STRUCTURE-BASED DESIGN AND SYNTHESIS OF TETRAHYDRO-1,3,5-TRIAZIN-2-AMINE DERIVATIVES FOR ANTI-MALARIAL DRUGS

A Dissertation Submitted to The School of Natural Sciences and Mathematics

Department of Chemistry

at

Chinhoyi University of Technology

In Fulfillment of The Requirements of

Master of Philosophy

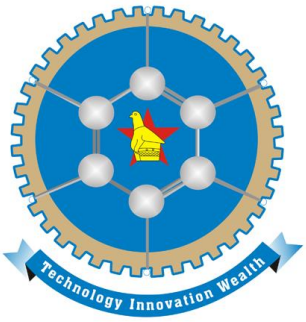
By

Brilliant Nyathi

C14124180G

Supervisor: Professor Grace Mugumbate

Co-Supervisors: Dr. Gadzikano Munyuki and Dr. Ryman Shoko



Department of Chemistry

School of Natural Sciences and Mathematics

Chinhoyi University of Technology

Private Bag 7724

Chinhoyi Zimbabwe

August 2023

Approval Page

Structure-Based Design and Synthesis of Tetrahydro-1,3,5-Triazin-2-Amine Derivatives for Anti-Malarial Drugs

By

Brilliant Nyathi

Declaration of Authorship

I, Brilliant Nyathi, hereby declare that this dissertation is wholly original to me. I can confirm that this work has not been presented at any other institution for any other degree or award. All sources have been properly acknowledged as references.

Acknowledgements

I would first like to thank Professor Grace Mugumbate for her help and support rendered in the duration of this study and for getting the project off the ground. Her guidance made the dissertation possible and I am forever grateful. I would also like to thank Dr. Gadzikano Munyuki and Dr. Ryman Shoko for their unwavering support and guidance throughout the practical aspects for this body of work.

I would also like to thank all the members of the Drug Discovery and Informatics research group for keeping me almost sane during the lockdown periods when everything else slowed down including research itself.

Studying during the COVID19 pandemic was a long and hard road, I am grateful for all the support I received from my supervisors, members of the research group, Kudakwashe Tsotsoro and members of the Department of Chemistry at the Chinhoyi University of Technology for all their support.

My gratitude goes towards The World Academy of Sciences (TWAS) which funded this work.

Finally, my family in Bulawayo, Harare, and friends from Woods, who have been supportive in all my endeavors, thank you all for your patience and support, I cannot thank you enough.

ABSTRACT

Malaria is a life-threatening disease, caused by the protozoan parasites of the *Plasmodium* genus, with 241 million cases recorded globally in 2021, and 619 00 deaths in that year. In 2018, the parasite *Plasmodium falciparum* accounted for 99.7% of estimated malaria cases in the WHO African Region, which carried a 94% load of all malaria cases worldwide. The emergence of drug resistant strains of the parasite, and the side effects of currently used drugs indicate the urgent need for the discovery of target specific anti-malaria drugs. Anti-*P. falciparum* activity targeting DHFR, the K1, W2, and D6 strains has been investigated using 1,3,5-triazine derivatives. It has also been demonstrated that 1,3,5-triazine-2-amine derivatives inhibit *Mycobacterium tuberculosis* DHFR. Triazines have been promising therapeutic options for treating resistant malaria and TB strains. The project aimed to design and synthesize 1,3,5-triazine-2-amine derivatives for use as inhibitors of *P. falciparum* heat shock protein 70-1 (*Pf*Hsp70-1). In this work, a highly diverse curated seed library of ninety-nine 1,3,5 triazine-2-amine derivatives was converted from fragments to lead like compounds using structure-based methods for *de novo* design and fragment-based methods. These compounds were designed within the ATP active site of *Pf*Hsp70-1 by ensuring synthetic accessibility using click chemistry and robust reactions for seed growth and recombination through mutations and crossovers respectively. The diversity of these newly designed compounds was evaluated through 3D chemical space mapping of the physicochemical properties. A library of 50 unique 1,3,5 triazine-2-amine derivatives was generated, all the compounds possess drug-like properties as defined by the Lipinski rule of 5. Two compounds with a VINA binding affinity above -10 kcal/mol were synthesized. The first compound synthesized was a carbamate through isocyanate-hydroxyl coupling with N-methylimidazole, the second compound synthesized was an amide through acid-amine coupling with N,N′-dicyclohexylcarbodiimide and 1-Hydroxybenzotriazole. The third compound synthesized was an ester derived from the recombination of the alcohol from the first compound with the acid from the second compound in a sulfuric acid coupled esterification. The three compounds namely BN_01, BN_02 and BN_03 were isolated as fine needle like clear crystals with yields of 77%, 94.8% and 96%, with melting points of 280 ^{o}C, 220 ^{o}C and 249 ^{o}C respectively. This study established a functional 3D model of *Pf*Hsp70-1 as well as a library of 50 synthetically accessible lead-like anti-malarial drugs, three of which were synthesized.

List of Abbreviations

ADMET	absorption, distribution, metabolism, and excretion–toxicity
ADP	adenosine diphosphate
ATP	adenosine triphosphate
BBB	blood-brain barrier
CDC	center for disease control
CNS	central nervous system
COVID-19	coronavirus disease of 2019
DCC	N,N′-dicyclohexylcarbodiimide
DCM	dichloromethane
DHFR	dihydrofolate reductase
DMF	dimethylformamide
DOPE	discrete optimized protein energy
DSSP	define secondary structure of proteins
ER	endoplasmic reticulum
FBDD	fragment-based drug design
GANDI	genetic algorithm-based *de novo* design of inhibitors
GIT	gastrointestinal tract
HBA	hydrogen bond acceptor
HBD	hydrogen bond donor
HIV	human immunodeficiency virus
HMDB	the human metabolome database
HOBt	hydroxy benzotriazole

HTS	high throughput screening
MDH	malate dehydrogenase
MDS	molecular dynamics simulation
MMGBSA	molecular mechanics/generalized born surface area
MMPBSA	molecular mechanics/poisson-boltzmann surface area
MSD	mean square displacement
MUSCLE	multiple sequence comparison by log-expectation
NBD	nucleotide binding domain
NCBI	national center for bioinformatics
NEF	nucleotide exchange factor
NIH	national institute of health
NMI	N-methylimidazole
NMR	nuclear magnetic resonance
PAINS	pan-assay interference substructures
PCA	principal component analysis
PDB	protein data bank
*Pf*Hsp70-1	*Plasmodium falciparum* heat shock protein 70-1
*Pf*Kelch13	*Plasmodium falciparum* Kelch 13
*Pf*PI3K	*Plasmodium falciparum* phosphatidylinositol 3-kinase
PtdIns3P	phosphatidylinositol-3-phosphate
PTEX	*Plasmodium* translocon of exported proteins
QSAR	quantitative structure activity relationship
RECAP	retrosynthetic combinatorial analysis procedure

RMSD	root mean square displacement
RMS	root mean square
ROSC	reactive oxidative stress complex
RotB	rotatable bonds
RSC	Royal Society of Chemistry
SBD	substrate binding domain
SBDD	structure-based drug design
SEED	solvation energy for exhaustive docking
SlogP	water/octanol partition coefficient
SPR	surface plasmon resonance
T-Coffee	Tree-based Consistency Objective Function for alignment Evaluation
TCP1	T-complex protein 1
TEA	triethylamine
THT	tetrahydro triazine
TPSA	topological polar surface area
TRAPP	Transient Pocket in Proteins
TRiC	T-ring complex
USFDA	United States Food and Drug Administration.
vdW	van der Waals
WHO	world health organization

Table of Contents

Table of Figures

Table of Equations

Table of Tables

CHAPTER 1: INTRODUCTION

Malaria is a potentially fatal disease caused by the *Plasmodium* genus protozoan parasite species (Talapko et al., 2019). This disease is caused by 5 species, namely the *Plasmodium Ovale*, *P. vivax*, *P. malariae*, *P. knowlesi*, and *P. falciparum* (Cox, 2010). The most lethal to humans is the *P. falciparum* parasite, capable of repeated replication over a 48-hour period in the erythrocytes, resulting in exponential growth, rapidly progressing the disease (Maier et al., 2019). The symptoms of malaria are fever and flu-like illness; these will include headache, tiredness, chills, and muscle ache all over the body. Diarrhea and nausea may also occur as the disease progresses. As a result of blood loss through the destruction of red blood cells, jaundice and anemia may be observed (Shetty & Steele, 1999). An infected female *Anopheles* mosquito is responsible for the transmission of the disease (Beck-Johnson et al., 2013).

There has been an increase of malaria cases from the year 2018 to 2022 as shows below in Figure 1. According to the World Health Organization (WHO), there were 228 million cases with 405 000 deaths in the year 2018, 229 million cases and 409 000 deaths in 2019, 241 million cases in 2020 with a spike in death of 627 000 cases in 2020, the spike in cases and deaths is attributed to the disruption in malaria control and treatment due to COVID19 (WHO, 2019, 2020, 2021) . A further 247 million cases of malaria were observed in 2021, and 619 000 people died from the disease and the increase continued in 2022 with 249 million cases and 609 000 deaths. Children of ages 5 and below were the most vulnerable group and accounting for 76% of all deaths globally in the year 2022. The WHO African Region was responsible for 95% of malaria cases and 593 000 of deaths from the disease in 2022 (WHO, 2022, 2023) .

Figure 1: Malaria cases and deaths from 2018 to 2022

The WHO has presented a classification of malaria through symptoms observed, these classes are separated as uncomplicated and severe malaria. For severe cases of malaria, the symptoms progress to impaired consciousness, prostration and respiratory distress (Ashley et al., 2018). The most severe type of the illness is cerebral malaria, a form of *P. falciparum* malaria that causes neurological abnormalities and may put a patient in a coma depending on their immune system. This type of malaria affects around 1% of children under the age of five. (Idro et al., 2010; Luzolo & Ngoyi, 2019a). Immune system and hormonal changes make pregnant women more susceptible to malaria infection, increasing the chances of miscarriage and anemia (Makenga et al., 2020; van Eijk et al., 2015). Likewise, children below the age of 5 face a higher burden of malaria infection which may lead to death, this has been attributed to a lack of immunity against malaria (Dao et al., 2021).

The subsequent section will briefly cover the *P. falciparum* parasite's life cycle in its hosts and the mutations that have caused its resistant to current treatment methods before moving on to the research topic, the goals of this research, questions, and objectives. Finally, the significance of the work will be addressed and capped off with the limitations of this research.

1.1 *P. falciparum* Life Cycle

The malaria sporozoite's lifecycle journeys through two hosts shown in Figure 2, the female *Anopheles* mosquito vector and the mammalian host (Ménard, 2000). During a blood meal, a malaria-infected *Anopheles* mosquito injects sporozoites into the human host. (Florens et al., 2002). The sporozoites infect the liver cells, where they develop into schizonts, the stage at which they undergo multiple nuclear divisions, producing daughter cells known as merozoites. The schizonts then rupture to release merozoites, this stage is known as the exo-erythrocytic schizogony, this first stage occurs in the liver of the mammalian host. Following this stage is the erythrocytic schizogony stage where merozoites infect the red blood cells and undergo asexual multiplication in the erythrocytes to trophozoites, a stage where they actively feed and grow (Beeson et al., 2016). Trophozoites now develop into schizonts, which burst to produce merozoites, continuing the cycle of the parasite in the human blood stage. The clinical signs and symptoms of the illness are brought on by blood stage parasites (CDC, 2018; Mohandas & An, 2012; U. Sahu et al., 2013; Yui & Inoue, 2021).

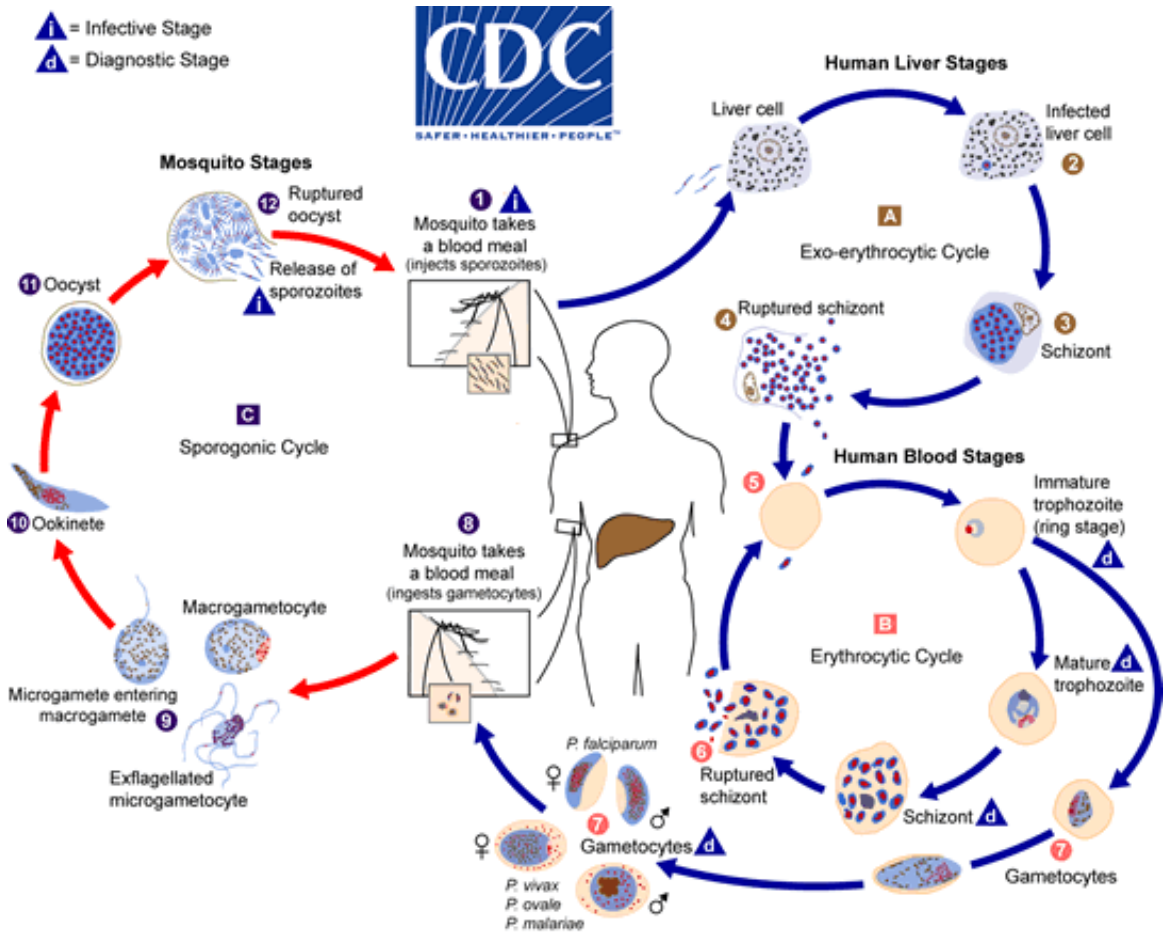


Figure 2:Malaria Parasite Life Cycle (CDC, 2018; Konopka et al., 2021)

The sporogony cycle is the process through which the parasites replicate within the mosquito, it starts with the *Anopheles* mosquito ingesting gametocytes during a blood meal. The parasite generates zygotes in the stomach of the mosquito by penetrating the macrogametes, in turn these will be elongated and motile allowing them to develop within the mid-gut wall into oocysts. Making their way from the mid-gut to the salivary glands are sporozoites which were released from the rapturing of oocysts after a successive growth period. Inoculation into the mammalian hosts perpetuates the cycle (CDC, 2018; Konopka et al., 2021).

1.2 Mechanisms of Resistance in *P. falciparum*

P. falciparum has developed drug resistance due to *P. falciparum* Kelch 13 (*Pf*Kelch13) mutations, identified as partial markers of partial artemisinin resistance. The widespread distribution of *P. falciparum* resistance to the commonly prescribed antimalarial medications, chloroquine and pyrimethamine/sulfadoxine had led to introduction of artemisinin-based combination therapy

(Haldar et al., 2018; Mita & Tanabe, 2012; Nandakumar et al., 2006; WHO, 2019). *Pf*Kelch13 is a substrate adapter for E3 ligase, a cullin with a hypothetical substrate of *P. falciparum* phosphatidylinositol-3-kinase (*Pf*PI3K). Increased levels of phosphatidylinositol-3-phosphate (PtdIns3P), which reduces susceptibility to artemisinin, and the parasite's response to unfolded proteins are two mechanisms of artemisinin resistance that are promoted by this mutation. (Barrett et al., 2019; Haldar et al., 2018; Toure et al., 2018).

Two mechanisms have been explored, the activation of the unfolded protein response (UPR) and proteostatic dysregulation of parasite phosphatidylinositol 3-kinase (*Pf*PI3K). Through inducing a transcriptional response to promote gene expression and improve redox conditions throughout the cell, the UPR process restores endoplasmic reticulum homeostasis through the reactive oxidative stress complex (ROSC) and T-complex protein 1 (TCP1), T-complex protein ring complex (TRiC). The TRiC restores normal protein folding, which controls the reduction of harmful protein aggregates in the endoplasmic reticulum of the cytoplasm. The ROSC proteasome controls their complete elimination. (Haldar et al., 2018; Mok et al., 2015). The second mechanism of proteostatic dysregulation of parasite phosphatidylinositol 3-kinase, works by clearing out artemisinin damaged, misfolded protein and toxic aggregates which would be found at any cellular location. The *Pf*Kelch13 mutation prevents *Pf*PI3K from binding to its substrate and from being ubiquitylated, which raises the levels of the kinase and its lipid product phosphatidylinositol 3-phosphate (PtdIns3P). *Pf*PI3K controls autophagy, which eliminates artemisinin-damaged, misfolded proteins. The apicoplast, food vacuole, and endoplasmic reticulum all contain PtdIns3P, and its presence has been associated with macro autophagy (Fairhurst & Dondorp, 2016; Haldar et al., 2018; Mbengue et al., 2015; Paloque et al., 2016).

1.3 Exploring new targets in *P. falciparum*

Strains of *P. falciparum* resistant to chloroquine were first observed in East Africa 1978 which caused an increased number of infections and deaths (Trape, 2001). WHO issued a decree supporting the use of artemisinin as a first-line drug for cases of *P. falciparum* resistant strain (Bosman & Mendis, 2007). Artemisinin is now facing resistance, reported as early as 2004 in Cambodia which has once again seen a new and charged drug rush looking into novel drugs and novel drug targets (Dondorp et al., 2010). *P. falciparum* heat shock proteins (*Pf*Hsp) are undergoing scrutiny as targets in the hope to find suitable drug candidates. This work has been on

the premise of their essentiality in the growth and survival of the parasite and implications in their potential as drug and vaccine targets (P. Acharya et al., 2007; Banumathy et al., 2003).

1.4 Research Significance

Artemisinin saved a lot of lives in developing countries when resistance to chloroquine was observed, a global health crisis was averted. However, the emergence of drug resistance from *Plasmodium* species has led to another global emergency requiring swift action to avert this crisis. This has pushed for a new approach to the design of anti-malaria compounds and possibly a reversal of artemisinin resistance. Artemisinin was not designed for a specific target, rather it works by the generation of free radicals to kill the parasite, affecting the heme pathway, the cell proteins and membrane. The damage done to the cells by free radicals can be completely averted by clearing out of toxins through the reactive oxidative stress complex in the unfolded protein response and the *Pf*Kelch13 mutation which upregulates *Pf*PI3K and PtdIns3P, responsible for autophagy and macro autophagy. Therefore, there is a need for a new approach of designing drug compounds with novel drug targets and modes of action.

1.5 Research Aim

The aim of this research was to design anti-malarial compounds with a different mode of action from artemisinin-based compounds, targeting a non-heme pathway and not affected by current resistance mechanisms, the *Pf*Kelch13 mutation and unfolded protein response.

1.6 Research Objectives

1. To generate a 3D structure of *P. falciparum* heat shock protein 70-1 (*Pf*Hsp70-1).
2. To design derivatives of 1,3,5-triazine-2-amine derivatives that inhibit *Pf*Hsp70-1.
3. To synthesize derivatives of 1,3,5-triazin-2-amine.

CHAPTER 2: LITERATURE REVIEW

2.1 Heat Shock Proteins as Drug Targets

Plasmodium falciparum undergoes a life cycle that spans two distinct physiological environments; the *Anopheles* mosquito vector and the human host. The parasite depends on the life support of a host environment because of its obligate intracellular parasitic nature (Daniyan et al., 2019). Within the human body, the parasite invades hepatocytes and erythrocytes, adjusting to the host environment, and specifically adapting to various stressors like temperature fluctuations, which are responsible for the febrile episodes observed in clinical malaria (J. M. Njunge et al., 2015; Shonhai, 2010). Stresses from heat shock increase proteins' propensity to unfold, which causes aggregates to develop (Pavithra et al., 2007a). As a consequence, organisms initiate a process called the heat shock response, prompting the mobilization of heat shock proteins (Hsps). This response enables them to cope with various physical stresses, including abrupt temperature fluctuations, exposure to toxic environments, and oxidative stress (Lilburn et al., 2014; Mabate et al., 2018).

*Pf*Hsps represent a widespread pervasive group of molecular chaperones responsible for promoting the parasite's viability and proliferation within host cells. They achieve this by aiding in the refolding of misfolded proteins and the unwinding of preformed aggregates. The effectiveness of these molecular chaperones is regulated by co-chaperones and co-factors (Muralidharan et al., 2012; Schulze et al., 2015; Zininga, Anokwuru, et al., 2017; Zininga et al., 2015). Classification of these proteins is based on their mass, measured in kilo Daltons (kDa), the most common ones are *Pf*Hsp40, 60s, 70s, 90, 100s, and 110s (Shonhai, 2010). Scientific research focuses on studying the functional pathways involving Hsps to identify potential druggable targets (Pavithra et al., 2007b; Shahinas, Folefoc, Taldone, et al., 2013).

Hsp40s play a crucial role as substrate scanners, responsible for enlisting Hsp70s to refold misfolded proteins. The client proteins for Hsp70 are recognized and their interaction is stabilized by Hsp40. This stabilization occurs when Hsp40 first binds to the substrate, which is then released to the substrate-binding domain (SBD) of Hsp70 for the refolding process (Botha et al., 2007; Daniyan et al., 2016). Hsp70 takes on the role of folding proteins, guiding them towards a state closely resembling their native conformation. Subsequently, these proteins are transferred to Hsp90, where they undergo a process of comprehensive folding to achieve functional maturation.

Hsp90's ability to recognize specific substrates, namely proteins in near-native conformations, is fundamental to this vital function. (Daniyan et al., 2019; Kravats et al., 2018; Shahinas, Folefoc, & Pillai, 2013). Hsp70s have an association with Hsp101, which itself is a part of the larger molecular machinery that facilitates protein trafficking via the *Plasmodium* translocon of exported proteins (PTEX). This 70kDa and 101kDa protein association in the PTEX complex suggests that exported proteins are kept in a translocation competent state by the 70kDa chaperone for passage through the translocon PTEX (Przyborski et al., 2015; Schulze et al., 2015; Q. Zhang et al., 2017).

