetic agents using integrated in silico approaches. *Journal of Biomolecular Structure and Dynamics*, 38, 5349-5361.

SHIRI, F., PIRHADI, S. & RAHMANI, A. 2018. Identification of new potential HIV-1 reverse transcriptase inhibitors by QSAR modeling and structure-based virtual screening. *Journal of Receptors and Signal Transduction*, 38, 37-47.

SHOICHET, B. K. 2004. Virtual screening of chemical libraries. *Nature*, 432, 862-865.

SILLAPACHAIYAPORN, C., RANGSINTH, P., NILKHET, S., MOUNGKOTE, N. & CHUCHAWANKUL, S. 2021. HIV-1 Protease and Reverse Transcriptase Inhibitory Activities of *Curcuma aeruginosa* Roxb. Rhizome Extracts and the Phytochemical Profile Analysis: In Vitro and In Silico Screening. *Pharmaceuticals*, 14, 1115.

SILVA, B., SILVA, R., FERNANDES, C., DE PINHO, P. G. & REMIÃO, F. 2020. Enantioselectivity on the absorption of methylone and pentedrone using Caco-2 cell line: Development and validation of an UHPLC method for cathinones quantification. *Toxicology and applied pharmacology*, 395, 114970.

T. FUJITA, J. I., C.J. HANSCH, 1964. A New Substituent Constant, π , Derived from Partition Coefficients. *J. Am. Chem. Soci.*, 86, 5175-5180.

UROOJ, S., DHARIWAL, A., SINGH, V. & ALROWAIS, F. In Silico Antituberculosis Drug Designing using UCSF Chimera.

VALENTINE, D., ZASLAVSKY, I., RICHARD, S., MEIER, O., HUDMAN, G., PEUCKER-EHRENBRINK, B. & STOCKS, K. 2021. EarthCube Data Discovery Studio: A gateway into geoscience data discovery and exploration with Jupyter notebooks. *Concurrency and Computation: Practice and Experience*, 33, e6086.

XU, L., ZHAO, L., CHE, J., ZHANG, Q., CAO, R. & LI, X. 2021. Identification of novel influenza polymerase PB2 inhibitors using virtual screening approach and molecular dynamics simulation analysis of active compounds. *Bioorganic & Medicinal Chemistry*, 116515.

YADAV, A. & MOHITE, S. 2020. ADME analysis of phytochemical constituents of Psidium guajava. *Asian J. Res. Chem.* 2020; 13 (5): 373, 375.

ZHANG, X., SUN, L., MEUSER, M. E., ZALLOUM, W. A., XU, S., HUANG, T., CHERUKUPALLI, S., JIANG, X., DING, X. & TAO, Y. 2021. Design, synthesis, and mechanism study of dimerized phenylalanine derivatives as novel HIV-1 capsid inhibitors. *European Journal of Medicinal Chemistry*, 226, 113848.

ZHAO, H., LIU, C., QUAN, P., WAN, X., SHEN, M. & FANG, L. 2017. Mechanism study on ion-pair complexes controlling skin permeability: effect of ion-pair dissociation in the viable epidermis on transdermal permeation of bisoprolol. *International journal of pharmaceutics*, 532, 29-36.

ZHONG, H.-J., MA, V. P.-Y., CHENG, Z., CHAN, D. S.-H., HE, H.-Z., LEUNG, K.-H., MA, D.-L. & LEUNG, C.-H. 2012. Discovery of a natural product inhibitor targeting protein neddylation by structure-based virtual screening. *Biochimie*, 94, 2457-2460.

ZUHRUFA, Z. & JULIANTO, T. S. 2021. Molecular Interaction Analysis of COX-2 Against Aryl Amino Alcohol Derivatives from Isoeugenol as Anti Breast Cancer using Molecular Docking. *Bulletin of Chemical Reaction Engineering & Catalysis*, 16, 581-587.

CHAPTER 4: IDENTIFICATION OF ZINC INHIBITORS FOR HIV-1 PROTEASE THROUGH STRUCTURE-BASED PHARMACOPHORE, VIRTUAL SCREENING, ADMET AND MOLECULAR DYNAMICS SIMULATION

ABSTRACT

Treating HIV is made difficult by the evolution of antimicrobial resistant mutants that occur in the life cycle of this virus. This has led to the evolution of numerous inhibition mechanisms that intends to challenge the replication of this virus. One such approach is the application of molecular modelling in drug discovery. In this study, the application of computer aided drug design was employed in finding an inhibitor that can contribute in finding a suitable drug candidate. Firstly, the ZINC database was screened using the Lipinski's rule of five, and compounds having similar biological features were classified and grouped together into a compound library. Secondly, the co-crystallized structure of 1rv7 wild type of HIV-1 protease was used to construct a pharmacophore model that was used to screen further, the ZINC database Ligand stored in the compound library. The structure-based pharmacophore was created by mimicking the three main active features of the co-crystallized structure from the Protein Data Bank, namely, 82 and 84, I50 and I150 residues. Thirdly, a total of 26 of the compounds that had less than 0,5 RMSD was identified and selected. Fourthly, based on the best docking ligands, ten of these compounds having high binding affinity were selected. These compounds are ZINC_001456687980 (-8.0 K_D), ZINC_001445792073 (-7.8 K_D) and ZINC_001461099137 (-7.4 K_D), successively. However, ZINC_001461099137 was eliminated based on unfavourable donor-donor interaction. The subsequent compound, ZINC_000015276352 was then placed on the top three leading compounds with a binding affinity of (-6.9 K_D). Fifthly, the ADMET property analysis was used on the top ten compounds, and the compounds that satisfied the analysis were, ZINC_001456687980, ZINC_000015276352, ZINC_001359888321 and ZINC_001460445290. The main properties that included the BBB permeation, the gastrointestinal absorption and pan assay interference structure (PAINS) were analysed. ZINC_001456687980 had two violations (mw >350, XLOGP3>3.5),

whiles ZINC_000015276352 had three violations ($mw > 350$, $XLOGP3 > 3.5$, $rotatable > 7$). Lastly, only two top compounds underwent molecular dynamics simulations. Molecular dynamics simulation revealed the formation of numerous bond interactions between ZINC_001456687980 and ZINC_000015276352 together with the target protein.

4.1 INTRODUCTION

Human immunodeficiency virus type 1 (HIV-1) acts as a cause of acquired immunodeficiency syndrome (AIDS) (Jette *et al.*, 2021). Since the beginning of this epidemic, the global public health has incurred serious economic threats, which has led to minimal supply of pharmaceutical care in low-income countries (Michael *et al.*, 2021). UNAIDS reported in 2020 that an estimated 37.7 million people were living with HIV, while there were 1, 5 million infections leading to at least 680 000 deaths related to HIV (UNAIDS, 2020). In the quest to find a cure for this disease, there has been an advent of more than 40 antiretroviral drugs that has targeted various places in the life cycle of this virus. HIV-1 viral particle consists of two copies of single stranded viral RNA and the viral enzymes, namely the integrase, the protease and the reverse transcriptase (Shin *et al.*, 2021) (Haaland *et al.*, 2021). These components of viral particles and enzymes are enveloped by capsid core (Guedán *et al.*, 2021).

In the first stages of viral intrusion into the host cell, the gp120 of the virus begins by attaching itself to the CD4 receptor of the host, on the T-cell membrane. The gp41 from the gp120 is then released into the host cell in order to bind with chemokine co-receptors that include CCR5, CCR4 and CXCR4. This process causes a fusion between the viral envelop and the co-receptors. The viral components will then be released inside the host's cell, this step is followed by other processes that leads to the maturation of a viral species that can repeat the cycle and infect another host cell (Andrianov *et al.*, 2021).

More than 1000 amino acids make up two subunits, namely the p66 and p51, whereby the binding region is found mainly within the p66. The p66 consists of sub domains, namely fingers, thumb, palm, connection and the RNase-H as proposed by Huang *et al.* (Huang *et al.*, 1998) (Huang *et al.*, 2014) (Zheng *et al.*, 2005). This therefore makes up the active sites where binding can occur. The

P66 subunit contains 560 amino acids while the p51 contains 440 amino acids. The functional mechanics of HIV-1 protease involves three main target structures, which are the HIV-1 reverse transcriptase, RNase- H and the integrase. HIV-1 reverse transcriptase and RNase- H are responsible for reverse transcription of the ssRNA into the dsDNA strands and consequently cleaves to the RNA strand from the RNA/DNA hybrid respectively, whereas, on the other hand, the HIV-1 integrase facilitates the integration of viral DNA into the hosts genome (Ippolito *et al.*, 2021).

In 1996, Dr David Ho proposed the High Active Antiretroviral Therapy (HAART), in the fight against HIV. This has yielded great success (Finzi *et al.*, 1997) (Ortiz *et al.*, 1999, Shafer and Vuitton, 1999). However, its success was short-lived by the evolution of cross-resistant strains of HIV severe side effects and in some cases, there were observations of poor patient compliance of pill administration (Murri *et al.*, 2000). This therefore paves a way for the need to find new and effective anti-HIV drug that has minimal side effects on patients and has the ability to improve pharmacokinetic properties.

Over the years, several inhibition approaches have evolved that intend to interrupt the life cycle of HIV-1 protease. This has been brought about by the contribution in the scientific fraternity, and one such contribution was made by Chen and his group who proposed that, the binding pocket of the HIV-1 protease undergoes some changes in the shape and conformation upon certain drug-resistant mutations. They further made an observation that as a result of this mutation there is a reduction in the binding affinity of the inhibitor. This has led to recent advancements about the flaps on the protein and to studies made to understand their morphology and physiological activities (Chen *et al.*, 2005) .

The recent findings performed through coarse-grained molecular dynamics simulation observed a spontaneous opening and reclosing of the flaps. These studies were performed by Tozzini *et al.*, who independently observed the fluctuation between semi-open, closed and fully open conformation at 30 ns and four different conformations between closed, semi-open, open and wide open, respectively. From the early studies of molecular dynamic simulation, which suggested a

closed ligand free conformation, to the recent studies that observed the four conformation of the HIV-1 protease, it was demonstrated that, there is a gap that needs to be filled by understanding the mechanism of the protease (Trylska *et al.*, 2007).

In this study, a computer aided drug design was used to screen and filter compounds from the ZINC database in an attempt to find a suitable drug candidate using pre-existing drugs. This procedure yielded great success in the past, with numerous ARVs being discovered using this approach (Leelananda and Lindert, 2016); (Zhao *et al.*, 2020). The search for a drug with similar features and biological properties to the drugs that are in the market is one the most fundamental tools to be used in the current period of drug discovery (Hussain *et al.*, 2021).

4.2 MATERIALS AND METHODS

Zinc database

The compounds in the ZINC database were screened using the Lipinski's rule of five. A total of 497 891 compounds were retrieved from the database. These compounds satisfied the application of the Lipinski's rule of five.

Compound library

A library of these ZINC compounds was created on the MOE software. All these compounds were screened with the pharmacophore.

Target identification

Through literature, the protein was identified and retrieved from the PDB database. This protein has a code of 1rv7.

Protein preparation

Discovery Studio was used to prepare the downloaded protein, as it was stored in the PDB as a co-crystallized structure. (Adeniji *et al.*, 2020)

Pharmacophore model

A pharmacophore model was developed based on the co-crystal compound with protein by MOE. The query was saved. The features from the co-crystallized structure retrieved from the database were employed in constructing the pharmacophore model that was used for virtual screening (Achiri *et al.*, 2021).

Virtual screening, docking with pharmacophore

The prepared library of the compounds obtained through the rule of five was screened based on a developed pharmacophore query (Tariq *et al.*, 2021).

Molecular docking study

Docking studies between the screened ZINC compounds were performed. In this analysis, a blind docking was executed with a grid box that was generated in X, Y and Z axis as follows: center_x = 19.1732 center_y = 41.6246 center_z = 0.0403 size_x = 37.819451558 size_y = 37.9149091911 size_z = 58.311979351. The binding affinity between the ligand and the protein was obtained after docking and their interaction was observed on discovery studio and PyMOL. The compounds that had the best hits from docking studies were analysed. Their 2D and 3D interaction was evaluated using the discovery studio.

ADMET calculation

For safety evaluation, this ZINC compound was employed using the SwissADME web server. This evaluation involves adsorption, distribution, metabolism, excretion and toxicity (ADMET) characteristic profiling (Daina *et al.*, 2017).

Molecular dynamics simulation

The Desmond package, which is part of Schrödinger LLC, was used to perform molecular dynamics and simulation at 200 nanoseconds (Gopinath and Kathiravan, 2021). This procedure

was performed after molecular docking studies, which predicted the ligand binding position in its static condition. The docking studies predicted the characteristic binding nature of the ligands together with the protein in the physiological environment. After it had been pre-processed, the ligand-protein complex was optimized and minimized using Maestro.

A 0.15 M salt (NaCl) was added in order to mimic the physiological environment which the simulation would take place. In addition, the following variables were ensembled as moles (N) for a complete simulation, temperature (T) and pressure (P). A 1 atm pressure and 300 K temperature were selected. The trajectory was saved at 100ps, with a simulation time of 200ns at the energy of 1,2. The pressure bar was at 1.01325 with following an estimated 25 390 number of atoms. The root mean square deviation (RMSD) and root mean square fluctuation (RMSF) were analysed and the stability of the simulation was evaluated (Xu *et al.*, 2012).

Below is a flowchart prescribed for the computer aided drug design and its methodology:

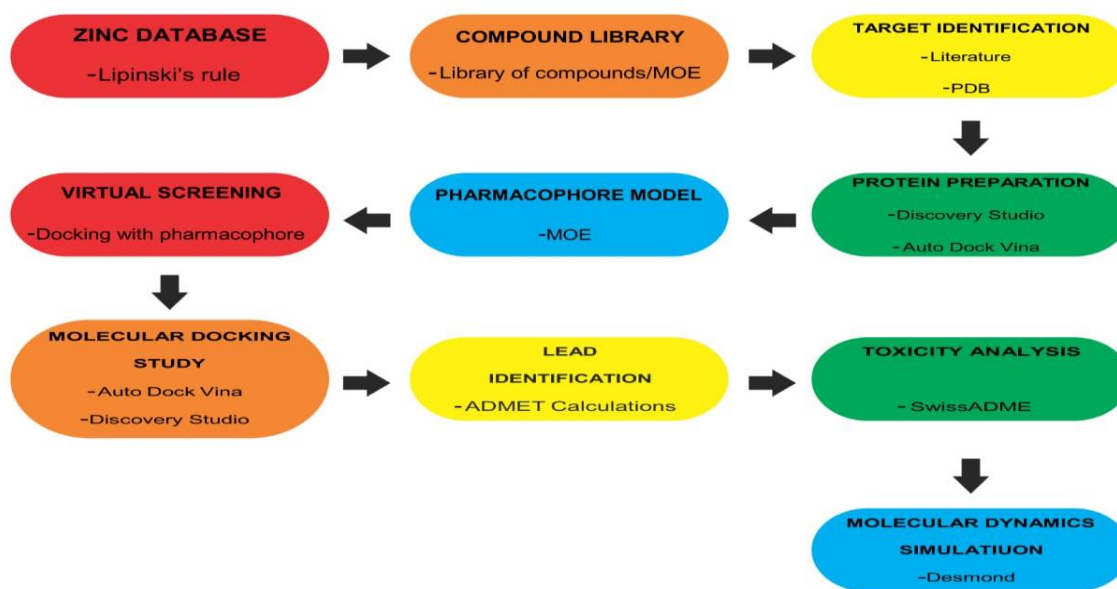


Figure 4. 1: The Flowchart of adopted computer-aided drug design (CADD) methodology

4.3 RESULTS

In Silico Analysis: Based virtual screening

4.3.1 Drug likeness analysis

A proposed lead compound was determined by following the flowchart through a virtual screening process, as shown by Figure 4.2. This approach is used in research, since it filters and incorporates various compounds that has common drug-likeness (Ouyang *et al.*, 2021). Through these processes, it was possible to mimic the co-crystallized structure through structure based pharmacophoric screening and later analyse the ligand-protein complexes through molecular dynamics simulation. Drug analysis follows a computational approach that includes the Lipinski's rule of five, that predicts that absorption is more likely when the compounds display, less than 5 hydrogen bonds, less than 10 hydrogen bonds, a molecular weight (mw) of less than 500 and a calculated LogP less than 5 (Lipinski *et al.*, 1997). In the process, molecular docking was applied in order to analyse the interactions between the ligand and the target protein. Ligand-protein contacts, root mean square deviation and root mean square fluctuation are some of characteristics that were analysed using molecular dynamics simulation, which is a computational approach that capture the behaviour of protein and the ligand (Hollingsworth and Dror, 2018).

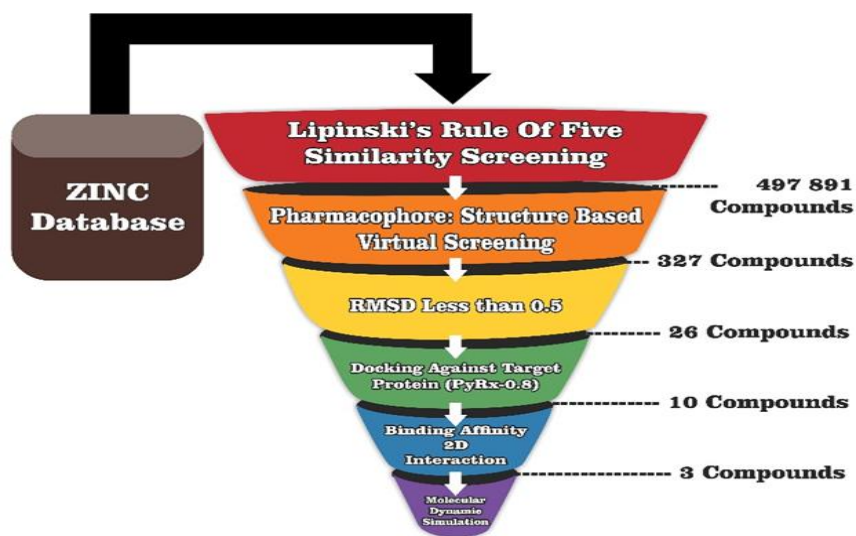


Figure 4. 2: The flowchart highlighting the filtering of ZINC compounds

4.3.2 Virtual screening of HIV-1 protease inhibitors using ZINC databank and AutoDock Vina applying the Lipinski's Rule of Five

Molecules were retrieved from the ZINC database. These ligands were searched based on their likeness in chemical composition. For the effectiveness of this search, the Lipinski's rule of five was employed and a total of 497 891 ligands was found on the ZINC database. The compounds in the ZINC database are a collection of commercially available compounds. These compounds are suitable for screening using the database.

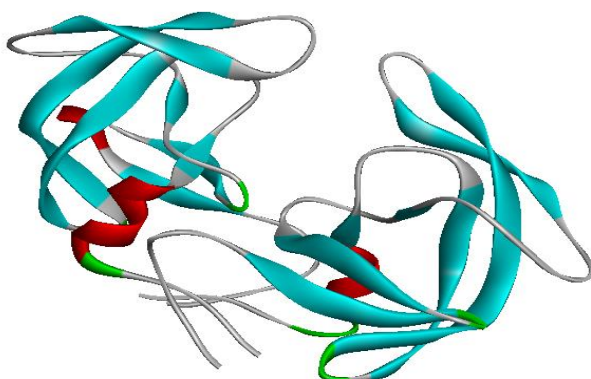


Figure 4. 3: An uncomplexed HIV-1 protease

4.3.3 Development of A pharmacophore model based on the co-crystal compound with protein by molecular operating environment. (*Structure based pharmacophore modelling*):

Pharmacophore technique based on the co-crystallized mimicry

A pharmacophore model was built using a structure-based modelling approach. This model is based on obtaining biological compounds that are available for inhibitor binding of HIV 1 protease using its co-crystallized structure. Three active pharmacological features were identified, namely, (F1: Don and Acc), (F2: Acc) and (F3: Don). This technique of creating the pharmacophore features is determined by the contacts and bonds found on the hot-spot residues. For this

determination, a pharmacophore query editor tool of MOE 2014.09 software was used to facilitate the process (Kobayakawa *et al.*, 2021).

These feature (F1: Don&Acc, F2: Acc & F3: Don) are the active parts of the 1vr7 receptor that has been mimicked as the original (F1: Don & Acc, F2: Acc & F3: Don) mimetic inhibitors that were used to design 1vr7 Co-crystal structure. A library for the 497 891 compounds was created on the MOE 2014.09 software and these compounds were docked with pharmacophore. Based on the pharmacophore-fit, a total of 327 compounds was retrieved.

Table 4. 1: Interaction between the trio hot-spot residues of 1rv7 co-crystal structure and the corresponding residues of the 1rv7 receptor at the binding site (PDB:1rv7)

1rv7 hot-spot residue	Type of interaction	Residue
F1 = Don/Acc	σ carbon-carbon bond	82 and 84
F2 = Acc	Van der Waals	I50
F3 = Don	Hydrogen bonds	I150

Table 4.1 above shows the trio hot-spot residue of 1rv7 co-crystalized structure. The hot-spot includes the F1, F2 and F3 residues which are the main points used to mimic and construct a pharmacophore, that will be used in screening. The F1 residues was suitable to be used as a hotspot since it allows a hybridization through the σ carbon-carbon bond. On the other hand, the Van der Waals forces, allows very large and long overlaps in the molecules, through the F2 residues. The attractive force between a hydrogen atom that is covalently bonded to a very electronegative atom is formed through the F3 residue. The position of F1, F2 and F3 residues are also shown on Table 4.1.

Through modeling studies, Logsdon *et al.* discovered a strange behaviour of a mutant strain HIV-1 MDR 769 that was isolated clinically from a patient who was failing extensive antiretroviral therapy. This outcome was as a result of viremia, a medical condition in which the viruses were present in the bloodstream. The isolated strain of an HIV-1 MDR 769 protease was examined through expression, purification and crystallization and it was discovered that, there was an expansion on the active site of the protease. This expansion was responsible for the decrease in the volume of amino acids leading to the loss of Van der Waals, hydrogen bonds and σ carbon-carbon

bond (Logsdon *et al.*, 2004), thus causing a decrease in the protein-ligand binding affinity. The derived strain of 1rv7 complex was deposited in the Protein Data Bank for further scientific studies and this is the very same co-crystallized structure used in this project to create a pharmacophore model.