The important role played by Hsps in the growth and survival of the parasite and specifically, the Hsp70 chaperone codependence and interplay has placed them at the center of drug targets (K. Liu & Houry, 2014). *P. falciparum* produces six proteins belonging to the HSp 70kDa family shown in

Table 1. These proteins are named PfHsp70-1 , 70-2, 70-3, 70-x, 70-y, and 70-z, with their PlasmoDB accession names as PF3D7_0818900, PF3D7_0917900, PF3D7_113400, PF3D7_0831700, PF3D7_1344200, and PF3D7_0708800 respectively (Daniyan et al., 2019). Grouped into two classes of the canonical Hsp70 and the Hsp110, *Pf*Hsp70-1, -2, -3, and *Pf*Hsp70-x exhibit features of Hsp70, whilst *Pf*Hsp70-y and *Pf*Hsp70-z have features of Hsp110 (Shonhai et al., 2007; Zininga, Ramatsui, et al., 2017).

Hsp110s are acknowledged as nucleotide exchange factors (NEFs) acting in conjunction with Hsp70 proteins across different species (Zininga, 2015; Zininga et al., 2015). NEFs enhance the ADP-to-ATP exchange, facilitating ATP binding to other proteins. This process expedites the dissociation of ADP from Hsp70, leading to ATP re-binding and subsequent dissociation of the substrate. NEFs play a role in regulating the longevity of the Hsp70-substrate complex (Q. Liu et al., 2019; Zininga, 2015).

Table 1: Function and subcellular localization of Hsp70s

Protein	**Localization**	**Functions**
*Pf*Hsp70-1	Cytosol, Nucleus	Promotion of proper protein conformation and prevention of protein aggregation.
*Pf*Hsp70-2	Endoplasmic Reticulum (ER)	Proteins are transported into the ER, then later, undergo retrograde translocation for degradation.
*Pf*Hsp70-3	Mitochondria	Proteins undergo translocation into the mitochondrion.
*Pf*Hsp70-x	Host, Parasitophorous Vacuole	Movement of specific virulence proteins.
*Pf*Hsp70-y	ER	Aids in the exchange of nucleotides for PfHsp70-2
*Pf*Hsp70-z	Cytosol	Aids in the exchange of nucleotides for *Pf*Hsp70-1. Prevention of the aggregation of proteins with asparagine repeats.

Several factors contribute to heat shock proteins as drug targets; Their function (Hsp70 and Hsp90) as ATPases is regulated by nucleotides, making them ideal druggable targets with the most notable inhibitors being ATP mimics (Banumathy et al., 2003; Zininga & Shonhai, 2014). Hsp70s display variations in sequence motifs responsible for interacting with their NEFs, resulting in different rates of nucleotide exchange. This functional characteristic offers potential drug design targets for these proteins (Zininga & Shonhai, 2014).

The existence of heat shock proteins in a functional network interplay for the refolding of denatured proteins implies that interrupting or rescinding this interplay will interfere and possibly cripple their function in promoting the parasite's viability and proliferation within host cells (E. R. Pesce & Blatch, 2014; Seraphim et al., 2019; Shahinas, Folefoc, Taldone, et al., 2013). Three observations have led to the selection of Hsp40 and Hsp70s as drug emerging drug targets in *P.*

falciparum, these have been: the varied distribution of heat shock proteins across species, with a much higher distribution of Hsp40s in the parasite than for humans; the lower cytotoxicity in humans for compounds set to target Hsp40 and Hsp70s in the parasite (J. Njunge et al., 2013; Zininga & Shonhai, 2014), and minor sequence variations in the parasite which have presented an opportunity for selective inhibition with effects in the human counterpart kept at a minimum (Lilburn et al., 2014; E.-R. Pesce et al., 2010; Zininga & Shonhai, 2014). These variations include the glycine-glycine-methionine-proline (GGMP) tetrapeptide repeats for *Pf*Hsp70-1 coupled with a GGMN/P variation, these repeats are only present for this protein and are postulated to have a marginal effect on the variation of *Pf*Hsp70-1 against all other heat shock protein 70s in the parasite and human (Daniyan et al., 2019; Makumire, 2019).

2.2 Anti-Malarial Drugs, Mechanisms of Action and Drug Resistance

With regards to their chemical features and modes of action, there are three types of medications to treat malaria. The first category consists of the artemisinin compounds, including Artemisinin, Artesunate, Dihydroartemisinin, Arteether and Artemether. These drugs are derived from the sweet wormwood plant and are known for their rapid and potent antimalarial activity, making them essential components of artemisinin-based combination therapies (ACTs) widely used to treat uncomplicated malaria. Artemisinin derivatives shown in Figure 3, are sesquiterpene trioxane lactone scaffold containing compounds with a unique endoperoxide moiety with a controversial history on its precise mode of action against *Plasmodium* species (Cui & Su, 2009a; Duffy & Mutabingwa, 2006). These compounds have exhibited several modes of action against the *Plasmodium* parasite, these are the inhibition of heme detoxification, disruption of the parasite's development and membrane integrity including parasite's metabolism. These rely on the endoperoxide moiety, core to artemisinin derivative's activity through the generation of reactive oxygen species. This feat is achieved through the reaction of the moiety with the iron in the parasite's vacuole, the generated reactive oxygen species damage several cellular components within the parasite, these being membranes and proteins (Counihan et al., 2021; O'Neill et al., 2010). The damage to membrane integrity has been shown to cause cell component leakage within the parasite, leading to cell lysis, a major cause of death to the parasite. The damage on cell proteins also caused by the reactive oxygen species has been shown to cause a major drawback in the parasite development, disrupting early ring forms and mature schizont stages in the parasite's lifecycle. Cell protein damage is also responsible for the disruption of hemozoin formation, this

protein is important in the clearing of the toxic free heme molecules generated through the digestion of the mammalian host hemoglobin as a nutrient source providing amino acids and energy from the iron molecule contained in heme molecule (Counihan et al., 2021; Matz, 2022).

Several experiments have been carried out to investigate parasite-specific mechanisms of action and non-specific mechanisms of action. A derivative of artemisinin, dihydroartemisinin was investigated on its activity against chloroquine-sensitive and chloroquine-resistant strains by comparison of half-maximal inhibitory concentration (IC_{50}) values for *P. falciparum* and eukaryotic cells. The findings from this work revealed no difference in IC_{50} values of dihydroartemisinin and artemisinin at 0.15 and 0.10 nM; 0.13 and 0.12 nM, respectively for chloroquine resistant and non-resistant strains. This suggests a different mode of action may be explored by the drugs in targeting the parasite. A comparison of dihydroartemisinin on *Leishmania major* and *P. falciparum* was also carried out, *L. major* had an IC_{50} value of 0.63 µM, this also suggest that the mode of action explored may be specific to *P. falciparum*. The interference of sarcoplasmic/endoplasmic calcium ATPase (SERCA) was also explored through investigating the similarities between artemisinin and thapsigargin, an inhibitor of SERCA which is involved in maintaining calcium ion concentrations within the parasite. SERCA. Artemisinin is steric sensitive, alterations of its steric nature showed a decline in activity within the parasite, suggesting the sensitivity of the compound to size and structure of the target molecule. This implied the activation of artemisinin occurs after binding to a specific site within the parasite. *Pf*ATP6 is the only SERCA-type Ca^{2+} ATPase in *P. falciparum* known to be inhibited by thapsigargin, given the striking similarities between thapsigargin and artemisinin and their steric sensitivity, inhibition of *Pf*ATP6 by artemisinin was investigated. The results revealed complete inhibition of the ATPase by artemisinin with a half-maximal inhibition constant (K_i) of 150 nM, specific to this ATPase such that at 50µM concentration of artemisinin no other transporters were affected including the including the non-SERCA Ca^{2+}-ATPase *Pf*ATP4. Affecting calcium levels within the parasite is also known to disrupt uptake of nutrients from the mammalian host through inhibition of endocytosis (Golenser et al., 2006; Moore et al., 2022; Nagasundaram et al., 2016).

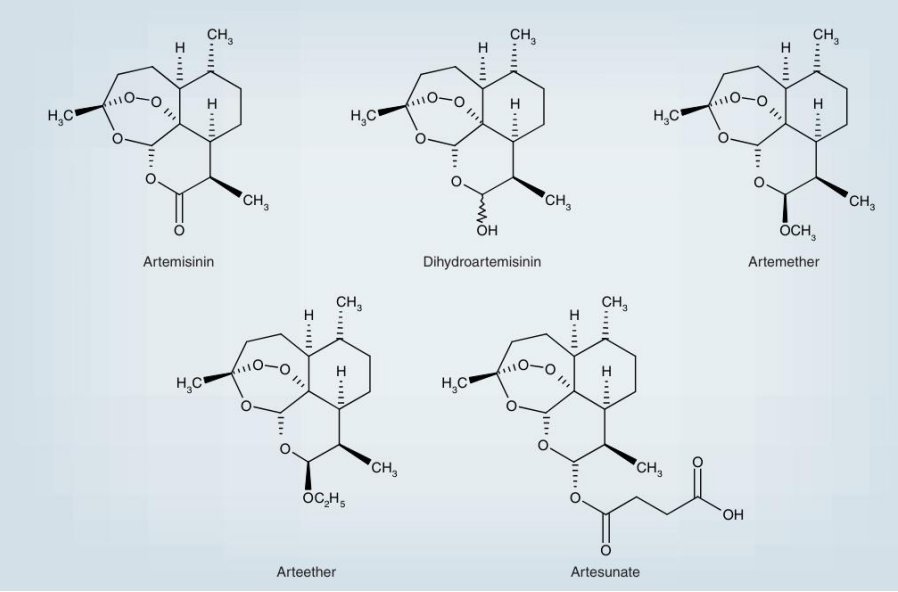


Figure 3: Artemisinin compounds for malaria treatment (Cui & Su, 2009b)

The second category encompasses the aryl amino alcohol compounds, which include chloroquine, quinine, quinidine, tafenoquine, amodiaquine, halofantrine, mefloquine, lumefantrine, and piperaquine, shown in Figure 4. These drugs act primarily by interfering with the heme detoxification process in the malaria parasite, leading to its destruction. While some of these drugs, like chloroquine, were once commonly used for malaria treatment, the emergence of resistance has limited their effectiveness in certain regions (Quiliano et al., 2016; Shibeshi et al., 2020).

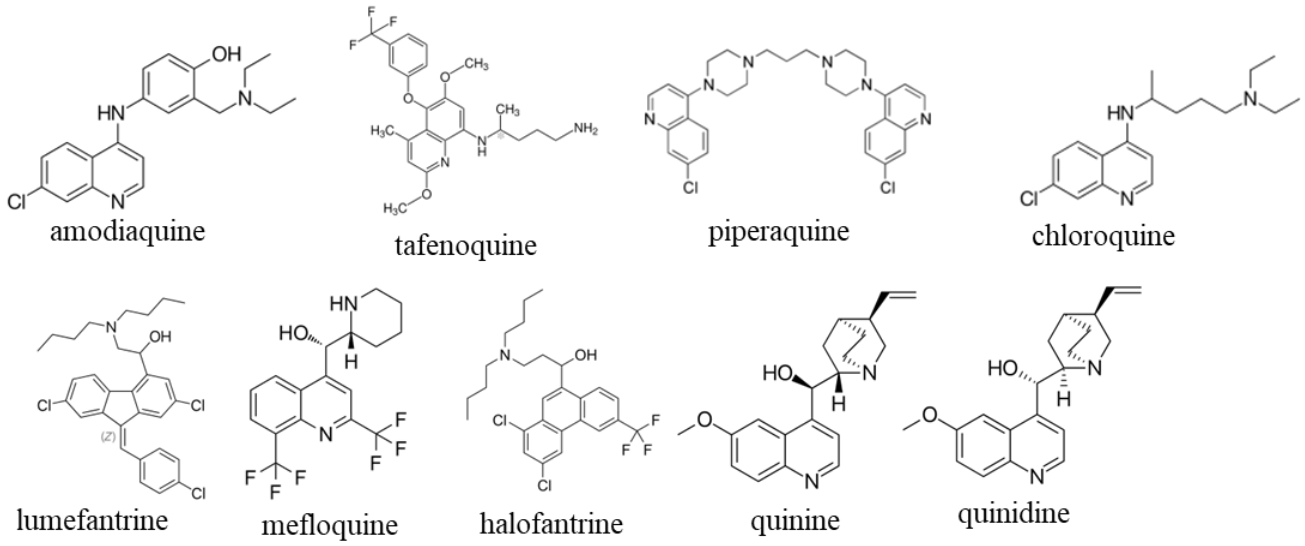


Figure 4: Aryl amino alcohol compounds for malaria treatment.

These compounds are known to disrupt major *P. falciparum* pathways, these include heme detoxification, nucleic acid metabolism and fatty acid biosynthesis aided by the production of reactive oxygen species leading to similar cell complications observed with artemisinin-based compounds (Alam et al., 2009; Egwu et al., 2021). Such action has been observed for mefloquine when used in combination with artemisinin and its derivatives, the generated reactive oxygen species are responsible for DNA degradation thus, inducing apoptosis (Gunjan et al., 2016) . Aminoquinoline compounds have been observed to specifically target the heme detoxification pathway by binding onto the highly toxic ferriprotoporphyrin IX which is released when the parasite degrades hemoglobin through oxidative denaturing or proteolytic degradation. Aminoquinoline compounds including chloroquine, amodiaquine, and piperaquine, bind tightly to ferriprotoporphyrin IX, diverting it from being converted to hemozoin, a non-toxic form of heme. Presence of these compounds during hemoglobin breakdown allows for the accumulation of chloroquine-ferriprotoporphyrin IX complex within the parasite, a slow release of ferriprotoporphyrin IX which accounts for cell lysis by destroying the membrane leading to the death of the parasite (Fitch, 2004; Fitch et al., 1984; Radfar et al., 2008) .

The third category comprises the antifolate compounds, which include trimethoprim, proguanil, pyrimethamine, and chlorproguanil. These drugs target the parasite's ability to synthesize folate, an essential molecule for its survival. By inhibiting folate synthesis, these antifolates disrupt the

parasite's DNA synthesis and cell division, ultimately leading to its demise (Müller & Hyde, 2010; Yuthavong, 2013).

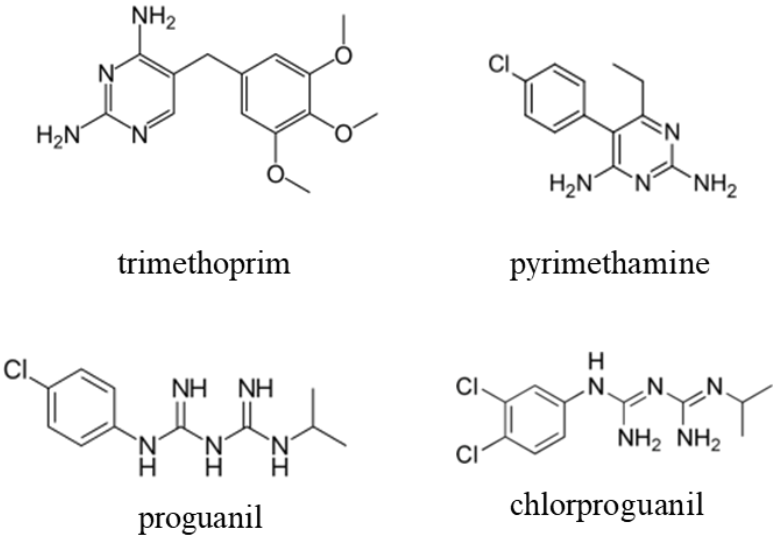


Figure 5: Antifolate compounds for malaria treatment

The current front-line antimalarial drugs are derived from the Artemisinin group, primarily, with a combination of other drugs. In comparison to quinine, artemisinin compounds show a faster clearance for the parasite and a reduction in fever in cases of severe malaria (Ouji et al., 2018; Saifi, 2014). The other class, rather new to these antimalarials, are the hydroxynaphthaquinones with Atovaquone being the only drug in the market as Malarone. Malarone contains Atovaquine and Proguanil (Blanshard & Hine, 2021; Saifi, 2014).

Three methods of treatment exist, *falciparum*, non-*falciparum* malaria treatment, and prophylaxis. Prophylaxis treatment is a preventive measure for individuals travelling to high malaria infection risk areas, *falciparum* method depends on the severity of disease manifestation (Al Khaja & Sequeira, 2021; Saifi, 2014). Myanmar has reported antimalarial drug resistance, characterized by significant delays in the clearance of *P. falciparum* parasites following artemisinin treatment. Myanmar has the highest burden of malaria from the *P. falciparum* parasite in the Southeast Asian region and resistance to artemisinin treatment had not been reported (Kyaw et al., 2013).

2.3 Use of 1,3,5-Triazine Derivatives as Therapeutic Agents

Tetrahydro-1,3,5-triazine-2-amine (THT) derivatives have been predicted and confirmed to inhibit dihydrofolate reductase (DHFR), a crucial gene in *Mycobacterium tuberculosis* for the conversion of dihydrofolate to tetrahydrofolate using NADPH as a cofactor (Mugumbate et al., 2015). Tetrahydrofolate is crucial for thymidylate and purine nucleotide synthesis, and amino acid metabolism serving as a one-carbon unit carrier (Tjong et al., 2023; Zarou et al., 2021)

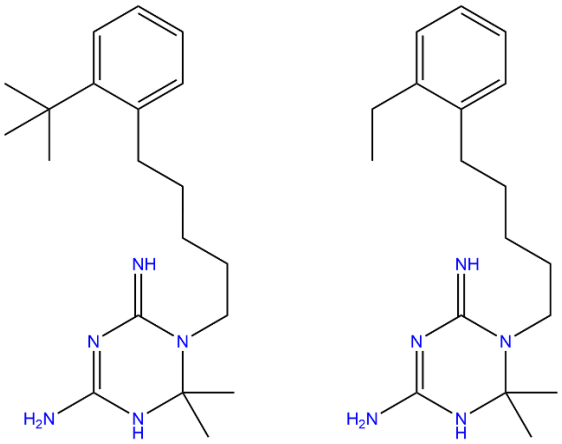


Figure 6: THT1 (left) and THT2 (right).

Similar work has been done on *Pf*-DHFR confirming phenylthiazolyl-1,3,5-triazine and thiazole-1,3,5-triazine derivatives as inhibitors for *Pf*-DHFR. The researchers created two Quantitative Structure-Activity Relationship (QSAR) models using a supervised machine learning technique. Anti-malarial compounds' 2D and 3D fingerprints were gathered in order to create a structure-activity model of the substance's biological action. A four-factor Kernel partial least squares regression (KPLS) 2D-QSAR model was generated from a training set of 47 compounds. The model showed a good internal correlation of R^2 =0.79 and a high predictive power of Q^2 = 0.78 from a test set of 10 compounds. The counterpart 3D-QSAR model was generated from a training set of 45 compounds with 5 latent variables and validated with the leave-one-out method on a test set of 12 compounds. This one also had an impressive internal correlation of R^2 = 0.82 and a score of 0.81 for external correlation. 1,3,5-triazine-2 amine derivatives predicted to have activity against drug-resistant strains of *P. falciparum* were synthesized and found to be active against the

parasite with 100% inhibition and IC_{50} values of 3.02 and 2.17 µM (Gahtori et al., 2012; S. Sahu et al., 2019, 2020).

Similar work on *P. falciparum* has been carried out with DHFR as the target protein. A set of 4,6-diamino-1,2-dihydro-1,3,5-triazine (Figure 7b) derivatives were synthesized and the activity of these compounds was compared against A16V+S108T mutant DHFR, a cycloguanil resistant strain. The activity of cycloguanil (Figure 7a) was tested against the wild-type *Pf*DHFR and A16V+S108T mutant DHFR to give a K_i value of 1.2 nM and 1314 nM respectively. A screening of 4,6-diamino-1,2-dihydro-1,3,5-triazine library against the two proteins gave a few comparable K_i values for the wild-type DHFR, however none of the compounds had better efficacy compared to cycloguanil. The K_i values for the 4,6-diamino-1,2-dihydro-1,3,5-triazine library reported for the A16V+S108T mutant DHFR gave three leads with activities of 35.5 nM, 146 nM and 316 nM which were 84-, 38 and 4-fold more effective than cycloguanil (Vilaivan et al., 2003)

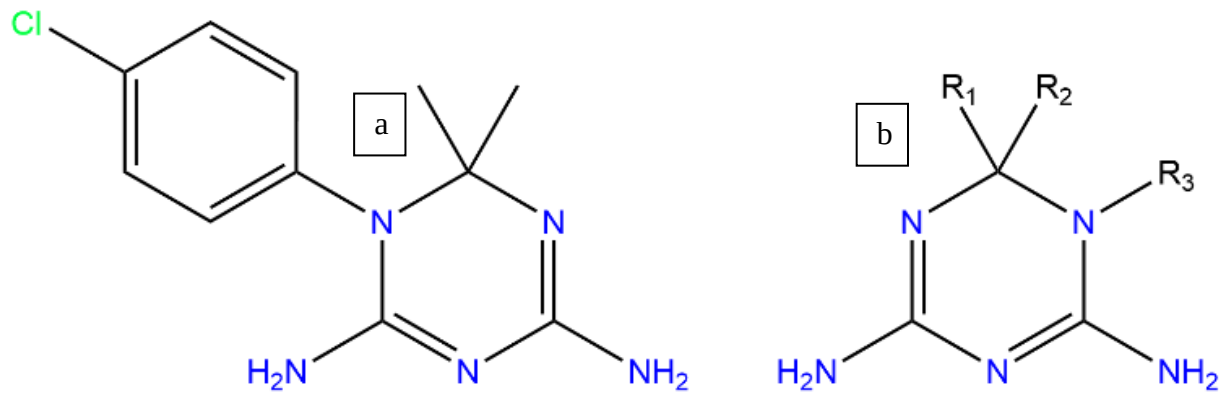


Figure 7: Cycloguanil (a) and 4,6-diamino-1,2-dihydro-1,3,5- triazine moiety (b).

Klenke and co-workers synthesized *s*-triazine substituted polyamines as agents for combating *Trypanosoma brucei*, a Human African Trypanosomiasis causative organism. This investigation looked at the chemical's potential to prevent the P2 transporter from absorbing adenosine. The output of this work saw most of the synthesized 1,3,5-triazine-2-amine compounds having a Ki range of 3-70 µM, thus indicating affinity towards the P2 transporter as moderate (Klenke et al., 2001). The same compounds were used in another body of work, targeting the malaria parasite *P. falciparum* at the erythrocyte stage. The *Pf*NF54 strain, which at the time was susceptible to well-

known antimalarial medications, and *Pf*K1, which was resistant to pyrimethamine and chloroquine, were the three strains utilized in the experiment. Chloroquine and artemisinin, two common antimalarial medications, were utilized to compare their activities. Overall, the research showed that 1,3,5-triazine derivatives were more effective against the *Pf*K1 strain than the *Pf*NF54 strain. A noteworthy finding was that compounds' activity increased when additional triazine units were added to them. (Klenke et al., 2003). 1,3,5-triazine-substituted polyamines have been found to exhibit anti-plasmodial action against *P. falciparum* in vitro by Klenke and colleagues.