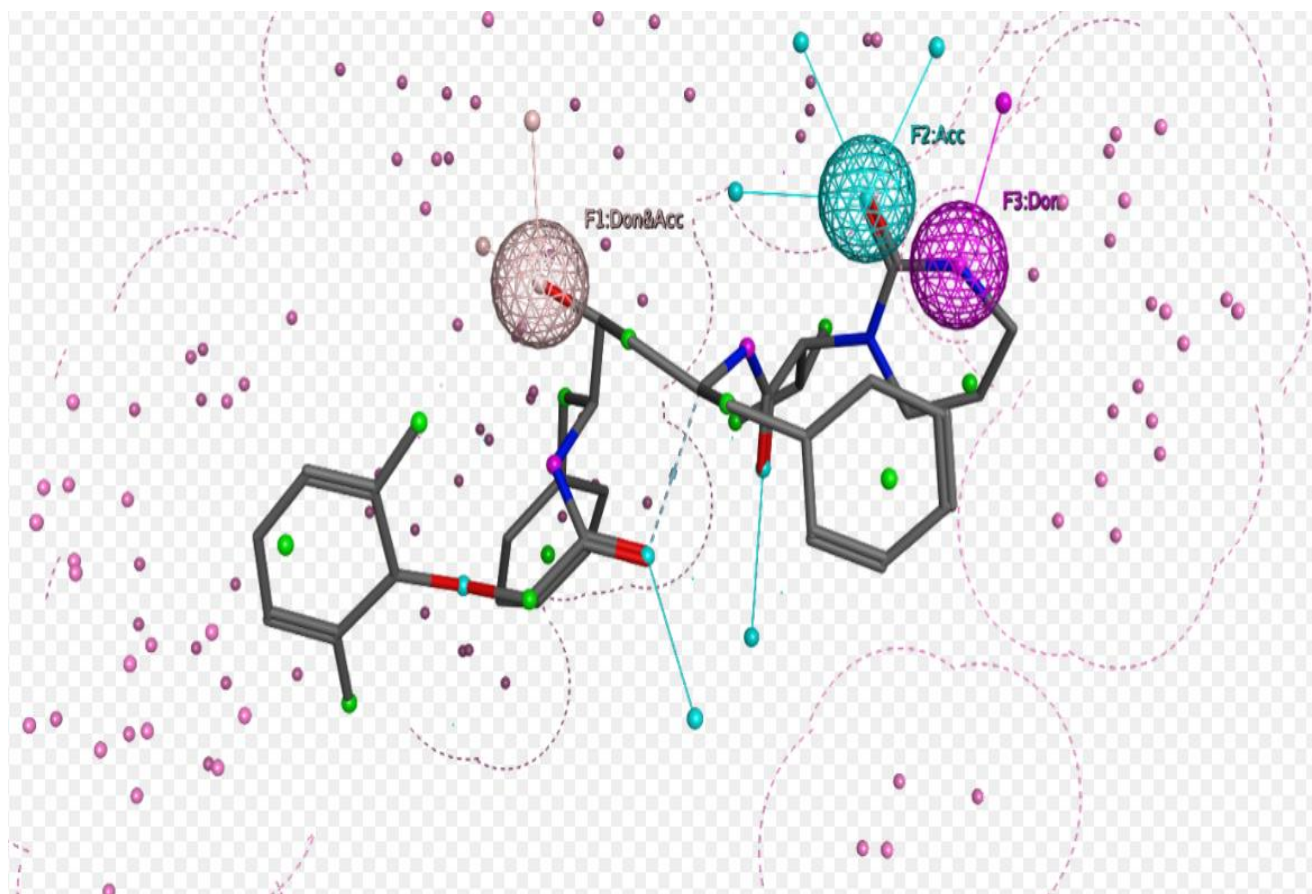


Figure 4. 4: A Pharmacophore model of 1rv7 co-crystallized structure

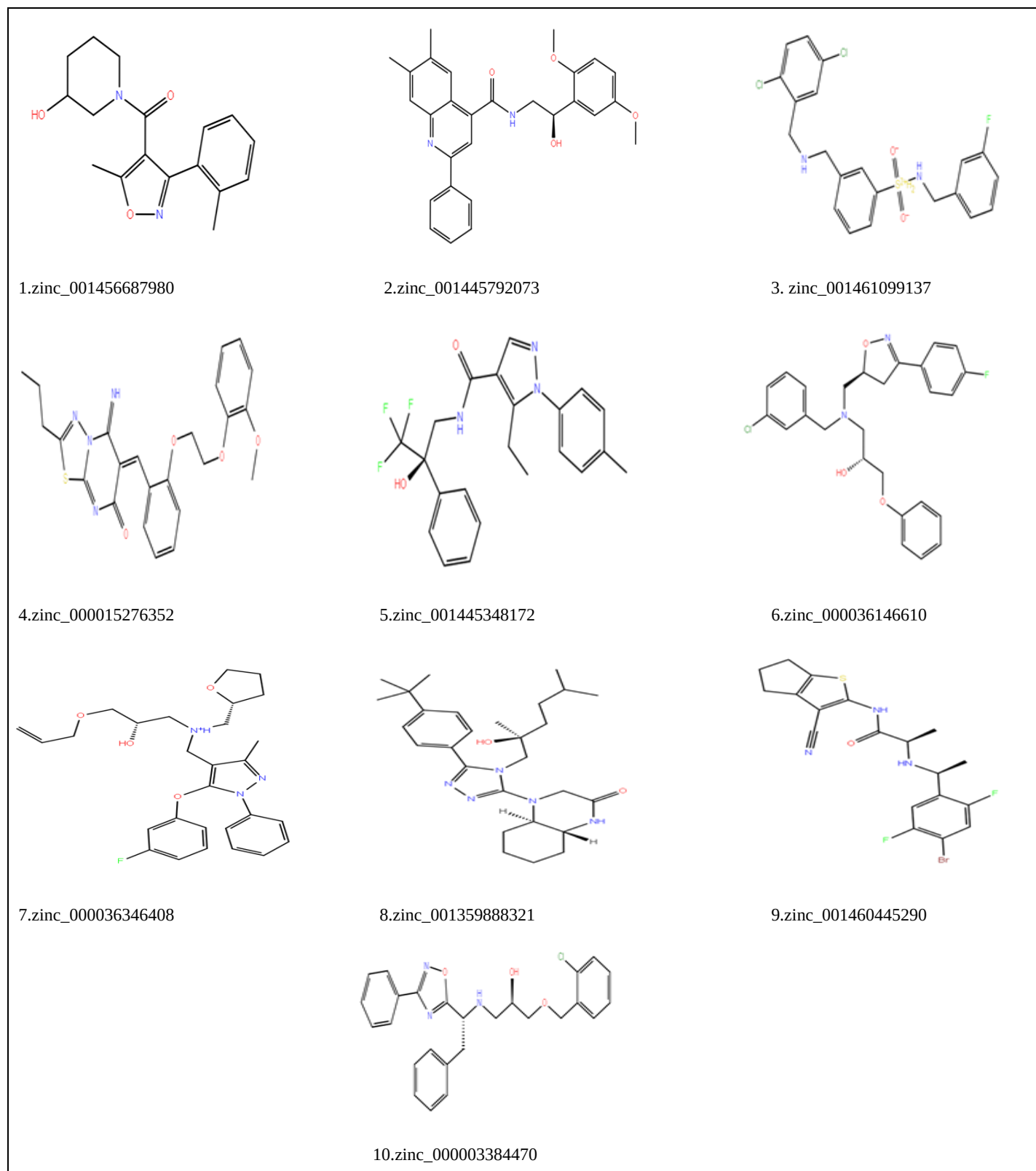


Figure 4. 5: Structures of the best 10 compounds from molecular docking

4.3.4 Ligand-Pharmacophore properties based on the RMSD

From the 327 ligands that were fitted with the pharmacophore, 26 compounds with the RMSD of less than 0,5 were selected.

4.3.5 Molecular docking against target protein

These 26 ligands were selected for molecular docking with the target protein. The features of the top 10 ligands are shown below and were performed through AutoDock Vina.

4.3.6 SwissADME studies

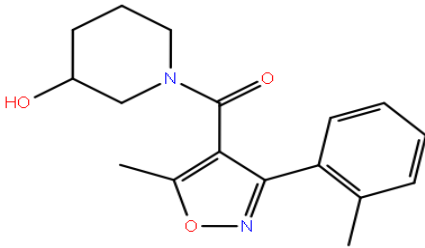
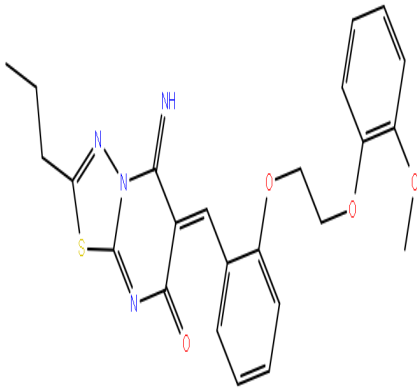
SwissADME/T web server was employed to predict the absorption, distribution, metabolism, excretion and toxicity of zinc_001456687980 and zinc_000015276352. Table 4.2 shows the predicted value for both compounds, for iLOGP, XLOGP, MLOGP, WLOGP and SILICOS-IT models. The partition coefficient between *n*-octanol and water ($\log P_{o/w}$) is a classical descriptor for lipophilicity, meaning ($\log P_{o/w}$) reflects the relative solubility of the drug in *n*-octanol and water. All the compounds were found to be more highly lipophilic.

Table 4. 2: SwissADME properties analysis for 10 best compounds

Zinc ID	M.W g/mol	Binding Affinity K _D	iLOGP	XLOGP3	WLOGP	MLOGP	SILICOS- IT	C/SUS LOG P o//w
*1.zinc_001456687980	455.34	-8	3.36	4.10	4.36	3.36	5.13	4.06
2.zinc_001445792073	456.53	-7.8	4.04	4.65	4.67	2.77	5.84	4.39
3.zinc_001461099137	453.36	-7.4	3.89	4.60	5.30	4.40	5.16	4.67
*4.zinc_000015276352	464.54	-6.9	3.93	4.69	3.22	2.13	5.82	3.96
5.zinc_001445348172	451.87	-6.9	3.97	4.77	5.73	3.63	4.95	4.61
6.zinc_000036146610	468.95	-6.8	4.05	5.16	4.80	3.80	5.94	4.75
7.zinc_000036346408	496.56	-6.8	4.76	4.38	3.51	-0.59	4.91	3.39
8.zinc_001359888321	481.67	-6.7	4.25	5.22	3.92	3.50	4.19	4.21
9.zinc_001460445290	454.33	- 6.7	3.33	4.86	5.15	3.43	6.34	4.62
10.zinc_000003384470	463.96	-6.6	3.80	4.55	4.36	2.91	5.51	4.23

From the 10 compounds with the best binding affinity, only two passed the Swiss ADME test (Highlighted with green). Four compounds, namely zinc_001456687980, zinc_000015276352, zinc_001359888321 and zinc_001460445290 were predicted to be suitable for oral administration. * = Two best compounds.

Table 4. 3: SwissADMET for two lead compounds

Ligand ID	QLogPo/v	QLogHERG	QPPCaco	QLogBB	QlogKhsa	2D Structure
Zinc_001456687980	4.797	-5.553	1802.65	-0.12	0.818	
Zinc_000015276352	4.901	-6.858	872.832	-1.269	0.544	

In Swiss ADME/T web server was used to determine various toxicity parameters, metabolizing enzyme interaction and absorption parameters. This was done based on the screened structures and molecules that were obtained for drug-likeness character. Table 4.4 gives a brief overview of some of the important parameters based on the following criterion: Mol. W 130 – 725: donorHB 0-6: accptHB 02 –20: QPlogPo/w (-2-6.5) below - : QPlogHERG 5: QPPCaco < 25 => poor >500 => great: QPlogBB (-3 – 1.2): QPlogKhsa (-1.5 – 1.5).

Table 4. 4: SwissADME/T properties for two lead compounds

Properties of lead compounds	Zinc_001456687980	Zinc_000015276352
Molecular weight (g/mol)	455.34	464.54
Toxicity	Non-toxic	Non-toxic
Rotatable bonds	4	9
Ring	3	4
Carcinogenetic	Non-carcinogenic	Non-carcinogenic
Solubility	Moderately Soluble	Moderately Soluble
Binding Energy	-8	-6.9
Molar refractivity	119.22	130.60
H-bond Acceptor	4	7
H- bond Donor	1	1
GI absorption	High	High
BBB permeation	yes	No
Abbott BA	0.55	0.55
Skin Permeation	-6.17 cm/s	- 5.80 cm/s
PAINS	0 Alert	0 Alert

Figure 4.6 shows the orientation and interaction of Zinc_001456687980 and Zinc_000015276352 at 1rv7 active site. Both Zinc_001456687980 and Zinc_000015276352 show binding in the active side and are located inside the binding pocket of the protein. Zinc_001456687980 displays the following hydrogen bond interactions Gly49, Gly48, Thr80, Val54, Asp30, Asp29, Asn25, Gly27. Hydrogen bonds are very important in the binding

of the drug to the receptor and their formation shows a certain potential that the complex can possess. As revealed by the interaction, the hydrogen bonds occurred between two strongly negative charged groups and this increases the strength and the stability of the drug-receptor significantly.

4.3.6 Binding Affinity: 2D interaction

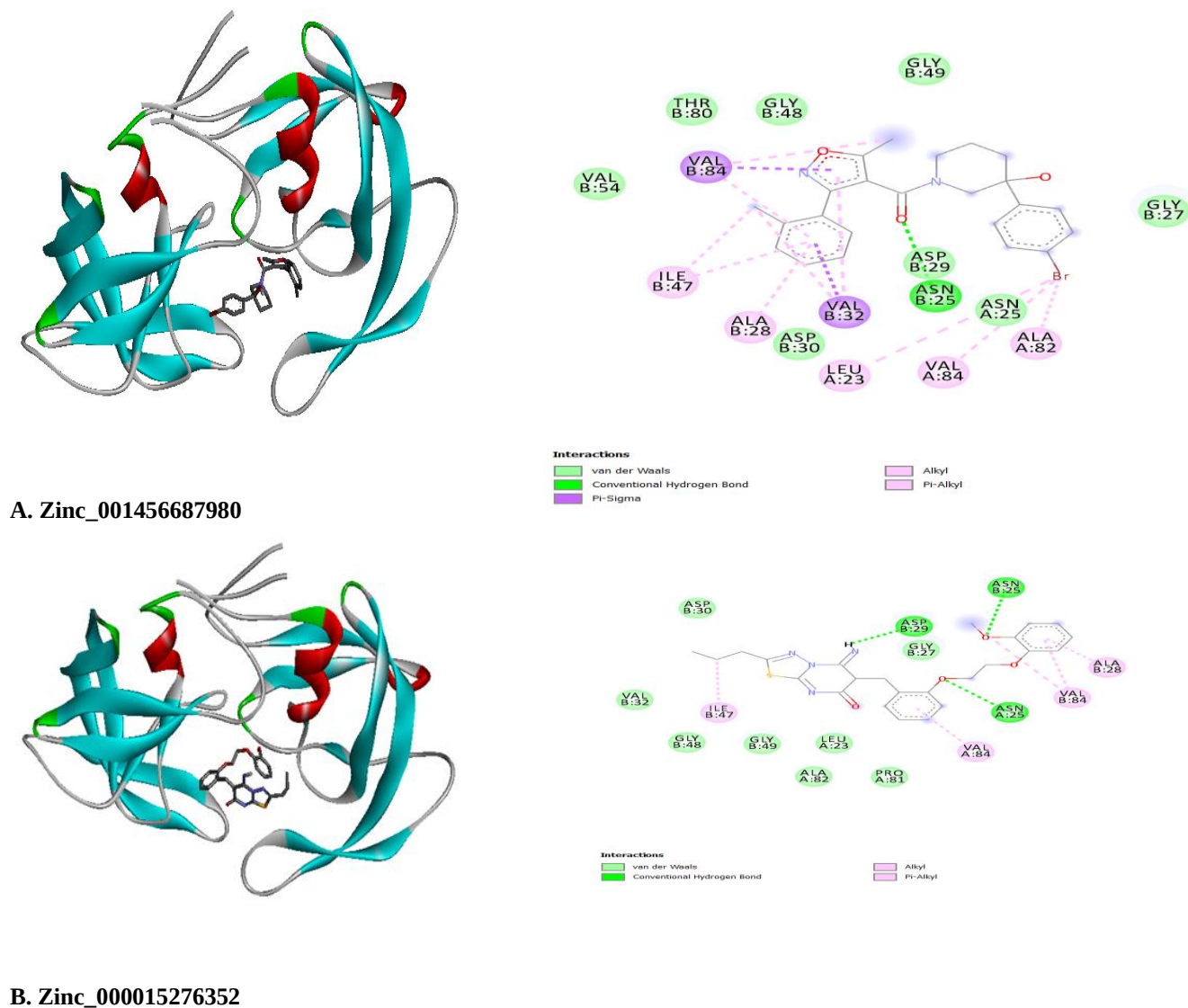


Figure 4. 6: Binding interaction of ZINC compounds and 1rv7

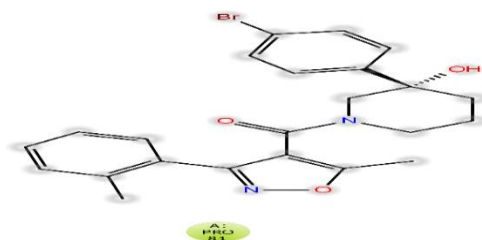
Table 4. 5: The docked Van der Waals, Conventional Hydrogen Bonds, Pi-Sigma, Alkyl, Pi-Alkyl

Compounds	Van der Waals	Conventional H bonds	pi-sigma	Alkyl	pi-alkyl
A) Z_001456687980	Gly49, Gly48, Thr80, Val54, Asp30, Asp29, Asn25, Gly27	Asn25	Val32, Val84	Ile47, Ala28	Leu23, Val84, Ala82
B) Z_000015276352	Asp30, Val32, Gly48, Gly49, Ala82, Leu23, Pro81, Gly27	Asn25, Asp29		Val84, Ala28	Ala47

The hydrogen bonds were formed by CH₂ of cyclohexane with Asn25 residue (Table 4.5). They also participated with conventional hydrogen bonding interaction with the ligand. Other participations included the alkyl interactions with Ile47, Ala28 and pi –alkyl interactions with Leu23, Val84, Ala82. Moreover, 1rv7 has a hydrophobic active site. In this active site, the ligand is scaffolded and positioned. In the case of Zinc_000015276352, three hydrogen bonds were formed namely, Asn_B25, Asn_A25 and Asp29. Based on the key residues and the interaction analysis between ligands and protein proposed, Zinc_ 001456687980 and Zinc_000015276352 possess binding properties that are suitable for protein ligand stability.

4.4 MOLECULAR DYNAMICS SIMULATION

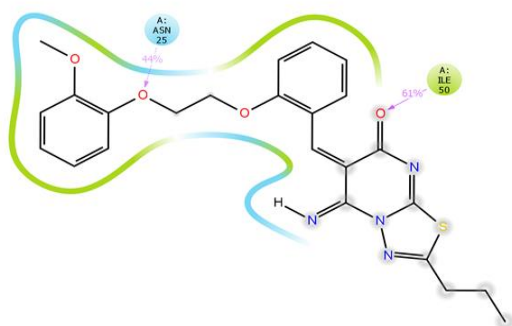
Using molecular dynamics simulation, the following Desmond simulation trajectories were analysed: the root mean square fluctuation, the root mean square deviation and the protein-ligand contact calculation. The protein-ligand contacts were also determined from the MD trajectory analysis.



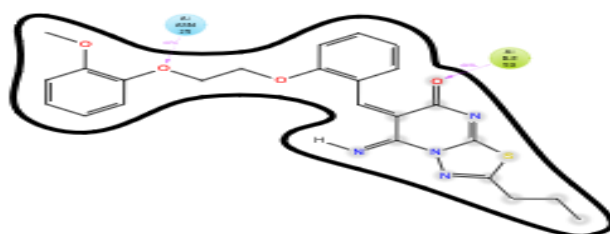
B. Z_001456687980



B. Z_001456687980



B. Z_000015276352



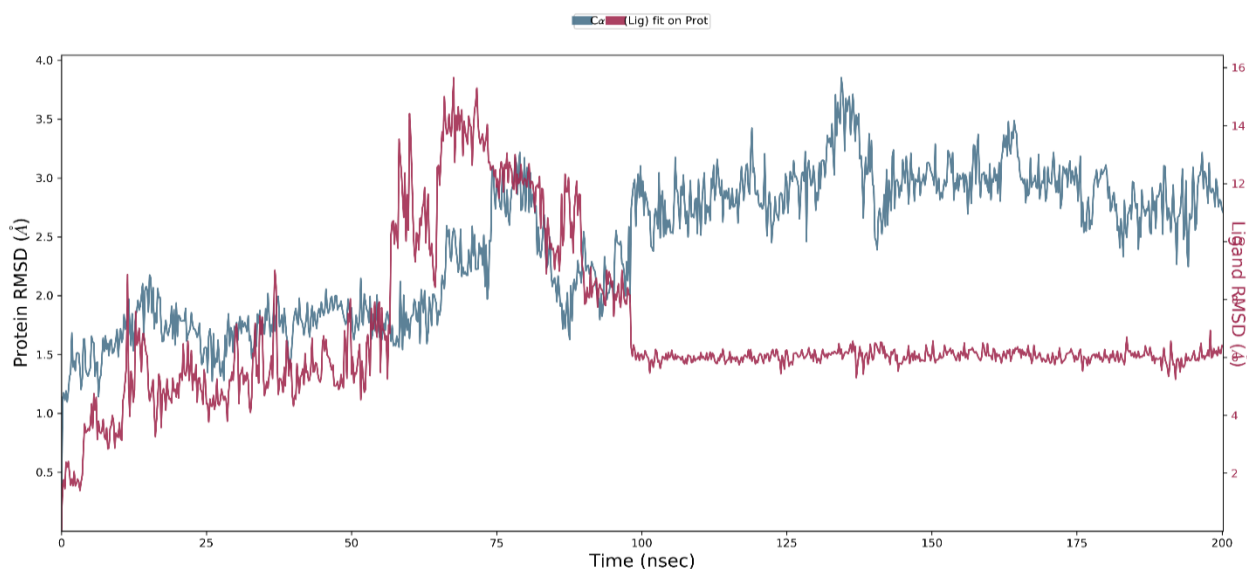
B. Z_000015276352

 Hydrophobic  Polar

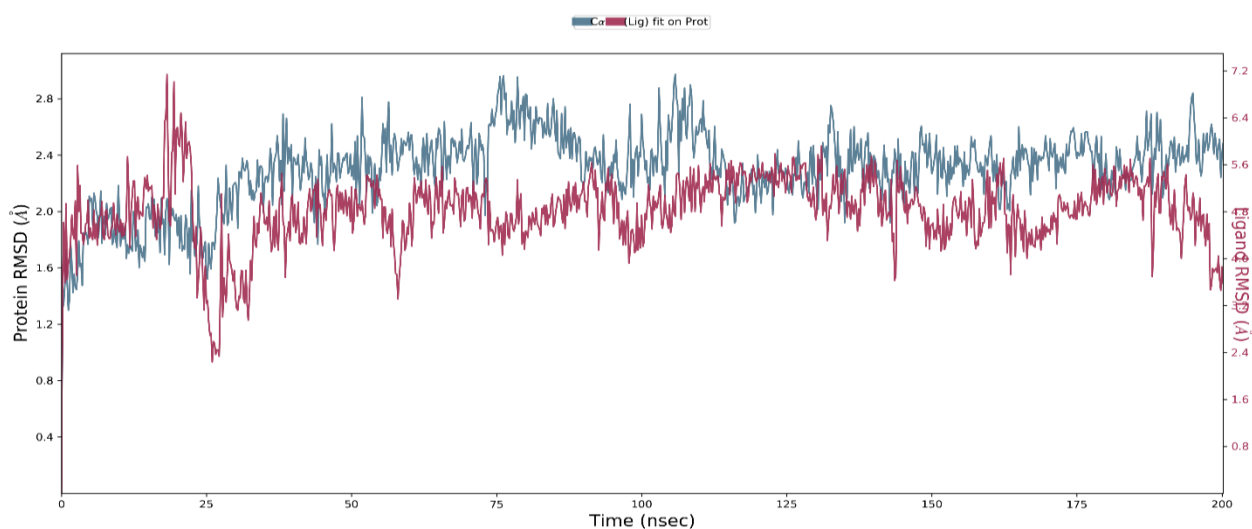
 Solvent exposure

Figure 4. 7: Ligand- Protein contacts

Figure 4.8 is the molecular dynamics Simulation diagrams and is of the root mean square deviation of the C- α atoms of the protein and the ligand with time in nano seconds (nsec). In this, diagram the left γ - axis shows the variation of protein RMSD through time whereby the right γ - axis shows the variation of the ligand RMSD through time.



A. Z_001456687980

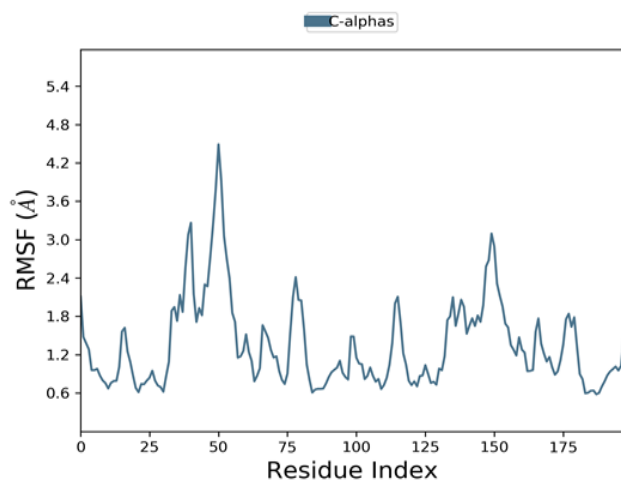


B. Z_000015276352

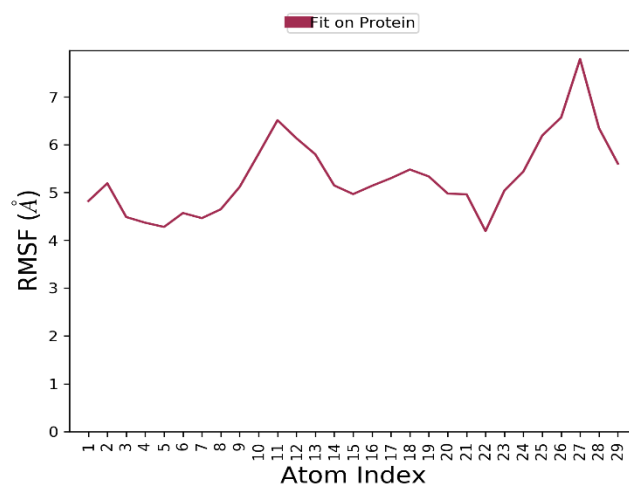
Figure 4. 8: Molecular dynamics Simulation diagrams at 200ns

Using the Desmond software, which is a package of the Schrödinger, the root mean square deviation (RMSD) trajectories were used to measure the scale of the distance between the ligand and the protein and takes place throughout the entire simulation. The interaction of the trajectories between the ligand and the protein was analysed using the C- α atoms at time 200 nanoseconds. The structural conformation between the ligand and the protein was also determined throughout the simulation by monitoring the RMSD. This simulation reveals interactions when it reaches equilibrium and monitors the deviation that occurs in the thermal average. RMSD indicated that zinc_001456687980, at time 100ns, reaches equilibrium until the end of the simulation 200ns. This complex remained at equilibrium with minimal fluctuations, and over a period of 100ns, there was a decrease in the conformation for the receptor-ligand complexed. Towards the end of the simulation, zinc_001456687980 and zinc_000015276352, for ligand interaction, showed values of 6 Å and 3.8 Å respectively, displaying zinc_000015276352 as a compound having the least value of 3.8Å.

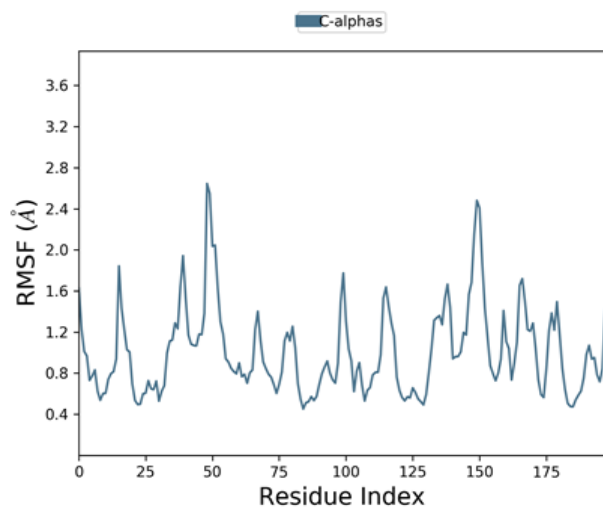
Figure 4.9 shows images of the protein root mean square fluctuation (RMSF) which analyses the C- α atoms with the peaks that show the areas that fluctuate the most, as indicated on the y-axis, whereas the x-axis indicates, the residue location that is found in the protein. The graphs for both compounds, shows major peaks at 50ns and 150ns overlaying the RMSF yields (Figure 4.9) which is a protein secondary structure. Secondary structures like beta and alpha helices are usually more rigid, meaning they are not flexible, as compared to the unstructured part of the protein. Therefore, this secondary structure fluctuates less. Figure 4.9 shows the protein-ligands contacts for zinc_001456687980 and zinc_000015276352. Zinc_000015276352 displays the following main interactions: Asn 25, Ile50, Ala82, while for zinc_001456687980 the interactions are Pro81, Gle50 and Asn25.



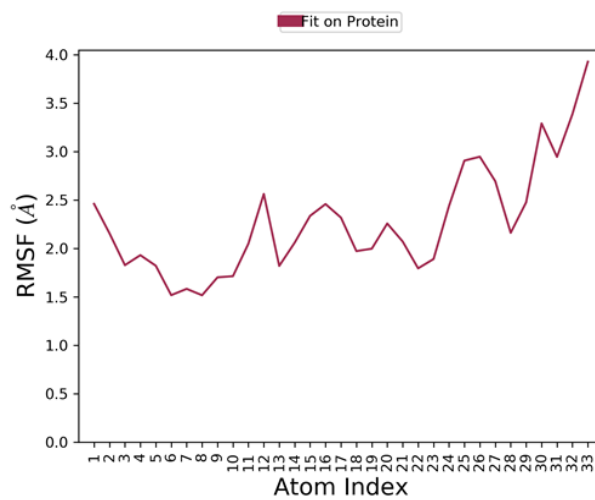
A. Z_001456687980/ P



B. Z_001456687980/ L



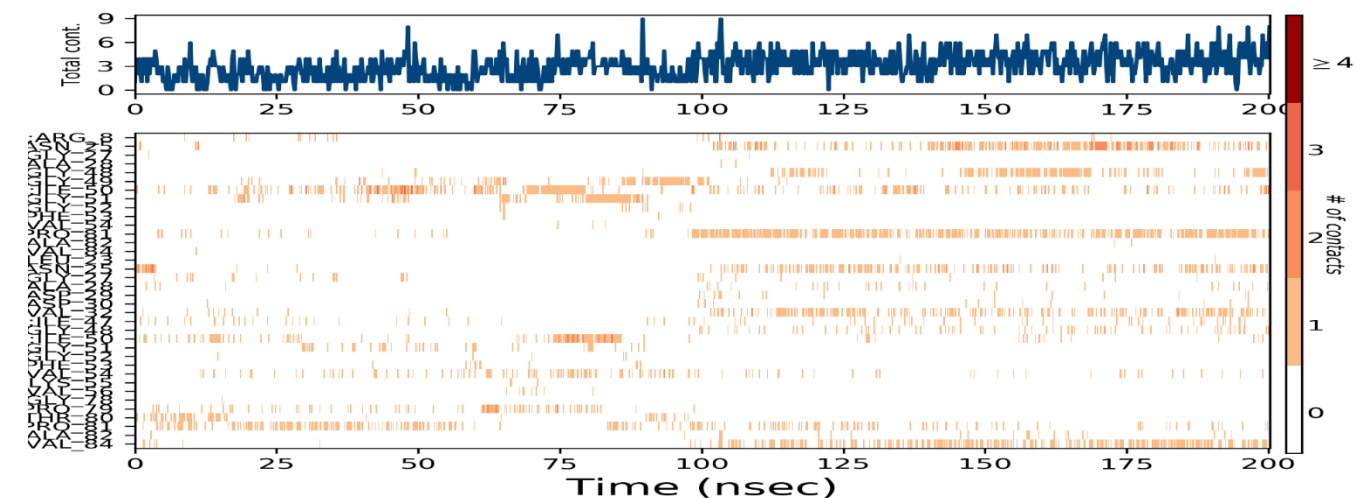
A. Z_000015276352/ P



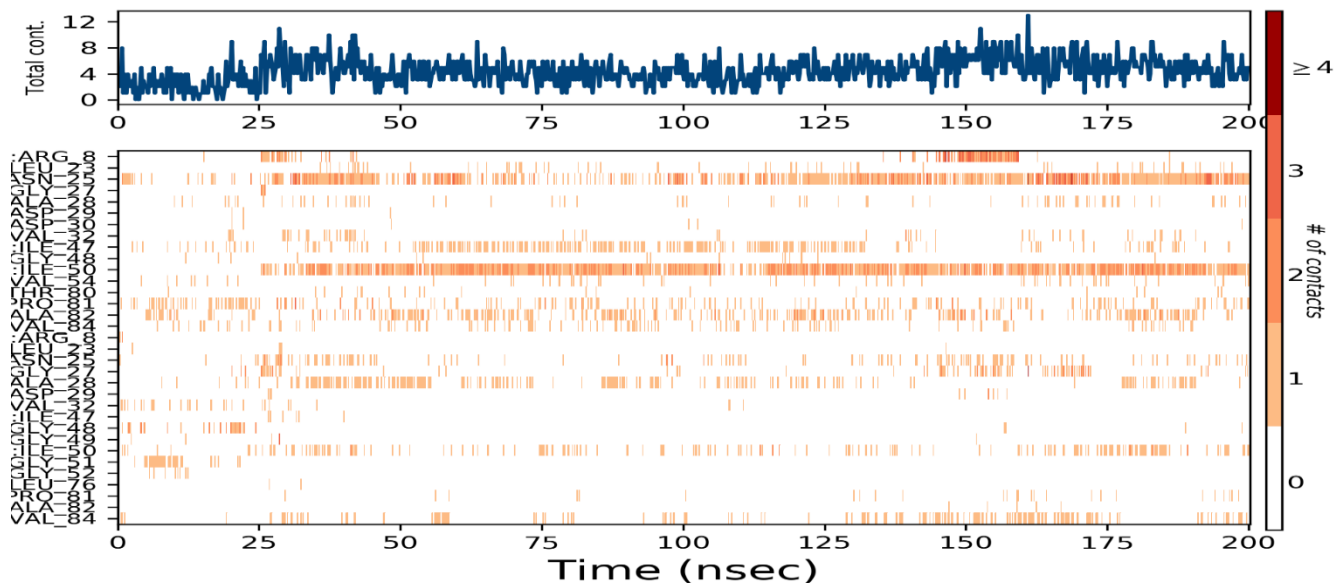
B. Z_000015276352/ L

Figure 4. 9: RMSF P/L for lead compounds

Figure 4. 10 shows the ligand-protein contacts for zinc_001456687980 and zinc_000015276352. The compounds that show more than 30% interactions are shown. Zinc_000015276352, Asn25 and Ile50 display 44 % and 50%, respectively. Zinc_001456687980 Pro81, displays hydrophobic interaction for Pro81. This compound shows solvent exposure on all carbon bonds, nitrogen and oxygen atoms.



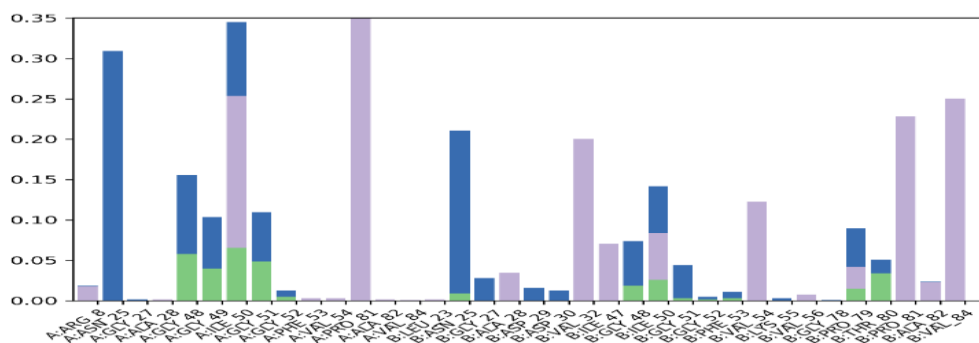
A. Z_001456687980



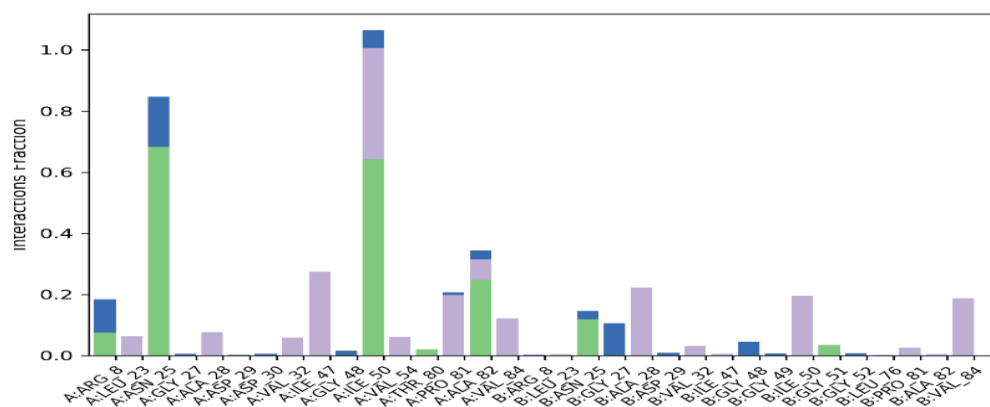
B. Z_000015276352

Figure 4. 10: Protein-Ligand contacts

Generally, protein-ligand interaction are facilitated by hydrogen bonds. This is essential since there is a need to monitor their strength in water molecule. Figure 4.11 shows the bond length 0.7 for Asn25 and 0,6 for Ile50. Ile50 also has 0,7 bond length in compound zinc_001456687980. Zinc_001456687980 formed strong hydrophobic interactions with Val54, Ala82, Thr80, Asp30, Phe53 and Val32. Zinc_001456687980 also interacts through water bridge bonds with Arg8, Leu23 and strong interactions with water bridge, hydrogen bond and hydrophobic bonds at Gly49. On the other hand, Zinc_000015276352 does not possess strong interaction compared to Zinc_001456687980. It does however have strong hydrogen bonds at Leu23 and Gly48. This attest the fact that inhibitors have unique features, which individually plays a role in biological and chemical interaction.



A. Z_001456687980



B. Z_000015276352

Figure 4. 11: Protein-ligand contact histogram

REFERENCE LIST

A PUBLICATION OF THE SOUTHERN AFRICAN HIV CLINICIANS SOCIETY. 2021. *HIV NURSING MATTERS* [Online]. Available: <https://sahivsoc.org/Files/Nursing%20matters%20May%202017> [Accessed].

ACHIRI, R., FOUZIA, M., BENOMARI, F. Z., DJABOU, N., BOUFELDJA, T., MUSELLI, A. & DIB, M. E. A. 2021. Chemical composition/pharmacophore modelling-based, virtual screening, molecular docking and dynamic simulation studies for the discovery of novel superoxide dismutase (SODs) of bioactive molecules from aerial parts of *Inula Montana* as antioxydant's agents. *Journal of Biomolecular Structure and Dynamics*, 1-22.

ADENIJI, S. E., ARTHUR, D. E. & OLUWASEYE, A. 2020. Computational modeling of 4-Phenoxy nicotinamide and 4-Phenoxy pyrimidine-5-carboxamide derivatives as potent anti-diabetic agent against TGR5 receptor. *Journal of King Saud University-Science*, 32, 102-115.

ANDRIANOV, A. M., NIKOLAEV, G. I., SHULDOV, N. A., BOSKO, I. P., ANISCHENKO, A. I. & TUZIKOV, A. V. 2021. Application of deep learning and molecular modeling to identify small drug-like compounds as potential HIV-1 entry inhibitors. *Journal of Biomolecular Structure and Dynamics*, 1-19.

CHEN, B., VOGAN, E. M., GONG, H., SKEHEL, J. J., WILEY, D. C. & HARRISON, S. C. 2005. Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature*, 433, 834-841.

DAINA, A., MICHIELIN, O. & ZOETE, V. 2017. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific reports*, 7, 1-13.

FINZI, D., HERMANKOVA, M., PIERSON, T., CARRUTH, L. M., BUCK, C., CHAISSON, R. E., QUINN, T. C., CHADWICK, K., MARGOLICK, J. & BROOKMEYER, R. 1997. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science*, 278, 1295-1300.

GOPINATH, P. & KATHIRAVAN, M. 2021. Docking studies and molecular dynamics simulation of triazole benzene sulfonamide derivatives with human carbonic anhydrase IX inhibition activity. *RSC Advances*, 11, 38079-38093.

GUEDÁN, A., CAROE, E. R., BARR, G. C. & BISHOP, K. N. 2021. The Role of Capsid in HIV-1 Nuclear Entry. *Viruses*, 13, 1425.

HAALAND, R. E., MARTIN, A., MENGESHA, M., DINH, C., FOUNTAIN, J., LUPO, L. D., HALL, L., CONWAY-WASHINGTON, C. & KELLEY, C. F. 2021. Evaluation of Antiretroviral Drug Concentrations in Minimally Invasive Specimens for Potential Development of Point-of-Care Drug Assays. *AIDS Research and Human Retroviruses*.

HOLLINGSWORTH, S. A. & DROR, R. O. 2018. Molecular dynamics simulation for all. *Neuron*, 99, 1129-1143.

HUANG, H., CHOPRA, R., VERDINE, G. L. & HARRISON, S. C. 1998. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science*, 282, 1669-1675.

HUANG, J., KANG, B. H., PANCERA, M., LEE, J. H., TONG, T., FENG, Y., IMAMICHI, H., GEORGIEV, I. S., CHUANG, G.-Y. & DRUZ, A. 2014. Broad and potent HIV-1 neutralization by a human antibody that binds the gp41–gp120 interface. *Nature*, 515, 138-142.

HUSSAIN, W., RASOOL, N. & KHAN, Y. D. 2021. Insights into Machine Learning-based approaches for Virtual Screening in Drug Discovery: Existing strategies and streamlining through FP-CADD. *Current Drug Discovery Technologies*, 18, 463-472.

IPPOLITO, J. A., NIU, H., BERTOLETTI, N., CARTER, Z. J., JIN, S., SPASOV, K. A., CISNEROS, J. A., VALHONDO, M., CUTRONA, K. J. & ANDERSON, K. S. 2021. Covalent Inhibition of Wild-Type HIV-1 Reverse Transcriptase Using a Fluorosulfate Warhead. *ACS Medicinal Chemistry Letters*, 12, 249-255.

JETTE, C. A., BARNES, C. O., KIRK, S. M., MELILLO, B., SMITH, A. B. & BJORKMAN, P. J. 2021. Cryo-EM structures of HIV-1 trimer bound to CD4-mimetics BNM-III-170 and M48U1 adopt a CD4-bound open conformation. *Nature communications*, 12, 1-10.

KOBAYAKAWA, T., YOKOYAMA, M., TSUJI, K., FUJINO, M., KURAKAMI, M., BOKU, S., NAKAYAMA, M., KANEKO, M., OHASHI, N. & KOTANI, O. 2021. Small-Molecule Anti-HIV-1 Agents Based on HIV-1 Capsid Proteins. *Biomolecules*, 11, 208.

LEELANANDA, S. P. & LINDERT, S. 2016. Computational methods in drug discovery. *Beilstein journal of organic chemistry*, 12, 2694-2718.

LIPINSKI, C. A., LOMBARDO, F., DOMINY, B. W. & FEENEY, P. J. 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, 23, 3-25.

LOGSDON, B. C., VICKREY, J. F., MARTIN, P., PROTEASA, G., KOEPKE, J. I., TERLECKY, S. R., WAWRZAK, Z., WINTERS, M. A., MERIGAN, T. C. & KOVARI, L. C. 2004. Crystal structures of a multidrug-resistant human immunodeficiency virus type 1 protease reveal an expanded active-site cavity. *Journal of virology*, 78, 3123-3132.

MICHAEL, H. U., NAIDOO, S., MENSAH, K. B., RAMLALL, S. & OOSTHUIZEN, F. 2021. The impact of antiretroviral therapy on neurocognitive outcomes among people living with HIV in low-and middle-income countries (LMICs): a systematic review. *AIDS and Behavior*, 25, 492-523.

MURRI, R., AMMASSARI, A., GALLICANO, K., DE LUCA, A., CINGOLANI, A., JACOBSON, D., WU, A. W. & ANTINORI, A. 2000. Patient-reported nonadherence to HAART

is related to protease inhibitor levels. *Journal of acquired immune deficiency syndromes* (1999), 24, 123-128.

ORTIZ, G. M., NIXON, D. F., TRKOLA, A., BINLEY, J., JIN, X., BONHOEFFER, S., KUEBLER, P. J., DONAHOE, S. M., DEMOITIE, M.-A. & KAKIMOTO, W. M. 1999. HIV-1–specific immune responses in subjects who temporarily contain virus replication after discontinuation of highly active antiretroviral therapy. *The Journal of clinical investigation*, 104, R13-R18.

OUYANG, Y., HUANG, J.-J., WANG, Y.-L., ZHONG, H., SONG, B.-A. & HAO, G.-F. 2021. In Silico Resources of Drug-Likeness as a Mirror: What Are We Lacking in Pesticide-Likeness? *Journal of Agricultural and Food Chemistry*, 69, 10761-10773.

SHAFER, R. & VUITTON, D. 1999. Highly active antiretroviral therapy (HAART) for the treatment of infection with human immunodeficiency virus type 1. *Biomedicine & pharmacotherapy*, 53, 73-86.

SHIN, Y. H., PARK, C. M. & YOON, C.-H. 2021. An Overview of Human Immunodeficiency Virus-1 Antiretroviral Drugs: General Principles and Current Status. *Infection & Chemotherapy*, 53, 29.

TARIQ, A., REHMAN, H. M., MATEEN, R. M., ALI, M., MUTAHIR, Z., AFZAL, M. S., SAJJAD, M., GUL, R. & SALEEM, M. 2021. A CADD Based Discovery of Lead-Like Compounds Against KDM5A for Cancers Using Pharmacophore Modeling and High-Throughput Virtual Screening. *Proteins: Structure, Function, and Bioinformatics*.

TRYLSKA, J., TOZZINI, V., CHIA-EN, A. C. & MCCAMMON, J. A. 2007. HIV-1 protease substrate binding and product release pathways explored with coarse-grained molecular dynamics. *Biophysical journal*, 92, 4179-4187.

UNAIDS. 2020. *HIV and AIDS in South Africa* [Online]. Available: <https://www.unaids.org/en/resources/fact-sheet>. [Accessed].

XU, C., CHENG, F., CHEN, L., DU, Z., LI, W., LIU, G., LEE, P. W. & TANG, Y. 2012. In silico prediction of chemical Ames mutagenicity. *Journal of chemical information and modeling*, 52, 2840-2847.

ZHAO, L., CIALLELLA, H. L., ALEKSUNES, L. M. & ZHU, H. 2020. Advancing computer-aided drug discovery (CADD) by big data and data-driven machine learning modeling. *Drug discovery today*.

ZHENG, W., BROOKS, B. R., DONIACH, S. & THIRUMALAI, D. 2005. Network of dynamically important residues in the open/closed transition in polymerases is strongly conserved. *Structure*, 13, 565-577.

CHAPTER 5: COMPUTATIONAL STUDIES FOR SELECTED MEDICINAL PLANTS AGAINST DOLUTEGRAVIR USING LIGAND-BASED PHARMACOPHORE, MOLECULAR DOCKING, ADMET PREDICTIONS AND MOLECULAR DYNAMICS SIMULATION

ABSTRACT

Human immunodeficiency virus type 1 (HIV-1) is the causative agent for acquired immunodeficiency syndrome (AIDS). In 2020, South Africa recorded an estimated 8,2 million people living with HIV. This extensive figure is a red flag to the country, as it causes serious economic burden to its health care system. In the quest for finding a suitable inhibitor for HIV-1 protease, computer aided drug design (CADD) approach stands out to be one of the leading fields of study in pursuit of a new drug for HIV. The Lipinski rule of five was applied in screening the 92 plant extracts from *Ocimum sanctum*, *Carica papaya*, *Persea Americana*, *Azadirachta indica* and *Spondias mombin* medicinal plants and 46 of the compounds complied with this rule. Afterwards, a ligand-based pharmacophore was constructed based on the active properties of the three (3) best binding compound obtained from the screened ZINC compounds (Zinc_001456687980, Zinc_001445792073 and Zinc_001461099137). The compounds with less than 0.5 RMSD were picked out. The best 10 compounds were docked and compared with a drug that is already in the market, namely: dolutegravir, 3, 5-di-O-galloyl-4- O-digalloylquinic acid 3,5,4,4 -Tetragqa (-8.6 K_D), epicatechin (-8 K_D) and quercitrin (-8 K_D) obtained the highest binding affinity. Although Tetragqa had the highest binding affinity, it failed the test because of its large molecular weight. The safety evaluation and other chemical parameters that included the lipophilicity and physicochemical properties of these compounds were performed through SwissADME/T web server. On the best top three binding compounds, only epicatechin (-8 K_D) had promising features of a drug candidate. However, the remaining compounds from the best 10 compounds were also analysed using the SwissADME/T tools, whereby two of them (juglanin and catechin) satisfied the ADMET prediction analysis. Catechin showed some promising features as well after it displayed good drug like properties, suitable for a novel compound. Lastly, molecular dynamics simulation were performed on the three lead compounds namely; epicatechin, juglanin and

catechin against dolutegravir. The ligand protein interaction between catechin and the protein displayed a minimal shift of the protein during the simulation that was performed over 100ns, signifying a strong complex association.

5.1 INTRODUCTION

The first case of human immunodeficiency virus was recorded in 1981 (Greene, 2007). Over the years and up to date, important scientific work has been done in attempting to eradicate this disease from humankind. In 2020, HIV claimed around 690 000 lives globally (UNAIDS, 2020). However, since the first case of HIV (AIDS) has been reported, a lot of progress has been made in the pharmaceuticals, with numerous computer aided drug design (CADD) approaches being applied in finding a cure for this disease (Stuart *et al.*, 2018). Subsequently, there is a strong need to manufacture antiretroviral drugs (ARVs) that are affordable, with minimal side effects and more effective than the heavy burden pills that were used some decade ago. South Africa, with the largest HIV treatment program in the world (AIDSmap), introduced dolutegravir to its patients with an estimate of over 4.8 million people on HIV treatment programs (Global information and education on HIV and AIDS, 2020). Dolutegravir, also known as “the game changer”, has proven that it has the capacity to suppress the patients’ viral load quicker than other ARVs, and, was used as a control in this study against selected medicinal plants (A publication of the Southern African HIV Clinicians Society, 2021).

Protease being an important structure found in the human immunodeficiency virus (Kräusslich *et al.*, 1989), helps in the processing of Gag-Pol and Pol polypeptide into mature functional proteins. It is also responsible for the maturation of virions that are released by the virus, and, it belongs to the class of aspartic acid (Oroszlan and Luftig, 1990). This class functions as a catalytic dimer. These virions only become non-infectious when the protease is blocked in order to mature the virions (Konvalinka *et al.*, 2015). This inhibits the replication of HIV-1 through the introduction of HIV-1 inhibitors that are competitive towards the binding site of the protease (Louis *et al.*, 1994) (Kohl *et al.*, 1988).

In this study, the application of CADD techniques was used in finding an inhibitor/s from five plant extracts, namely: *Ocimum sanctum* (Rajinikanth *et al.*, 2013), *Carica papaya* (Yogiraj *et al.*, 2014), *Persea Americana* (Owolabi *et al.*, 2005), *Azadirachta indica* (Mahapatra *et al.*, 2012) and *Spondias mombin* (de Lima *et al.*, 2016). The screened phytochemicals were then compared with dolutegravir, which is an ARV that is currently being administered to HIV patients in South Africa (Hauser *et al.*, 2020). The active biological activity of dolutegravir was compared with the lead compounds retrieved from PubChem database. The application of literature was an important tool in finding respective biological composition for each plant (a number of 92 phytochemicals were identified) (Mulligan *et al.*, 2018).