Manohar studied chloroquine hybrid molecules of 1,2,3-triazole and 1,3,5-triazine for their antimalarial activity on the *P. falciparum* W2 and D6 strains which are resistant and sensitive to chloroquine respectively. The chloroquine-resistant strains mean activities ranged from 0.73 µM to 5.97 µM, while those of the chloroquine-sensitive strain ranged from 0.58 µM to 5.25 µM. This research showed that adding a 1,3,5-triazine moiety to antimalarial drugs increased their effectiveness, as shown by comparable IC_{50} values for resistant and susceptible strains of *P. falciparum* (Manohar et al., 2011).

According to reports, 1,3,5-triazine derivatives are being used as anti-cancer agents in the twenty-first century (Cascioferro et al., 2017). The extensive study of the *s*-triazine moiety has been motivated by its immense potential for use, with an array of biological properties ranging from the anti-inflammatory, anti-viral and anti-bacterial activity. In their review, the researchers report on several *s*-triazine compound with high potency and selectivity as antitumor agents. Figure 8a show a di-substituted molecule with an IC_{50} activity of 0.02-0.007 µM in an *in vitro* assay against cyclin dependent kinases and glycogen synthase kinase, whereas Figure 8b shows a medication that has reached the clinical development stage for the treatment of acute myeloid leukemia (AML). The researchers also reported that a tri-substituted 1,3,5-triazine derivative has significant in vivo and in vitro action against Hsp90 (Figure 8c).

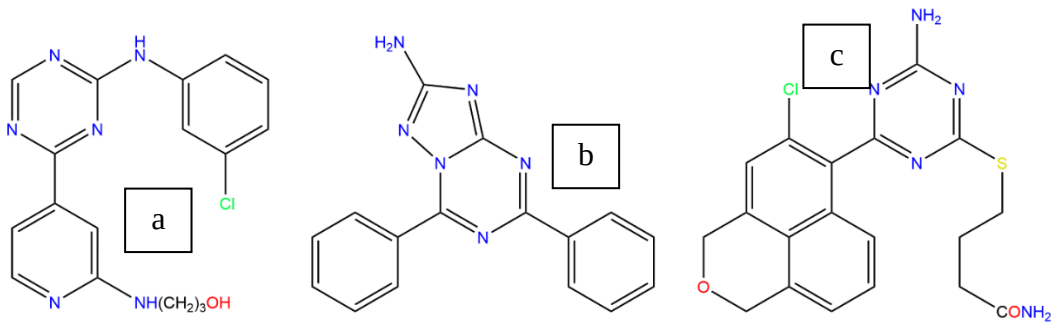


Figure 8: 1,3,5 triazine derivatives as anti-cancer agents a, b and c.

Bahar explores the use of 1,3,5-triazine derivatives in peptidomimetic studies to combat the rise of multidrug resistant bacteria. The purpose of this work was to use host defense peptides that are naturally occurring antimicrobial peptides. These have a broad spectrum of microbes to target, but because mammalian hosts are particularly harmful to them, peptidomimetics with the 1,3,5-triazine moiety at the parent molecule are being prioritized. Biological assays with *E. coli* and *Pseudomonas aeruginosa* showed complete inhibition of biofilm formation at 20µM. These studies also revealed that *E. coli* RP437, PDO300 and *P. aeruginosa* PAO1 and growth were completely inhibited at 12.8 µM. The results of the trials demonstrate the enormous potential of 1,3,5-triazine derivatives (Figure 9) as antibacterial agents (Bahar et al., 2015).

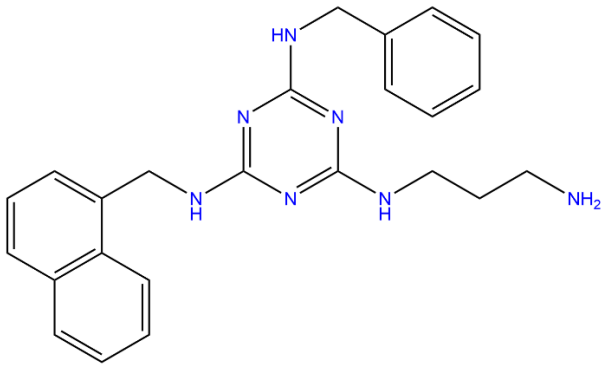


Figure 9 : 1,3,5-triazine derivative active against *E. coli* and *P. aeruginosa*.

Postulations have been made about PS-15 shown in Figure 10, a pro-drug with a triazine metabolite and their action against *P. falciparum* being of a non-folate mechanism. Experimental work carried out by Kinyanjui revealed weak in-vitro activity of PS-15 in *P. falciparum* compared

to its metabolite WR99210 which is a triazine compound. Four strains of *P. falciparum* were used for this work, VI/S, W282, M24 and K39 and the expressed IC_{50} values in ng/mL were 379 and 0.039, 445 and 0.034, 209 and 0.032, 206 and 0.031 for PS-15 and WR99210 respectively. WR99210 is 10,000 more potent than its parent compound PS-15, revealing the antagonist action of WR99210 primarily as an antifolate (Kinyanjui et al., 1999). PS-15 has higher activity against *P. falciparum* compared to its close analog proguanil which has a triazine metabolite known as cycloguanil, a known antimalarial drug with proven antifolate inhibition targeting DHFR. Recent work has also shown that cycloguanil has a synergistic affect went used in combination with atovaquone (Thapar et al., 2003; Watkins & Sixsmith, 1984). These results point out to triazine analogs being antifolate while their prodrugs act with non-antifolate mechanism.

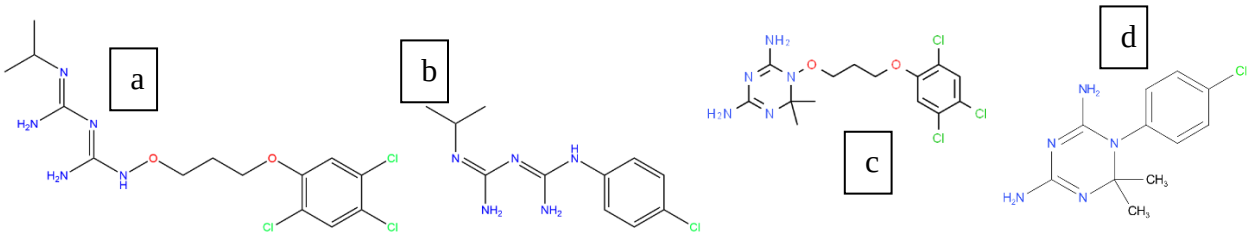


Figure 10 : (a) PS-15 (b)Proguanil (c) WR99210 and (d) cycloguanil.

2.4 Privileged Scaffolds in Medicinal Chemistry

One of the most ubiquitous N-heterocyclic compounds abundantly present in nature are purine based, shown in Figure 11. The core structure of guanine and adenine is a purine moiety, purine nucleotides such as adenosine triphosphate, guanine triphosphate, cyclic guanosine monophosphate, cyclic adenosine monophosphate, flavin adenine dinucleotide nicotinamide adenine dinucleotide are central in the function of many proteins as substrates, co-factor or mediators in biological functions (Legraverend & Grierson, 2006). This makes purines one of the most privileged scaffolds of interest to medicinal chemists for drug discovery, a huge number of marketed drugs have specific enzyme targets, thus the potential for development of new drugs based on this moiety is high (Cohen & Alessi, 2013).

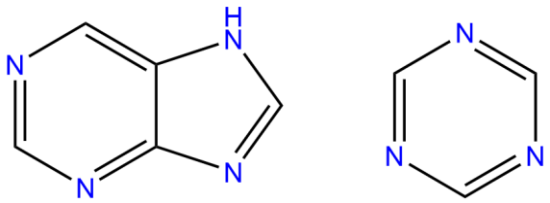


Figure 11 : Purine (left) and *s*-Triazine (right).

The fusion of 1,3,5-triazine with a 5-membered ring gives off a 7-atom bicyclic ring system which has a huge potential for being a purine bio isostere. Using nitrogen containing 5-membered rings like pyrazole, pyrrole, 1,2,4-triazole, 1,2,3-triazole, or imidazole is central to producing a purine bio isostere such as [1,2,4]triazolo[4,3-a][1,3,5]triazine catapulting 1,3,5-triazine into the class of privileged scaffolds in medicinal chemistry (Lim & Dolzhenko, 2014).

Welsch and colleagues provide a systematic approach for developing new therapeutic medicines that uses scaffolds that are already in existence, whether they are found natively in plants or as naturally occurring protein inhibitors in humans. This method comes about as a creative approach to drug design as several analogues of the molecule may be created to tackle different diseases with compounds with a high bio-specificity toward their target (Welsch et al., 2010).

The use of purine has seen a steady growth in the pharmaceutical industry over the last two decades which has seen compounds shown in Figure 12, abacavir, an antiviral for HIV, cafaminol which is a nasal decongestant, and pentoxifylline which is a hemorheological agent, making it to the market. Among them is caffeine, a stimulant that naturally contains the purine scaffold (Welsch et al., 2010).

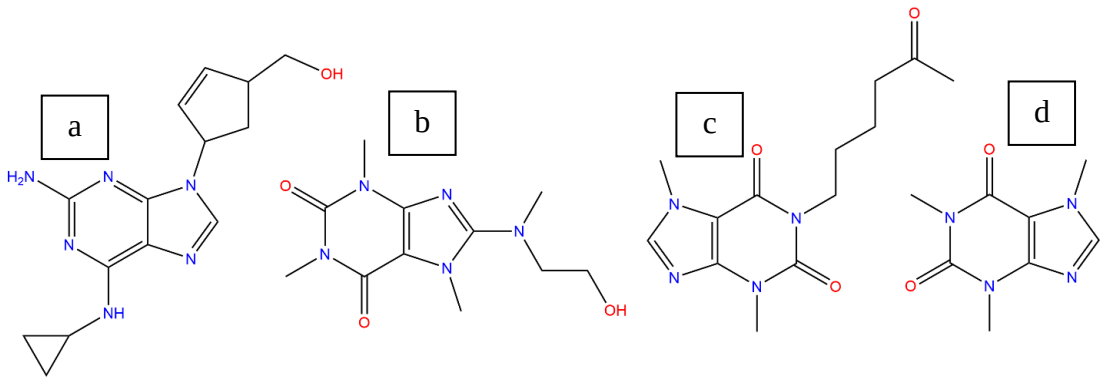


Figure 12 : Abacavir (a), Cafaminol (b), Pentoxifylline (c), and Caffeine (d).

The extensive use of 1,3,5-triazine derivatives as potential therapeutic agents in the antimicrobial, anti-tumor, and anti-inflammatory across an array of species including *M. tuberculosis, P. falciparum P. aeruginosa, E. coli* and *T. brucei* is a fascinating observation in medicinal chemistry. Their proven track record in specifically targeting the folate pathway has without doubt catapulted them into the realm of privileged scaffolds in medicinal chemistry. 1,3,5-Triazine purine bioisosteres are formed by combining this scaffold with nitrogen-containing rings such as pyrazole, pyrrole, 1,2,4-triazole, 1,2,3-triazole, or imidazole. This combination is made possible by the scaffold's unique functionality.

Central to creativity in medicinal chemistry, is the ability to rapidly manipulate chemical compounds as seen with the generation of purine bio isosteres and rapidly testing them against an array of biological targets in high throughput screening. Such efforts have been made possible by the growth of a multidisciplinary field of science concerned with collection, analysis and interpretation of biological data. This data driven field was coined bioinformatics by Paulien Hogeweg and Ben Hesper, and has since grown to include physics, computer programming, mathematics, information engineering and chemistry, all with the aim of pattern recognition (Hogeweg, 2011). The following sections will delve deeper into the methods, techniques and computer resources used in carrying out experimental work *insilico*.

2.5 Comparative/ Homology Modelling

The comprehension of biological systems depends upon a clear understanding of protein functionality. An in-depth knowledge of the protein structures involved in the complex networks in biological systems is required for this. Due to recent advances in technology and instrumentation, scientists can now determine the 3D structure of proteins with high accuracy and resolution. Recent advances have seen a human apoferritin 3D structure being determined at a resolution of 1.25 Å with single-particle electron cryo-microscopy techniques. When compared to the 1.54 Å structure of the same protein, which was the previous record holder for the world record of reconstruction, the structure recovered had twice as much information in its three-dimensional form (Kato et al., 2019; Yip et al., 2020).

Aside from electron microscopy, nuclear magnetic resonance (NMR) also provides great insight into elucidating the 3D structure of proteins (Clore & Gronenborn, 1998). The protein system is dynamic hence it is constantly changing from one state to another, several techniques are used to significantly reduce the time required to elucidate the 3D structures of proteins. Using paramagnetic NMR spectroscopy, (Pan et al., 2016) substantially shortened the time needed for structure elucidation to just 2 hours. Long-range structural information is produced throughout the NMR experiment with the help of paramagnetic relaxation enhancements, pseudo contact shifts, and residual dipolar couplings (Marius Clore & Iwahara, 2009; Müntener et al., 2020). These are invaluable tools giving us structural information that allows scientists to study structure and function relationships through local motifs and molecular folding championing protein evolution (Dorn et al., 2014).

With the use of X-ray crystallography, the crystallized structures of expressed proteins in their purest form can also be clarified. Due to recent developments in ultrabright electron and X-ray sources, this method is also capable of determining crystallized structures in the femtosecond time period, enabling researchers to record atomic motions necessary for protein structural transformations (Garman, 2014; R. J. D. Miller, 2014).

With such advances in instrumentation, databases like the Protein Data Bank (PDB) have been created for the storage of structural data collected through these structure determination experiments. A closer look at these structures from different species has revealed some similarities within these structures and this is the breakthrough that had led to structure prediction methods (Dutta & Berman, 2005). It is noted that certain structural domains are conserved among homologous complexes, providing an abundant resource of templates for the modeling of almost all proteins through known protein-protein interactions (Kundrotas et al., 2012; Q. C. Zhang et al., 2010).

When a 3D structure of a homologous protein (template) is available, using the primary sequence from the target protein to be modelled, the quaternary structure from the template is transferred onto the sequence to mold the target protein accurately. This is achievable due to the presence of conserved domains on the protein in both the target and template (Waterhouse et al., 2018). Minuscule changes in the protein's primary sequence barely affect the resultant structure of protein in a significant manner, as this has been noted in mutant proteins. Evolution has seen the mutation

of proteins overtime; however, these mutations are constrained to retain protein functionality by conserving intermolecular interactions and intramolecular interactions responsible for mediating family and superfamily functions by proteins (T. Liu et al., 2011; Mehlhoff et al., 2020; Yu et al., 2015).

Comparative modelling predicts an accurate structure of an unknown compound from information obtained from a homologous partner, 40% sequence identity can produce X-ray or NMR resolutions (Lam et al., 2017; Sánchez & Šali, 1997). Sequence similarity is crucial for assessing the precision and caliber of the model that will be applied, it is possible for one to make an estimate *a priori*, an indication of the usefulness of the model for application in further work as seen with COVID19 (Cozzetto & Tramontano, 2005; Rabaan et al., 2020; Tahir ul Qamar et al., 2020; Tilocca et al., 2020; Townsend et al., 2021) Automated systems being developed address simplicity in the modelling pipeline, which can be summed up in a few steps. These may be split into (i) template structure identification, (ii) sequence alignment for the query and template, (iii) modelling the query structure from the template, and (iv) model evaluation (Lam et al., 2017).

Template structure identification encompasses comparison of target to homologous proteins with known 3D structural features, more often the function of the template and its sequence similarity in conserved domains is of paramount importance. The most common search tool for this task is the basic local alignment search tool (BLAST) which is hosted by the National Center for Biotechnology Information (NCBI) http://www.ncbi.nlm.nih.gov/BLAST/. This tool works by identifying hot spots on the sequences, which are short matches between two sequences, in a process called seeding. Common three-letter sequences and repetitions are exploited to get the best match. It is at this important stage that estimates of model quality may be deduced (Boratyn et al., 2019; Chakravarty et al., 2020; Rayan, 2009; Ye et al., 2006).

The next step after identifying the template involves overlaying the primary sequence of the target on the template. This method establishes an accurate correspondence between the target and template by inferring evolutionary homologous sites on the templates, these sites may include motifs or conserved regions through ancestral profiling. The alignments may be split into two, global alignment which attempts all residue matching, where sequences are estimated to be of equal size, or local alignments useful for sequences of different sizes hence the methods seeks regions with high similarities (Notredame, 2007; Webb-Robertson et al., 2008). Global alignment

assumes sequential identity of the amino acid sequences whilst local alignment assumes smaller regions conserved throughout the homologous partner are similar, to such effect, the global alignment is seldom accurate in its assumptions due to genome reshuffling (Gong et al., 2009). Multiple sequence alignment is a sophisticated method of aligning more than two sequences from homologous sequences, it is more sophisticated compared to pairwise alignment (González-Domínguez, 2021; Notredame, 2007). The most common multiple sequence alignment algorithms are Clustal Omega and W, Tree-based Consistency Objective Function for Alignment Evaluation (T-Coffee), MUltiple Sequence Comparison by Log-Expectation (MUSCLE) (Daugelaite et al., 2013).

Building of the model after sequence alignment involves atomic coordinate prediction using the defined equivalent residues from sequence alignment with three methods being preferred, segment matching, rigid body assembly, and spatial restrains satisfaction. MODELLER has seen success through its accuracy in building structures by the use of spatial restraints satisfaction. Homology derived spatial restrains are generated from the alignment of the target and the template based on the notion that similar residues always conserve structural features of a protein. The satisfaction of restrains is upheld in the optimization of the structure through molecular dynamics with simple annealing and a conjugate gradient algorithm, all while stereochemical realism of the amino acids is retained (Janson et al., 2019; Wallner & Elofsson, 2005; Webb & Sali, 2016). Segment matching relies on the conservation of conformational space (motifs and folds) to map out the backbone of conserved residues on the protein, the whole protein is then completed based on knowledge based and X-ray refined known building blocks of peptides obtained from known 3D structures, to produce the final structure. An amino acid sequence is used in choosing the correct segment to build the model, including conformational similarity. Both should also satisfy the van der Waals' interactions to ensure there are no clashes within the amino acids (Levitt, 1992; Pitman & Menz, 2006). In rigid body assembly there is the super-positioning of 3D structures to obtain a framework characteristic of the set of atoms that lies within the known structures cluster, the atomic coordinates from the conserved regions of the target are then mapped on the generated topologies to form the core of the protein, structural databases are then probed for loops that will satisfy the conformation of atoms in the core. The structure is then generated with constraints meant to satisfy residue conformation equivalent to the template. This method is fragment-based, and takes advantage of structurally conserved regions like α helices, regions that are typical of active sites

to an enzyme, thus the target sequence is aligned to those important regions. Using several templates to construct the target structure proved to be accurate (T. Liu et al., 2011; Pitman & Menz, 2006; Sutcliffe et al., 1987).

Model evaluation or validation can be carried out with several external tools, the crystal structure of human heat shock protein 70 PDB code 4IO8 will be used for illustration. Protein Structure Analysis ProSA as shown in Figure 13 from left to right illustrating a plot for overall model quality, local model quality, and visualization, is available at https://prosa.services.came.sbg.ac.at/. This is a commonly used method for the structural validation of 3D protein structures derived from comparative modeling. The entire model quality is assessed through a z-score using a database of experimentally established structures from NMR or X-ray crystallography. The total energy of the model is evaluated against random confirmation derived energy distributions. A plot is generated from known data about experimentally determined structure to indicate the size (number of residues) to the z-score, thus, the model is added to the plotted graph to indicate whether it falls within acceptable experimentally derived z-score or not.

The other criteria for local model quality with ProSA illustrate the relationship between energy and amino acid, statistical parameters derived from known amino acid and protein properties are used to guide the knowledge-based energies. To produce a smooth figure, the average of each amino acid is determined over 40 residues; positive values suggest model flaws. To visualize all this data in a simple way, a protein structure is shown in 3D with colored regions for unusually high energies, this shows more detail on where the energetically unfavorable regions are located (Dasgupta et al., 2020; Papathoti et al., 2020; Wiederstein & Sippl, 2007).

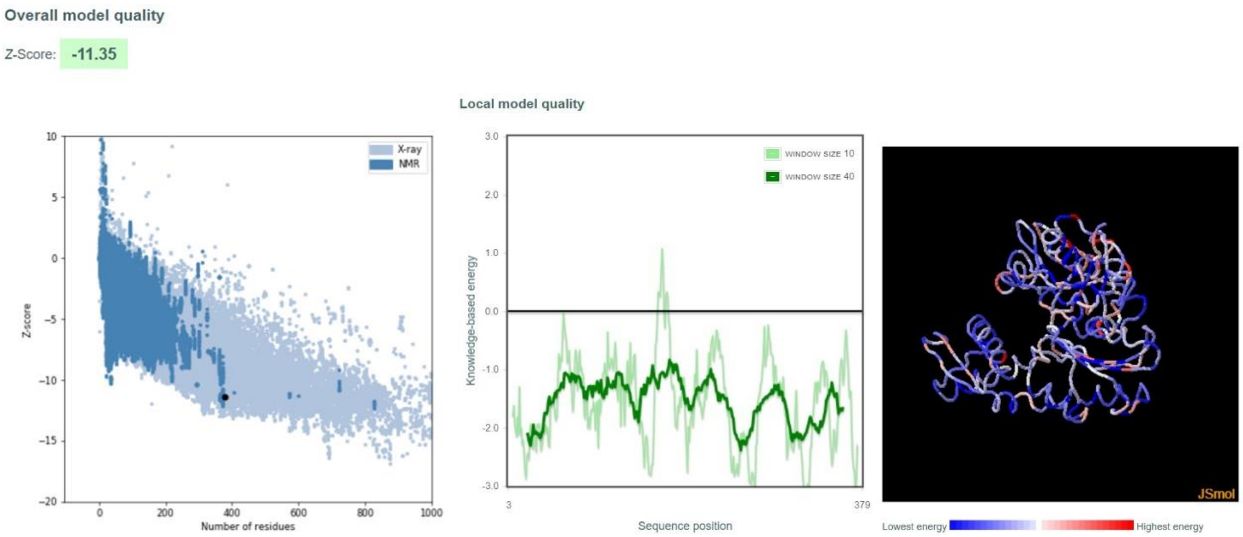
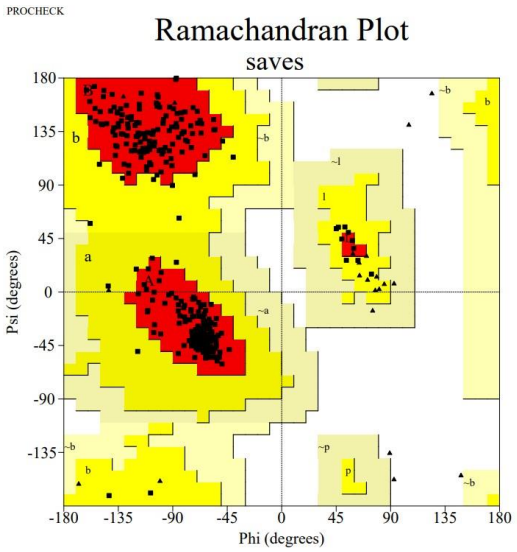


Figure 13 : ProSA plot for overall model quality, local model quality, and visualization

The Ramachandran plot as shown in Figure 14, examines the Psi and Phi backbone dihedral angles to identify the energetically permissible backbone areas. To assess the model's performance on the permitted structural geometries, the angles and are displayed against one another in the plot. There are authorized regions and banned regions because stereochemical collisions prevent some amino acid geometries from being used in produced structures.