Computer Aided Drug Design (CADD) is a molecular modelling approach that can be used to search for new drugs that can cure diseases. This approach has yielded much success over the years with some of the ARVs being discovered through CADD (Surabhi and Singh, 2018), (Veselovsky and Ivanov, 2003). With the administration of highly active antiretroviral therapy (HAART), there is a lot of progress that has been made in an attempt to curb this epidemic. However, some of these medications create a lot of pill burden to patients who are placed on such therapeutic processes (Shafer and Vuitton, 1999). Other challenges like antimicrobial resistance make it difficult to treat some patients who are failing extensive antiretroviral therapy (Rackal *et al.*, 2011), hence a need for a permanent solution to this disease.

Docking studies between virtually screened ligands from the ZINC database (for pharmacophore modelling) and PubChem (for selected phytochemicals) played an important role in finding potential drug compounds that can be used to contribute to the fight against this HIV. ZINC and PubChem being the largest databases of small ligand compounds with approximately more than 1, 3 billion compounds (Ton *et al.*, 2020) and 111 million structures (Gaulton *et al.*, 2013) respectively, marks a good start for the screening of suitable novel compounds (Sabe *et al.*, 2021).

Auto Dock Vina is one of the tools used to dock such compounds with the protein. It has yielded great success in an attempt to understand the protein-ligand interactive properties and also in the quest of finding novel compounds (Grinter and Zou, 2014) (Karplus and McCammon, 2002). One

of the best tools of filtering these compounds is pharmacophore modelling, which is a process of mimicking already existing biologically active properties on known structures, either in complexed or uncomplexed structures, in order to filter compounds in a class based on their similarity and likeness (Walters and Wang, 2020).

Using MD as an approach to gain more insight pertaining to prominent molecular interactions, it is also possible to analyse dynamic features that are being formed by such interactions. A simulation process entails a post-docked protein-ligands compound, used in order to analyse the different orientation that might have raised in the process. Only the best-docked compounds, which were predicted to satisfy the health and safety administration analysis using ADMET, were simulated in time of nanosecond scale. These simulated results include statistical parameters such as the RMSD, RMSF and Hydrogen contacts (Borhani and Shaw, 2012) (Vora et al., 2019).

In this study, we followed these CADD processes to find a possible drug candidate for HIV-1 protease. ADMET property analysis was performed on the lead compounds to predict the bioavailability of those compounds. In drug discovery, determining the toxicity level of compound is one of the fundamental routines that is encouraged to be performed early in the search for a new drug.

In this regards, SwissADME as a web server, is an approach that predicts the different parameters of compounds. It saves time and it limits the risks of putting a patient's life in danger (Daina et al., 2017) (Brenk et al., 2008). At a later stage, these multiple chemical parameters are very essential in determining the efficacy of a novel compound (Baell and Holloway, 2010).

5.2 MATERIALS AND METHODS

PubChem database

Phytochemicals were retrieved from the PubChem database. This database contains unique chemical structure. Eighty-nine compounds that were used in this investigation were retrieved, except for three, which were drawn from the MOE software.

Phytochemicals library

A library in the MOE was created for all the retrieved phytochemical. In addition, the Lipinski's rule was used to screen all 92 stored phytochemicals.

Target identification

HIV-1 protein (1rv7) was also retrieved from the Protein data bank (O'Leary *et al.*, 2016). This structure is one of the five wild type structures stored in the Protein Data Bank and it has a resolution of 2.7 Å (Sillanpää *et al.*, 2008). 1rv7 is a protease with two chains (A and B) with a sequence length of 99. It was isolated from a patient who was failing antiretroviral therapy and subsequently deposited into the database (Liu *et al.*, 2015).

Protein preparation

The retrieved protein was downloaded from the PDB (Liu *et al.*, 2015). Water molecules and hetatoms were then deleted using Biovia discovery studios. Since protein was retrieved as a co-crystallized complex with the ligand, the present ligand was removed by the “Define and Edit binding site” then “Current selection” and “Delete ligand”. Water molecules were then added by applying Chemistry hydrogen then adding polar (Adeniji *et al.*, 2020).

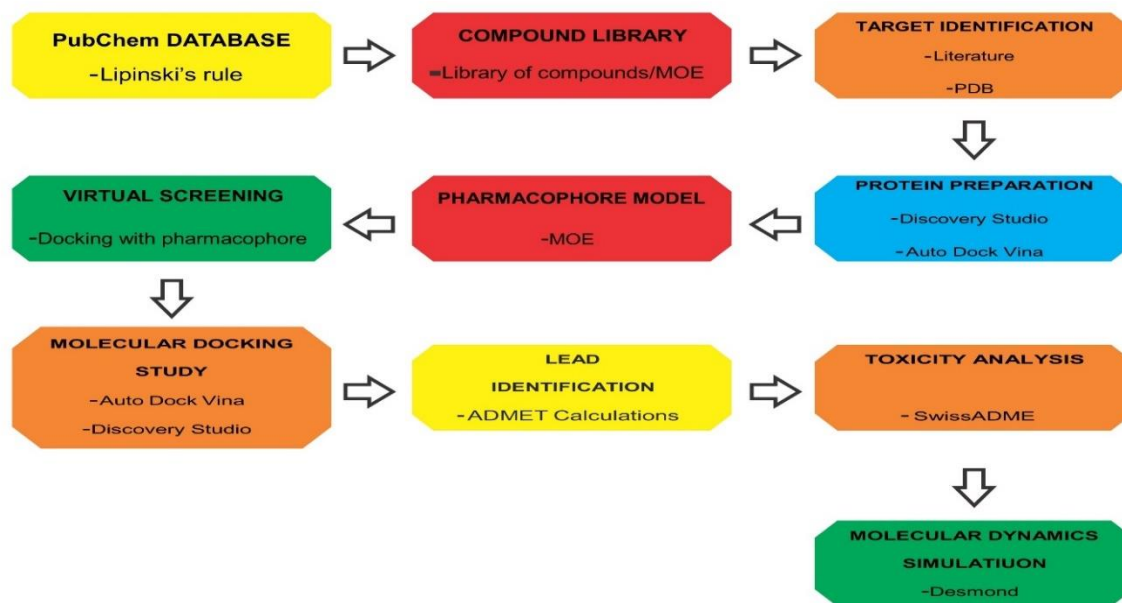


Figure 5. 1: The Flowchart of adopted computer-aided drug design (CADD) methodology

Pharmacophore model

Using molecular operating environment (MOE 2015.08) software, a pharmacophore model was constructed using the structures of the best binding ZINC compounds (ZINC_001456687980, ZINC_001445792073 and ZINC_000015276354) (from the first stages of this investigation). 46 compounds were energy minimized the MMFF94 force field, and 0.01 RMSD was maintained. The main features of the three ligands were isolated in order to mimic the compound likeness so that the phytochemicals could be further screened using this pharmacophore model. This method is helpful in categorizing structures, since there are many compounds with a range of structural diversity that has unique biochemical activities (Mannhold *et al.*, 2006).

Virtual screening

The phytochemicals were screened from the created MOE library using the constructed pharmacophore model, based on the developed pharmacophore query.

Docking procedures

A blind docking analysis between the phytochemicals and the protein was performed applying the AutoDock Vina v.1.2.0 using the following axis: center = 19.1732 center_y = 41.6246 center_z = 0.0403 size_x = 37.819451558 size_y = 37.9149091911 size_z = 58.311979351. The interactions between the ligand and the protein were observed on the Biovia discovery studio.

Lead compound identification

Compounds with the most active inhibition capability were discovered based on the best binding affinity and ligand-protein interactions (Kim and Skolnick, 2008). These compounds displayed high biochemical activity, since they derived from a pharmacophoric culture and were also compared with a control drug, Dolutegravir, a drug that is already in the market and in use in the South African health care sector (Plewczynski *et al.*, 2011).

Toxicity analysis

The drug like descriptions and toxicity characteristics were identified through SwissADME (Daina *et al.*, 2017). This calculation includes (ADMET calculation) adsorption, distribution, metabolism, excretion and toxicity (Xu *et al.*, 2012). The mutagenicity and the carcinogenicity were also determined for the lead compounds. The determination of these lead compounds was based on the pharmacophore the docking score and the ligand-protein interactions. Other related studies include, mol mw, donorHB, accptHB, QPlogPo/w, QPlogHERG, QPPCaco, QPlogBB and number of rings determination etc. (Low *et al.*, 2011); (Mulliner *et al.*, 2016, Alavijeh *et al.*, 2005).

Molecular dynamic simulation

After performing molecular docking interaction between the protein and the ligands was analysed. Molecular dynamics simulation was carried out using Desmond, a Package of Schrödinger LLC (Alavijeh *et al.*, 2005). The simulation was performed at 200ns in order to predict the ligand binding properties in the physiological environment. In this system that was prepared by the System Builder tool, the Solvent Model with an orthorhombic box was selected as Transferable

Intermolecular Interaction Potential 3 Points (TIP3P) using the OPLS_2005 force field throughout the simulation (Al-Shabib *et al.*, 2018), . The models were then relaxed before the simulation could resume and the counter ions were added in order to neutralize the model. The following parameters were used to mimic the physiological environment: a) Isothermal-Isobaric: Moles (N), pressure (P), and temperature (T) were conserved at 300K temperature, (these are The NPT ensemble) b) a 0.15 M salt (NaCl). c) (Deganutti *et al.*, 2020). In this simulation, trajectories were saved after every 100 ps for analysis.

5.3 RESULTS

5.3.1 Drug likeness analysis

Molecular properties that include the Lipinski rule, pharmacophore model, molecular weight, molecular docking and molecular dynamic simulation are some of the guided computational approaches that were applied in this study. The prediction using the Lipinski rule states that, a compound having either poor permeation or absorption is more likely to fail when it violates the Lipinski rule of 5 (Lipinski *et al.*, 1997). Using a free web SwissADME server <http://www.swissadme.ch/>, different chemical parameters were predicted. This includes, the pan assay interference compounds (PAINS), which predicts the structural alerts that has unstable, reactive and toxic elements in their biological components (Ferreira and Andricopulo, 2019). The essence of searching for a novel drug candidate relies on using the features of an existing drug that has been approved by the food and drug association (FDA) (Ekins *et al.*, 2014). One such tool is the application of pharmacophore modelling, which is a tool that mimics the features of the active biological components of a compound. This is done in order to have a library of drug likeness, which can be screened to find a suitable drug candidate (Ritchie *et al.*, 2011).

In Silico Analysis: Based virtual screening

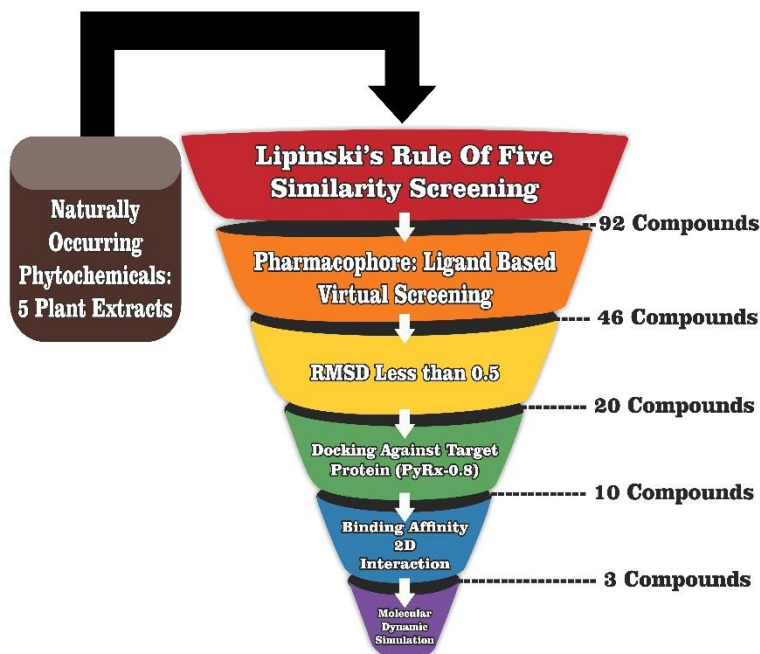


Figure 5. 2: Flowchart of the results obtained from computer aided drug design (CADD)

5.3.2 Construction of A ligand-based pharmacophore model

Pharmacophore technique based on the active ZINC ligands mimicry;- Ligand based pharmacophore modelling

The three best ZINC compounds which are zinc_1456687980, zinc_1445792073 and zinc_15276352 were used to construct a ligand-based pharmacophore on the molecular operating environment where a conformation database and the pharmacophore model were built. These compounds are the most active compounds screened from the ZINC database. A pharmacophore query was also built. This query was searched to show the features that would determine the likeness from the phytochemicals library.

Table 5. 1: Interaction between the trio hot-spot residues of three lead ZINC ligand's structure (from previous investigation): PDB ID: 1rv7

ligands hot-spot residue	Type of interaction	Residue
F1 = Aro	Alky/Pi Alkyl	23, 28, 82, and 84,
F2 = Acc	Van der Waals	32 and 84
F3 = Don/Acc	Pi Sigma	32

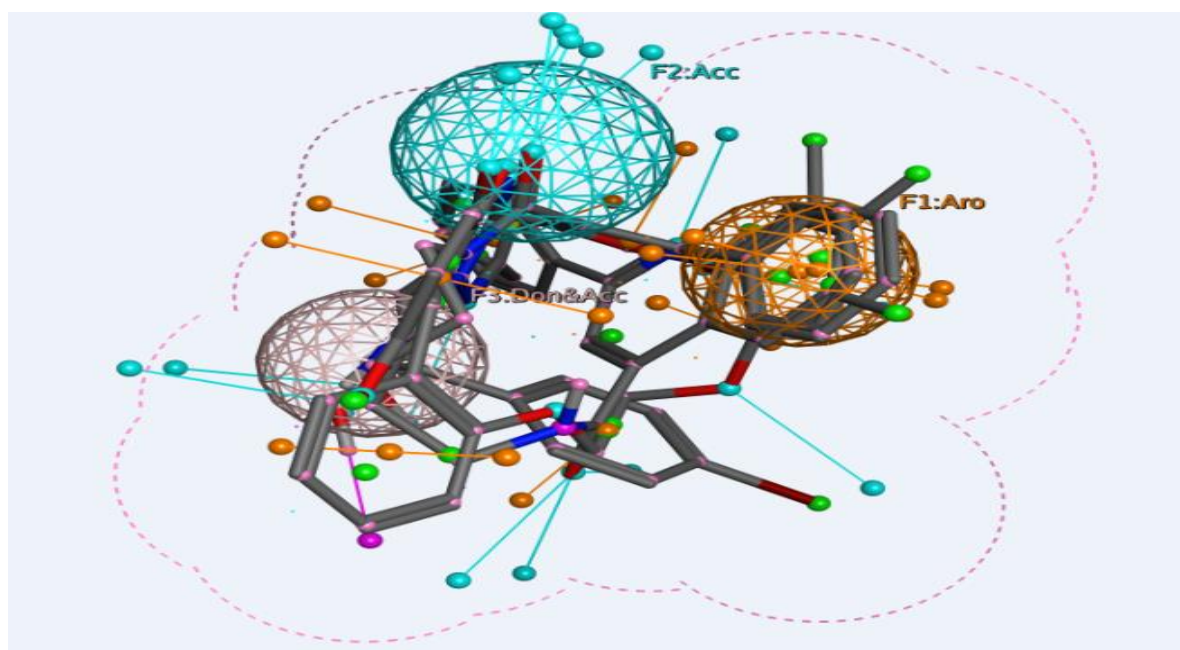


Figure 5. 3: A Pharmacophore model of three ZINC Ligands structure.

5.3.3 Virtual screening of HIV-1 Protease inhibitors using PubChem Databank, AutoDock Vina and the Lipinski's Rule of Five

Virtual Screening of some phytochemical plant extracts

Below is a list of the phytochemical extracts that were screened.

Table 5. 2: Phytochemical extracts that were screened

Name of Medicinal plant	Phytochemicals
a) <i>Ocimum santum</i>	1 α-Pinene 2 α-thujene 3 α-Camphene 4 β-Pinene 5 α-Myrcene 6. D-Limonene 7. Eucalyptol 8. cis-α-Terpineol 9. Sabinene 10. Borneol 11. Bornyl acetate 12. Camphor 13. Eugenol 14. Selinene 15. Caryophyllene oxide 16. Veridifloro 17. Cubenol 18. Caryophyllene oxide 19. Selinene 20. β guaiene 21. Phytol 22 Oleic acid 23. Aromadendrene oxide 24. β-gurjunene 25. Eicosane 26. Ethyl cyclohexenal ketone 27. n - butyl benzoate 28. 3 – Furaldehyde 29 Benzaldehyde 30 Heptanol 31 1-Octen-3-ol
b) <i>Carica papaya</i>	1. Decylene 2 Trans-Geranylacetone 3.Methyl tridecanoate 4.Palmitic acid 5.Myristic acid, methyl ester 6. Myristic acid 7.Palmitic acid, methyl ester 8.Hexadecanoic acid 9.Linolelaidic acid, methyl ester 10. Methyl cis-6-octadecenoate 11. Stearic acid, methyl ester 12. Oleic acid 13.Stearic acid 14. 15-Tetracosenoic acid 15. Methyl heptacosanoate 16. Trans-13-Docosenoic acid 17. Methyl erucate 18. Methyl behenate 19. Heneicosanoic acid, methyl ester 20. Farnesyl cyanide 21.Quercetin 22. Chlorogenic acid 23. 5,7 dimethoxycoumarin 24. Caffeic acid 25. Kaempferol 26. P-Coumaric acid 27. Protocatehuic acid.
c) <i>Persea americana</i>	1. C₂₁H₂₀O₁₀ 2. Afzelin 3. Catechin 4.Juglanin 5. Quercitrin 6. Epicatechin 7.

	Astragaline 8. Trans-tiliroside 9. Senecin 10. Juglalin 11. Quercetin 12. (6R,9R)-3-oxo- α -ionol-9-O- β -D-glucopyranoside 13. Roseoside 14. Ficumegasoside 15 Icariside B ₁ 16. (+)-lyoniresinol 17. (+)-isolariciresinol 9-O- β -D-xylopyranoside 18 Flavonol glycoside 19. Megastigmane glycosides 20 Lignans.
d) <i>Azadirachta indica</i> (neem)	1. Azadirachtin 2. Nimbolinin A. 3. Nimbin 4. Quenrcetin 5 Sitosterol 6. Salannin 7. Gedunin
e) <i>Spondias mombin</i>	1.) Geraniin 2.) 3, 5-di-O-galloyl-4- O-digalloylquinic acid 3.) 3-O-digalloyl-4,5-di-O-galloylquinic acid 4) 1,3,4,5-tetra-Ogalloylquinic acid 5) 2-O-Caffeoyl-(+)-allohydroxycitric acid. 6) 6-(8'Z,11'Z, 14'Zheptadecatrienyl)- salicylic acid 7) 6-(10'Zheptadecenyl)- salicylic acid.

Control drug = Dolutegravir (A recommended ARV in South African healthcare system) **The three phytochemicals that were not found in the ZINC Data Base:** All the phytochemicals were found in the ZINC Database except for three, which are: *Senecin*, *(6R,9R)-3-oxo- α -ionol-9-O- β -d-glucopyranoside* and *Megastigmane glycosides* from the plant *Persea Americana*. Their structures were constructed using the MOE.

Phytochemical library

A library of all 92 phytochemicals was created on the molecular operation environment. A total of 46 phytochemicals was identified after employing the Lipinski's rule of five. The retrieved extracts were compounds that met the following criteria: a) H-Bond Donor > 5, Molecular weight >500, Hydrogen Bond acceptors >10, LogP >5 (Lipinski *et al.*, 1997). This method is often employed in finding a category of inhibitors that has the same pharmacological properties.

Phytochemicals-Pharmacophore properties based on the RMSD

The compounds that had the best pharmacophoric hit were identified and filtered using an RMSD score. Twenty of the 46 compounds had a score less than that of 0.5 RMSD. These compounds were then chosen as candidate compounds for molecular docking studies.

5.3.4 Molecular docking studies

The compounds that were docked displayed different binding affinities and the one with the least binding affinity was picked and analysed further using AutoDock Vina and MGL tools. Their interaction were viewed on Biovia discovery studio. This observation tool assisted in visualizing the different poses and the residue interaction between the protein and the ligands. The chain structures of important active sites were also displayed. From the hierarchy Window the amino acid sequence for the complexes was determined. The Amino acids that make up the active site were identified and analysed, using discovery studio.

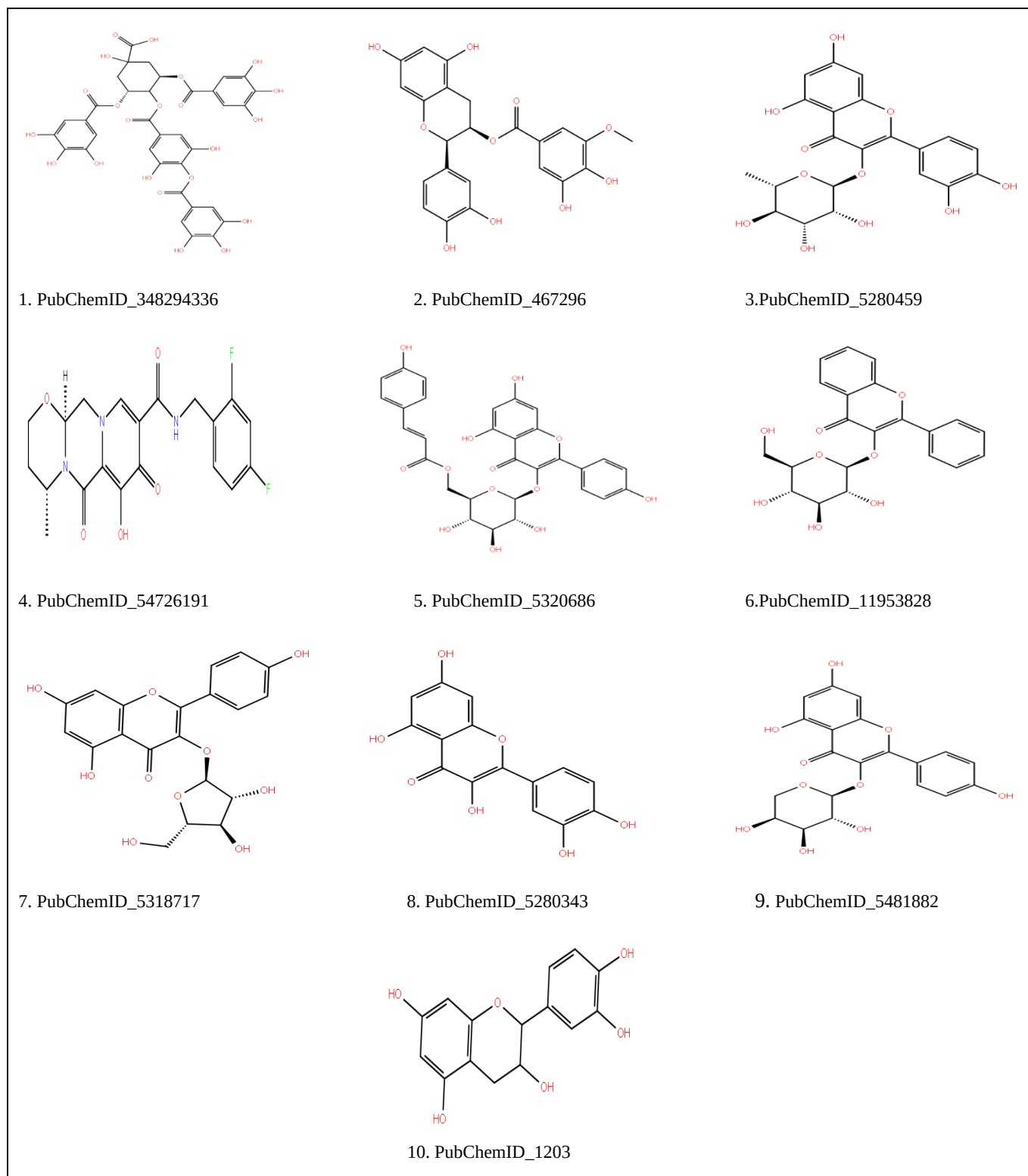


Figure 5. 4: The best 10 compounds from molecular docking

5.3.5 SwissADME studies

The results of the ligand properties were obtained from SwissADME server as shown on Table 5.2 and 5. 3. SwissADME is a web server that can be used to predict the properties of the compounds. Using SwissADME web server, lipophilicity was applied in predicting the physico-chemical properties for the ten best compounds. The five models of this parameter were determined as shown in Table 5.3. These models are, iLOGP, XLOGP, MLOGP, WLOGP and SILICOS-IT. The partition coefficient between *n*-octanol and water ($\log P_{o/w}$) is a classical descriptor for lipophilicity, meaning ($\log P_{o/w}$) reflects the relative solubility of the drug in *n*-octanol and water. All the compounds were found to be highly lipophilic. Based on these parameters, three compounds, namely epicatechin, juglanin and catechin were identified as the lead compounds. The lead compounds were identified as epicatechin, juglanin and catechin

Table 5. 3: Lipophilicity of the 10 best phytochemicals

PubChemID	M.W	Binding Affinity K _D	RMSD Score	iLOGP	XLOGP3	WLOGP	MLOGP	SILICOS-IT	C/SUS LOG P _{o/w}
1. PubChemID_348294336/3, 5-di-O-galloyl-4- O-digalloylquinic acid 3,5,4,4-Tetragga	800.59	-8.6	0.1377	1.15	1.83	1.25	-2.08	-1.17	0.20
*2.PubChemID_467296/epicatechin	456.40	-8	0.1800	2.35	1.85	2.51	0.53	1.59	1.77
3.PubChemID_5280459/quercitrin	448.28	-8	0.1819	1.60	0.86	0.49	-1.84	0.01	0.22
4.PubChemID_54726191/Dolutegrair	419.38	-8	contro l	2.10	2.44	1.66	1.05	1.85	1.82
5.PubChemID_5320686/transtiliosie	594.52	-7.7	0.1919	2.68	2.47	1.62	-1.04	1.56	1.46
6.PubChemID_11953828/flavonol g.	400.38	-7.6	0.2840	2.33	1.23	0.64	-0.60	1.30	0.98
*7. PubChemID_5318717/juglanin	418.35	-7.6	0.1820	1.76	1.33	0.39	-1.57	0.54	0.49
8. PubChemID_5280343/Quercetin	302.24	-7.2	0.2055	1.76	1.33	0.39	-1.57	0.53	0.49
9. PubChemID_5481882/juglalin	418.35	-7.1	0.1819	1.76	0.78	0.39	-157	0.38	0.38
*10. PubChemID_1203/catechin	290.27	-6.9	0.2055	1.46	0,36	1.22	0.24	0.98	0.85

* = The three lead compounds

The lipophilicity studies was used to analyze if the drug is suitable for polar or non-polar activity. Uncharged and nonpolar drugs dissolve in oily substances and in contrast, charged and polar drugs dissolves well in water, hence hydrophilic. In the body biomembranes are fatty acids, which are lipophilic barriers. When drugs are in an active state, they are usually lipophilic, hence the need to know the nature of the drug as either lipophilic or hydrophilic (Olasupo et al., 2020; Ejeh et al., 2021; Vivek and Swapna, 2021). These drug features have a great effect on the pharmacokinetic properties that include the adsorption, distribution, metabolism and its excretion (Chowdhury *et al.*, 2021) and therefore, there is a great need to analyse these properties in the quest of finding a drug with suitable properties. SwissADME shows two compounds with similar lipophilic properties, namely juglanin and quercetin. One passed the test and the other failed. The reason for this lies in the fact that quercetin has no rotational bonds.

Table 5. 4: Swiss ADME properties

Ligand ID	QPLogHERG	QPPCaco	2D Structure
PubChemID_467296	-6.139	36.507	
PubChemID_5318717	-5.391	69.995	
PubChemID_1203	-5.223	107.351	
PubChemID_54726191	-0.968	1341.852	

All the properties of the drugs that are well known and connected to the drug were retrieved from the server. These properties are adsorption, distribution, metabolism and excretion and these results are mainly based on the following parameters: lipophilicity, physicochemical, water solubility, pharmacokinetics, drug likeness and medicinal chemistry. The lead compounds were analysed, and their results matched against dolutegravir, a drug already in the market, and was used as a control compound for this study. The canonical smiles for epicatechin, juglanin, catechin and dolutegravir were retrieved from PubChem in order to run each compound to obtain their properties. According to the SwissADME bioavailability coloured radar, all the lead compounds were structure within the box, showed predictions of compounds that can used as suitable drugs candidate, hence a further analysis of results was carried out. Catechin with a molecular weight of 290.27 g/mol follows the Lipinski's rule with no violation as compared to epicatechin and juglanin that both compounds having 6 hydrogen donors and hence violating the Lipinski's rule. Epicatechin and catechin revealed the presence of a toxic substance called catechol, which when found in small amounts, has no dangerous effects, since it occurs naturally in fruits and vegetables (Harel and Mayer, 1971).

Table 5. 5: SwissADME/T properties for four Lead Compounds

Properties of lead compounds	Epicatechin	Juglanin	Catechin	Dolutegravir
Molecular weight	456.40	418.35	290.27	419.38
Toxicity	Catechol	Non-toxic	Catechol	Non-toxic
Rotatable bonds	5	4	1	4
Ring	4	4	3	4
Carcinogenetic	0.1800	0.1820	0.2055	Control
Solubility	Soluble	M/S	Soluble	Moderately Soluble
Binding Energy (K_D)	8	-7.6	-6.9	-8
Molar refractivity	114.51	102.17	74.33	104.48
H-bond Acceptor	10	10	6	7
H- bond Donor	6	6	5	2
GI absorption	Low	Low	High	High
BBB permeation	No	No	No	No
Abbott BA	0.55	0.55	0.55	0.55
Skin Permeation	-7.77	- 7.91 cm/s	-7.82 cm/s	-7.13 cm/s
PAINS	1 alert: catechol	0 alert	1 alert: catechol	0 alert

MW: Molecular Weight; HBA: Num. H-Bond Acceptors; HBD: Num. H-Bond Donors; MR: Molar Refractivity; P-M: Poor-Moderate; P: Poor; GI: Gastrointestinal; BBB: Blood–Brain Barrier; Average of five prediction, 1 violation: MLOGP > 4.15, PAINS: Pan-assay Interference Compounds. PPB (Plasma Protein Binding); BBB (Blood–Brain Barrier).

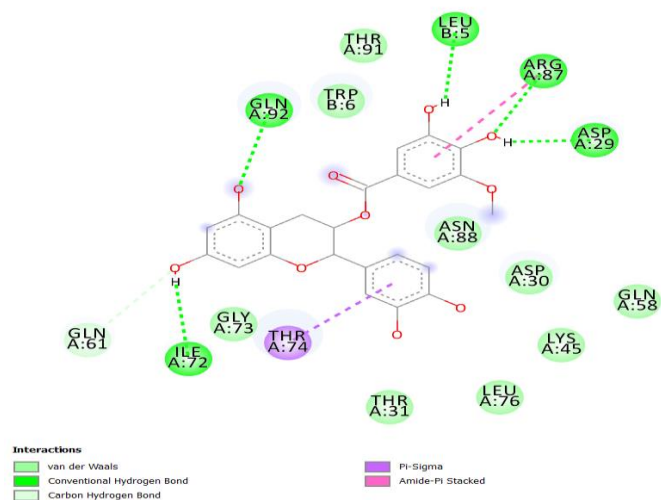
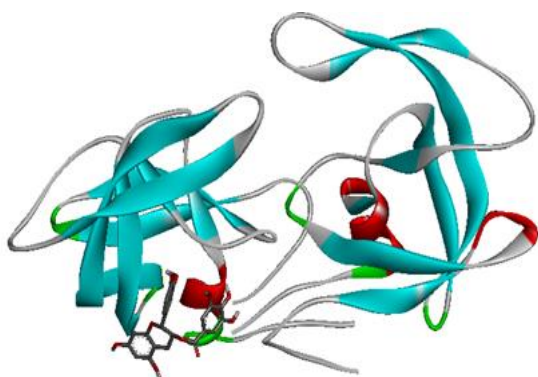
Catechin and dolutegravir demonstrates a high gastrointestinal absorption. Having size as its advantage, the distribution of catechin through the blood vessels can be predicted a compound that can possess minimal resistance to its target, bearing in mind the intravascular blood pressure. Capillaries are tiny blood vessels, permitting passage of blood cells, hence the need to find a suitable drug that can pass through the blood vessels without blocking, damaging or breaking the

cell membrane. Compared to dolutegravir, the lead compound in this study have no blood brain barrier (BBB), meaning their movement through the a highly selective semipermeable membrane of the brain is not restricted, whilst catechin shows low molar refractivity as compared to other compounds predicting how easy it be to infuse the dipole. Because these compounds show the ability to penetrate to form part of the digestive metabolites from the mouth to the gastrointestinal (GI) tract through the blood brain barrier, these are suitable for oral intake. The drug likeness of these drugs showed some similarities mainly for catechin and juglanin compared to certain components of dolutegravir. However, irrespective of epicatechin binding outside of the active site of the protein, it still possess important feature in some instances, which are similar to that of the control. Epicatechin and dolutegravir have the same binding affinity of ($-8 K_D$), and they also have the same number of rings. In some instances epicatechin falls within the range of dolutegravir than the rest of the other compounds. Some of its properties might be of great importance in understanding the binding affinity against this disease. Water-soluble compounds plays a major role in determining the outcome of drug development and its activities. Since some of these drugs are administered orally, they their success rely mainly on the manner in which they dissolve when administered orally. Epicatechin and catechin predicted a soluble state when dissolved in water. The decimal logarithms of the molar solubility in water ($\log S$), was predicted for all the four compounds, with juglanin and dolutegravir showing a moderate solubility when dissolved in water.

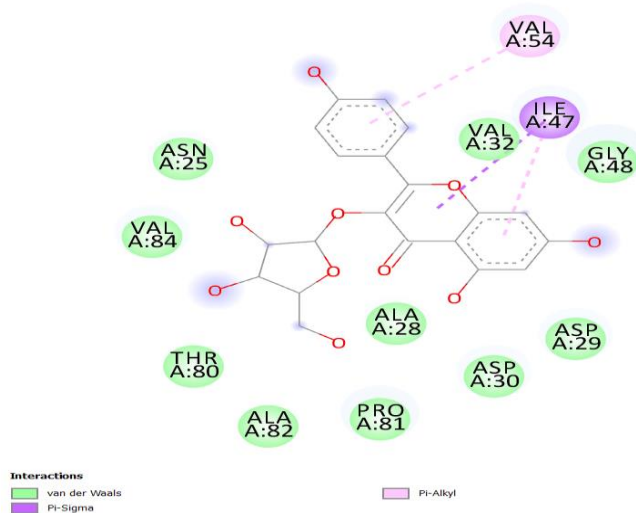
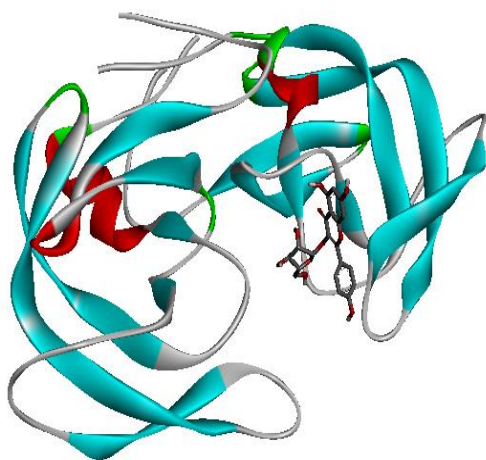
5.3.6 Docking analysis for lead compounds

A detailed interaction analysis between the inhibitors and the protein residue was made. Table 5.6 above shows the orientation and interaction of these lead phytochemicals compared with dolutegravir. It can be observed that, juglanin and dolutegravir binds inside the protein at the same binding site together with dolutegravir. These drugs displays a good pose inside the catalytic pocket of the receptor. Juglanin displays the following hydrophobic interactive with the following residues: Asn25, Ala82, Val84, Pro81, Thr80, Ala28, Gly48 Asp30, Asp29, and Val32, whiles dolutegravir shows the following hydrophobic interaction Pro79, GlyB27, GlyA27, Ala28, Pro81, ThrB80, ThrA80, and Ile47. The interaction between juglanin and the protein shows no formation

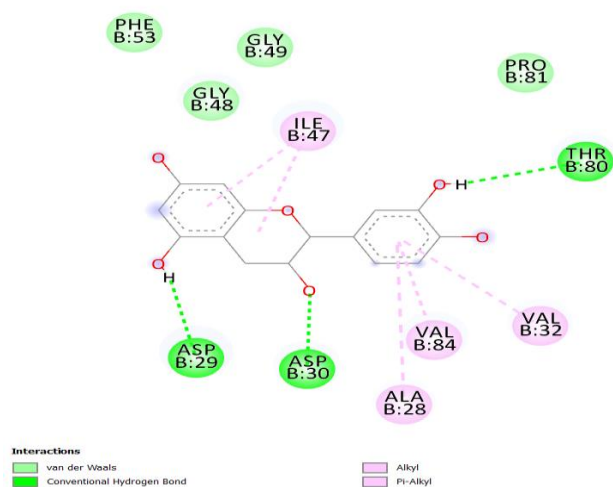
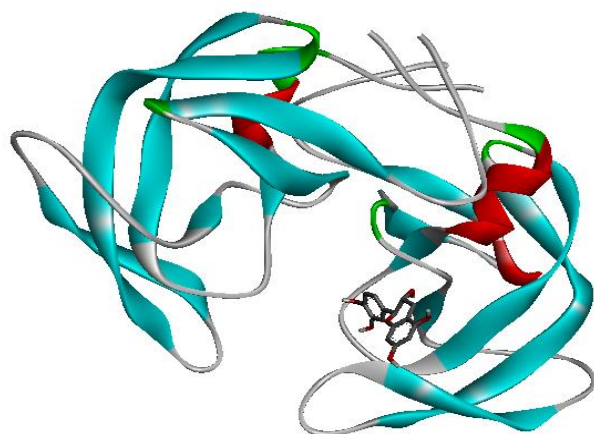
of hydrogen bonds, while dolutegravir forms two hydrogen bond with the protein with residue: Asn25, Asn25. The efficacy of the drug is evaluated based on the formed hydrogen bonds (Raj *et al.*, 2020). The Van der Waals, which includes the hydrogen bonds, are one of the most important binding forces in protein-ligand interaction.



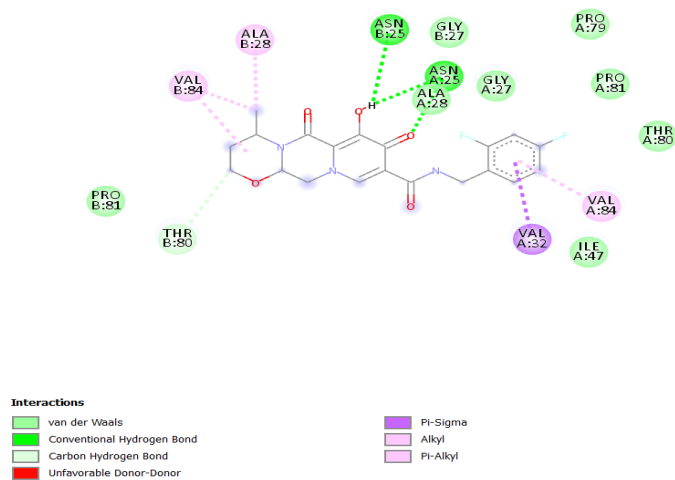
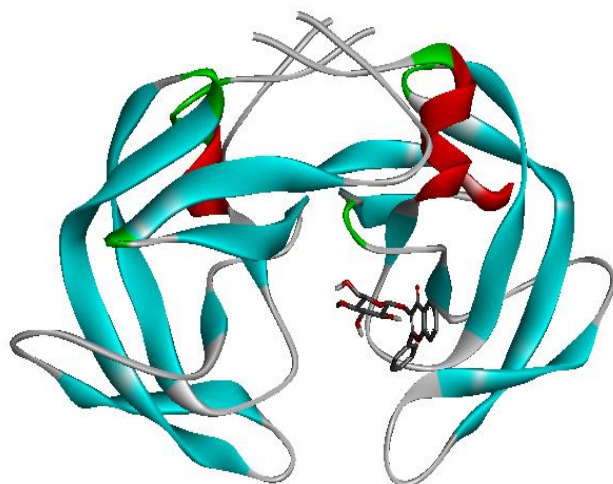
A) PubChemID_467296/ Epicatechin



B) PubChemID_5318717/ Juglanin



C. PubChemID_1203/ Catechin



D) PubChemID_54726191/Dolutegravir

Figure 5. 5: Binding interaction of epicatechin, juglanin and dolutegravir

Catechin displays and the protein form three hydrogen bond on residue Thr80, Asp30, Asp29 with the following hydrophobic interaction Phe53, Gly48, Gly49, Val32. With Epicatechin binding outside of the binding pocket it has hydrophobic interactions with the following residues: Thr91, Trp6, Gln61, Gly73, Thr31, Leu76, Lys48, Gln58, Asn30, and Asn88. Interestingly, epicatechin forms five hydrogen bonds with the following residue: Leu5, Gln92, Ile72, Asp29, and Arg87, and has a single pi (π) bond on residue Thr74, and Val54 for juglanin. In most instances, these pi bonds are not stronger than sigma (σ) bonds, they are usually weaker. In cooperative stability contribution, the stability added by pi bonds is lesser than that of sigma bond (Mhatre *et al.*, 2021), (Iheagwam *et al.*, 2019). Catechin and dolutegravir, with four pi bonds and residue Ile47, Ala28 Val84 Val32 and a single pi bond on residue Val84, respectively has contributes independently to their association with the protein.

Table 5. 6: The docked Van der Waals, Conventional Hydrogen Bonds, Pi-Sigma, Alkyl, Pi-Alkyl between 1rv7 protein and best three compounds

Compounds	Van der Waals	Conventional H bonds	pi-sigma	Alkyl	pi-alkyl
A) Epicatechin	THR91, TRP6, GLN61, GLY73, THR31, LEU76, LYS48, GLN58, ASN30, ASN88	LEU5, GLN92, ILE72, ASP29, ARG87	THR74	None	None
B) Juglanin	ASN25, ALA82, VAL84, PRO81, THR80, ALA28, GLY48 ASP30, ASP29	None	VAL54	None	None
C. Catechin	PHE53, GLY48, GLY49, PRO81	THR80, ASP30, ASP29		ILEB47, VALB84, AIB28, VALB32	

D) Dolutegravir	PRO79, GLY27, GLY27, VAL84, ALA28, PRO81, ALA28, PRO81, ALA28, THR80, ILE47, THR80, THR80, ILE47, THR80	ASN25, ASN25	VAL32	VALA84	None
------------------------	--	--------------	-------	--------	------

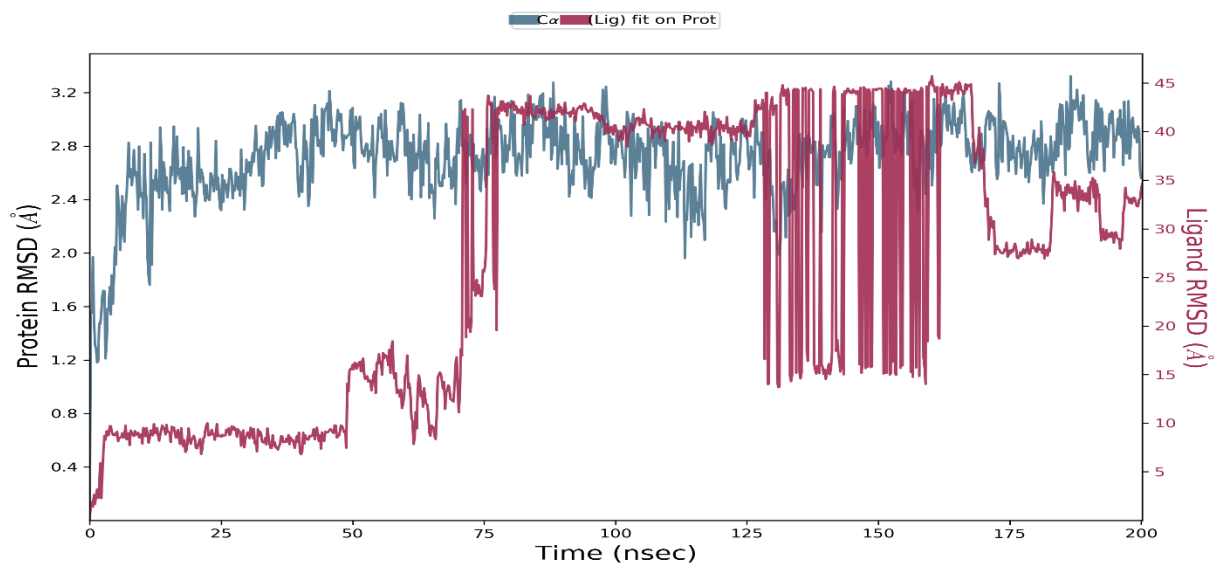
5.4 MOLECULAR DYNAMICS SIMULATION

A package of the Schrödinger, called Desmond LLC, was used for molecular dynamics simulation for the three lead phytochemical against the control Dolutegravir. Dolutegravir is a drug that is been administered to HIV patients in South African health care sector. The simulations for Epicatechin, Juglanin, Catechin and Dolutegravir were performed for 200 nanosecond to predict a ligand binding status along with their trajectories that were analysed using the C- α atom. In molecular dynamics simulation, the RMSD was measures between the ligand and the protein. It was used to analyse the protein-ligand association and movements of all the simulated complexes in order to observe the structural stability during the simulation. For the complex to be regarded as functional its movements within the hydrated environment plays a vital role in the determination of the stability of the complex and catechin displayed such interactions as compared to other compounds.

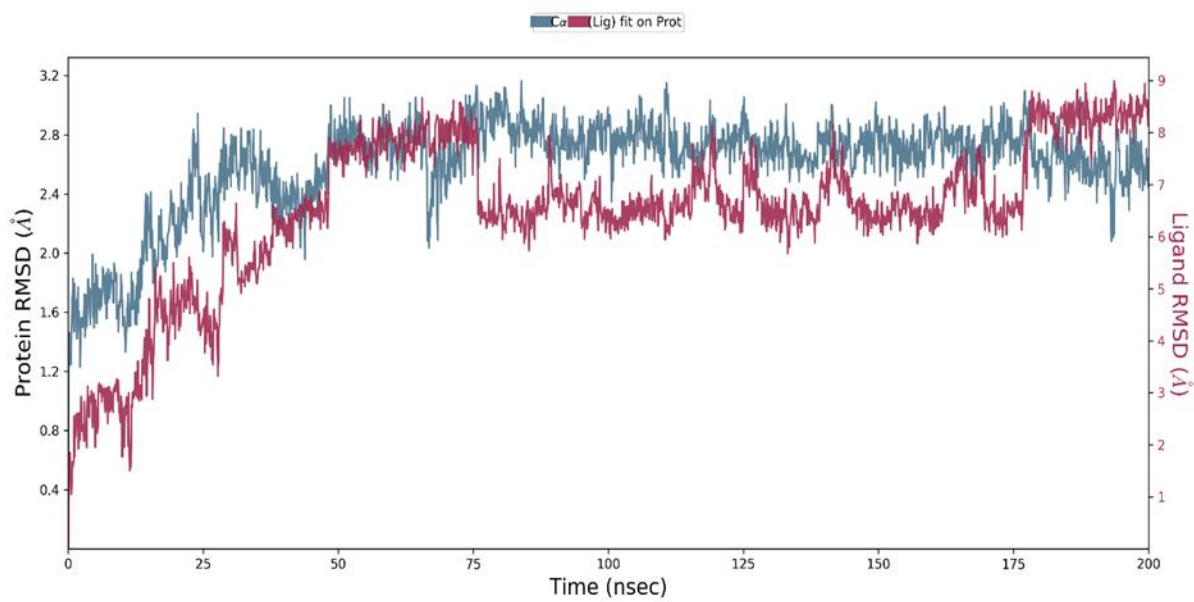
Figure 5.6 shows the root mean square deviation for epicatechin, juglanin, catechin and dolutegravir. Catechin at 25ns attained a value of 1.3 Å, while epicatechin displayed a value 0,7 Å from 5- 50 ns. Looking into the RSMD pattern of dolutegravir, a three-phase pattern is observed with an increase of the RMSD value, which one trait the compounds exhibit. This pattern can be observed on the other the other compounds in study, as either a three-phase pattern or a four-phase pattern, usually in an increasing RMSD value.