Plot statistics

Residues in most favoured regions [A,B,L]	311	92.8%
Residues in additional allowed regions [a,b,l,p]	24	7.2%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	335	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	12	
Total number of residues	377	

Figure 14: PROCHECK Ramachandran output.

The white areas are sterically disallowed regions where polypeptide atoms are closer causing steric clashes, except for glycine, red corresponds to allowed regions with no steric clashes and yellow corresponds to additionally allowed regions where atoms may be closer to each other, and the light off-white are generally allowed regions (Hollingsworth & Karplus, 2010; K. Gopalakrishnan et al., 2007; Laskowski et al., 1993; Zhou et al., 2011).

2.6 Molecular Dynamics

There is a huge disparity in who can identify new drug leads because it can cost upwards of a billion dollars to develop a drug that is market-approved. Computer-aided drug design has included a lot more players in the sphere of lead identification, however, there is still need to address the flexibility of proteins. Molecular docking aims to position a flexible ligand optimally into a rigid active site. To address this challenge of rigidity, a computational technique that simulates particle motion within the system is utilized. Molecular dynamics simulation (MDS) is the utilized technique for simulating the motion of particles in a well-defined system. The evolution of a system in molecular dynamics is probed by allowing atoms to interact for a fixed period. Determining atom trajectories is done by solving Newton's laws of motion (Ganesan et al., 2017; Karplus & Petsko, 1990). Simulations are valuable for assessing the free energy of a system. They achieve this by tracking the individual motion of particles over time, thereby revealing crucial properties of the system under investigation. The energy of the investigated system is represented based on atomic coordinates, enabling the determination of metastable or stable structures and their respective relative stabilities from the potential energy surfaces. The basics of a molecular dynamic simulation would ideally involve a protein in water surrounded by a lipid bilayer, each atom exerts a force on other atoms. By applying Newton's laws of motion, the positions of individual atoms in the system are calculated over time, and this calculation is iteratively repeated at each time step to update the velocities of the atoms. The result is a movie of the motion of every atom within the system (Hollingsworth & Dror, 2018).

The ability to capture the motion of every atom during a simulation, including a precise control of simulation conditions like the voltage across the membrane, the initial conformation of the protein, temperature, molecules surrounding the protein, and bound ligands make the use of MDS useful in studying complex structural changes (Fu et al., 2018; Venable et al., 2019). Biological structural changes usually occur in the nanoseconds to microsecond timeframes hence the time steps of

molecular dynamic simulations typically fall in the femtosecond range (Acun et al., 2018). At its most fundamental level, MDS is employed to study the flexibility and dynamics of proteins, as well as the movement of solvation molecules (such as ions and water) both inside and around the protein. This approach provides valuable information on how proteins behave in contrast to experimental methods, which often offer insights into the proteins' average structure (Docampo-Álvarez et al., 2016).

Simulations are also applied in fixing of biological structures modelled from templates, or elucidated from experimental work (Hospital et al., 2015; McGreevy et al., 2016). X-ray and NMR crystal structures often have structural abnormalities because of a non-existent lipid bilayer for membrane proteins or lattice packing of the crystals being formed in a non-relaxed environment because the protein was purified in an environment which does not resemble the conditions it exists in within the original host ecosystem. Due to such artifacts existing, the binding poses of co-crystallized ligands may be affected, including for those binding poses from molecular docking. More-often freely accessible computer packages for docking studies (AutoDock and Vina) are not complex and do not address solvation within the binding pocket (Bain et al., 2010; Burg et al., 2015; Jaghoori et al., 2016). In such cases, water molecules can play a significant role in facilitating interactions between the ligand and protein. Additionally, ions present may influence the ionization state of residues or the ligand itself, making them important factors to consider (Smith et al., 2000; Yadav et al., 2020). In such a case, the binding affinities obtained through molecular docking are not accurate representations of the protein’s affinity for the ligand. MDS play a significant role by relaxing the protein within the membrane and adding water and ions normally present for the protein in the original host, the interactions between the ligand and protein may be accurately assessed as the system resembles what may be normally observed within the host organism (Bertalan et al., 2020; Jong et al., 2017).

Under ordinary temperatures, the vibrational energies of small molecules yield comparable representation of bonding energy functions because the angles and bond lengths of the polypeptide chain are near their equilibrium state (Hansson et al., 2002; Karplus & Petsko, 1990). In a 3D structure of a globular protein, the close proximity of atoms due to the folds and motifs leads to significant interactions that greatly influence the system's potential energy. Modelled with Lennard-Jones potentials, the energy function of the system includes both non-bonding terms, such

as van der Waals (vdW) and electrostatic interactions, and bonding terms like torsional angles (expressed as a cosine function), bond length, and angles (modelled using harmonic potentials) (Ganesan et al., 2017).

Classical MDS regard bonds as springs and atoms as solid spheres, this allows the use of Newton's equation of motion as the system allows atoms to move within a constrained distance that is energetically favorable providing insight into the structural changes of a biological system which is a time dependent variation related to its function. Newton's equation of motion is the basis for molecular dynamics, and it is given as (Ganesan et al., 2017):

$$m_i \frac{\delta^r r_i}{\delta t^2} = F_i \qquad \text{Equation 1}$$

Equation 1: Newton's equation of motion.

F_i is a part of the net force that exerts an action on the ith atom with a particular mass of m_i, where r_i is the symbol for the position of the given atom at time t:

$$F_i = -\frac{\delta U(r_1, r_2, \ldots, r_n)}{\delta r_i} \qquad \text{Equation 2}$$

Equation 2: The part of Newton's equation describing motion's net force.

Where $U(r_1,r_2\ldots,r_n)$ is specific for the conformation, giving us the potential energy function described by predefined parameters of a mathematical expression comprised of potential energy from non-bonded and bonded interaction terms. This mathematical expression is what is called a force field.

Over time, the movements of atoms in a system are calculated by integrating Newton's equation of motion (Equation 1) which is derived by discrete numerical approximation of time which has been partitioned into time steps for propagating the system forward. The acceleration of a system is then computed for the forces that act on the atom (Salmaso & Moro, 2018).

Structure equilibration and the dynamic interactions in a biological system are studied widely using MDS as the golden computational technique (Anwar-Mohamed et al., 2014; Barakat & Tuszynski, 2011). Most experiments are carried out with the control of ionic concentration, pH, temperature,

moles of reacting compounds and pressure. The ability of MDS to mimic real experimental condition gives this computational technique an edge in understanding biological systems (Ganesan et al., 2017; Hollingsworth & Dror, 2018).

After performing comparative modelling, the protein does not immediately reach the equilibrium phase space, leading to artifactual trajectories that impact its dynamics. To address this, experimental constraints from NMR spectroscopy and X-ray crystallography can be utilized to refine the 3D structures of macromolecules and protein complexes. Additionally, MDS can be employed to model ligand interactions with macromolecules in ligand docking experiments (Salmaso & Moro, 2018b).

MDS may be summed up in three steps; these are the creation of an initial state of the system, introducing interaction potentials, and the prediction of particle movement in the system created. In the GROMACS software suite, the MDS pipeline involves several key steps to study the behavior of biomolecules and their interactions within a defined system. Firstly, the simulation begins by specifying the initial state of the system, where the simulation box is defined, and the positions of different particles (atoms) within the box are set. Additionally, the presence of solvents and ions in the system is accounted for, and parameters such as temperature and pressure are specified to create a realistic environment for the simulation. Next, the interaction potentials that dictate the nature of particle interactions are introduced. GROMACS employs a force field, which consists of three types of interaction potentials: non-bonded, bonded, and external terms. Long-range forces between particles, such as electrostatic interactions and vdW forces, are accounted for by the non-bonded interactions. The bonded interactions describe covalent bonds, angles, and dihedral angles between connected atoms within molecules. The external terms include interactions with external fields, such as electric or magnetic fields. Subsequently, the simulation predicts the movement of particles using Newton's laws of motion. GROMACS computes the positions and velocities of particles over short time steps by solving the equations of motion. This dynamic movement enables the particles to respond to the forces acting on them. This process iterates through multiple time steps, effectively simulating the system's behavior over time, providing insights into its dynamic properties, structural changes, and thermodynamic behavior. The combination of these steps in the GROMACS pipeline allows researchers to study complex biological systems and gain valuable information about their interactions, conformational changes,

and function. Through MDS, GROMACS has become essential in computational biology, aiding in the understanding of biomolecular mechanisms and contributing to the development of novel therapeutics (Van Der Spoel et al., 2005).

New particle velocities and locations are produced after computing the forces acting on each particle in the system based on their potential for interaction. The positions and velocities of the particles within the system are altered in accordance with the control of parameters like temperature and pressure, and the procedure is repeated numerous times to produce the particle trajectories. The behavior of the system is inferred through the movements of the particles within the whole system (Hollingsworth & Dror, 2018; Kandt et al., 2007). The MDS packages usually used are CHARMM, GROMACS, AMBER, NAMD and OpenMM (Brooks et al., 2009; Case et al., 2005; Eastman et al., 2013; Phillips et al., 2005; Van Der Spoel et al., 2005).

Existing experimental techniques lack the capability to thoroughly examine spatial resolutions in complex biological systems, which is a crucial aspect of comprehending the dynamics of protein-ligand interactions. Pairing these experimental techniques with MDS provides insights into the mechanisms driving the behavior of the systems which occur in the millisecond range (Borhani & Shaw, 2011). The control of ion channels for membrane proteins is one example that can be extensively studied with MDS compared to experimental methods. The mechanisms of ligand-gated channels are well understood through simulation studies by controlling the voltage across the membrane, pressure, temperature, and ion concentration within the system, such has been demonstrated by work carried out by (Beckstein & Sansom, 2005).

Monte Carlo grand canonical simulations significantly improved binding site identification and potential ligand prediction for the target. The method involves gradually adding small fragments to the target and then removing them from the binding cavity. This process identifies tightly bound ligands, leading to more accurate and reliable free energy calculations. This is a p38 Kinase inhibitor prediction approach using fragment-based drug design (FBDD). (Moffett et al., 2011).

In order to calculate binding free energies, Abbott Laboratories used molecular mechanics with Poisson-Boltzmann surface area (MM-PBSA). A protein-ligand simulation was carried out to obtain the trajectories of atoms within the system, thermal ensembles obtained from this step are represented by the structures used in the next step of analysis. Energies of the solvation phase and

gas-phase are analyzed in this step, with the subsequent free binding energies obtained through the combination of all the ensembles in the equation below:

$$\langle \Delta G_{\text{bind}}^{\text{MM-PBSA}} \rangle = \langle \Delta E_{\text{MM}} \rangle + \langle \Delta G_{\text{PBSA}} \rangle - T \langle \Delta S_{\text{solute}} \rangle \qquad \text{Equation 3}$$

Equation 3: MM-PBSA ligand binding affinity

Where ΔE_{MM} is for molecular mechanics energy change upon binding, the internal entropy change that occurs when solutes bind is ΔS_{solute}, T is the system's absolute temperature and ΔG_{PBSA} is for the binding-induced change in solvation free energy (Brown & Muchmore, 2007, 2009).

Course-grained MDS have made it possible to gain relevant insight into biological systems by applying realistic timescales which allow for processes to occur, all at the same time spatial details have been sacrificed. An example is a one-millisecond simulation that has identified a biophysics phenomenon of *de novo* protein folding for WW domain for FiP35 protein, the fastest known WW domain. The output of this was a design of a faster variant of FiP35 (Ensign & Pande, 2009; Krivov, 2011; Mori & Saito, 2015).

Another important aspect of MDS is the ability to identify allosteric binding sites and cryptic binding sites, which a difficult to determine with X-ray crystallography and NMR experiments (Durrant & McCammon, 2011). A trench on the HIV integrase was identified using MDS, which drove the development of an approved HIV integrase inhibitor by the USFDA. Raltegravir , was the first antiretroviral of its kind developed by Merck & Co (Grant et al., 2011; Reddy et al., 2012; Schames et al., 2004).

2.7 Structure-Based and Fragment-Based Drug Design

The computer-aided drug design technique known as "structure-based drug design" (SBDD) manipulates and quantifies the features of compounds that will make them potential therapeutic candidates. The impetus behind SBDD is the accessibility of macromolecules' 3D structures. These structures are available in bioinformatics resources like the Protein Data Bank (PDB) (Bernstein et al., 1977). These structures deposited into these databases are obtained experimentally through NMR or X-ray crystallography, or through homology modelling. The availability of these

structures allows researchers to visualize the binding cavities and the binding poses of ligands within the protein. The process of SBDD follows through several steps which are:

2.7.1 Protein Structure Preparation and Binding Site Identification

This is the stage where the conditions for macromolecule are set, these include adding hydrogens, setting the pH, adding charges. Proteins are usually represented without explicit hydrogen atoms in structural databases, adding these hydrogens improves the accuracy of the molecular docking predictions. Biological systems exhibit variability, including in their physiological environments. Setting the appropriate pH ensures accurate protonation states of amino acids, thereby enhancing the experiment's precision. The accuracy of docking relies on electrostatic interactions between proteins and ligands. Correctly assigning the charge states of ionizable residues like lysine, arginine, histidine, aspartate, and glutamate enhances this accuracy (Madhavi Sastry et al., 2013). Clear information can be obtained which helps the researchers to understand the phenomena of protein-ligand interaction, the potential hydrogen bonds to be formed and in turn identify pharmacophores for the binding site. Co-crystallized ligands provide insight into the binding site, and site-directed mutagenesis are also experimental methods that may determine binding sites. Online servers are available to determine binding site, these are DEPTH, MSpocket, PockDrug. Further information on the volume may be obtained by TRAnsient Pockets in Proteins (TRAPP) and POVME (Borrel et al., 2015; Durrant et al., 2011; Hussein et al., 2015; Kalidas & Chandra, 2008; Stank et al., 2017; Wagner et al., 2017; Zhu & Pisabarro, 2011).

2.7.2 Ligand Library Preparation

Filtering chemicals to suit the researcher's preferred criteria may be a part of this stage. It may be more accurate to screen ligands for drug-like compounds using the Lipinski rule of five filter; logP (octanol-water partition coefficient) < 5, Mw (Molecular weight) < 500 Da, HBD (Hydrogen Bond Donors) < 5, and HBA (Hydrogen Bond Acceptors) < 10. A modified version of the rule with logP < 3, MW < 300 Da, HBD < 3, HBA < 3, and RotB (Rotatable Bonds) < 3 particularly filters for lead-like compounds in FBDD (Lipinski, 2004). The ADMET risk score is an additional filter that can be used to weed out substances that are unsuitable for usage as drugs. Risk parameters such as inhibition of 3A4 oxidation of midazolam, hepatotoxicity, and carcinogenicity are important considerations for the removal of compounds that may be troublesome further down the design process (S. Sahu et al., 2019). These procedures are designed to preserve substances with perfect

pharmacokinetic characteristics and compliance with medication safety requirements. The addition of hydrogens, adding Kollman, or computing Gasteiger charges is done, similar to protein preparation (Rabal et al., 2011).

2.7.3 Molecular Docking and Scoring Functions

The initial ligand conformation greatly affects the binding pose to be observed in the end, thus generation of multiple starting confirmations of the ligand greatly improves the accuracy of the results. LigPrep and ConfGen are major software that can carry out the task of generating multiple starting conformations of the same ligand (I. J. Chen & Foloppe, 2010).

The ligands undergo an extensive rearrangement of their conformations to find one conformation with the lowest energy within the binding pockets, this will become the docking pose which will undergo scoring. Scoring functions provide the protein-ligand affinity with speed and accuracy by calculating the interactions between the protein and ligand that are non-covalent, these are hydrogen bonding, salt bridges, vdW and electrostatic forces (Li et al., 2019; Zhao et al., 2021).

The fastest of the scoring functions, the empirical scoring function optimizes hydrogen bonding energy, desolvation energy, vdW interaction energy, hydrophobicity, and electrostatic energy to compute binding affinities between the protein surface area and binding molecule interacting atoms. A trusted scoring function validated through 14 targets is TS-Chemscore (Guedes et al., 2018; W. J. Wang et al., 2015). AutoDock Vina is a docking programs that heavily employs empirical scoring for docked compounds, it includes a finite repulsion term, Gaussian interaction term, entropic term, and hydrogen bond interaction term (Gaillard, 2018). AutoDock has a semi-empirical scoring function that is physics-based and includes the Lennard-Jones 12-6 potential, an entropic term, Coulomb potential, a desolvation term which is volume-based, and a hydrogen-bond term based on the 12-10 potential (Bitencourt-Ferreira & de Azevedo, 2018).

The other scoring function that has seen less success because of its complex mechanism to predict the free energy of binding of protein-ligand complexes is the force field analysis often used by OPLS, CHARMM and GoldScore (S. Wang et al., 2021). The terms included are vdW and hydrophobicity. The knowledge-based scoring function uses experimentally determined geometries to predict free binding energy (Q. Shen et al., 2011).

Relevant success has been seen with SBDD in the development of therapeutic agents. Dorzolamide had been developed targeting carbonic anhydrase by Merck Sharp and Dohme. In Basel, Norvatis developed a compound for targeting BCR-ABL named imatinib. Similar success against the HIV protease target was seen with the development of saquinavir by a firm based in the United Kingdom, Roche. Agouron developed three compounds codenamed AG337, AG331 and AG85 to target thymidylate synthase (X. Wang et al., 2018).

2.8 Bioinformatics and Chemoinformatic Resources

Computer-driven drug design relies on collating data and its organization into specific databases catering to specific information type. Likewise, big data requires software for the analysis and processing of this data to draw conclusions from Insilco experiments, a list of databases and software packages is presented below in Table 2 and Table 3.

Table 2: List of public databases for biological and chemical data resources.

Database	Information contained	Website
Protein Data Bank	3D structures of macromolecules	https://www.rcsb.org/
CheMBL	Bioactive small molecules extracted from literature	https://www.ebi.ac.uk/chembl/
PubChem	Small molecules and their bioactivities	https://pubchem.ncbi.nlm.nih.gov/
ZINC	Commercially available chemical compounds	https://zinc15.docking.org/
ChemSpider	RSC published chemical structures	http://www.chemspider.com/
DrugBank	Pharmacological and biochemical information on drugs	http://www.drugbank.ca/
HMDB	Data on human metabolism and metabolites	http://www.hmdb.ca/
ChEBI	Small molecule entities	https://www.ebi.ac.uk/chebi/

TOXNET	Chemical structures and their toxicological data	http://toxnet.nlm.nih.gov/

Table 3: List of molecular docking software.

Software	Applications	Scoring Function
AutoDock 4.2	SBDD, Virtual Screening, Protein-protein docking and lead optimization	Empirical, Force Field
GLIDE	Prediction of binding mode, virtual screening	Empirical
Dock 6.6	Protein-protein docking, small molecule-protein binding mode prediction	Force Field
Flex 2.1.3	High-throughput screening, prediction of binding mode	Empirical
Fred 2013	Virtual screening, non-stochastic and systematic evaluation of protein-ligand poses	Empirical
Affinity	Model flexibility for the target molecule Binding mode prediction	Empirical
AutoDock Vina	Virtual screening Prediction of binding mode.	Empirical + knowledge based

2.9 Advances and Automation in FBDD

The ultimate goal for technology development is the automation of processes (Jämsä-Jounela, 2007; Manyika et al., 2017). In the identification of small molecules with biological activity, high throughput screening (HTS) which is a hit or lead discovery method, has been effective in dealing with millions of compounds. However, the hit output is low, thus resulting compounds face

difficulties in optimization to drug-like compounds (Gribbon & Andreas, 2005; Mayr & Bojanic, 2009). FBDD utilizes bioinformatics and chemoinformatic resources to tackle the problems and challenges still faced in HTS techniques. To process large amounts of data obtained from compounds and macromolecule databases, powerful computational techniques have been developed to systematically curate data in the drug design process. A general semi-automated pipeline for FBDD is as follows:

2.9.1 Fragment Library Design

Filter algorithms enable researchers to efficiently sort through vast amounts of data to identify compounds that have the potential to be optimized into lead compounds. These filters include physicochemical filters based on the rule of three (modified rule of five), chemical diversity analysis using molecular fingerprints, and synthetic accessibility using retrosynthetic combinatorial analysis procedure (RECAP) to fragment molecules based on chemical knowledge. Solubility prediction with QSAR models and fragment informatics, which includes fragment bioactivity data, are also utilized in the filtering process. These techniques streamline the identification of promising compounds for further development in drug discovery (Arabi, 2017; Kralj et al., 2023)

2.9.2 Fragment Screening with Molecular Docking

The library of fragments is further filtered using a different method, fragment docking, considering the filters applied at the first stage. The hit compounds are identified at this stage from fragment docking results and similar compounds which match the substructure or superstructure are added onto the library and examined for binding affinities (Bon et al., 2022; Y. Chen & Shoichet, 2009).

2.9.3 Optimization of Ligand Binding

At this stage, the hit fragments are built into novel ligands with *de novo* design; computational growth algorithms are employed in designing chemical structures that fit set constrains, an example being synthetic accessibility. The main advantage of these methods is the exploration of a broader chemical space. At each step in building the ligands from the fragments, the binding pose of the built ligand is explored through MDS and molecular docking, including the use of scoring functions to predict binding affinity (Dey & Caflisch, 2008; Hoffer et al., 2018; Spiegel & Durrant, 2020)

2.9.4 Lead to Candidate Optimization

SBDD techniques are employed to evaluate the novel ligands that have been built by molecular docking and scoring of the compounds, then MDS is employed to equilibrate the complex and obtain a more accurate binding affinity. Ligand-based drug design techniques utilize quantitative structure-activity relationship (QSAR) analysis and pharmacophore modeling. QSAR explores the relationship between known inhibitors' chemical fingerprints and their activity to score designed ligands (Pal et al., 2019). Pharmacophore modeling identifies essential interactions in protein-ligand complexes based on preexisting data on steric and electronic features triggering a biological response. These approaches aid in designing new ligands with improved activity and specificity for drug discovery (C. Acharya et al., 2011). All this leads to a library of lead compounds which may be further optimized as per the researcher’s objectives (Giordano et al., 2022).