The protein RMSD maintains equilibrium at the ranges of 50ns, for epicatechin, juglanin and catechin. At time 50ns the protein RMSD for the interaction between juglanin and catechin reaches

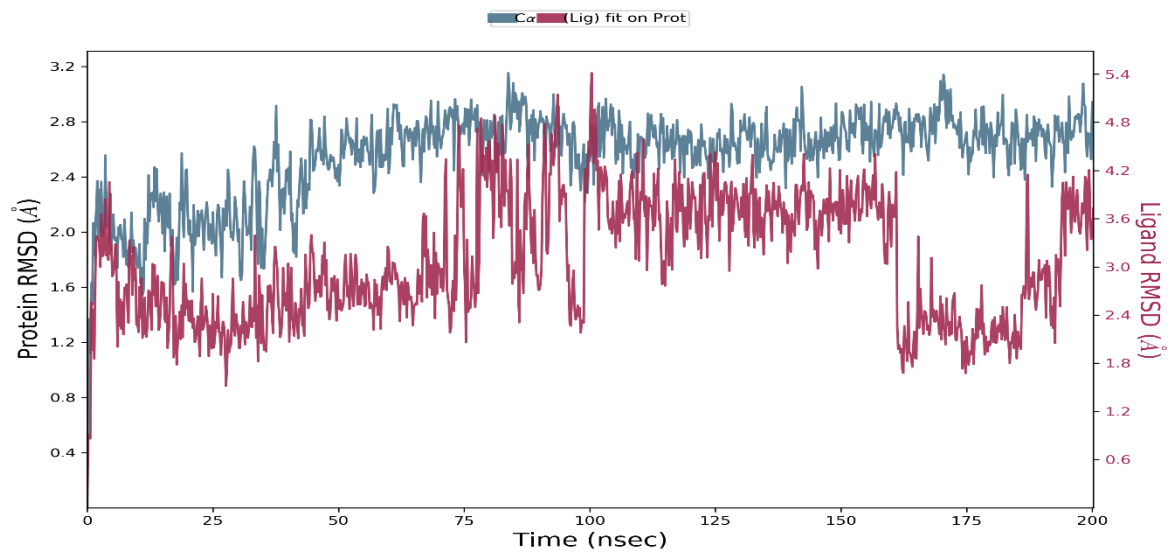
equilibrium until the end of the simulation, while equilibrium is reached at 115ns and 25ns for dolutegravir and epicatechin, with vast deviation between epicatechin and the protein between 0 – 75ns. Protein RMSD reached equilibrium at 2.8 Å at time 50ns for both epicatechin and juglanin. Epicatechin and catechin shows the lowest RMSD of 3,5 Å and 3,6 Å respectively, at the end of the simulation, with two main observations picked from the molecular docking of these compounds, which are: epicatechin binding outside of the binding pocket and catechin binding at a location similar to that of dolutegravir. Dolutegravir had an RMSD of 10,5 Å at the end of the simulation. Figure 5.6 also shows the hydrogen bonds and the bond length that were formed by the compounds under study and the control. Catechin in one of the having the longest bond length of 1.6 at residue Asp30 and the second highest bond length that follows, is 0.9 at residue Asp29, while dolutegravir display its highest bond length of 0.2 at residue Asn25.



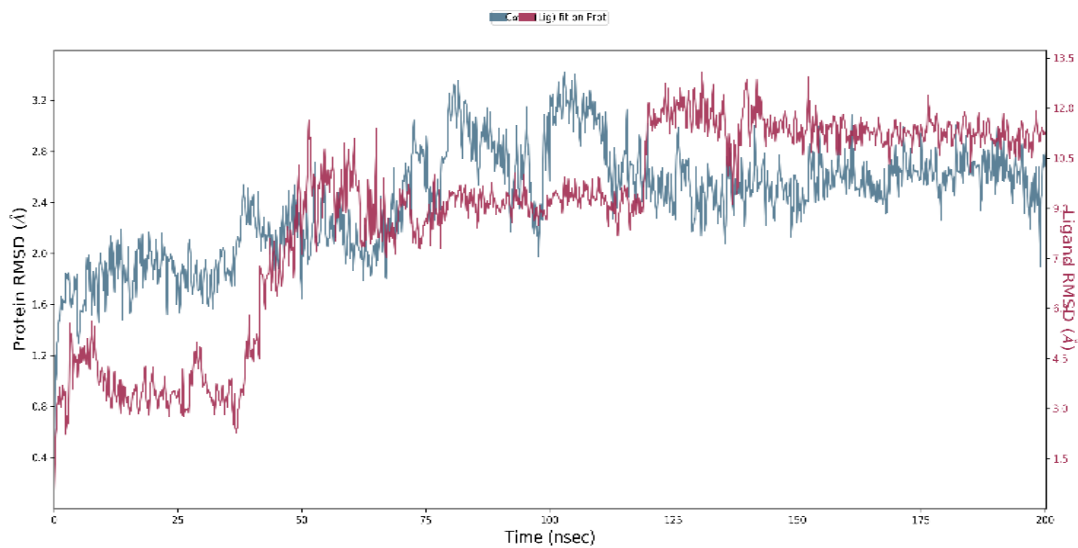
Epicatechin



Juglanin



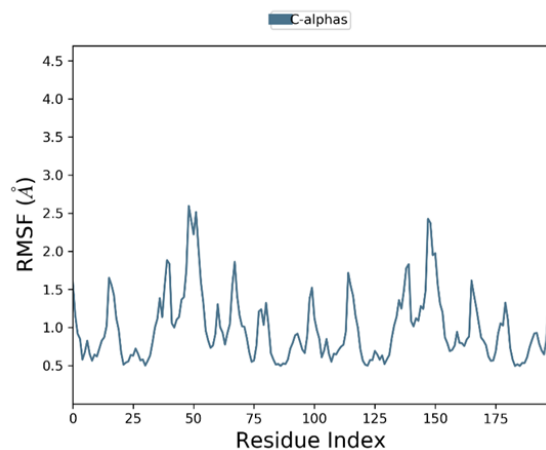
Catechin



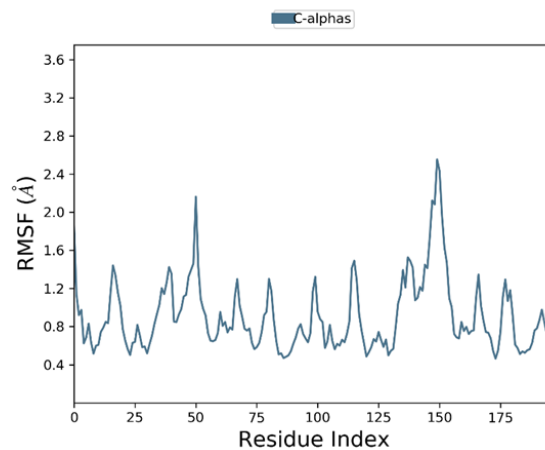
Dolutegravir

Figure 5. 6: Molecular dynamics Simulation diagrams at 200ns

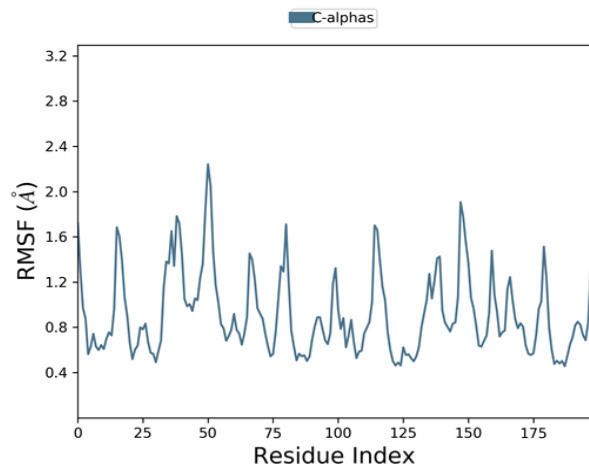
In Figure 5.6 the diagram/s A, B and C: is obtained from Molecular dynamics Simulation and is of the root mean square deviation of the C- α atoms of the protein and the ligand with time in nano seconds (nsec). In this diagram the left γ - axis shows the variation of protein RMSD through time whereby, the right γ - axis shows the variation of the ligand RMSD through time. Interestingly, Figure 5.6 shows similar trajectories between juglanin and dolutegravir when the trajectories are saved after every 10ps. This attest to the prediction that juglanin has some properties that resemble that of the control. However, the fluctuations between catechin are not as high as the one observed in the complex between dolutegravir and 1rv7. Despite the similarity of trajectories between the novel compounds and dolutegravir, the fluctuations between the control and the protein remains a bit high when compared to the rest of the compounds in study. In order to assess which residues are involved in the protein or complex changes in the structure, we analyze the root mean square fluctuation (RMSF), Figure 5.7. The low RMSF residue are the ones that are more stable. This is because the motions of the residue atoms during modelling are limited. The four compounds are bound to the protein with a range between 2,2 Å RMSF value and 3,5 Å and an average of 2,7 Å displaying their interactive stability. These phytochemicals also exhibit similar patterns of fluctuation at specific residue during time period of 50n and 150ns. Figure 5. 8 is the plots that summarizes the SSE distribution by residue position through the protein structure. In this plots, the potein SSE was monitored throughout the simulation, categorized as beta-strands, alpha-strands and random coils. Mainly, inside the core of the protein there are Alpha-helices, which have hydrophobic residues. These proteins are sometimes referred to as transmembrane proteins. The Beta strands have polar amino acids and parttens of hydrophobic. A statistical distribution of all chains is indicated by the white space in all of the plots. These is due to the fact that a random coil is a polymer. There are more fluctutions of the beta-strands as compared to the alpha-strands. And, the SSE is more rigid than area where there is an unstructured part of the protein, therefore they fluctuate less than the loop region. The protein-ligand contacts (Figure 5.9), shows the multiple ligand contacts through various inteaction. These inteactions include the hydrogen bonding interaction (which is very important in drug design due to their influnce in adsorption, metabolization and drug specificity), hydrogen interaction, ionic interaction and water bridges (which is seen in high number). The same protein residues can be observed during the simulation.



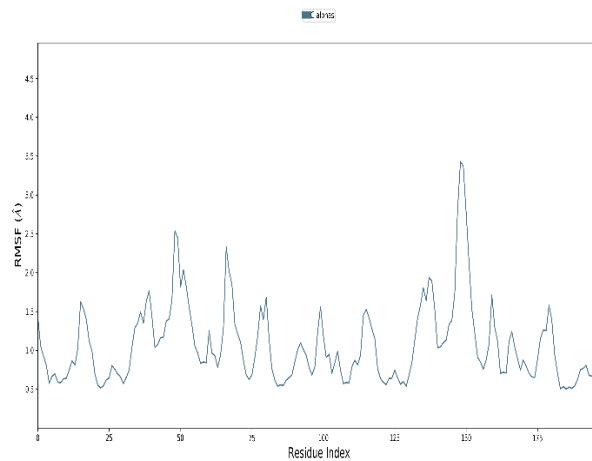
Epicatechin



Juglanin

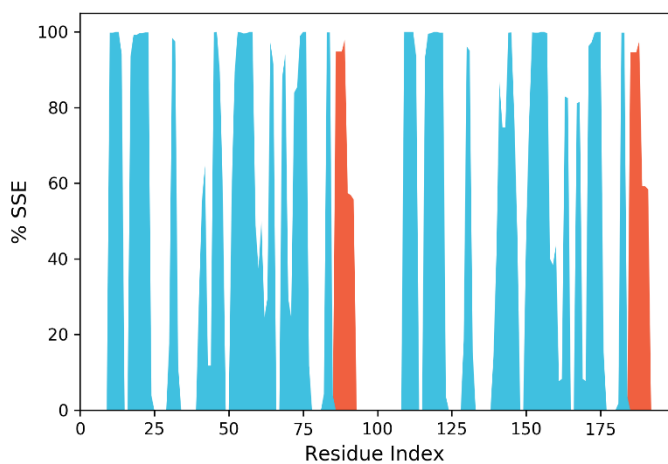


Catechin

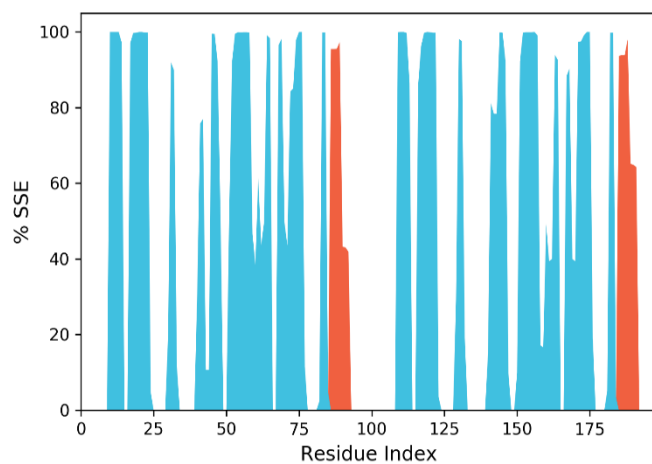


Dolutegravir

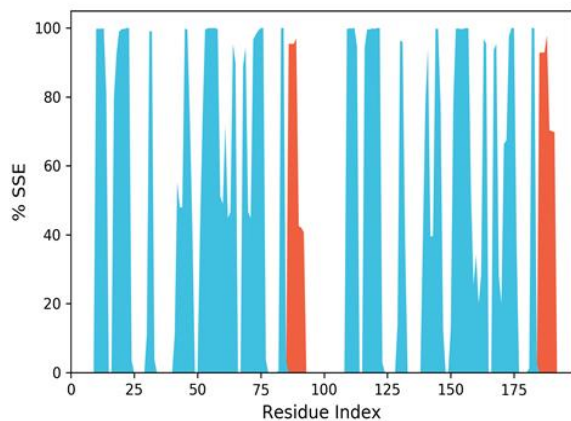
Figure 5. 7: Residue wise Root Mean Square Fluctuation (RMSF)



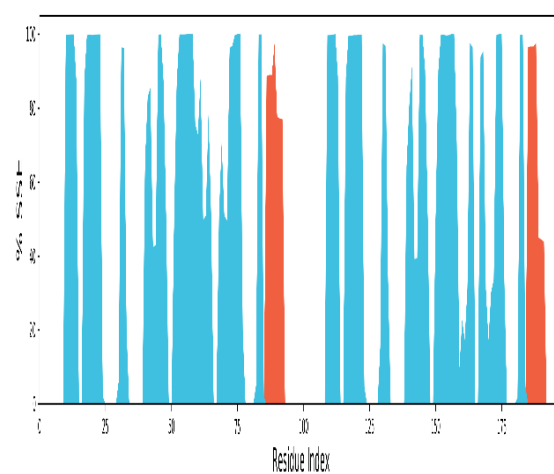
Epicatechin



Juglanin

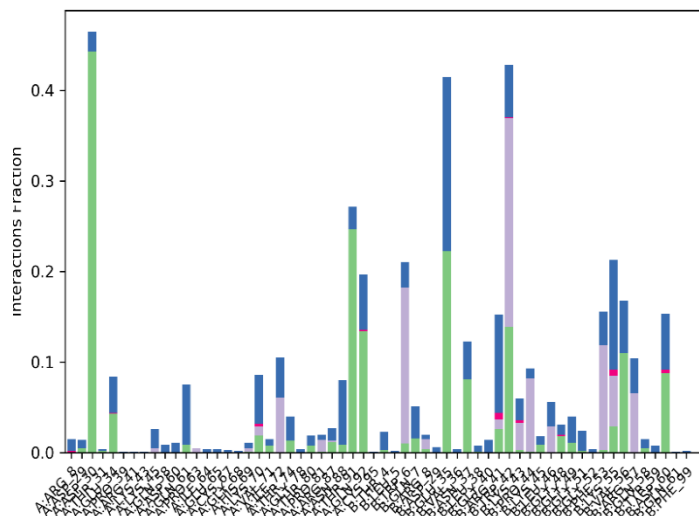


Catechin

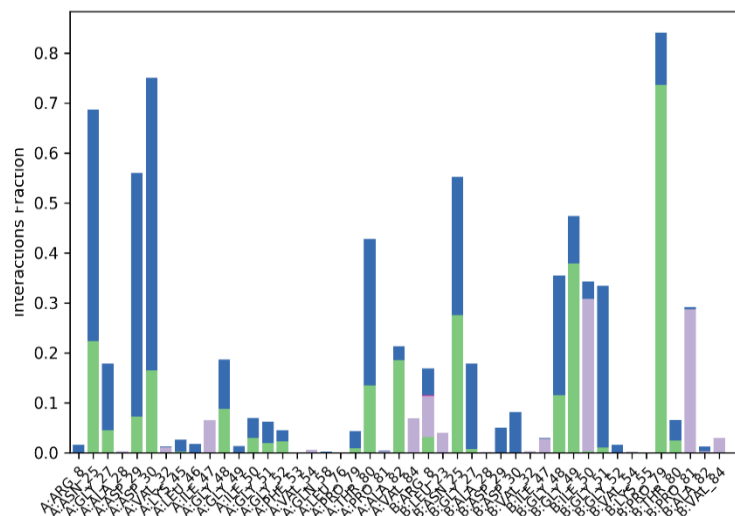


Dolutegravir

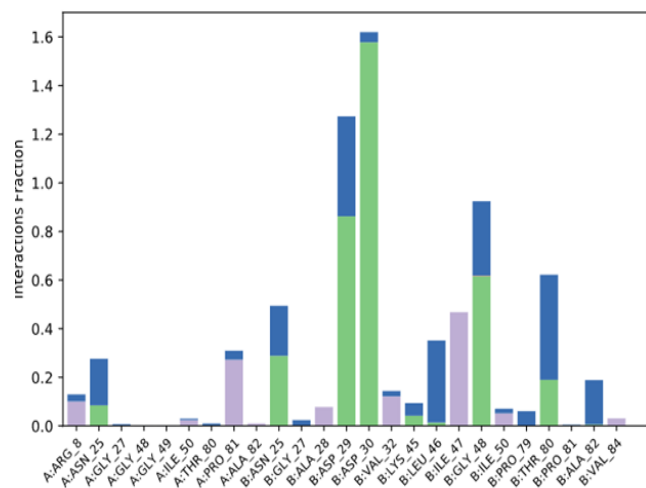
Figure 5. 8: Protein Secondary Structures element distribution by residue index throughout the protein structure



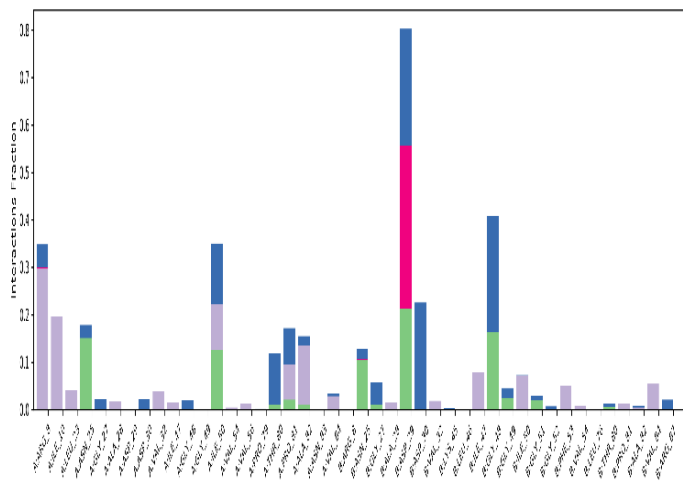
Epicatechin



Juglanin



Catechin



Dolutegravir

Figure 5. 9: Protein-ligand contact histogram

REFERENCE LIST

- A PUBLICATION OF THE SOUTHERN AFRICAN HIV CLINICIANS SOCIETY. 2021. *HIV NURSING MATTERS* [Online]. Available: <https://sahivsoc.org/Files/Nursing%20matters%20May%202017> [Accessed].
- ADENIJI, S. E., ARTHUR, D. E. & OLUWASEYE, A. 2020. Computational modeling of 4-Phenoxynicotinamide and 4-Phenoxypyrimidine-5-carboxamide derivatives as potent anti-diabetic agent against TGR5 receptor. *Journal of King Saud University-Science*, 32, 102-115.
- AL-SHABIB, N. A., KHAN, J. M., MALIK, A., ALSENAIDY, M. A., REHMAN, M. T., ALAJMI, M. F., ALSENAIDY, A. M., HUSAIN, F. M. & KHAN, R. H. 2018. Molecular insight into binding behavior of polyphenol (rutin) with beta lactoglobulin: Spectroscopic, molecular docking and MD simulation studies. *Journal of Molecular Liquids*, 269, 511-520.
- ALAVIJEH, M. S., CHISHTY, M., QAISER, M. Z. & PALMER, A. M. 2005. Drug metabolism and pharmacokinetics, the blood-brain barrier, and central nervous system drug discovery. *NeuroRx*, 2, 554-571.
- BAELL, J. B. & HOLLOWAY, G. A. 2010. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *Journal of medicinal chemistry*, 53, 2719-2740.
- BORHANI, D. W. & SHAW, D. E. 2012. The future of molecular dynamics simulations in drug discovery. *Journal of computer-aided molecular design*, 26, 15-26.
- BRENK, R., SCHIPANI, A., JAMES, D., KRASOWSKI, A., GILBERT, I. H., FREARSON, J. & WYATT, P. G. 2008. Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. *ChemMedChem*, 3, 435.

CHENG, F., LI, W., ZHOU, Y., SHEN, J., WU, Z., LIU, G., LEE, P. W. & TANG, Y. 2012. admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties. ACS Publications.

CHOWDHURY, S. A., KIBRIA, S. M. & ALAM, M. R. 2021. DFT Based Pharmacokinetic, Molecular Docking, and ADMET Studies of Some Glucopyranoside Esters. *Journal of Applied Science & Process Engineering*, 8, 671-683.

DAINA, A., MICHIELIN, O. & ZOETE, V. 2017. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific reports*, 7, 1-13.

DE LIMA, E. Q., DE OLIVEIRA, E. & DE BRITO, H. R. 2016. Extraction and characterization of the essential oils from *Spondias mombin* L.(Caj), *Spondias purpurea* L.(Ciriguela) and *Spondias* ssp (Cajarana do sertão). *African Journal of Agricultural Research*, 11, 105-116.

DEGANUTTI, G., MORO, S. & REYNOLDS, C. A. 2020. A Supervised Molecular Dynamics Approach to Unbiased Ligand–Protein Unbinding. *Journal of chemical information and modeling*, 60, 1804-1817.

EJEH, S., UZAIRU, A., SHALLANGWA, G. A. & ABECHI, S. E. 2021. Computational insight to design new potential hepatitis C virus NS5B polymerase inhibitors with drug-likeness and pharmacokinetic ADMET parameters predictions. *Future Journal of Pharmaceutical Sciences*, 7, 1-13.

EKINS, S., FREUNDLICH, J. S. & COFFEE, M. 2014. A common feature pharmacophore for FDA-approved drugs inhibiting the Ebola virus. *F1000Research*, 3.

FERREIRA, L. L. & ANDRICOPULO, A. D. 2019. ADMET modeling approaches in drug discovery. *Drug discovery today*, 24, 1157-1165.

GAULTON, A., KALE, N., VAN WESTEN, G. J., BELLIS, L. J., BENTO, A. P., DAVIES, M., HERSEY, A., PAPADATOS, G., FORSTER, M. & WEGE, P. 2013. The ChEMBL bioactivity database: an update. *Scientific Data*, 2, 150032.

GLOBAL INFORMATION AND EDUCATION ON HIV AND AIDS. 2020. *ABOUT HIV & AIDS* [Online]. Available: <https://www.avert.org/professionals/hiv-around-world/sub-saharan-africa/south-africa>. [Accessed].

GREENE, W. C. 2007. A history of AIDS: looking back to see ahead. *European journal of immunology*, 37, S94-S102.

GRINTER, S. Z. & ZOU, X. 2014. Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design. *Molecules*, 19, 10150-10176.

HAREL, E. & MAYER, A. 1971. Partial purification and properties of catechol oxidases in grapes. *Phytochemistry*, 10, 17-22.

HAUSER, A., KUSEJKO, K., JOHNSON, L. F., GÜNTARD, H. F., RIOU, J., WANDELER, G., EGGER, M. & KOUYOS, R. D. 2020. Impact of scaling up dolutegravir on antiretroviral resistance in South Africa: A modeling study. *PLoS medicine*, 17, e1003397.

IHEAGWAM, F. N., OGUNLANA, O. O., OGUNLANA, O. E., ISEWON, I. & OYELADE, J. 2019. Potential anti-cancer flavonoids isolated from *Caesalpinia bonduc* young twigs and leaves: molecular docking and in silico studies. *Bioinformatics and Biology Insights*, 13, 1177932218821371.

KARPLUS, M. & MCCAMMON, J. A. 2002. Molecular dynamics simulations of biomolecules. *Nature structural biology*, 9, 646-652.

KIM, R. & SKOLNICK, J. 2008. Assessment of programs for ligand binding affinity prediction. *Journal of computational chemistry*, 29, 1316-1331.

KOHL, N. E., EMINI, E. A., SCHLEIF, W. A., DAVIS, L. J., HEIMBACH, J. C., DIXON, R., SCOLNICK, E. M. & SIGAL, I. S. 1988. Active human immunodeficiency virus protease is required for viral infectivity. *Proceedings of the National Academy of Sciences*, 85, 4686-4690.

KONVALINKA, J., KRÄUSSLICH, H.-G. & MÜLLER, B. 2015. Retroviral proteases and their roles in virion maturation. *Virology*, 479, 403-417.

KRÄUSSLICH, H.-G., INGRAHAM, R. H., SKOOG, M. T., WIMMER, E., PALLAI, P. V. & CARTER, C. A. 1989. Activity of purified biosynthetic proteinase of human immunodeficiency virus on natural substrates and synthetic peptides. *Proceedings of the National Academy of Sciences*, 86, 807-811.

LIPINSKI, C. A., LOMBARDO, F., DOMINY, B. W. & FEENEY, P. J. 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, 23, 3-25.

LIU, Z., LI, Y., HAN, L., LI, J., LIU, J., ZHAO, Z., NIE, W., LIU, Y. & WANG, R. 2015. PDB-wide collection of binding data: current status of the PDBbind database. *Bioinformatics*, 31, 405-412.

LOUIS, J. M., NASHED, N. T., PARRIS, K. D., KIMMEL, A. R. & JERINA, D. M. 1994. Kinetics and mechanism of autoprocessing of human immunodeficiency virus type 1 protease from an analog of the Gag-Pol polyprotein. *Proceedings of the National Academy of Sciences*, 91, 7970-7974.

LOW, Y., UEHARA, T., MINOWA, Y., YAMADA, H., OHNO, Y., URUSHIDANI, T., SEDYKH, A., MURATOV, E., KUZ'MIN, V. & FOURCHES, D. 2011. Predicting drug-induced hepatotoxicity using QSAR and toxicogenomics approaches. *Chemical research in toxicology*, 24, 1251-1262.

MAHAPATRA, S., YOUNG, C. Y., KOHLI, M., KARNES, R. J., KLEE, E. W., HOLMES, M. W., TINDALL, D. J. & DONKENA, K. V. 2012. Antiangiogenic effects and therapeutic targets

of *Azadirachta indica* leaf extract in endothelial cells. *Evidence-Based Complementary and Alternative Medicine*, 2012.

MANNHOLD, R., KUBINYI, H. & FOLKERS, G. 2006. *Pharmacophores and pharmacophore searches*, John Wiley & Sons.

MHATRE, S., NAIK, S. & PATRAVALE, V. 2021. A molecular docking study of EGCG and theaflavin digallate with the druggable targets of SARS-CoV-2. *Computers in biology and medicine*, 129, 104137.

MULLIGAN, N., BROOKIE, M., JIAJIA, W., CAPPARELLI, E. V., ALICE, S., EMILY, B., BUSCHUR, S. L., ACOSTA, E. P., SMITH, E. & CHAKHTOURA, N. 2018. Dolutegravir pharmacokinetics in pregnant and postpartum women living with HIV. *AIDS (London, England)*, 32, 729.

MULLINER, D., SCHMIDT, F., STOLTE, M., SPIRKL, H.-P., CZICH, A. & AMBERG, A. 2016. Computational models for human and animal hepatotoxicity with a global application scope. *Chemical research in toxicology*, 29, 757-767.

O'LEARY, N. A., WRIGHT, M. W., BRISTER, J. R., CIUFO, S., HADDAD, D., MCVEIGH, R., RAJPUT, B., ROBERTSE, B., SMITH-WHITE, B. & AKO-ADJEI, D. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic acids research*, 44, D733-D745.

OLASUPO, S. B., UZAIRU, A., ADAMU, G. S. & UBA, S. 2020. Computational modeling and pharmacokinetics/ADMET study of some arylpiperazine derivatives as novel antipsychotic agents targeting depression. *Chemistry Africa*, 3, 979-988.

OROSZLAN, S. & LUFTIG, R. 1990. Retroviral proteinases. *Retroviruses*, 153-185.

OWOLABI, M. A., JAJA, S. I. & COKER, H. A. 2005. Vasorelaxant action of aqueous extract of the leaves of *Persea americana* on isolated thoracic rat aorta. *Fitoterapia*, 76, 567-573.

PLEWCZYNSKI, D., ŁAŻNIEWSKI, M., AUGUSTYNIAK, R. & GINALSKI, K. 2011. Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *Journal of computational chemistry*, 32, 742-755.

RACKAL, J. M., TYNAN, A. M., HANDFORD, C. D., RZEZNIKIEWIZ, D., AGHA, A. & GLAZIER, R. 2011. Provider training and experience for people living with HIV/AIDS. *Cochrane Database of Systematic Reviews*.

RAJ, S., SASIDHARAN, S., BALAJI, S. & SAUDAGAR, P. 2020. An overview of biochemically characterized drug targets in metabolic pathways of Leishmania parasite. *Parasitology Research*, 119, 2025-2037.

RAJINIKANTH, R., GOVARTHANAN, M., PAUL, A., SELVANKUMAR, T. & SENGOTTAIYAN, A. 2013. Antioxidant potential and secondary metabolites in Ocimum sanctum L. at various habitats. *Journal of Medicinal Plants Research*, 7, 706-712.

RITCHIE, T. J., ERTL, P. & LEWIS, R. 2011. The graphical representation of ADME-related molecule properties for medicinal chemists. *Drug Discovery Today*, 16, 65-72.

SABE, V. T., NTOMBELA, T., JHAMBAL, L. A., MAGUIRE, G. E., GOVENDER, T., NAICKER, T. & KRUGER, H. G. 2021. Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review. *European Journal of Medicinal Chemistry*, 113705.

SHAFER, R. & VUITTON, D. 1999. Highly active antiretroviral therapy (HAART) for the treatment of infection with human immunodeficiency virus type 1. *Biomedicine & pharmacotherapy*, 53, 73-86.

SILLANPÄÄ, J., NALLAPAREDDY, S. R., PRAKASH, V. P., QIN, X., HOOK, M., WEINSTOCK, G. M. & MURRAY, B. E. 2008. Identification and phenotypic characterization of a second collagen adhesin, Scm, and genome-based identification and analysis of 13 other

predicted MSCRAMMs, including four distinct pilus loci, in *Enterococcus faecium*. *Microbiology (Reading, England)*, 154, 3199.

STUART, R. M., FRASER-HURT, N., KERR, C. C., MABUSELA, E., MADI, V., MKHWANAZI, F., PILLAY, Y., BARRON, P., MUZAH, B. & MATSEBULA, T. 2018. The City of Johannesburg can end AIDS by 2030: modelling the impact of achieving the Fast-Track targets and what it will take to get there. *Journal of the International AIDS Society*, 21, e25068.

SURABHI, S. & SINGH, B. 2018. Computer aided drug design: an overview. *Journal of Drug delivery and Therapeutics*, 8, 504-509.

TON, A. T., GENTILE, F., HSING, M., BAN, F. & CHERKASOV, A. 2020. Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds. *Molecular informatics*, 39, 2000028.

UNAIDS. 2020. *HIV and AIDS in South Africa* [Online]. Available: <https://www.unaids.org/en/resources/fact-sheet>. [Accessed].

VESELOVSKY, A. & IVANOV, A. 2003. Strategy of computer-aided drug design. *Current Drug Targets-Infectious Disorders*, 3, 33-40.

VIVEK, A. & SWAPNA, T. 2021. Evaluation of Pharmacokinetics and Drug-likeness of Taxifolin from *Spatholobus parviflorus* (DC.) Kuntze: An in-silico Approach.

VORA, J., PATEL, S., ATHAR, M., SINHA, S., CHHABRIA, M. T., JHA, P. C. & SHRIVASTAVA, N. 2019. Pharmacophore modeling, molecular docking and molecular dynamics simulation for screening and identifying anti-dengue phytocompounds. *Journal of Biomolecular Structure and Dynamics*.

WALTERS, W. P. & WANG, R. 2020. New trends in virtual screening. ACS Publications.

XU, C., CHENG, F., CHEN, L., DU, Z., LI, W., LIU, G., LEE, P. W. & TANG, Y. 2012. In silico prediction of chemical Ames mutagenicity. *Journal of chemical information and modeling*, 52, 2840-2847.

YOGIRAJ, V., GOYAL, P. K., CHAUHAN, C. S., GOYAL, A. & VYAS, B. 2014. Carica papaya Linn: an overview. *International Journal of Herbal Medicine*, 2, 01-08.

CHAPTER 6: A DFT-BASED QSAR ANALYSIS OF EPICATECHIN, JUGLANIN, CATECHIN AND ZINC COMPOUNDS AGAINST DOLUTEGRAVIR AS POTENTIAL INHIBITORS FOR HIV-1 PROTEASE

ABSTRACT

Quantitative Structure-Activity Relationship (QSAR) is the study that is based on the hypothesis that changes the molecular structure that reflects proportional changes in the biological activity of such compounds. AutoQSAR (automated quantitative structure activity relationship) was used to build a model for the prediction of compounds bioactivities. Based on the analysis of the 2D-QSAR model parameters, statistical analysis, and regression diagrams, the lead inhibitors compounds were retrieved. Using epicatechin, juglanin, catechin, two Zinc Compounds namely, Zinc_001456687980 and Zinc_000015276352 forty-four associates' structures were retrieved from ChEMBL. QSAR was used to predict biological response between chemical structures and biological activities. Through Density Functional Theory (DFT) calculations Juglanin from the phytochemicals and Zinc_001456687980 were discovered as the most chemically reactive compound. This compound proved to be more electron deficient, having a rich region of the compounds accepted and donated hydrogen to HIV-1 amino acids residues during intermolecular interactions. The knowledge obtained in this study may help reveal natural compounds that can become potent inhibitors for HIV-1 protease. The pIC₅₀ values for the inhibitors were predicted by machine learning-based model generated by AutoQSAR algorithms. The predicted inhibitory prowess of the lead compounds from the medicinal plants for Catechin, Epicatechin, Juglanin, and Dolutegravir had predicted pIC₅₀ M of 8.982 M, 9.009 M, 8.709 M and 8.843 M, respectively. In addition, Zinc_001456687980 had the pIC₅₀ value of 8.6 M with Zinc_000015276352 of pIC₅₀ of 8.80 M.

6.1 INTRODUCTION

Quantitative structure-activity relationships (QSAR) is a method used in the development of relationships between physicochemical properties of chemical substances and their biological activities to obtain a reliable statistical model for prediction of the activities of new chemical entities. The techniques employed on the QSAR model, rely on a simple physico-chemical descriptors (Gaulton et al., 2012). These descriptors are not sufficient for generation comprehensive structure-activity relationship. QSAR studies may involve other biological relationship end points, atomic and groups, affinities of ligands to their binding sites, inhibition constants, rate constants, or molecular properties such as lipophilicity, polarizability, steric and structural features have in correlation (Dixon et al., 2016).

One such descriptors is the Electronic descriptors, which is obtained by calculations and can describe defined molecular activities (Roy et al., 2016). Recent studies have revealed that, using QSAR studies, descriptors calculated by employing the density functional theory method, improves the accuracy and lead to more reliable QSAR models. QSAR methods are preferred other than employing semi-empirical methods (Halder and Dias Soeiro Cordeiro, 2021). Its fundamental principle involves underlying the formalism is that the difference in structural properties is responsible for the variations in biological activities of the compounds (Daoui et al., 2022).

The frontier molecular orbitals (FMOs), entails HOMO and LUMO that determine the reactivity of chemical compounds based on quantum mechanics. HOMO maps the specific region of compounds that can donate electrons, whereas LUMO shows the specific atoms of the compounds that can accept electrons. However, the reactivity and stability of a chemical compound is calculated by the difference between the HOMO and LUMO values, the bandgap (Eno et al., 2022; Iwaloye et al., 2022). The selection of the inhibitors in this study were performed according to the various models that includes determining the IC_{50} values for the lead compounds. The IC_{50} value is the value that indicate the concentration required to inhibit biological process by half, thus providing a measure of the potency of the antagonist drug (Dibia et al., 2022). The selected structures were organized according to the IC_{50} (nM) values and the structure with the lowest value (IC_{50} of 0.96 μ M) was chosen to be the pivot structure.

6.2 MATERIALS AND METHODS

6.2.1 Calculation of Density Functional Theory

Density functional theory (DFT) research is based on the Hohenberg-Kohn theorem, which claims that the energy of a system may be expressed as a functional that is entirely dependent on the electron density. The Jaguar panel of the Material Science interface of the Schrödinger suite calculated quantum mechanics for the top scoring compounds, using a pseudo-spectral technique to ensure a quick DFT computation as described in previous study (Olawale et al., 2021a; Olawale et al., 2021b). All atoms of the molecule were optimized geometrically using density functional theory, with Becke's three parameter exchange–correlation functional hybrid (B3LYP) as the basic function and 6-31G** as the basis set. The ability of the molecular to transport or receive electrons during chemical reactions was determined using frontier molecular orbital parameters such as highest occupied molecular orbital (HOMO) energy, lowest unoccupied molecular orbital (LUMO) energy, and band gap energy. Electrophilicity, hardness, softness, electron affinity, and ionization potential were also calculated as global descriptive indices for chemical reactivity.

6.2.2 QSAR Modelling

A dataset (excel file) containing forty-four known inhibitors of HIV-1 protease was retrieved from ChEMBL database. This dataset contains the pIC₅₀ of the inhibitors, alongside its canonical smiles and drug-like properties. This file was exported to DATAWARIOR software to convert it to SDF format. Consequently, these compounds were exported to Maestro graphical interface, and prepared using ligprep. The QSAR model was built using AutoQSAR panel of Schrodinger using 2D descriptors and binary fingerprint of the compounds (Olawale et al., 2022). The program randomly split the compounds into training and test set in the ratio 3 to 1.

6.2.3 External validation QSAR model

The QSAR model was later verified using the characteristics external validation plus tool built by Roy et al. (2016). The program can measure the accuracy of the model by evaluating errors between pair observations describing comparable occurrences. The test set from the model was utilized for the external validation, and the correctness of the model was stated using R² based

Metrics and error based metrics. Before error based metrics validation, the test sets of the QSAR were originally inspected for error owing to systematic bias (Kolmar and Grulke, 2021).

6.3 RESULTS

6.3.1 DFT Calculation (Phytochemicals)

The results in Table 6.1 shows the results complied for the electronic behavior of the investigated compounds. They include the HOMO, LUMO, and energy gap of the compounds ($E_{\text{HOMO}} - E_{\text{HOMO}}$). The reactivity of chemical compounds is determined by the frontier molecular orbitals (FMOs), which include HOMO and LUMO. Compounds that can give electrons are shown in HOMO, whereas those that can accept them are shown in LUMO (Elekofehinti et al., 2021). Bandgap, which measures the difference between HOMO and LUMOS, governs the reactivity and stability of any given chemical molecule (Iwaloye et al. 2022). The results shows that Catechin, Epicatechin, Juglanin, and Dolutegravir had energy gap of 5.431380384eV, 4.555175733eV, 4.542386411eV, and 3.364136057eV, respectively. Based on the description of energy gap, Juglanin is the most reactive compound among the lead compounds among the hits, while Catechin is the most thermally stable compound. Figure 6.1 is a schematic illustrating the HOMO and LUMO portions of lead compounds. The global descriptive parameters which include electronegativity, ionization potential (IP), electron affinity (EA), hardness, and softness are also shown in Table 6.1.

6.3.2 Electronic behavior of the compounds (ZINC compounds)

The DFT calculations determine the behavior of the compounds (Zinc_000015276352, Zinc_001456687980 and Dolutegravir) in gas phase, and the comprehensive results are shown in Table 6.1. The frontier molecular orbitals (FMOs), which include the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) is often utilized to determine the reactivity and stability of a chemical compound. The phrase "HOMO–LUMO gap" refers to the difference in energy that exists between the HOMO and LUMO levels. In certain contexts, such as the border molecular orbital theory, the HOMO and LUMO orbitals are grouped together and referred to as the frontier orbitals. The energy gap that exists between these two frontier

orbitals can be utilized to make predictions on the reactivity and stability of a chemical compound (Elekofehinti et al., 2021; Iwaloye et al. 2022). According to previous reports (Iwaloye et al. 2022; Olawale et al, 2021) compounds with smaller energy gap tend to have high reactivity but low stability. The energy gap results show that Zinc_000015276352, Zinc_001456687980 and Dolutegravir have values of 3.303726916eV, 5.449611972eV, and 3.364136057eV. Zinc_000015276352 has low energy gap than dolutegravir. These denotes that Zinc_000015276352 is the most reactive compound. The low energy gap for Zinc_000015276352 may be attributed to the NH group that has a charge on a nitrogen atom. Figure 6.1 is a schematic illustrating the HOMO and LUMO portions of lead compounds. The global descriptive parameters which include electronegativity, ionization potential (IP), electron affinity (EA), hardness, and softness are also shown in Table 6.1.

Table 6. 1: DFT calculations of the investigated compounds

Compounds name	HUMO	LUMO	Energy gap	IP	EA	hardness	Softness
Catechin	-5.51573549	-0.084355106	5.431380384	5.5157355	0.084355106	2.715690192	0.184115258
Epicatechin	-5.48471458	-0.929538847	4.555175733	5.4847146	0.929538847	2.277587867	0.219530499
Juglanin	-5.765807563	-1.223421153	4.542386411	5.7658076	1.223421153	2.271193205	0.220148598
Zinc_000015276352	-5.497776016	-2.1940491	3.303726916	5.497776	2.1940491	1.651863458	0.302688456
Zinc_001456687980	-6.305680243	-0.856068271	5.449611972	6.3056802	0.856068271	2.724805986	0.183499303
Dolutegravir	-5.213417674	-1.849281618	3.364136057	5.2134177	1.849281618	1.682068028	0.297253138

6.3.3 QSAR Modelling

The QSAR model parameters, statistical analysis, and regression diagram are shown in Table 6.2, Table 6.3 and Figure 6.2. The QSAR model was built using molprint2D (binary fingerprint) as 2D descriptor and kernel partial least squares regression (KPLS) as the statistical analysis. The constructed 2D QSAR with model code kpls_molprint2D_13 has standard deviation (S.D) of 0.1894, R^2 of 0.7416, root mean square error (RMSE) of 0.2043, and Q^2 value of 0.627. The Figure 6.2 shows the scatter plot of the observed pIC50 against the predicted pIC50. Along with the benchmark score, the parameters for the external validation of the predictive QSAR model are shown in Table 6.3. The predicted inhibitory prowess of the lead compounds from the medicinal plants and as well as the standard compound are enlisted in Table 6.4. Catechin, Epicatechin, Juglanin, and Dolutegravir had predicted pIC50 of 8.982 M, 9.009 M, 8.709 M and 8.843 M. On the other hand Zinc_000015276352 and Zinc_001456687980 had predicted pIC50 of 8.689 M, and 8.803 M respectively. The automated module split the 44 HIV protease inhibitors randomly into 75% training set and 25% test set. The observed activities and predicted activities of the training set and test set in the negative logarithm of inhibitor concentration (pIC50, Table 6.4) are represented in Appendix 2. The algorithm generated ten best models and the result of the best model is shown in Table 6.1. The best model kpls_linear_9 recorded a standard deviation (SD) of 0.1894, R^2 of 0.7416, root mean square of the docking score for hits and reference compounds (Table 4) depicted that Zinc_001456687980 and epicatechin had the most favorable binding energy (-8 kcal/mol), as discussed in chapter 4 and 5 with 1rv7. The second and third top docked compounds were Juglanin and Zinc_000015276352/ Catechin with a docking score of $-7,6$ kcal/mol and $-6,9$ kcal/mol, respectively. Interestingly, dolutegravir, which is the reference compound showed the same binding affinity of (-8 kcal/mol) for docking score. Because docking score predicts a very low correlation between experimental activity and the predicted binding free energy (Singh and Muthusamy, 2013), and many proteins (receptor) undergo side chain movement upon binding to ligand (Mezei, 2013), this study made steadfast effort to use a more sophisticated docking protocol, to further validate the docking results (Table 6.4). These methods offers a fast and accurate prediction of ligand-induced conformational changes in the receptor-active sites

(Venkatachalam et al., 2003). This study reported to have epicatechin having the most favorable ligand-induced conformational changes.

Table 6. 2: The QSAR model

	Training set		Test set	
Model code	S.D	R ²	RMSE	Q ²
kpls_molprint2D_13	0.1894	0.7416	0.2043	0.627

The constructed model (KPLS_Molprint2D_9) fall within the required metrics for the robustness of a model – it falls within the accepted value for classical metric (for 100% data and after removing 5% data with high residuals), error-based metrics, threshold values utilized to judge the model predictions and MAE-based criteria applied on 95% data. The predicted pIC50 showed that the compounds have significant bioactivity against HIV-1 protease, which shows their tendency to inhibit the functional activity of HIV under in vitro conditions. It is noteworthy that beyond this study, the model generated could serve as an essential guide in the development of novel HIV inhibitors for future research.

Table 6. 3: External validation parameters of the constructed QSAR model

Test	Data	Observed value	Acceptable value
Model biasness test	SystematicErrorResult	Absent	-
	nPE / nNE	1.1000	< 5
	nNE / nPE	0.9091	< 5
	MPE / MNE	1.0131	< 2
	MNE / MPE	0.9870	< 2
	AAE - AE	0.5295	> 0.5
	R ² (Residuals; serial correlation)	0.0226	
	R ² (Residuals and Yobs values)	0.1492	< 0.5
Classical Metrics (for 100% data)	R ² Test(100% data)	0.6919	Close to 1.0
	R0 ² Test(100% data)	0.6797	Close to 1.0
	R0 ² Test(100% data)	0.6418	Close to 1.0
	Q2F1(100% data)	0.6814	> 0.6
	Q2F2(100% data)	0.6796	> 0.6
	Scaled Avg.Rm ² (100% data)	0.5788	> 0.6
	Scaled DeltaRm ² (100% data)	0.0734	-
	CCC(100% data)	0.8299	> 0.85
	R ² Test(95% data)	0.7562	Close to 1.0

Classical Metric (after removing 5% data with high residuals)	R0^2Test(95% data)	0.7459	Close to 1.0
	R0'^2Test(95% data)	0.6382	Close to 1.0
	Q2F1(95% data)	0.7466	> 0.6
	Q2F2(95% data)	0.7458	> 0.6
	ScaledAvgRm2(95% data)	0.6608	-
	ScaledDeltaRm2(95% data)	0.0364	-
	CCC(95% data)	0.8687	> 0.85
Error-based metrics (for 100% data)	RMSEP(100% data)	0.7127	Close to 0.00
	SD(100% data)	0.4432	Close to 0.00
	SE(100% data)	0.0432	Close to 0.00
	MAE(100% data)	0.5598	Close to RMSEP
Error-based metric (after removing 5% data with high residuals)	RMSEP(95% data)	0.6127	> 0.85
	SD(95% data)	0.3636	Close to 0.00
	SE(95% data)	0.0365	Close to 0.00
	MAE(95% data)	0.4945	
	MAE+3*SD(95% data)	1.5853	
Threshold values utilized to judge the model predictions RESULT (MAE-based criteria applied on 95% data)	(0.1*TrainingSetRange)		MAE ≤ (0.15*Training set range) MAE+3*SD ≤(0.25*Training set range)
	(0.15*TrainingSetRange)	105	
	(0.2*TrainingSetRange)	7.8195	
	(0.25*TrainingSetRange)	6.7604	
	Prediction Quality	MODERATE	

Table 6. 4: pIC₅₀ and IC₅₀ of the lead compounds

Compound Name	Catechin	Epicatechin	Juglanin	Zinc_00001 5276352	Zinc_0014 56687980	Dolutegravir (Control)
pIC ₅₀ (M)	8.982	9.009	8.709	8.689	8.803	8.843
IC ₅₀ (μM)	1.042	0.979	1.954	2.046	1.573	1.435

The reactivity of the compounds are based on the quantum mechanics, these are the frontier molecular orbitals (FMOs), which entails HOMO and LUMO. LUMO is where atoms can accept electrons and HOMO is a region of compound that donates electrons. It is, however, the bandgap, which calculate the reactivity and stability of chemical compounds. The red colour regions signify the atomic portion with a negative charge, and the blue colour highlights the area with a positive charge. Based on these calculations, epicatechin appeared to be the most chemically reactive

compound. Most of the bioactive compounds, the HOMO mapped sites act as hydrogen donors while interacting with the amino acids of 1rv7. The LUMO of the compounds is shown to accept hydrogen upon interacting with the residues at the active site of 1rv7. Other parameters such as electronegativity, ionization potential, electron affinity, hardness and softness, with Zinc_000015276352 (Table 6.1), having promising results when compared to the control. This is known as the global descriptors of chemical reactivity.