GANDI (Genetic Algorithm-Based *de novo* Design of Inhibitors) is one software that joins SEED (solvation energy for exhaustive docking) pre-docked fragments using user supplied linkers, genetic algorithm is used for global optimization of the ligands (Dey & Caflisch, 2008; Majeux et al., 1999). The building blocks are obtained from mol-inspiration cheminformatics, fragments are examined for clashes and individuals with unfavorable interactions are removed from the fragment library. A Tabu search algorithm is used for efficient linking of the fragments, the force field energy is then optimized whilst also enforcing the 2D Tanimoto similarity to known inhibitors and 3D similarity to known binding modes and poses. The phenomena of scaffold hopping may also be promoted if desired (Dey & Caflisch, 2008).

LigBuilder does *de novo* design, ab initio, and heat-to-lead optimization through growing of the fragment or linking two or more fragments. It uses drug-likeness filters for ligands built and ensures the synthetic accessibility of compounds. In docking of compounds, the software uses a genetic algorithm and pharmacophore modelling for heat to lead optimization. Compounds are scored on their binding affinities, physicochemical properties, synthetic accessibility and ligand-protein lock-key match. LigBuilder has been successfully used in the discovery of potent cyclophilin A inhibitors (Ni et al., 2009; Yuan et al., 2011, 2020).

AutoGrow in an evolutionary algorithm for *ab initio, de novo* design and hit-to-lead optimization. The software employs principles of biological evolution, which are selection, crossover, and the third one being mutation. For each generation there is evolution, however the generation is first

populated by selected compounds fit from the previous generations. Variation in a new generation is introduced by crossing over in an operation of combining the characteristics of two compounds to create a third varied compound. This acts as the contributor to internal variation, whereas external validation will be brought about by random minor changes (mutation) to a member of the present population. AutoDock Vina is the default engine for conformational sampling of all ligands in every generation and returns binding affinities of the ligands, the ones with favorable affinities are selected for the next generation. The ligand grows by the random mixing of parent compounds advanced from a previous generation; this process follows click chemistry rules to ensure synthetic accessibility of the ligands. Physicochemical filters may also reject non-drug-like compounds in each generation (Durrant et al., 2009, 2013; Sheng & Zhang, 2013).

2.10 Conclusion of Literature Review

This chapter has explored the extensive work carried out by several researchers on heat shock proteins and their high potential as druggable targets. The work carried out by Lilburn and Mabate revealed the parasite's survival mechanism in response to physiological stress within the mammalian body being the overexpression of a group of molecular chaperones known as heat shock proteins (Lilburn et al., 2014; Mabate et al., 2018). These chaperones exist in six classes as laid out by Shonhai and central to the interplay of these classes, is *P. falciparum* heat shock protein 70-1 playing the central role of folding denatured proteins (Shonhai et al., 2007). This evidence underscores the importance of targeting heat shock protein 70s, specifically *P. falciparum* heat shock protein 70-1 as a druggable target for antimalarial drug design and development.

Also reviewed in this chapter was the classes of present antimalarial drugs on the market and their mode of action in destroying the parasite. Two classes of compounds, artemisinin based and aryl amino compounds have been shown to possess similar modes of action through the generation of reactive oxygen species targeting the heme pathway which is a source of food and nutrients. Another class of antimalarial drugs has been shown to target the folate pathway. With the emergence of resistance against the current class of antimalarial drugs, the artemisinin-based compounds through a mutation in the *Pf*Kelch13 there is a need for development of a new class of compounds that target a new pathway not yet explored before, further underscoring targeting heat shock proteins in new drug development. The rationale for choosing heat shock proteins as the new target for antimalarial drugs stems from two observed methods by which the malaria parasite

survives within the human host, that is the refolding of misfolded proteins through the heat shock protein functional interplay and the autophagy and macro autophagy aided by the *Pf*Kelch13 mutation and the unfolded protein response through the reactive oxidation stress complex. The initial approach aims to restore damaged proteins, while the latter aims to eliminate them. However, the autophagy method has been compromised by artemisinin resistance due to the *Pf*Kelch13 mutation. Therefore, the former approach remains unexplored as a viable pathway for maintaining cellular homeostasis. *Pf*Hsp70-1 has been selected from the heat shock proteins 70 family because of its distinctive glycine-glycine-methionine-proline (GGMP) tetrapeptide repeats, which are exclusive to this protein and not found in any other proteins in either the parasite or humans. Hence, with a novel target in mind, a novel category of 1,3,5-triazine derivatives has been selected as the foundation for designing new drugs. This new class of compounds, built upon a privileged scaffold in medicinal chemistry, has been investigated for its potential in targeting various species including *M. tuberculosis, T. brucei, E. coli, P. aeruginosa* and *P. falciparum*. These compounds, derivatives of 1,3,5-triuazine have exhibited remarkable activity as anti-bacterial, anti-inflammatory and anti-tumor agents. They have shown micromolar to sub-micromolar activity against the *P. falciparum* W2, D6, K1 and NF54 strains, underscoring their importance as a new class of compounds with higher efficacy. Evidence to this has been observed in a 10,000-fold activity profile against *P. falciparum* for a PS-15 metabolite WR99210, which has a 1,3,5-triazine moiety.

The search for effective antimalarial compounds has driven the adaptation of cutting-edge computational techniques in the field of drug design. SBDD uses the 3D structures of target proteins. In cases where the 3D structure is not available, comparative modeling (homology modeling) plays a vital role in predicting the 3D structures of these proteins from their homologous proteins (Krieger et al., 2003; X. Wang et al., 2018)

Molecular docking is essential in SBDD, allowing for the virtual screening of small molecules by precisely fitting tetrahydro-1,3,5-triazin-2-amine derivatives into the nucleotide binding domain of *Pf*Hsp70-1. MDS complements molecular docking by providing a dynamic view of protein-ligand interactions over time, offering insights into stability and flexibility of the complex. FBDD is a strategy used to identify small molecule binders for the protein's active site. These fragments are the foundation for building larger, more potent molecules, guiding the design of novel

tetrahydro-1,3,5-triazin-2-amine derivatives with improved pharmacological properties (Agamah et al., 2019).

Bioinformatics and chemoinformatic speed up drug discovery by managing and analyzing large biological and chemical datasets. Automation advancements have also improved the process by enabling high-throughput screening. QSAR modeling is another computational method that can predict the biological activity of drug candidates. GANDI, Ligbuilder, and Autogrow are examples of de novo design packages that serve as a creative force in drug design. These programs generate novel compounds with tailored chemical features and optimized interactions with target proteins, providing a unique opportunity to discover novel anti-malarial drugs (Hoffer et al., 2018; S. Sahu et al., 2020).

In summary, the multidimensional approach of structure-based design, coupled with advances in automation and the support of bioinformatics and chemoinformatic, holds tremendous promise for the discovery and synthesis of effective tetrahydro-1,3,5-triazin-2-amine derivatives as potential anti-malarial drugs. These computational methods accelerate the drug discovery process, potentially bringing us closer to the development of more efficient and safer drugs to combat malaria and save lives.

CHAPTER 3 METHODOLOGY

In this chapter, methods used to achieve the set-out objectives will be described in detail. The first objective focusing on the generation of thew 3D structure of *Plasmodium falciparum* heat shock protein 70-1 was carried out through comparative modelling with a software tool MODELLER 9v5 hosted on the interactive webserver PRIMO. This work was validated on an independent webserver ProSA hosted on the University of California, Los Angeles servers. The second objective focusing on the design of 1,3,5-triazine analogs was also carried out with the aid of a standalone package AutoGrow 4.2 from the Durrant lab of computational biology and drug discovery at the University of Pittsburgh. Chemical structures used in this objective were obtained from the ZINC15 database. The third objective focusing on synthesis, validated work done on the second objective through molecular dynamics simulation, and chemical compounds were then sourced from ChemSpace in Ukraine for synthesis.

3.1 Objective 1: Comparative Modeling of *Pf*Hsp70-1

The FASTA file for the primary sequence of *Pf*Hsp70-1 (PlasmoDB accession number: PF3D7_0818900) was downloaded from UniProt under the entry accession of Q8IB24_PLAF7 in FATSA format. An interactive web server PRIMO, was used for the steps of modeling the target protein (Hatherley et al., 2016). The template 3D structure was found by running the target sequence through BLAST to probe the National Center for Biotechnology Information (NCBI) database for PDB files. From the BLAST output, two PDB files were selected as templates upon which the target structure was modeled.

In this experimental approach, a template was initially selected from *P. falciparum*, specifically the nucleotide-binding domain (NBD) of heat shock protein 70-x (PDB ID: 6RZQ), which exhibited an 85% sequence identity with the target protein and covered 56% of its sequence. Additionally, a second template was chosen from the heat shock protein 70 of *E. coli K12* (PDB ID: 2KHO), which showed a sequence identity of 48% and a sequence coverage of 89%. These templates are available in the Protein Data Bank. To cover both the nucleotide-binding domain and the substrate-binding domain, we used 6RZQ and 2KHO, respectively. The alignment of the target and templates was performed using Clustal Omega 1.2.4, a multiple sequence alignment (MSA) algorithm. Finally, the 3D structure of the target protein was modeled using MODELLER 9v5 through PRIMO. Subsequently, the structures underwent refinement using a GalaxyWeb

service for refining protein structures known as GalaxyRefine2 (Ko et al., 2012). Evaluation and validation of the 3D homology structure was performed using web server tools, including ProSA, PROCHECK, and Verify3D, all accessed from the SAVESv6.0 - Structure Validation Server (ucla.edu) (Bowie et al., 1991; Colovos & Yeates, 1993; Laskowski et al., 1993, 1996; Lüthy et al., 1992; Wiederstein & Sippl, 2007).

3.2 Objective 2: *De Novo* Design

3.2.1 Defining the Target

The generated 3D structure of *Pf*Hsp70-1 was subsequently analyzed for binding sites and pockets suitable for compound binding. The focus was placed on the NBD, which is a known cavity for ATP binding, a natural substrate of the protein. This probe was carried out through blind docking ATP onto the protein with no search area defined. The binding site was identified through cluster docking of ADP. ADP was chosen in this case over ATP because of the protein-ligand stability associated with *Pf*Hsp70-1-ADP complex over *Pf*Hsp70-1-ATP complex, this is seen in the residence time of the ligands within the protein, ADP has a longer residence time over ATP. Compound clustering based on the distance matrix allowed for the identification of numerous binding sites that exist on the protein. Analysis of each cluster, the binding poses, and binding energies revealed the active site of the protein. Calculations were carried out with Blind Docking Server, available at: http://bio-hpc.eu/software/blind-docking-server/ (Sánchez-Linares et al., 2012). The active site dynamics were further analyzed with the TRAPP webserver, this tool can predict binding site flexibility and can detect transient pockets in proteins (Stank et al., 2017b).

3.2.2 Generating Ligand Fragment Library

The primary focus of this library was to generate a unique set of fragments containing the 1,3,5-triazine substructure. This systematic curation method was adapted from the work done by the Durrant Lab to ensure practicality and ease of synthesis, compounds were carefully curated from ZINC15 database and subjected to specific filter parameters available on the database (Spiegel & Durrant, 2020).

1. Molecular weight of fragments kept below 250 Da, adhering to Lipinski's rule of 5:
2. Selected compounds were commercially available:
3. Reactivity level of fragments considered, ranging from reactive to highly reactive:
4. LogP value of fragments limited to below 5.0:

In total, the library consisted of 99 seed compounds, each possessing a diverse array of functional groups suitable for use with the AutoGrow software. The diversity of the seed compounds was investigated using the Principal Component Analysis (PCA) on Mw, logP, HBA, HBD, RotB, and Topological Polar Surface Area (TPSA). Tanimoto distance matrix based on Morgan fingerprints was also employed with 50 epochs at a learning rate of 1.0 for output in 3D. These calculations were carried out on KNIME Analytics Platform v4.6.3 (Berthold et al., 2006) This diverse and well-curated set of fragments provides a promising foundation for the design and optimization of novel compounds with the desired anti-malarial activity.

3.2.3 Molecule Assembly

The step of fragment exploration with molecular docking was carried out with an autodock_vina_1_1_2 plugin within the AutoGrow 4.2 package. This step evaluated the fragments' abilities to bind to the protein and evaluate the fitness of each fragment in use for growing the compounds. The Assemble of the molecules based on fragment fitness was also carried out with several plugins within the AutoGrow package. Compound assembly was carried out by combining favorable characteristics of two fragments using a crossover operator and reacting two different groups with a mutation operator. The reactions carried out by the mutation operator, as curated by AutoGrow, are the Click Chemistry Reactions and the Robust Reaction.

The number of mutants and the number of cross-overs for the first generation were chosen to be 50 for each. The same number was maintained for compounds to seed the next generation, the number of mutations, and the number of crossovers to occur in each generation, this also included elite molecules to be advanced to the next generation. The diversity of molecules to seed the next generation and the diversity of seed to depreciate in each generation was maintained at 10 molecules, which also accounted for molecules promoted to the next generation via elitism characterized by the best-fitting compounds. The design was carried out for 5 generations and at each generation compounds were evaluated through molecular docking.

3.2.4 Design Refinement and Evaluation

Multiple compound variations were created through molecule assembly. Subsequently, three compounds underwent further modifications to enhance their synthetic accessibility. These refined compounds were then reassessed by estimating free binding energies using protein-ligand pair molecular dynamics simulations and molecular mechanics. The GROMACS package was used for

protein-ligand pair simulation in ensuring complex stability. The design refinement process involved finding and removing unwanted groups that bring about challenges in synthesis, like the imine functional group. Further on, scaffolds were confirmed for availability for purchasing and non-available scaffolds were substituted for readily available highly similar scaffolds.

Starting from the predicted docking pose of each ligand, the length of the simulation was set at 100ns. The complex was solvated with TIP3 water molecules in a dodecahedron periodic boundary condition with a 1.0 Å margin. After neutralization of the total charge in the system with the addition of sodium and chlorine ions, the CHARMM36 modeled system had its energy minimized with the steepest descent algorithm while restrained at the protein backbone. The system underwent a simulated annealing procedure, where both the ligand and protein were restrained, and the simulation was carried out under NPT conditions. The annealing process involved gradually increasing the system's temperature from 0.1 K to 300 K over 100 ps.

Following the annealing step, the system transitioned to the NVT ensemble while still retaining position restraints on the ligand and protein. The simulation was performed at a constant temperature of 300 K. During this 100 ps equilibration, temperature and pressure were controlled using a modified Berendsen thermostat at 300 K and a Berendsen barostat at 1 atmosphere. Particle Mesh Ewald was employed for probing long-range electrostatic potential with cut-off at 1.2 nm, the Verlet algorithm was used for probing short-range vdW interactions at 1.2 nm, LINCS algorithm was also applied for covalent bond constraints. For the production run with a length of 100 ns, the Parrinello-Rahman barostat was used at a pressure of 1 atmosphere.

A molecular mechanics with generalized Born and surface area solvation (MM-GBSA) free energy calculation was carried out with gmx_MMPBSA software package on the most stable time frame from MDS studies, the stable binding residues and overall free binding energies were estimated over a 10 ns window with low RMSD fluctuations (B. R. Miller et al., 2012; Valdés-Tresanco et al., 2021).

3.3 Objective 3: Synthesis of Compounds of Interest

3.3.1 Synthesis of BN_01 4-hydroxy-6-(morpholin-4-yl)-1,3,5-triazin-2-yl N-(3-fluoro-4-methylphenyl) carbamate.

The reaction, adapted from (Yoganathan & Miller, 2013), is a one-pot synthesis that couples an alcohol with an isocyanate under N-Methylimidazole (NMI) to create a carbamate compound as shown in Figure 15. For this reaction, the alcohol, and the isocyanate were obtained from ChemSpace and the coupling agent alongside the solvents and all other materials used were obtained from Sigma-Aldrich. 2mmol (1 equivalent) of the alcohol (BN_01_TF) (EN300-27101457) was dissolved in 4M NaOH and the organic compound was extracted with dichloromethane (DCM) and filtered through $MgSO_4$ to remove any water present. To the resultant organic extract, 0.2 equivalent of NMI was added to the solution and mixed with swirling. To the resultant mixture, 1 equivalent of the isocyanate (BN_01_AF) (CSC011215732) which was in liquid form was added, and the round-bottomed flask swirled to mix the contents. The mixture was allowed to stand undisturbed for 24 hours until crystals began to appear. An additional 48 hours were given for the crystals to develop. Subsequently, the formed crystals were filtered and washed with DCM, followed by thorough drying under a vacuum. The reaction was performed three times in total. The average yield obtained from the three runs was 77%.

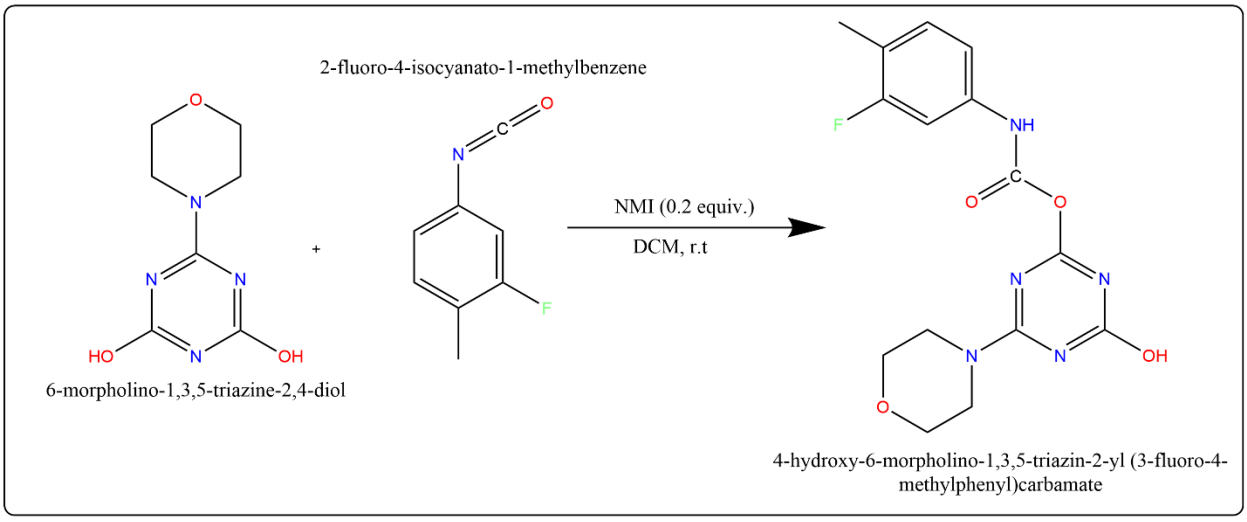


Figure 15: BN_01 Overall Reaction.

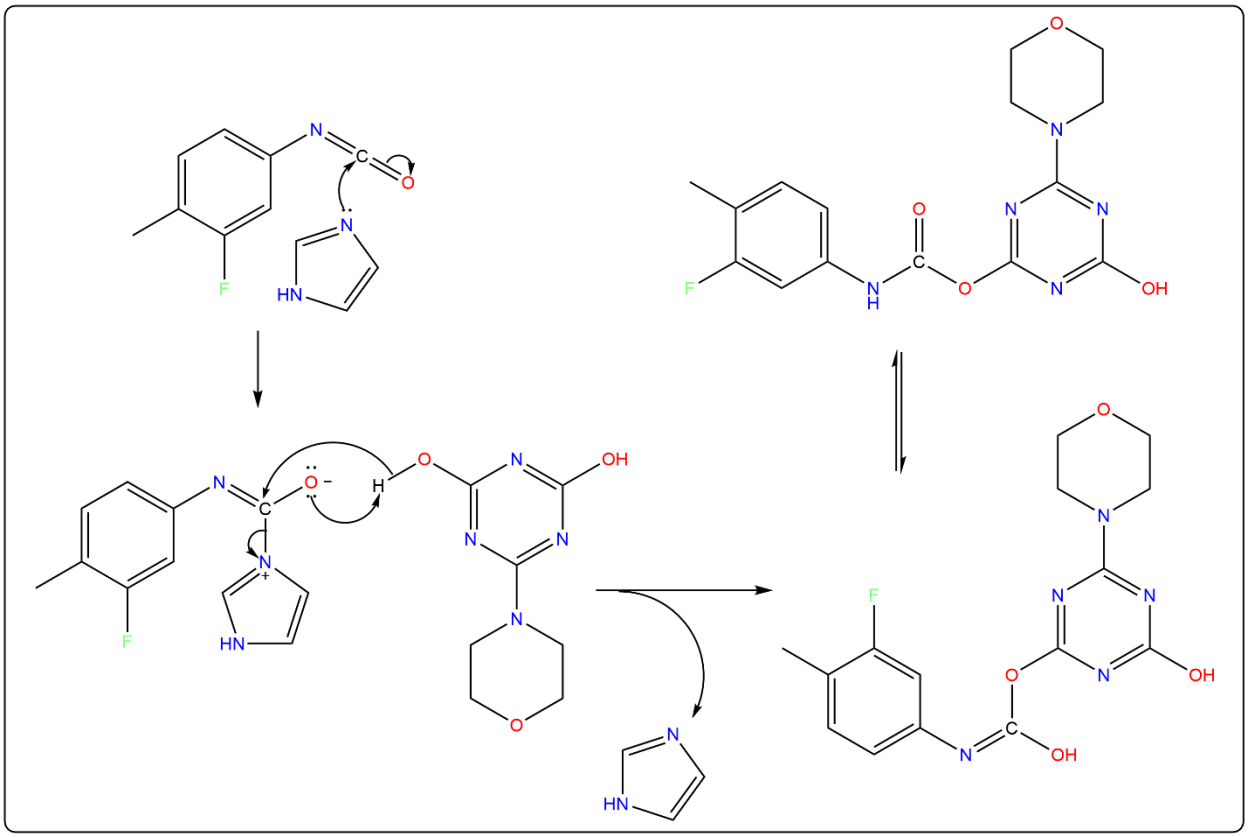


Figure 16: Proposed reaction mechanism for BN_01

3.3.2 Synthesis of BN_02 4-({7-amino-3-methyl-[1,2,4]triazolo[4,3-α][1,3,5]triazin-5-yl}carbamoyl)-1-hydroxypyridin-1-ium.

This was a one-pot synthesis for the coupling of an acid with an amine adapted from (El-Faham & Albericio, 2011; Jaradat, 2018) with N,N′-dicyclohexylcarbodiimide (DCC) and 1-Hydroxybenzotriazole hydrate (HOBt) as coupling agents, shown in Figure 17. The acid and amine were obtained from ChemSpace and all other reagents were obtained from Sigma Aldrich. 3.2mmol (1 equivalent) of the acid (BN_02_AF) (CSC000211280) was dissolved in 4.0M NaOH, and the organic compound was extracted with DCM. The resultant was filtered through $MgSO_4$ to remove any water present. 3 equivalents of triethylamine (TEA) were added to BN_02_AF and the reaction was cooled to 0^oC in an ice bath. The mixture was then treated with 2 equivalents of DCC and 2 equivalents of HOBt previously dissolved in DMF. Finally, 1.5 equivalents of the amine (BN_02_TF) (CSC009965197) was added to the mixture and the vessel was continuously stirred till a colloidal suspension started to form.

The vessel was left in the fume hood for 6 days, allowing the solvent to evaporate and facilitate crystal formation from a colloidal suspension. As a result, crystals were formed. Afterward, the crystals were filtered, washed with DCM, and dried under vacuum conditions. The entire reaction was repeated three times. The average yield obtained from the three repetitions was recorded as 94.8%.

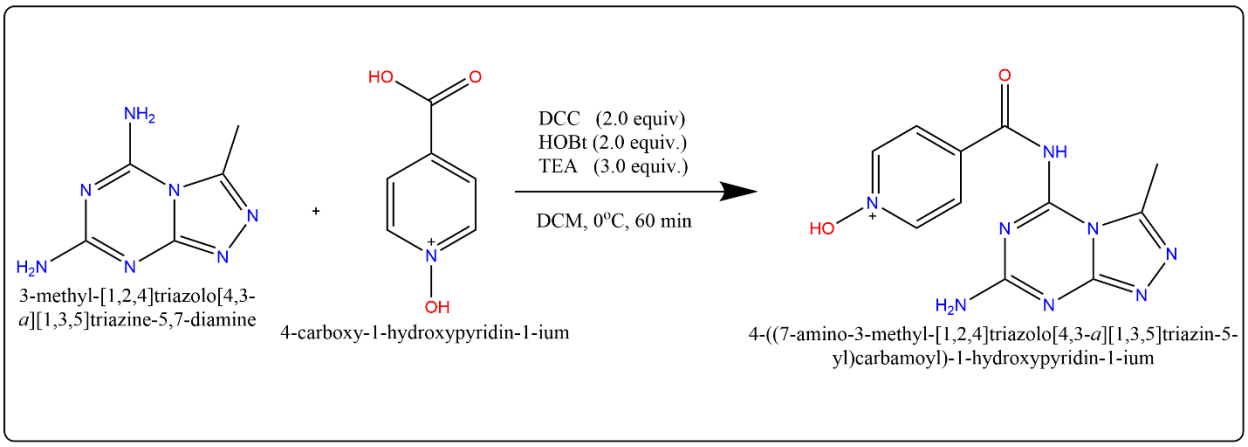


Figure 17: BN_02 Overall Reaction.

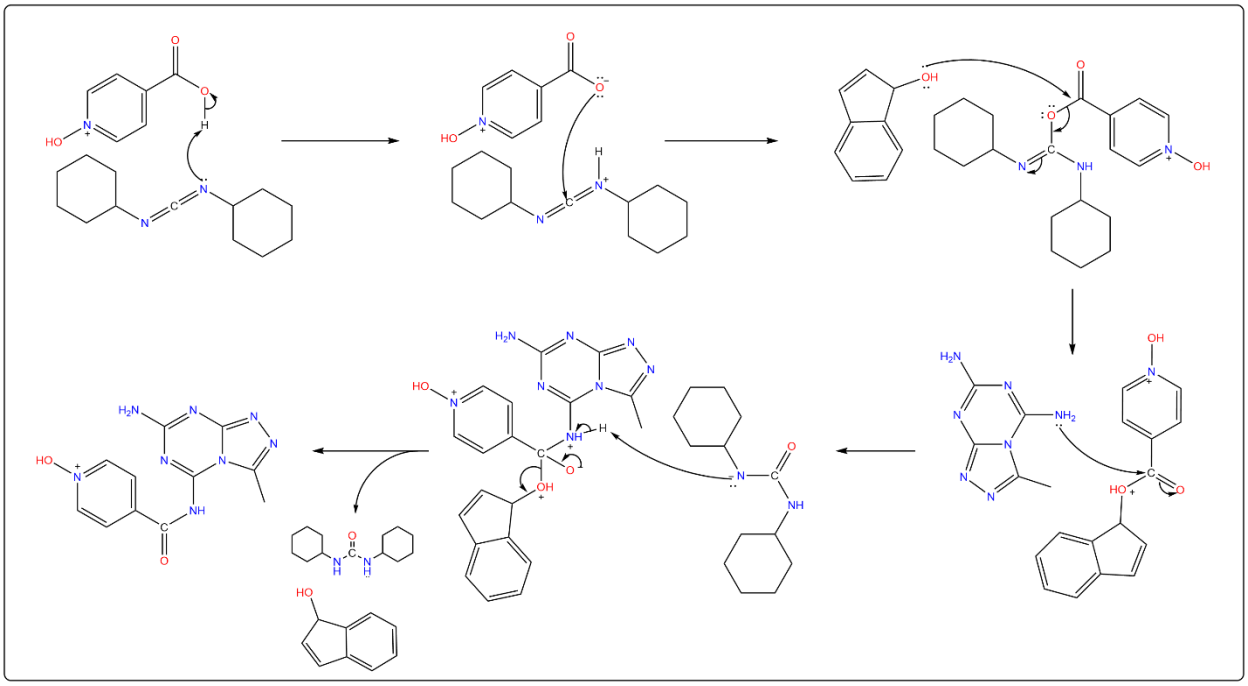


Figure 18: BN_02 Reaction Mechanism.

3.3.3 Synthesis of BN_03 1-hydroxy-4-({[4-hydroxy-6-(morpholin-4-yl)-1,3,5-triazin-2-yl]oxy}carbonyl)pyridin-1-ium.

The reaction was a one-pot fisher synthesis for coupling an acid with an alcohol, shown in Figure 19, catalyzed by sulfuric acid. 0.085mmol (1 equivalent) of the acid (BN_02_AF) was dissolved in 4.0M NaOH and the organic compound was extracted with DCM and filtered through $MgSO_4$ to remove any water present. The same treatment was done to 1 equivalent of the alcohol (BN_01_TF) which was later mixed with the acid solution. The solution was then treated with 2 drops of concentrated H_2SO_4 and boiled under reflux, an additional drop of acid was added after 30 minutes and the total reflux time was 2 hours. The solution was cooled in an ice bath and left in a fume hood. After 48 hours, crystals began to form, and within an additional 24 hours, they had fully grown. The solution was then filtered, and the resulting crystals were washed with DCM and dried under vacuum. The reaction was performed three times. The yield obtained from the three runs was an impressive 96%.

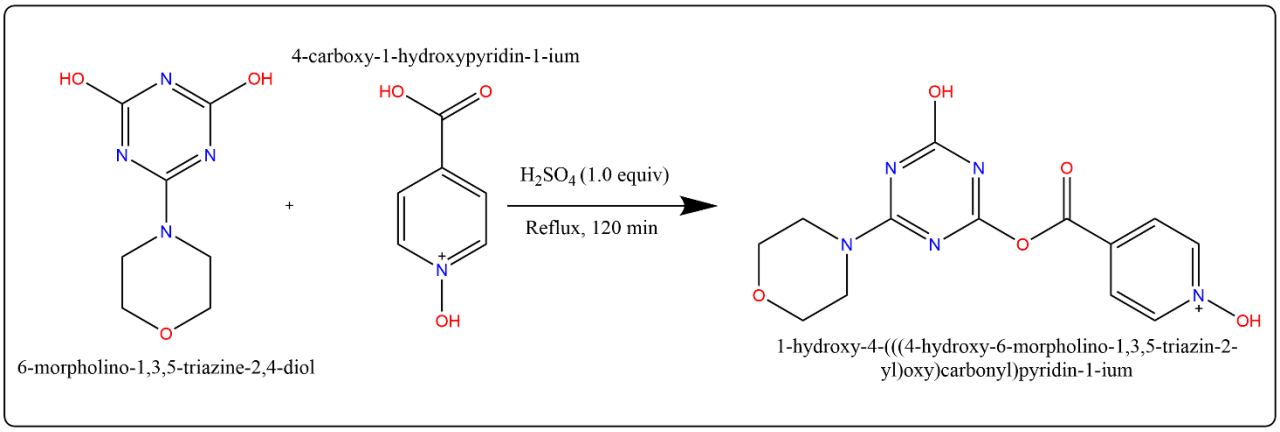


Figure 19: BN_03 overall reaction

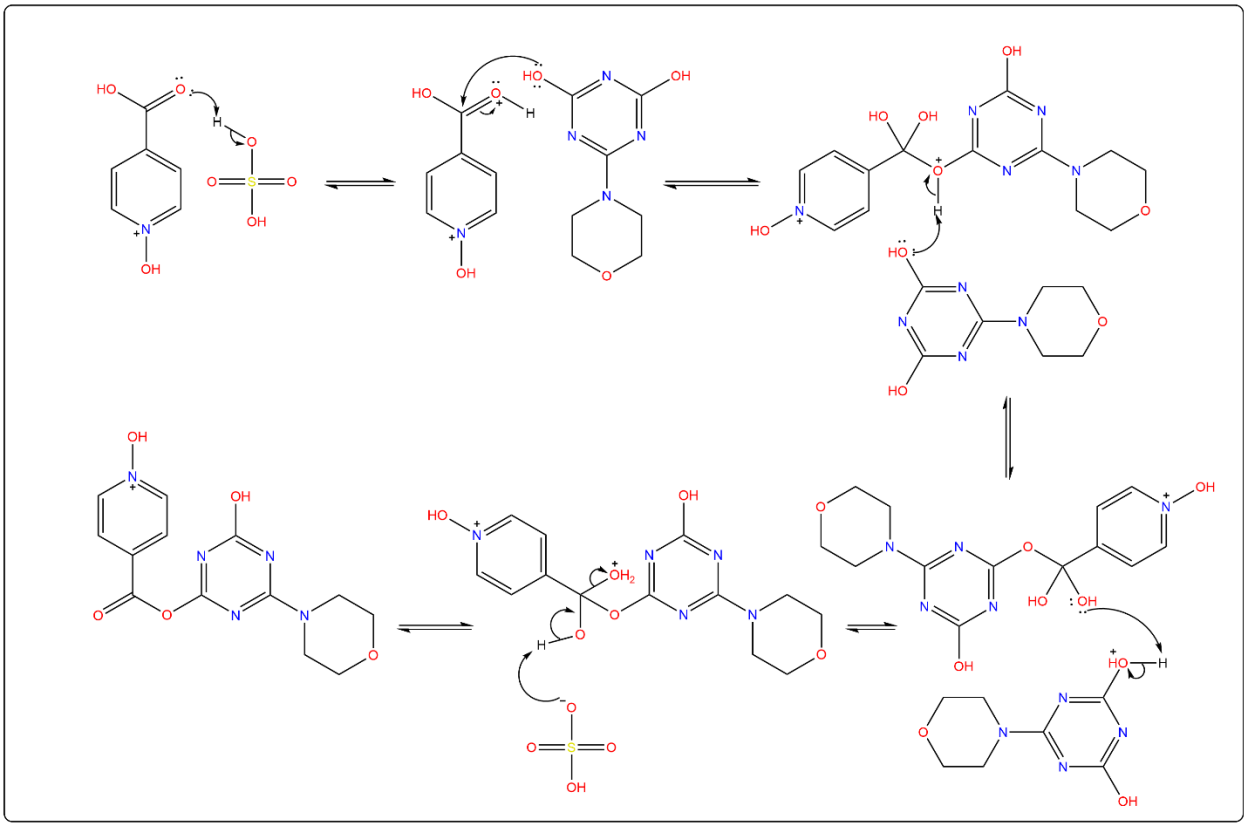


Figure 20: BN_03 Reaction Mechanism

CHAPTER 4 RESULTS

4.1 Comparative Modelling

The 3D structure of *Pf*Hsp70-1 was successfully generated, showcasing the nucleotide binding domain in green and the substrate binding domain in shades of maroon and purple, as illustrated in Figure 21.

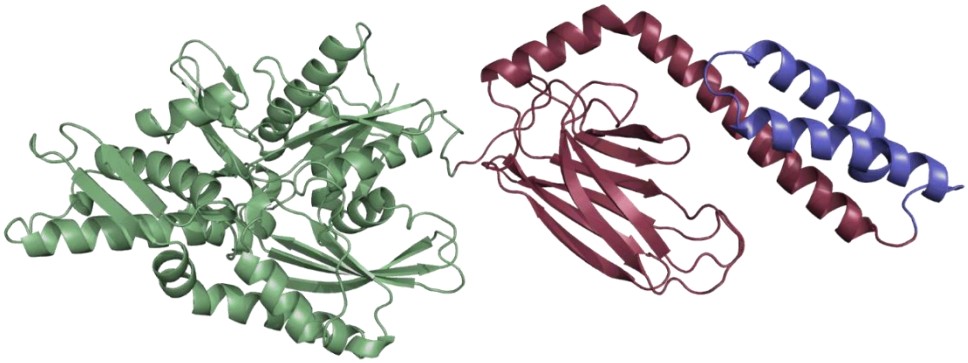


Figure 21: The 3D structure of PfHsp70-1

The structure was validated through several tools which will be discussed below. The structure was passed through ProSA (Protein Structure Analysis) to examine the quality of the model, a Discrete Optimized Protein Energy (DOPE) Z-score of -10.6 indicating a good score for the model. In Figure 22a below, the z-score for structures experimentally determined through NMR spectroscopy and X-Ray crystallography are plotted against the number of residues they contain. The generated structure was plotted on the graph as a black dot, falling in the region of structures experimentally determined through X-Ray crystallography. A typical -1 z-score and below indicates the likeliness of the generated structure to be near the native structures and in the overall model quality, comparing the z-scores of experimentally determined structures and the modelled structure estimates of the protein structure quality in comparison with all experimentally determine structures.

The other indicator of model quality from ProSA was the local model quality score (Figure 22b), which is a plot of the structure's residue energies. The average energy of a sliding window (which is 40 residues for default settings and 80 residues for bigger structures, as in this case), is plotted as a function of the residue at the center to get a smoother plot with minor fluctuations. Parts of a

model that are erroneous and most likely to be problematic possess positive energy values, and negative values correspond to parts with a good near-native resemblance. The average energy calculated over an 80 residue fragment $s_{i,i+79}$ is assigned to the central residue of the fragment at position i + 39.

All this information is further simplified by a 3D visualization (Figure 23) of the structure. In the visualization, red marks high-energy, potentially erroneous residues, while blue denotes low-energy residues (Wiederstein & Sippl, 2007). The overall model quality of *Pf*Hsp70-1 showed high degree of similarity to known X-ray elucidated structures as can be seen in Figure 22a, the local model quality had a few regions with positive energies (Figure 22b), which were mostly clustered in the substrate binding domain, as indicated in the 3D visualization of the protein. An overall good quality *Pf*Hsp70-1 3D structure was generated through comparative/homology modeling.

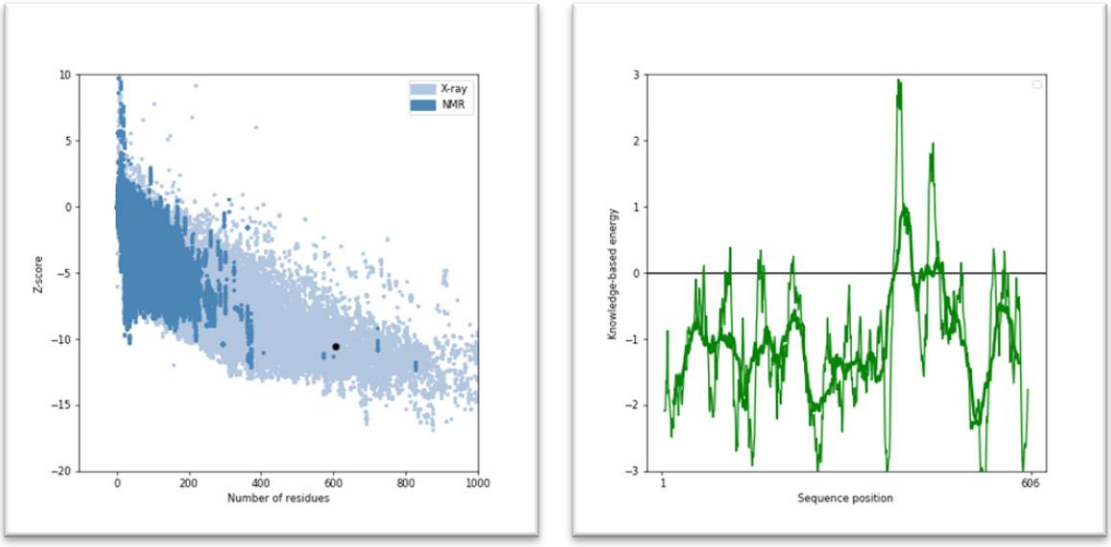


Figure 22: (a) Z-Score and (b) Local model quality score for the *Pf*Hsp70-1 structure.

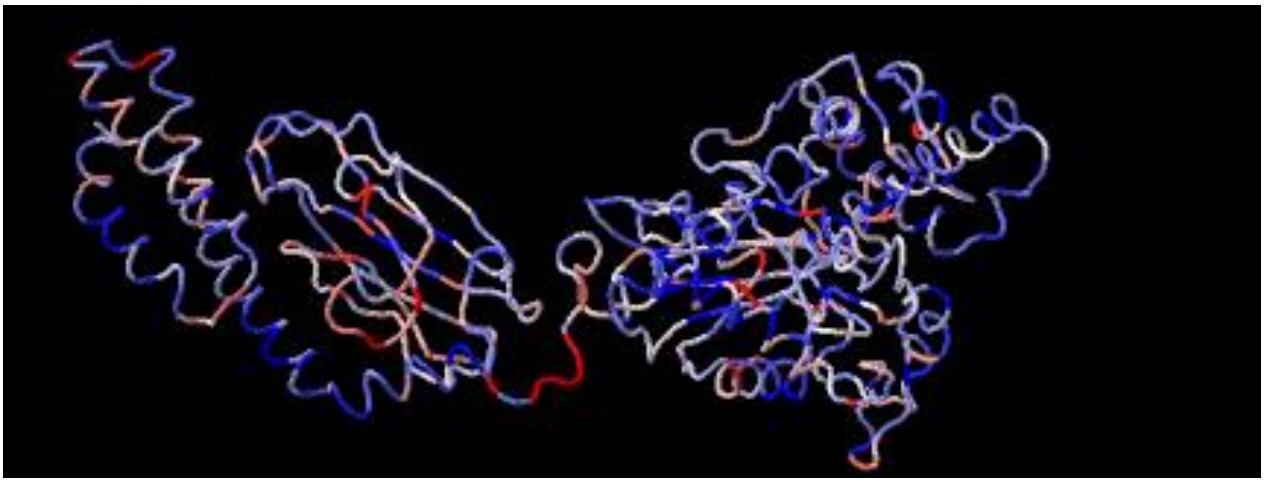


Figure 23: Model 3D visualization of local model quality score, red marks high-energy, potentially erroneous residues, while blue denotes low-energy residues.

Verify3D was then used to further evaluate the quality of the structure. The 3D model of the structure was compared to the amino acid sequence (1D) to determine compatibility. Structural classes of the protein based on their environment are assigned to sections of the protein, these include alpha and beta sheets, loops, and polar and non-polar regions, it is these regions where the 3D nature is compared to the 1D primary sequence for compatibility and the output is compared to good structures already known (Bowie et al., 1991; Lüthy et al., 1992).

This method uses the Define Secondary Structure of Proteins (DSSP) algorithm, which is a hydrogen bond estimation algorithm that assigns the secondary structure of a protein to its amino acid sequence considering the atomic coordinates and atomic resolution (Kabsch & Sander, 1983).

In Figure 24 below, a plot of the protein structural clashes to the 3D-1D score is shown, generally showing all residues to be in the positive range. 92.57% of the residues averaged a 3D-1D score >= 0.2, at least 80% for the residues averaging >= 0.2 is required for a good modelled protein.

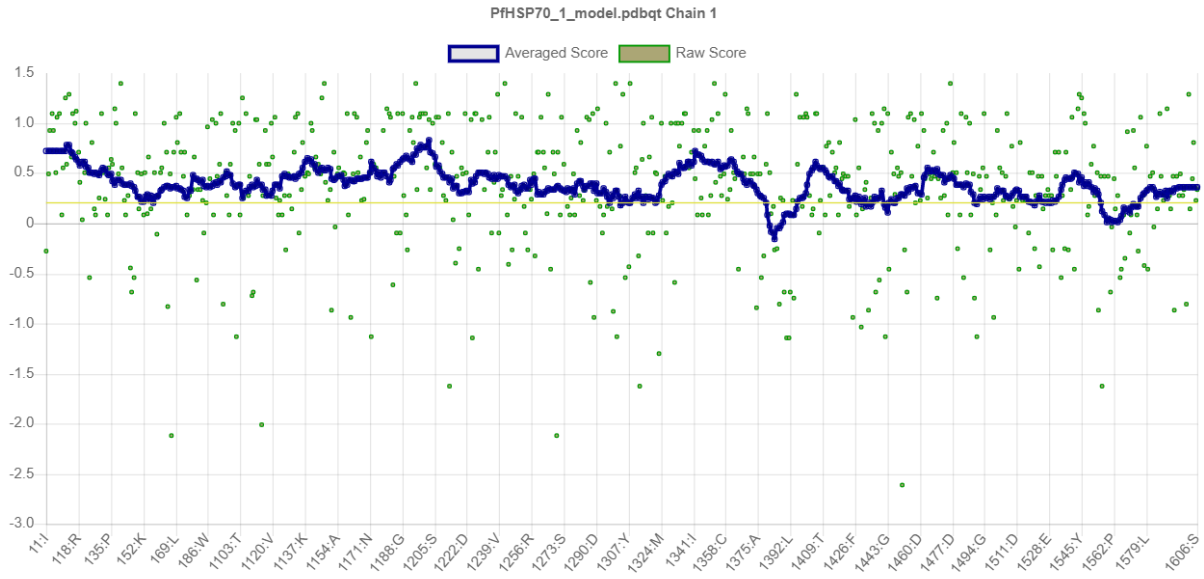


Figure 24: 3D-1D score of *Pf*Hsp70-1revealing the relationship of the 3D positioning of amino acids in a protein relative to expected outcomes from the 1D amino acid primary sequence.