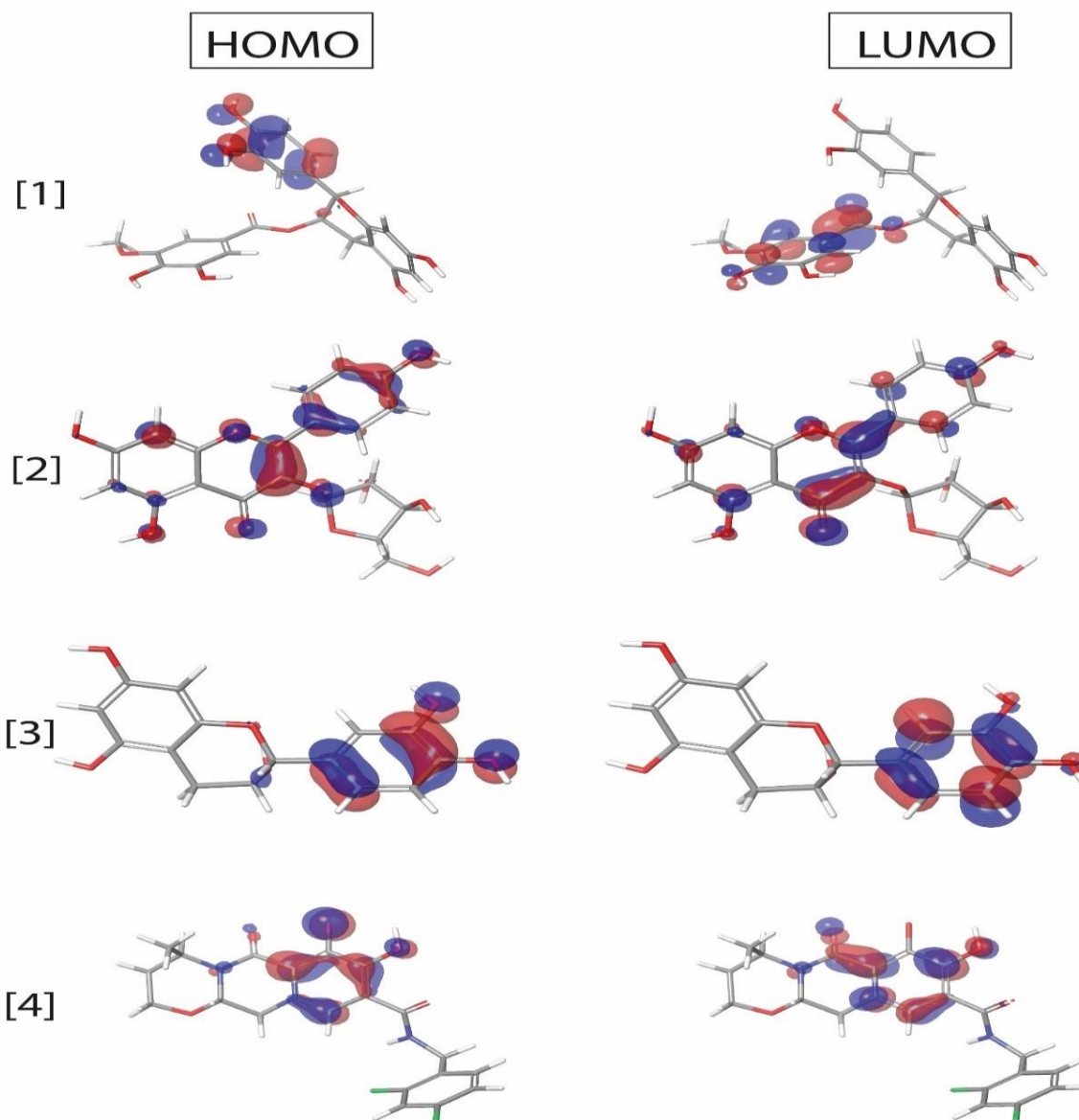


Figure 6. 1: HOMO and LUMO diagram of (1). Epicatechin (2). Juglanin (3). Catechin and (4). Dolutegravir

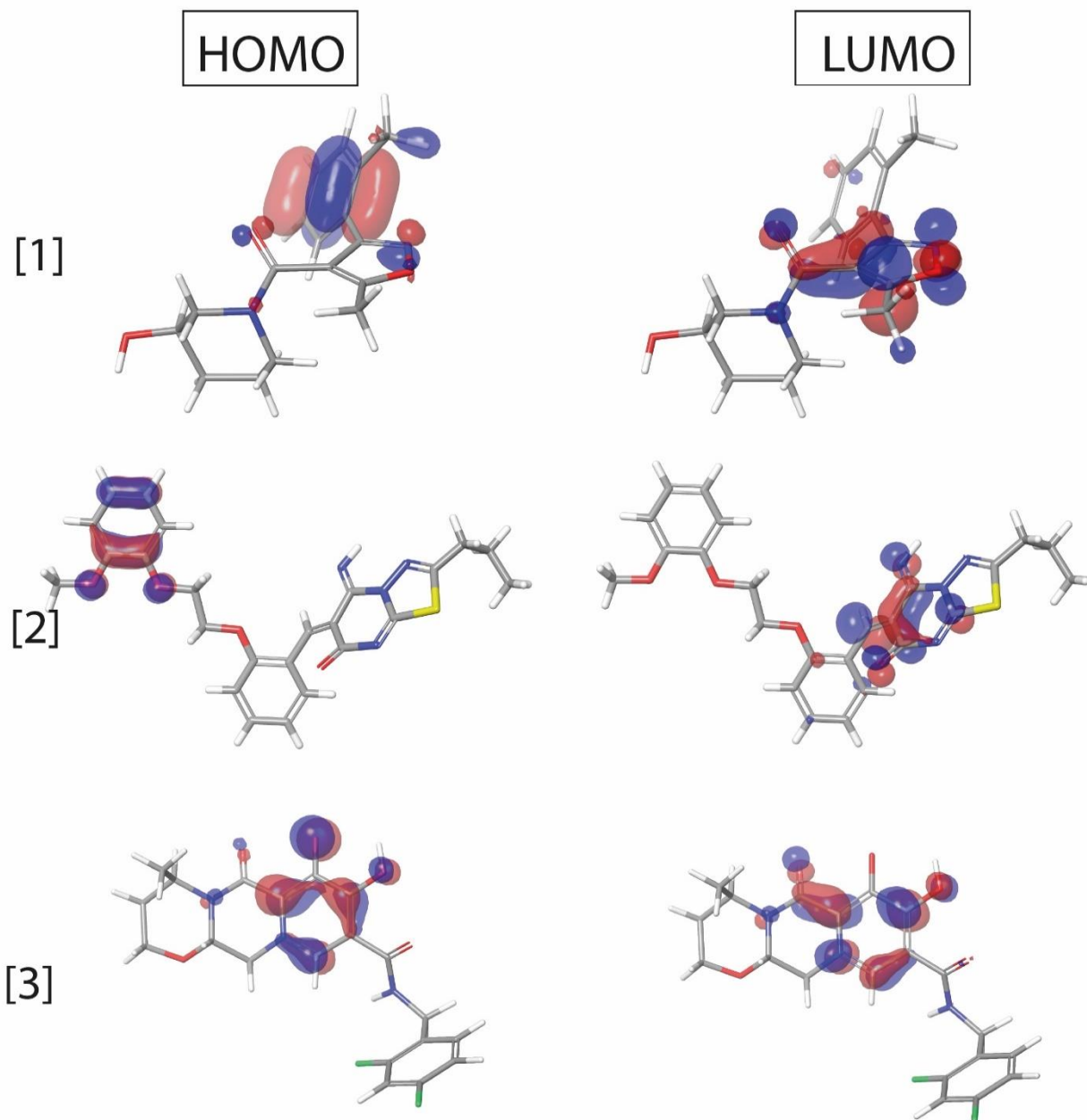


Figure 6. 2: HOMO and LUMO diagram of ZINC (1) Zinc_000015276352 (2) Zinc_001456687980 and (3) Dolutegravir

REFERENCE LIST

DAOUI, O., NOUR, H., ABCHIR, O., ELKHATTABI, S., BAKHOUCHE, M. & CHTITA, S. 2022. A computer-aided drug design approach to explore novel type II inhibitors of c-Met receptor tyrosine kinase for cancer therapy: QSAR, molecular docking, ADMET and molecular dynamics simulations. *Journal of Biomolecular Structure and Dynamics*, 1-18.

DIBIA, K. T., IGBOKWE, P. K., EZEMAGU, G. I. & ASADU, C. O. 2022. Exploration of the quantitative Structure-Activity relationships for predicting Cyclooxygenase-2 inhibition bioactivity by Machine learning approaches. *Results in Chemistry*, 4, 100272.

ELEKOFEHINTI, O. O., IWALOYE, O., OLAWALE, F., CHUKWUEMEKA, P. O. & FOLORUNSO, I. M. 2021. Newly designed compounds from scaffolds of known actives as inhibitors of survivin: computational analysis from the perspective of fragment-based drug design. *In Silico Pharmacology*, 9, 1-16.

ENO, E. A., LOUIS, H., UNIMUKE, T. O., GBER, T. E., AKPANKE, J. A., AMODU, I. O., MANICUM, A.-L. E. & OFFIONG, O. E. 2022. Modeling of Re (I) tricarbonyl complexes against SARS-CoV-2 receptor via DFT, in-silico molecular docking, and QSAR. *Chemical Physics Impact*, 5, 100105.

EROGLU, E. & TÜRKMEN, H. 2007. A DFT-based quantum theoretic QSAR study of aromatic and heterocyclic sulfonamides as carbonic anhydrase inhibitors against isozyme, CA-II. *Journal of Molecular Graphics and Modelling*, 26, 701-708.

HALDER, A. K. & DIAS SOEIRO CORDEIRO, M. N. 2021. QSAR-Co-X: An open source toolkit for multitarget QSAR modelling. *Journal of cheminformatics*, 13, 1-18.

IWALOYE, O., ELEKOFEHINTI, O. O., OLAWALE, F., CHUKWUEMEKA, O., BABATOMIWA, K. & FOLORUNSO, I. M. 2022. Fragment-Based Drug Design, 2D-QSAR

and DFT Calculation: Scaffolds of 1, 2, 4, triazolo [1, 5-a] pyrimidin-7-amines as Potential Inhibitors of Plasmodium Falciparum Dihydroorotate Dehydrogenase. *Letters in Drug Design & Discovery*, 19, 000-000.

KOLMAR, S. S. & GRULKE, C. M. 2021. The effect of noise on the predictive limit of QSAR models. *Journal of cheminformatics*, 13, 1-19.

MEZEI, J. & WIKSTRÖM, R. Aggregation operators and interval-valued fuzzy numbers in decision making. *Advances in Information Systems and Technologies*, 2013. Springer, 535-544.

MRACEC, M., MRACEC, M., KURUNCZI, L., NUSSER, T., SIMON, Z. & NARAY-SZABO, G. 1996. QSAR study with steric (MTD), electronic and hydrophobicity parameters on psychotomimetic phenylalkylamines. *Journal of Molecular Structure: THEOCHEM*, 367, 139-149.

OLAWALE, F., IWALOYE, O., OLOFINSAN, K., OGUNYEMI, O. M., GYEBI, G. A. & IBRAHIM, I. M. 2022a. Homology modelling, vHTS, pharmacophore, molecular docking and molecular dynamics studies for the identification of natural compound-derived inhibitor of MRP3 in acute leukaemia treatment. *Chemical Papers*, 76, 3729-3757.

OLAWALE, F., OLOFINSAN, K., IWALOYE, O., CHUKWUEMEKA, P. O. & ELEKOFEHINTI, O. O. 2022b. Screening of compounds from Nigerian antidiabetic plants as protein tyrosine phosphatase 1B inhibitor. *Computational Toxicology*, 21, 100200.

ROY, K., DAS, R. N., AMBURE, P. & AHER, R. B. 2016. Be aware of error measures. Further studies on validation of predictive QSAR models. *Chemometrics and Intelligent Laboratory Systems*, 152, 18-33.

SINGH, K. D. & MUTHUSAMY, K. 2013. Molecular modeling, quantum polarized ligand docking and structure-based 3D-QSAR analysis of the imidazole series as dual AT1 and ETA receptor antagonists. *Acta Pharmacologica Sinica*, 34, 1592-1606.

VENKATACHALAM, C. M., JIANG, X., OLDFIELD, T. & WALDMAN, M. 2003. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *Journal of Molecular Graphics and Modelling*, 21, 289-307.

ZHU, M., GE, F., ZHU, R., WANG, X. & ZHENG, X. 2010. A DFT-based QSAR study of the toxicity of quaternary ammonium compounds on *Chlorella vulgaris*. *Chemosphere*, 80, 46-52.

CHAPTER 7: CONCLUSION AND RECOMMENDATIONS

7.1 CONCLUSIONS

Computer aided drug design was employed in the discovery of the lead inhibitors for HIV-1 protease. This approach was based on screening the Zinc database using pharmacophore modelling that was mimicked from a pre-existing protein-drug interaction of HIV strain and lopinavir. The active biological components of the discovered lead compounds were then used to model a pharmacophore, which was used in the screening process of the phytochemicals. A co-crystallized structure of a multidrug-resistant (MDR), an MDR 769 co-structure of human immunodeficiency virus type-1 (HIV-1) protease together with lopinavir, that was clinically isolated from a patient was used to model this pharmacophore. The objective of this approach was to maintain the structural composition of the known protein-ligand interaction compound in the search for other class of compounds having similar biological and chemical properties. Plant extracts which are traditionally known medicinal plants were screened using the mimicked pharmacophore. These medicinal plants are known to have biological properties that can treat diseases like cancer, hypertension, HIV and numerous chronic illnesses.

After screening the Zinc database, Zinc_000015276352 and Zinc_001456687980 became the suitable lead compounds from the ZINC database, while; juglanin, catechin and epicatechin were the discovered phytochemical inhibitors that were suitable from the library of phytochemicals. The discovery of these compounds set a notable and remarkable milestone in the fight against HIV.

Furthermore, QSAR modelling was performed and a dataset contains the pIC₅₀ of the inhibitors, alongside its canonical smiles and drug-like properties and the inhibition potential for these compounds were determined. It was then observed that, Zinc_001456687980, with an IC₅₀ of 1.573 μ M displayed a very close IC₅₀ value compared to the control (Dolutegravir = IC₅₀ = 1.435 μ M), a drug called a “game changer” against the fight for HIV/AIDS in South African health care system. The results obtained from this approaches attest to the relevance of CADD. Using CADD, it is possible to find a suitable compound by filtering the databases that comprises of millions of compounds.

Astounding results concerning the findings that are based on a hit of a suitable inhibitor for a particular disease. In this process, epicatechin demonstrated a lower IC_{50} of $0.979\ \mu\text{M}$ compared to dolutegravir, portraying a good signal for an effective inhibitor, as a control.

ADMET was also employed in predicting the adsorption, distribution, metabolism, excretion, and toxicity of these lead compounds. This enabled the prediction of sufficient efficacy against the therapeutic target. One such properties that was mainly observed, was the toxicity of compounds, which was within the range of human consumption.

Through ADMET analysis, it was also possible to reveal toxicity range of inhibitors. Looking again into epicatechin, it was possible to detect the presence of a natural compound called catechol. Catechol is not always a problem in medicinal sciences since it is widely used to produce antioxidants. In addition, since we do not know the quantity that disqualifies catechol not to be used, its quantity in epicatechin must be determined and evaluated in order to determine its effects when used as an inhibitor of HIV.

Epicatechin is widely used as an antioxidant therapy in diabetes, cancer and HIV. The investigation of the interaction patterns of all these lead compounds showed many favorable interactions that agrees with previous studies in literature. Among these interactions, during docking, the inhibitors bound inside the active site of the protein. One of the major discoveries was the revelation that, these compounds possess favorable ADMET properties and they very stable when analyzing molecular dynamic simulation and there is no doubt that further experimental examination will affirm a promising drug candidate for HIV. CADD is indeed a very useful tool for pharmaceutical companies and academic research groups to search for suitable drugs.

7.1.1 Identified Zinc Compound (ZINC_001456687980)

One of the promising inhibitor discovered in this study is, ZINC_001456687980. It is identified as Salicin, a compound that was discovered in the 1800's. This compound was modified with an acetyl group to create aspirin (Cary and Peterlin, 2018). It has a molecular formula of $C_{17}H_{20}N_2O_3$. It has good biological properties and is a suitable antioxidant. It has a lower HOMO-LUMO energy gap which displays its high reactivity and its susceptibility to react with other molecular species such as protein (Akintemi et al., 2022), in this case, 1rv7. It is non-toxic, non-carcinogenic, with an affinity of $-8 K_D$ and an $IC_{50} = 1.573 \mu M$. Its IUPAC name is

(2R,3S,4S,5R,6S)-2-(hydroxymethyl)-6-[2-(hydroxymethyl)phenoxy]oxane-3,4,5-triol

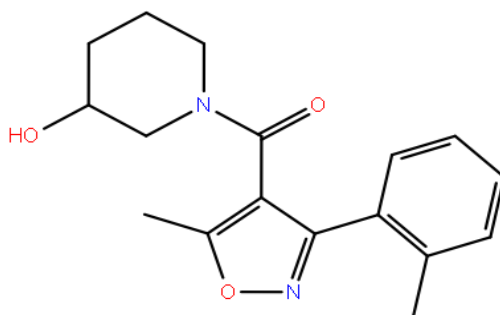


Figure 7.1: The structure of Zinc_001456687980

7.1.2 Identified Zinc compound (ZINC_000015276352)

The other lead compound that demonstrated promises HIV protease inhibition is, ZINC_000015276352. This compound was identified as Otamixaban, IUPAC name, methyl (2R,3R)-2-[(3-carbamimidoylphenyl)methyl]-3-[[4-(1-oxidopyridin-1-ium-4-yl)benzoyl]amino]butanoate and molecular formula $C_{25}H_{26}N_4O_4$. It was used as experimental injectable anticoagulant direct factor Xa inhibitor, however its development was terminated after it performed poorly in a Phase III clinical phase. It has an $IC_{50} = 2.046 \mu M$, non-toxic and has a binding affinity of $-6.9 K_D$.

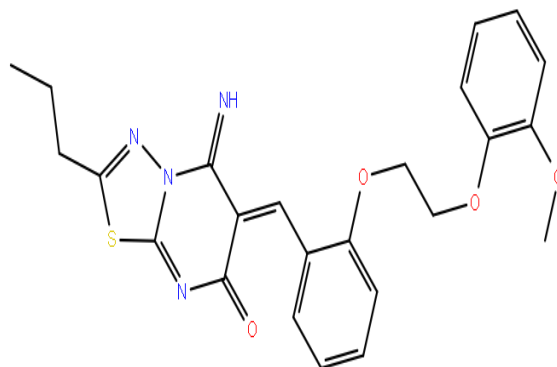


Figure 7.2: The structure of ZINC_000015276352.

7.2 RECOMMENDATIONS

Drug discovery is a preliminary stage of finding a suitable drug candidate for a particular disease. With the information attained from this study, a wet lab experiments will be performed for all lead compounds. The lead compounds will be synthesized, tested and analysed as a potential inhibitor. Further modifications, if necessary, with different substituents, will be explored in the laboratory to advance the drug effectiveness. This will create further analysis for these compounds in the area of HIV treatment.

The lead compounds have a unique and yet suitable biological properties for a potential drug candidate, and in some instances, have closeness and similarities in structure. The success of a candidate drug, using any of the proposed approaches, will lead to the drug being tested on animals, allowing necessary processes of drug development to take place.

In every stage of this research, dolutegravir was used as a control in finding these compounds. It was used to compare the properties and the pharmacokinetics of the discovered drugs. This led to the outcome of finding similar drugs, hence the need to further analyse the biological and chemical compositions of these drugs. The discovered drugs may be used independently, either as a multiple-drug or with the control. These processes are aimed at improving and repurposing the existing drugs that are in the market, especially dolutegravir. The findings obtained from this

research highlighted the effectiveness and the relevancy of computer aided drug design in this time and age.

Computational tools have some advantages in the saving of costs and in the lead optimization phase of the drug development. The compounds discovered in this project are highly recommended as targets in pharmaceutical industries and some medicinal centres in developing new drugs and in pursuit of finding a cure for HIV/AIDS.

REFERENCE LIST

AKINTEMI, E. O., GOVENDER, K. K. & SINGH, T. 2022. A DFT study of the chemical reactivity properties, spectroscopy and bioactivity scores of bioactive flavonols. *Computational and Theoretical Chemistry*, 1210, 113658.

CARY, D. C. & PETERLIN, B. M. 2018. Natural products and HIV/AIDS. *AIDS research and human retroviruses*, 34, 31-38.

SCHAMES, J. R., HENCHMAN, R. H., SIEGEL, J. S., SOTRIFFER, C. A., NI, H. & MCCAMMON, J. A. 2004. Discovery of a novel binding trench in HIV integrase. *Journal of medicinal chemistry*, 47, 1879-1881.

APPENDIX 1

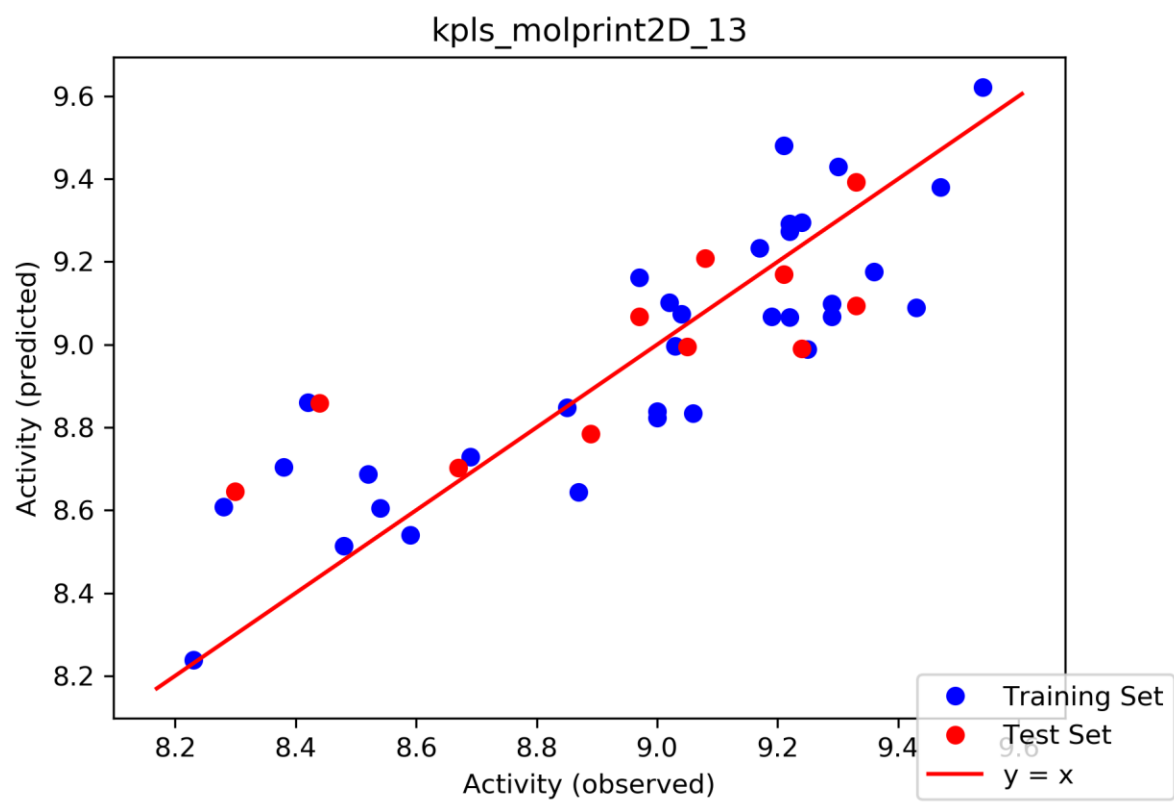


Figure A: Scatter plots of kpls_molprint2D_13 auto QSAR model generated from HIV-1 protease

APPENDIX 2

Quantitative parameters used for generating QSAR model

S/N	COMPOUND NAME	Set	Y(Obs)	Y(Pred)	Residue Error
1	CHEMBL3670886	train	9.2200	9.2928	0.0728
2	CHEMBL3670887	train	9.5400	9.6217	0.0817
3	CHEMBL3670903	train	8.4800	8.5139	0.0339
4	CHEMBL3670891	test	8.4400	8.8593	0.4193
5	CHEMBL3670867	train	9.3600	9.1765	0.1835
6	CHEMBL3670881	train	9.2500	8.9891	0.2609
7	CHEMBL3670898	train	8.8700	8.6438	0.2262
8	CHEMBL3670863	Test	8.3000	8.6457	0.3457
9	CHEMBL3670890	train	9.2200	9.2729	0.0529
10	CHEMBL3670902	train	8.5400	8.6057	0.0657
11	CHEMBL3670892	test	9.0500	8.9956	0.0544
12	CHEMBL3670880	test	9.2400	8.9909	0.2491
13	CHEMBL3670878	train	9.4700	9.3806	0.0894
14	CHEMBL3670889	train	8.6900	8.7293	0.0393
15	CHEMBL3670870	train	9.1900	9.0678	0.1222
16	CHEMBL3670870	train	9.2900	9.0678	0.2222
17	CHEMBL3670899	train	8.3800	8.7039	0.3239
18	CHEMBL3670883	train	9.3000	9.4301	0.1301
19	CHEMBL3670885	train	9.0400	9.0732	0.0332
20	CHEMBL3670895	train	8.4200	8.8599	0.4399
21	CHEMBL3670871	train	9.0000	8.8236	0.1764
22	CHEMBL3670888	train	9.0000	8.8381	0.1619
23	CHEMBL3670897	train	8.8500	8.8479	0.0021
24	CHEMBL3670896	test	8.6700	8.7029	0.0329
25	CHEMBL3670894	train	8.5200	8.6868	0.1668
26	CHEMBL3670900	train	9.0300	8.9972	0.0328
27	CHEMBL3670879	test	9.2100	9.1702	0.03989
28	CHEMBL3670866	train	9.2200	9.0655	0.1545
29	CHEMBL3670874	test	8.9700	9.0678	0.0978
30	CHEMBL3670876	train	9.2100	9.4810	0.2710
31	CHEMBL3670877	train	9.4300	9.0892	0.3408
32	CHEMBL3670829	test	8.8900	8.7853	0.1047
33	CHEMBL3670893	train	9.0200	9.1021	0.0821
34	CHEMBL3670882	test	9.0800	9.2078	0.1278
35	CHEMBL3670884	train	9.1700	9.2333	0.0633
36	CHEMBL3670864	train	9.0600	8.8334	0.2266
37	CHEMBL3670865	train	8.2300	8.2388	0.0088
38	CHEMBL3670868	test	9.3300	9.0945	0.2355
39	CHEMBL3670873	train	9.2400	9.2950	0.0550

40	CHEMBL3670869	train	9.2900	9.0983	0.1917
41	CHEMBL3670875	test	9.3300	9.3931	0.0631
42	CHEMBL3670901	train	8.2800	8.6079	0.3279
43	CHEMBL3639697	train	8.5900	8.5409	0.0491
44	CHEMBL3670872	train	8.9700	9.1617	0.1917