Another important protein validation tool used was the Ramachandran plot from PROCHECK. This method evaluates the stereo-chemical quality of the structure under investigation. The requirement for a structure to be evaluated as a good model is to have over 90% of the residues in the most allowed region. This model had a 92.1% of the residues in the most favored region and 0.2 % of residues in the disallowed region. An additional 0.2% and 7.5% of the residues were in the generally allowed region and additionally allowed region, respectively. A total of 99.8% of the residues were in allowed regions, indicating a good spatial arrangement of atoms and residues giving way to an acceptable stereochemical quality of the structure.

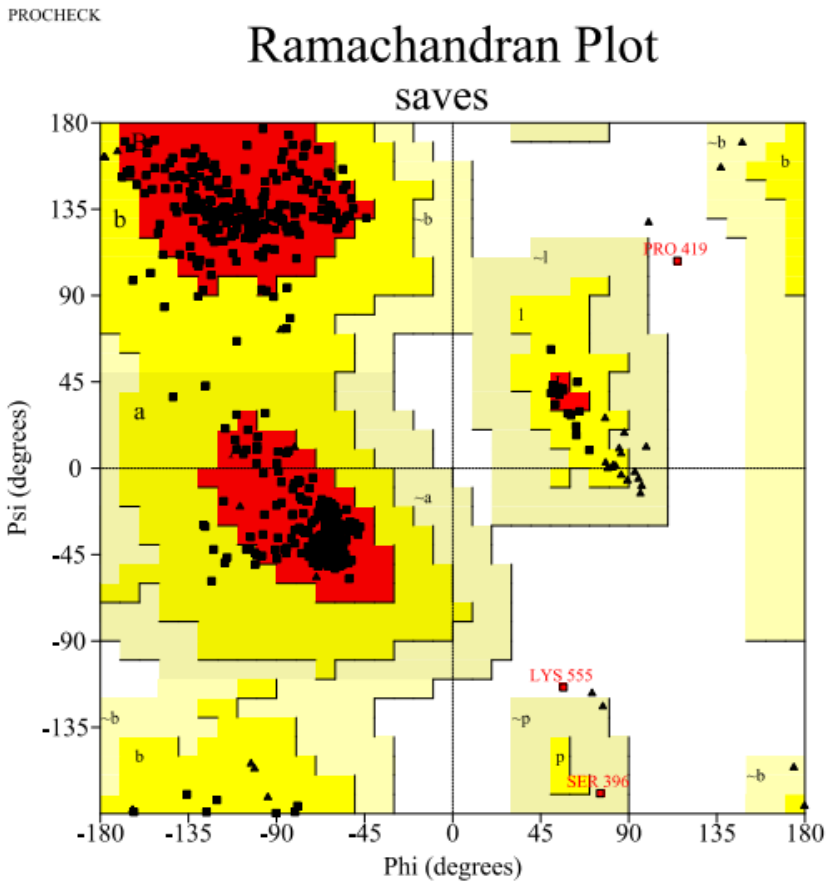


Figure 25: Ramachandran plot for *Pf*Hsp70-1 revealed Proline and Lysine as the only critical residues that did not fall within the accepted regions for their Psi-Phi angles.

92.1% of the residues were in the most favored region, 7.5% of the residues were in the additionally allowed region and 0.2% of the residues fell into the generously allowed regions and 0.2% of the residues fell in the disallowed region.

The data obtained from ProSA, Verify3D, and PROCHECK all indicate that the quality of the model was good, with a Z-score of -10.6 plotted within the X-Ray crystallography determined structures, with a 3D-1D score of above 0.2 for 92.57% of the residues, and 92.1% of residues in the most favored region.

The active site of the *Pf*Hsp70-1 was determined as the residues within 5 Å of ADP, it had a druggability probability of 0.74 based on PockDrug and a polar proportion of 0.42 on a scale of zero to one, one favoring high polarity. The aromatic proportion stood at 0.08 with Kyte hydrophobicity of +0.06. A visual representation of the pocket is shown in Figure 26 below where

the residue sidechains are color coded, red is for positively charged side chains, green is for negatively charged side chains, blue is for polar uncharged side chains, yellow is for hydrophobic side chains and orange is for special cases (cysteine, glycine, and proline).

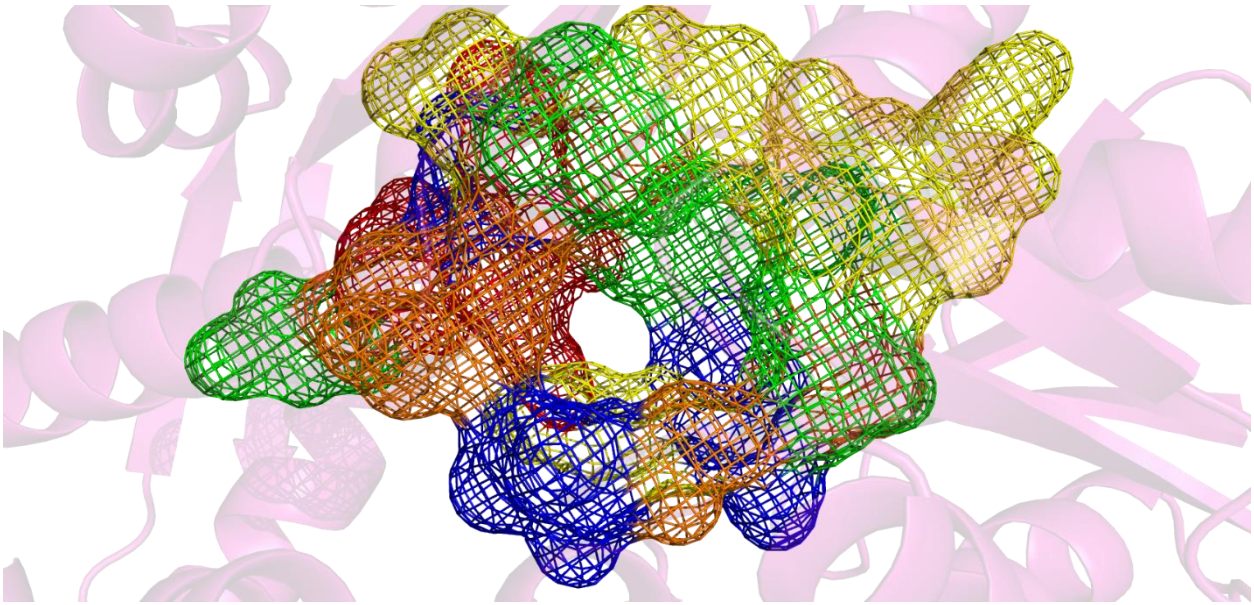


Figure 26: Active site for *Pf*Hsp70-1

The residues with positively charged side chains located within 5 Å of ADP were Lys269, Arg270 and 340. The residues with negative side chains were Asp6, 196, 231, 364 and Glu365. The residues with polar uncharged side chains were Thr9, 10, 33 and Ser12 and 338. The residues with hydrophobic side chains were Tyr11 and 369, Leu197, Ala366 and 368, Val335 and 367. The special cases were Gly8, 198, 199, 200, 227, 336 and 337, and Cys13.

4.2 *De Novo* Design

The *de novo* design went through for 5 generations, at the fifth-generation compounds converged at a global minimum, with the 2 most prevalent scaffolds (Figure 28) being returned for each generation. The diversity of compounds within the seed library can be observed in Figure 27 below. The seed compounds comprised of 99 unique variations of the 1,3,5-triazine-2-amine compound, of which all of them were highly similar to 1,3,5-triazine-2-amine.

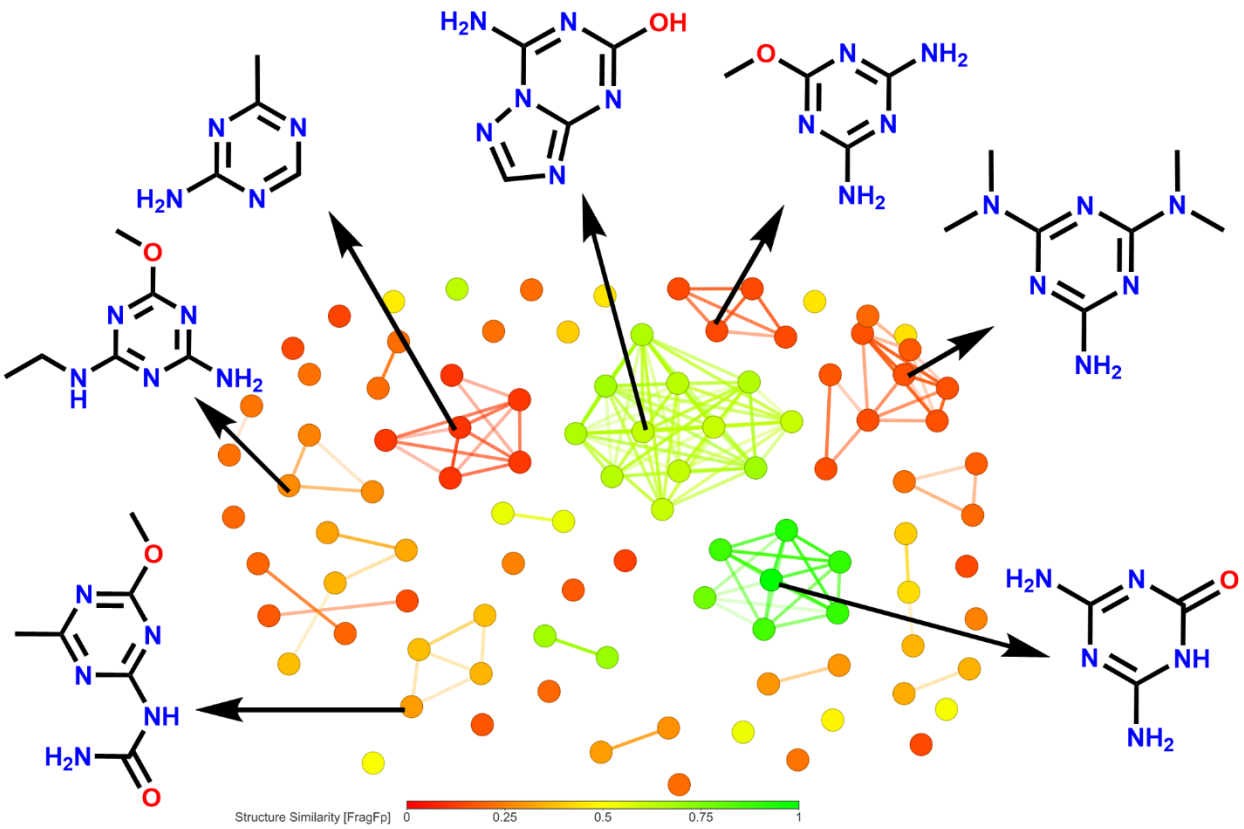


Figure 27: Activity cliffs scattering within the seed library

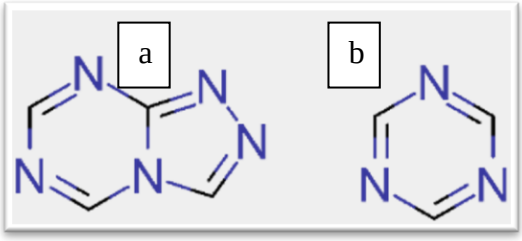


Figure 28: (a) [1,2,4]triazolo[4,3-a][1,3,5]triazine and (b) 1,3,5-triazine.

Two variations of the triazine core fragment present in the seed library were [1,2,4]triazolo[4,3-a][1,3,5]triazine and 1,3,5-triazine, the same core fragments returned in the library of designed compounds. The scattering of activity cliffs for the seed library compounds can be observed in Figure 27 with the largest group of dissimilar compounds containing the [1,2,4]triazolo[4,3-a][1,3,5]triazine core fragment. Further analysis of the nature of similarity for the seed library compounds involved the Tanimoto distance matrix carried out on Morgan fingerprints as descriptors of the compound (

Figure 29). The analysis of the compounds in 3D space revealed that they are scattered without forming specific clusters based on their similarity to the 1,3,5-triazine-2-amine moiety, which is the primary fragment for the *de novo* design process. The compounds in the seed library showed high similarity to 1,3,5-triazine-2-amine, with the least similar compound having a similarity index of 0.08 (where 0 represents high similarity and 1 represents high dissimilarity). The diversity in the seed library mainly arises from substitutions at positions 2, 4, and 6 on the triazine scaffold.

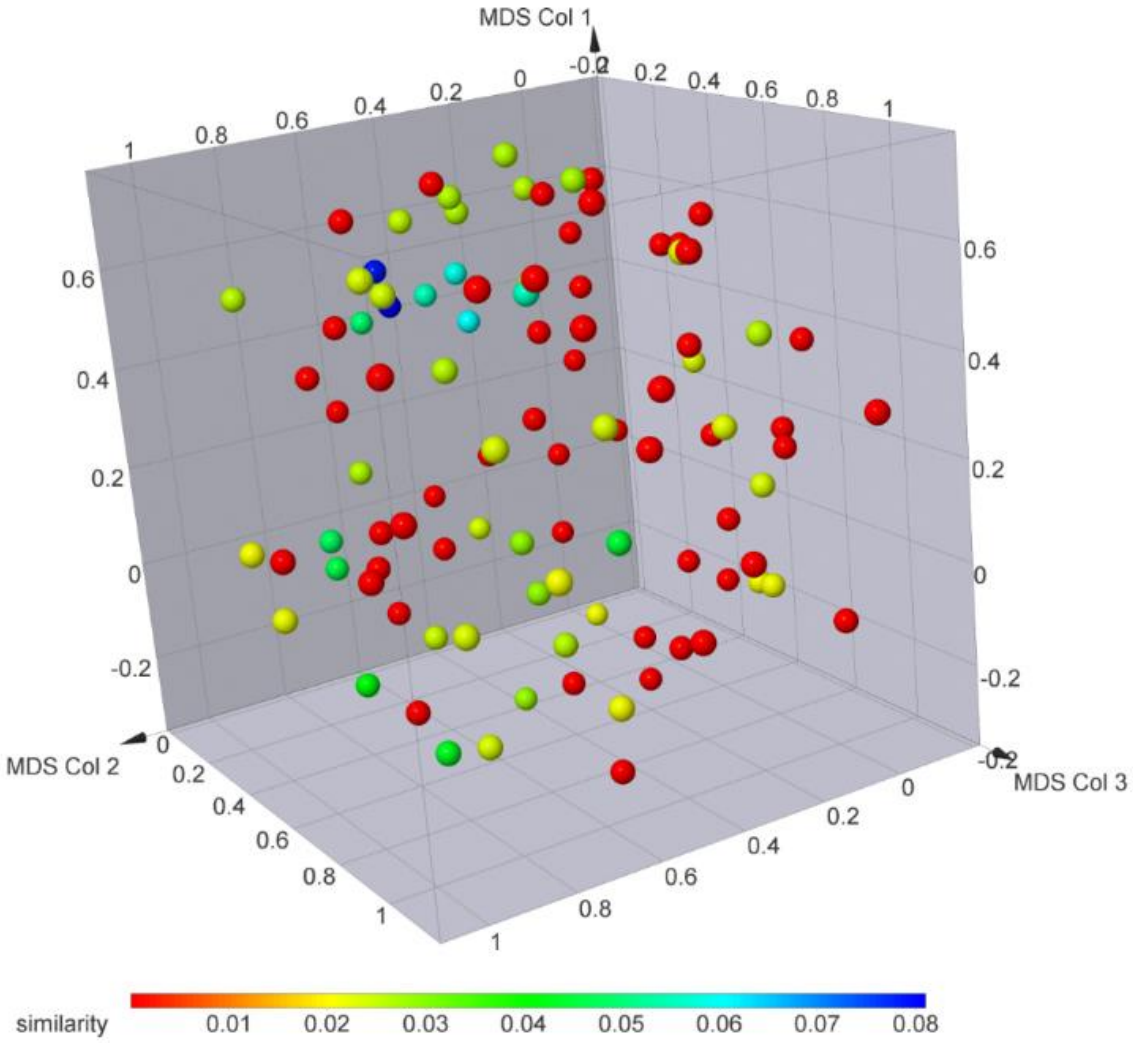


Figure 29: Library diversity calculated on the Tanimoto Distance Matrix for the Morgan fingerprints of compounds in the seed library all possessing close similarity to 1,3,5-triazine-2-amine.

Table 4: Average values for the physicochemical properties of the seed library

SlogP	TPSA Å^2	MW (g/mol)	RotB	HBD	HBA
-0.99	98.05	162.23	1	2	6

PCA (Figure 30) based on the seed library's physicochemical properties SlogP, TPSA, Mw, RotB, HBA and HBD confirmed the diversity of the compounds within the library. One observation of note for the library was the correlation of logP and the molecular weight of the compounds with an R^2 coefficient of 0.8361 (Figure 31).

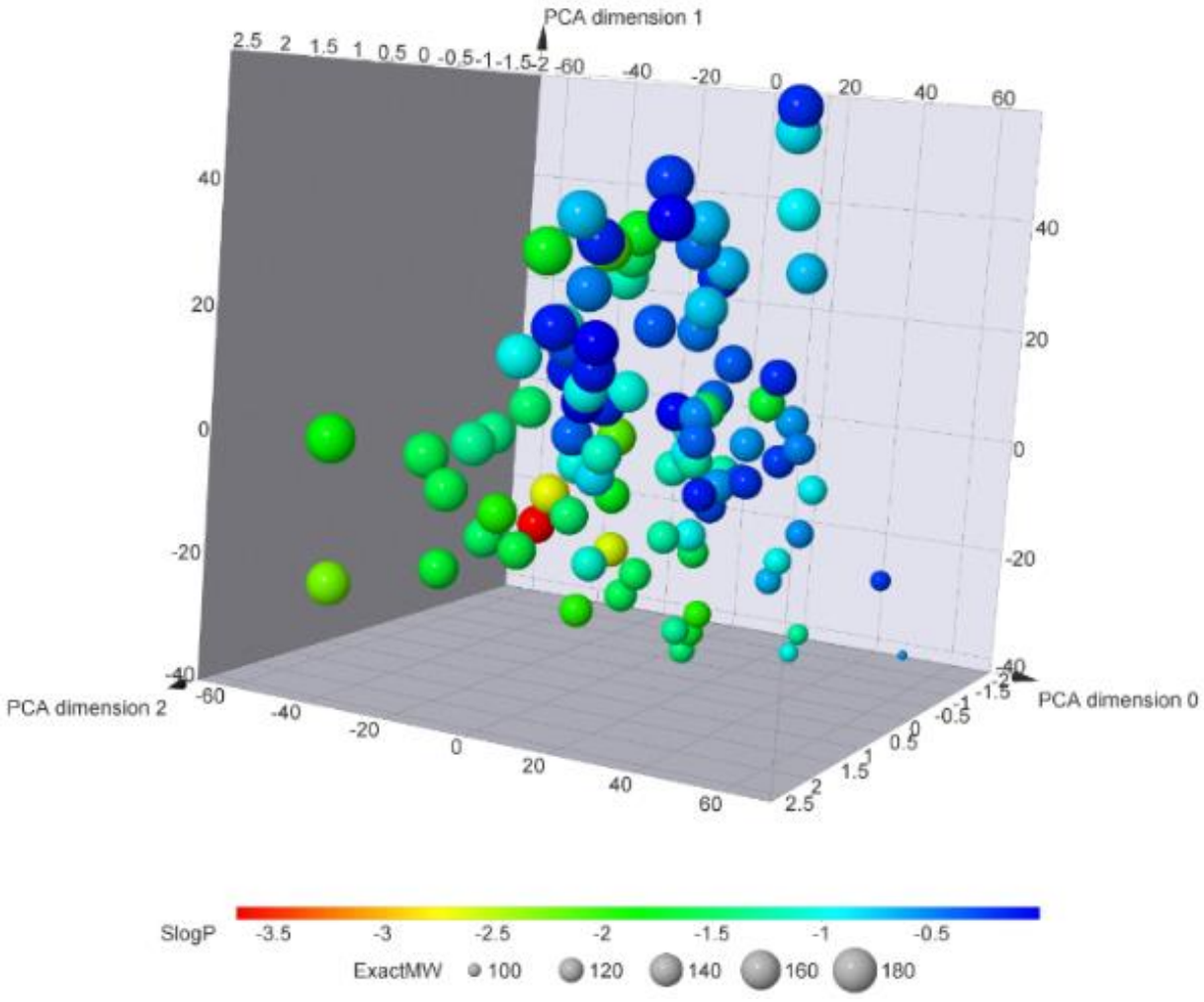


Figure 30: PCA analysis of the seed library

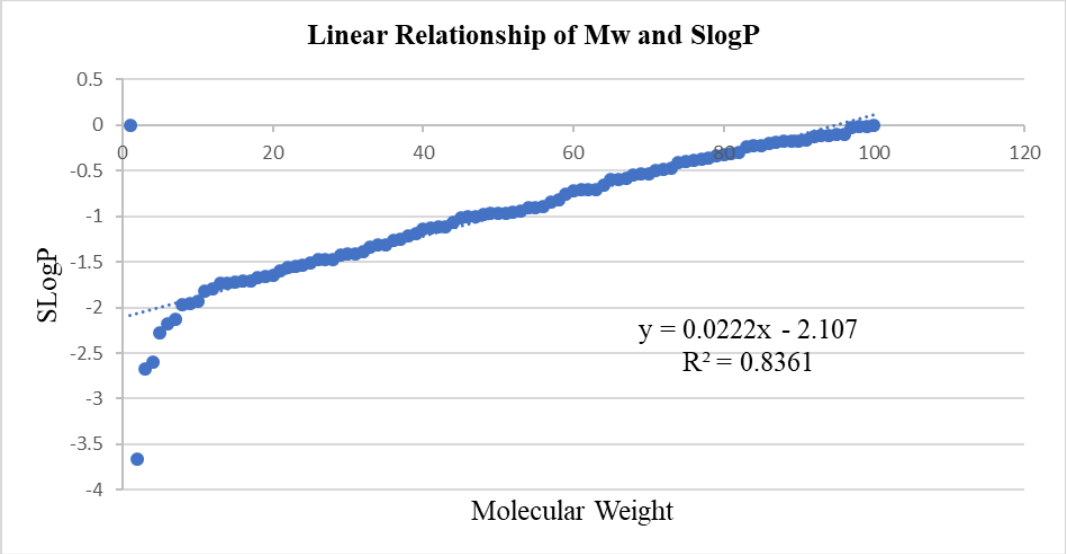


Figure 31: Linear relationship of Molecular weight and SlogP of seed library

Table 5: Loading values for the first three Principal Components

Property	PC1	PC2	PC3
SlogP	0.45	-0.28	0.26
TPSA	-0.56	0.20	-0.003
MW	0.22	-0.62	-0.21
RotB	-0.17	-0.59	0.57
HBD	-0.45	0.27	0.60
HBA	-0.45	-0.27	-0.45
Proportion / %	44.31	28.18	12.40

Principal component analysis utilized machine learning techniques to find patterns in multivariate high dimensional data, reduction of dimensions in this data and revealed variance in chemical space exploration. The plot analysis revealed that the first three principal components (PC) accounted for 84.89% of all variance, indicating that these principal components could be used to define the property space of this seed library.

PC1 explains the largest proportion of variance (44.31%). It is a linear combination of the original variables: SlogP, TPSA, MW, RotB, HBD, and HBA. SlogP has a positive loading (0.45), indicating a strong positive relationship with PC1. This suggests that as SlogP increases, PC1 also increases. TPSA has a negative loading (-0.56), indicating a negative relationship with PC1. As TPSA increases, PC1 decreases. MW (-0.22), RotB (-0.17), HBD (-0.45), and HBA (-0.45) also have negative loadings in PC1indicating a negative relationship. These loadings represent the weights of each variable in PC1. The magnitude of the loading indicates the strength of the contribution. In this case, SlogP has the highest loading magnitude, suggesting it has the strongest influence on PC1. The negative loadings for TPSA, MW, RotB, HBD, and HBA indicate that they have an inverse relationship with PC1.

PC2 explains 28.18% of the total variance. It is also a linear combination of the original variables. SlogP has a negative loading (-0.28) in PC2, indicating a negative relationship. As SlogP increases, PC2 decreases. MW (-0.62) and RotB (-0.59) have negative loadings in PC2, indicating a strong negative relationship. HBA also has a negative loading (-0.27) in PC2. TPSA (0.20) and HBD (0.27) have positive loadings in PC2. PC2 captures relationships between these properties, with MW and RotB having the most significant negative influence, and HBD having the most significant positive influence.

PC3 explains 12.40% of the total variance. It is again a linear combination of the original variables. Analyzing the PC3 column in the table, we can see the contributions of each property:SlogP (0.26), RotB (0.57), and HBD (0.60) have positive loadings in PC3, with HBD and RotB having strong positive correlations to PC3. TPSA (-0.003), MW (-0.21), HBA (-0.45) have negative loadings in PC3. PC3 captures additional relationships among the properties, with RotB, HBD, and SlogP having positive influences.

A loading plot for components one and two shows the description of physicochemical property space of the seed library. Principal component 1 describes the property space of the seed library on the lipophilicity of the compounds shown by a huge positive load of SlogP. The third principal component describes the chemical space of the compounds based on the largely positive loaded components, hydrogen bond donors and rotatable bond count. This has been observed and demonstrated in Figure 27 for activity cliffs, whereby the increase in the rotatable bond count is

expected to correspond with an increase in oxygen and nitrogen containing subgroup onto the already existing substituents at position 2,4 and 6 for the 1,3,5-triazine core moiety.

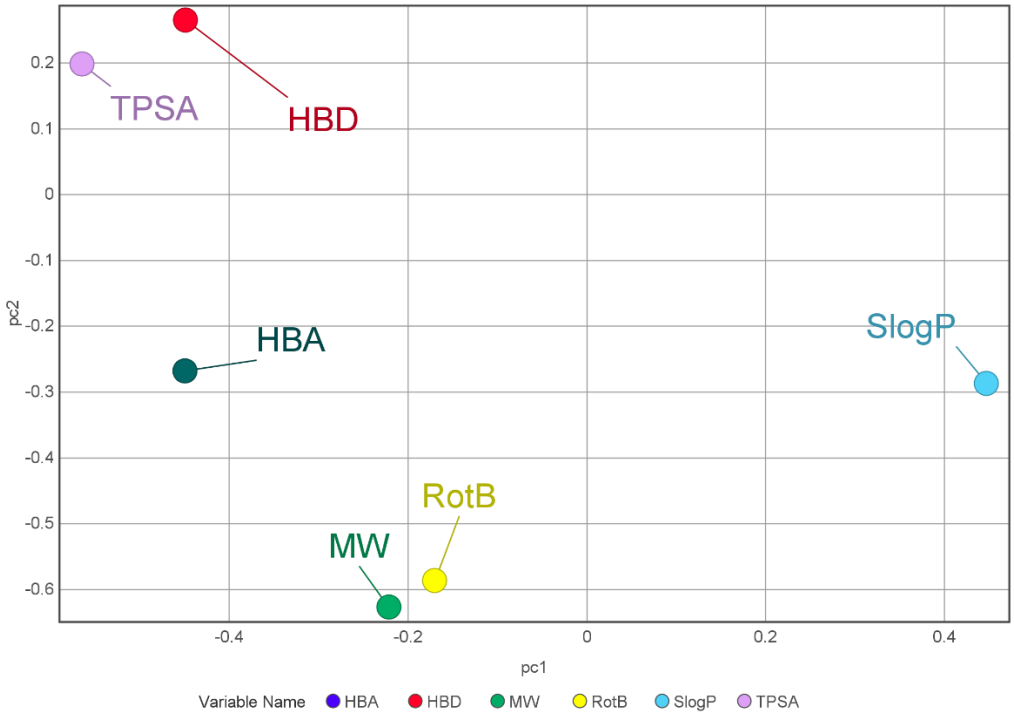


Figure 32: PC1 and PC2 loading plot for the physicochemical properties of the seed library.

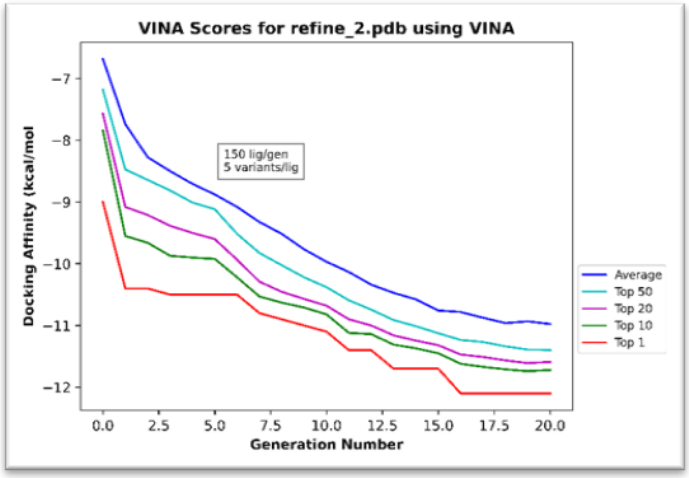


Figure 33: Docking affinities per generation

Seventy-one compounds were designed through the *de novo* design method which presented a highly diverse library as calculated on Morgan fingerprints identifiers with the Tanimoto distance matrix and PCA based on the physicochemical properties logP, TPSA, Mw, RotB, HBA and HBD

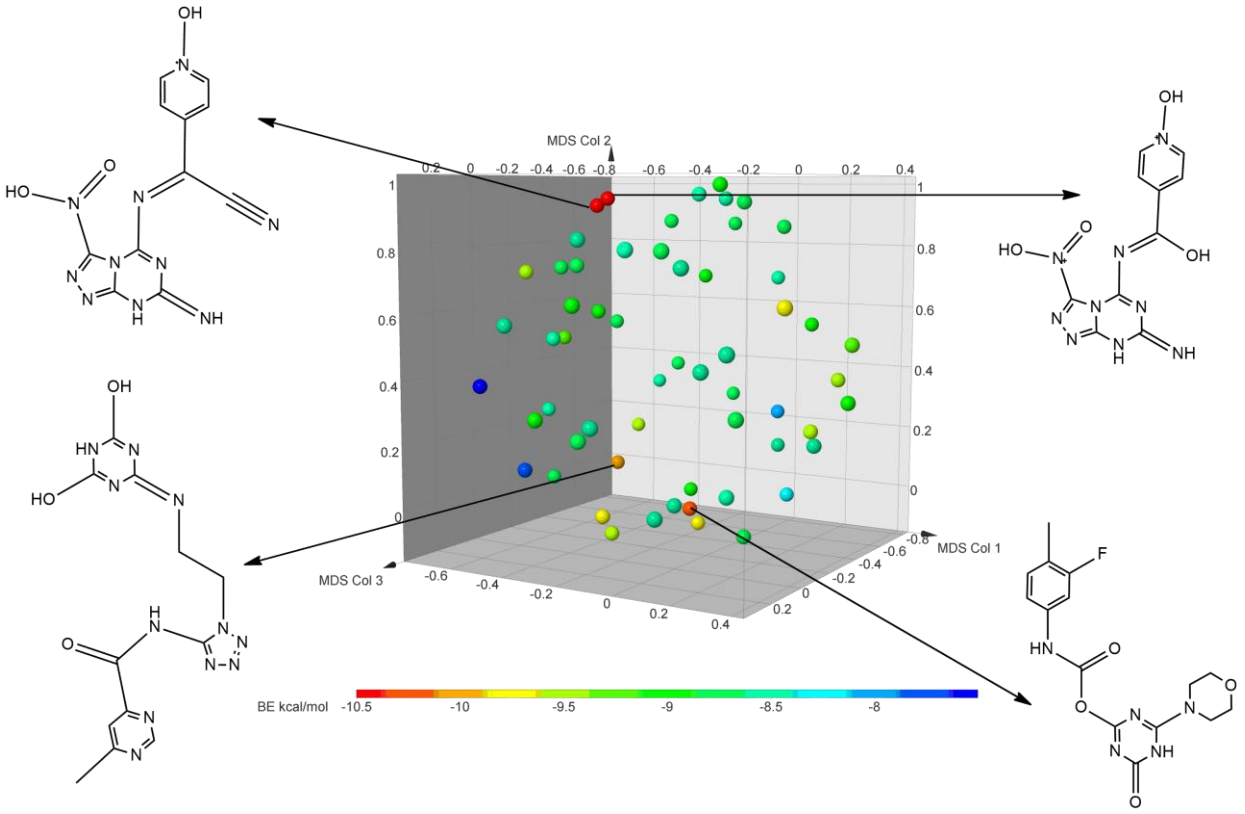
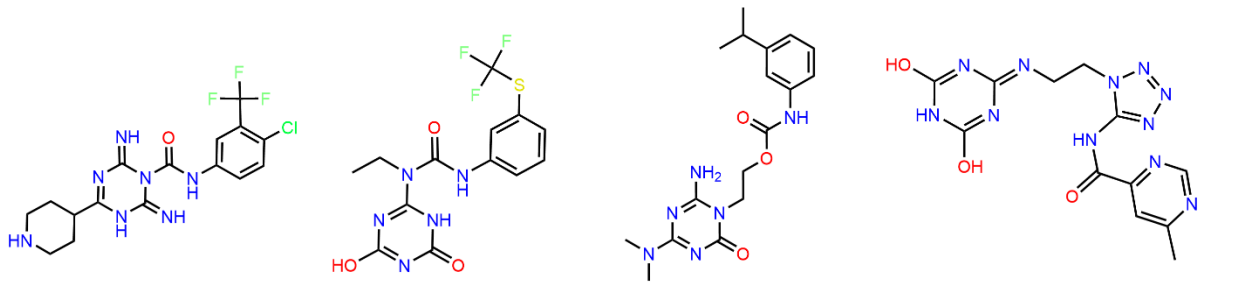


Figure 34: Chemical space visualization of the designed compounds using the Tanimoto Distance Matrix on Morgan Fingerprints

Table 6: Average values for the physicochemical properties of the designed compounds.

SlogP	TPSA Å^2	MW (g/mol)	RotB	HBD	HBA
-0.97	143.50	256.56	2	3	7

The lead-like compounds should meet specific criteria, including a molecular weight between 250 g/mol and 350 g/mol, less than 7 rotatable bonds, and an octanol/water partition coefficient lower than 3.5. The library successfully adhered to lead-like requirements, as indicated by the average physicochemical attributes of the generated compounds shown in Table 6. Careful analysis for these two parameters revealed only 4 compounds in the database had a mass above 350 g/mol, ranging from 359 to 415g/mol, giving them sufficient room for optimization withing the Rule of Five criterion.



Gen_1_Mutant_31_461033__3
415 g/mol

Gen_1_Mutant_31_975975__3
375 g/mol

Gen_1_Mutant_35_728042__2
360 g/mol

Gen_1_Mutant_75_26962__3
359 g/mol

Figure 35: Four compounds violating the lead-like criteria upper bound for molecular weight.

Molecular weight ranged from 187 g/mol to 415 g/mol, total polar surface area ranged from 90 to 193 $Å^{2,}$ SlogP ranged from -3.3 to 2.5, rotatable bond count ranged from 0 to 6, hydrogen bond acceptors ranged from 4 to 12 and lastly hydrogen bond donors ranged from 1 to 8. This information is represented in a box and whisker plot in Figure 36 and Figure 37.

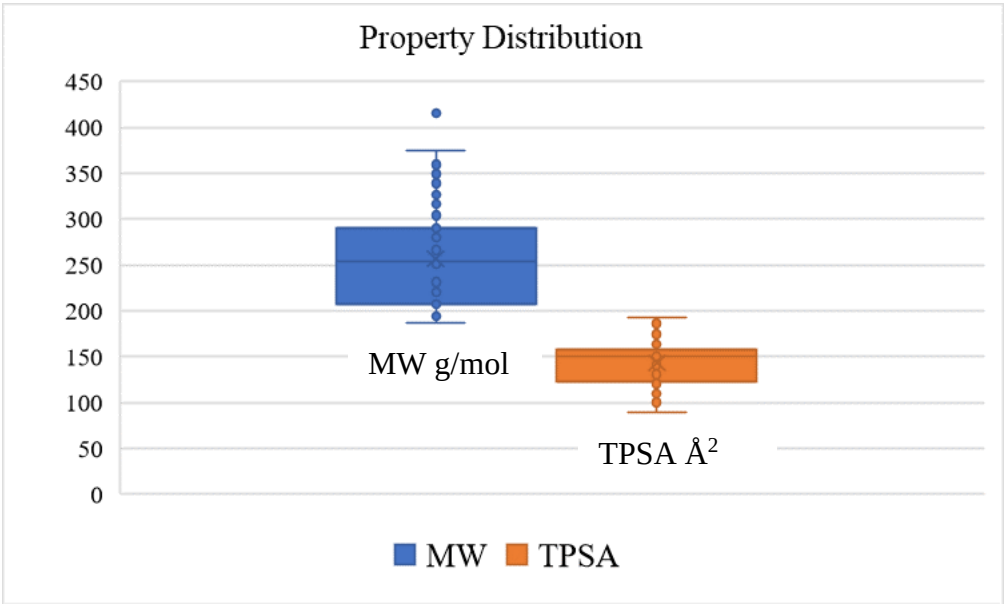


Figure 36: Distribution of Molecular Weight and Total Polar Surface Area for The Generated Compounds.

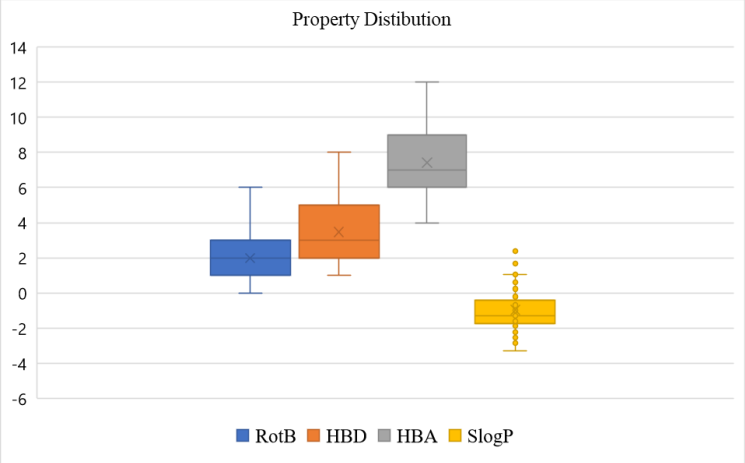


Figure 37: Property Distribution for Rotatable Bonds, Hydrogen Bond Donors and Acceptors; And Enhanced Atomic Logarithm of The Octanol/Water Partition Coefficient for the Generated Compounds.

Principal component analysis on the designed compounds was carried out for the six physicochemical properties as shown in Table 7

Table 7: Loading values for the first three Principal Components

Property	PC1	PC2	PC3
SlogP	0.49	-0.37	0.04
TPSA	0.23	0.64	-0.23
MW	0.54	0.09	-0.31
RotB	0.44	0.14	-0.53
HBD	-0.37	-0.11	-0.69
HBA	0.27	0.65	0.28
Proportion / %	40.71	25.50	20.17

PC1 explains the largest proportion of variance (40.71%) in the dataset. It is influenced by all the properties: SlogP, TPSA, MW, RotB, HBD, and HBA. SlogP has a positive loading (0.49) in PC1, indicating a positive relationship. Compounds with higher SlogP values contribute more to PC1 scores. TPSA (0.23), MW (0.54), RotB (0.44), and HBA (0.27) also have positive loading in PC1 indicating a positive relationship and their contribution to PC1. HBD has a negative loading (-0.37) in PC1, indicating an inverse relationship. Compounds with lower HBD values contribute more to PC1 scores. The loadings indicate the relative importance of each property in capturing the structural variations in PC1. In this case, SlogP, MW, RotB, and HBA have relatively higher loadings, suggesting they play a significant role in driving the structural variations captured by PC1.

PC2 explains 25.50% of the total variance in the dataset. It is influenced by SlogP, TPSA, MW, RotB, HBD, and HBA. TPSA (0.64) and HBA (0.65) have positive loading in PC2 indicating a strong positive relationship. Compounds with higher TPSA and HBA values contribute more to the PC2 scores. MW (0.09) and RotB (0.14) also have positive loading in PC2. HBD (-0.11) and SlogP (-0.37) have negative loading in PC2, indicating an inverse relationship. Compounds with lower SlogP values contribute more to PC2 scores. The loadings reveal that TPSA and HBA have relatively higher contributions to PC2. This suggests that PC2 captures variations related to the balance between TPSA and the presence of HBA, with additional influences from the other properties.

PC3 explains 20.17% of the total variance. It is influenced by SlogP, TPSA, MW, RotB, HBD, and HBA. SlogP (0.04) and HBA (0.28) have positive loading in PC3 indicating a positive relationship. TPSA (-0.23), MW (-0.31), RotB (-0.53), HBD (-0.69) have negative loading in PC3, indicating an inverse relationship. The loadings suggest that PC3 captures variations related to the balance between TPSA, MW, RotB, HBD, and HBA. The loadings suggest an inverse relationship for HBA and HBD, RotB, TPSA, and MW.

The 3D chemical space mapping of the compounds is presented in Figure 38 below to illustrate the scattering of designed compounds along the 1,3,5-triazine chemical space.

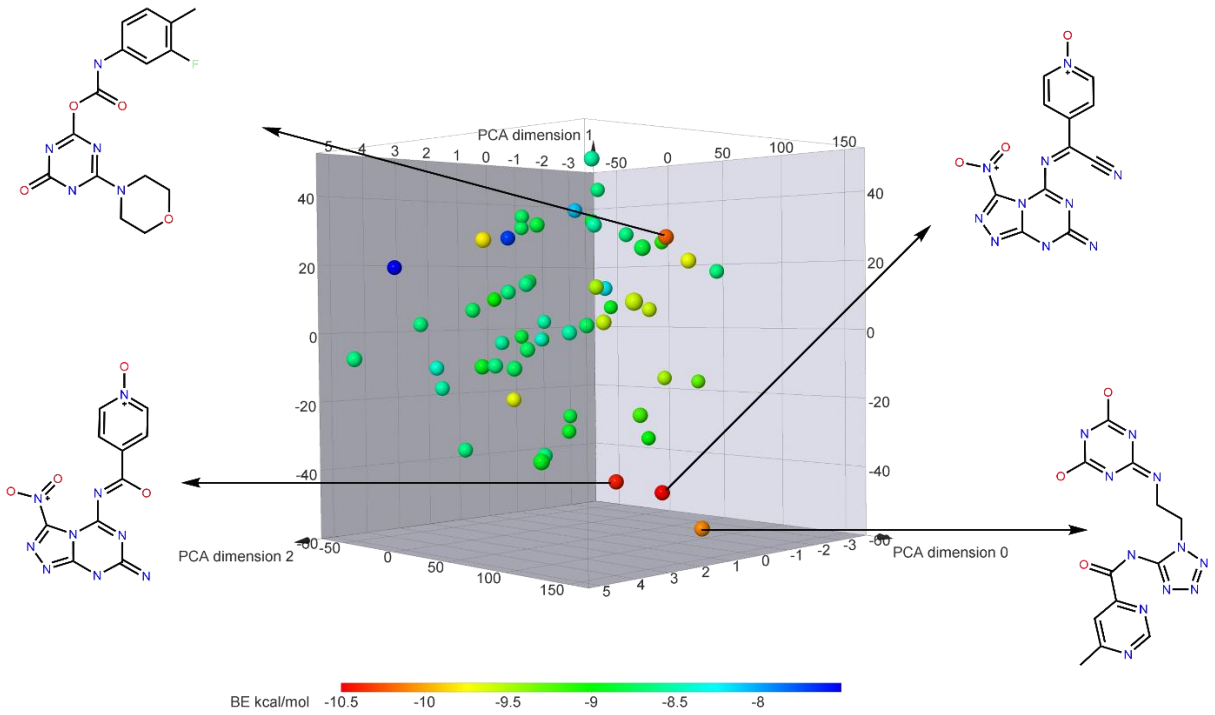


Figure 38: 3D Principal Component Analysis of designed compounds based on the six physicochemical properties Mw, SlogP, TPSA, RotB, HBA and HBD.

Comparing the PCA for the seed library and the generated compounds; In both cases, PC1 captures the largest proportion of variance. However, the loadings of the properties differ between the seed compounds and the generated compounds. For example, in the seed compounds, SlogP, TPSA, MW, RotB, HBD, and HBA all contribute to PC1. In the generated compounds, the contribution of SlogP is still significant, but TPSA, MW, and RotB show different loadings.

The loadings and proportions of variance explained by PC2 and PC3 also differ between the seed compounds and the generated compounds. These differences indicate that the structural variations between the two sets of compounds might be distinct. This is an expected occurrence since the compounds have undergone mutations and crossovers. The loadings of the molecular properties in each principal component shown in Figure 39, provided insight into the relationships between these properties and their contributions to the structural variations captured by the PCA. In both the seed compounds and the generated compounds, SlogP consistently showed a significant influence on the first principal component (PC1). Higher SlogP values contributed positively to PC1 scores, indicating that lipophilicity played a key role in the structural variations of the compounds. TPSA had different loadings in PC1 and PC2 between the seed compounds and the

generated compounds. This suggested that the balance between lipophilicity (SlogP) and polar surface area (TPSA) might have changed during the de novo design process. These properties may be improved in the synthesis of derivatives through the addition of alkyl chains capped off by a hydroxyl or amine group as this would greatly improve TPSA and lipophilicity of the compounds Both MW and RotB show varying loadings in different principal components across the seed compounds and the generated compounds. These properties contribute to capturing different aspects of the structural variations. The loadings of HBD and HBA vary across the principal components in both sets of compounds. This indicates that the presence of hydrogen bond donors and acceptors contributes to different structural variations captured by the PCA.

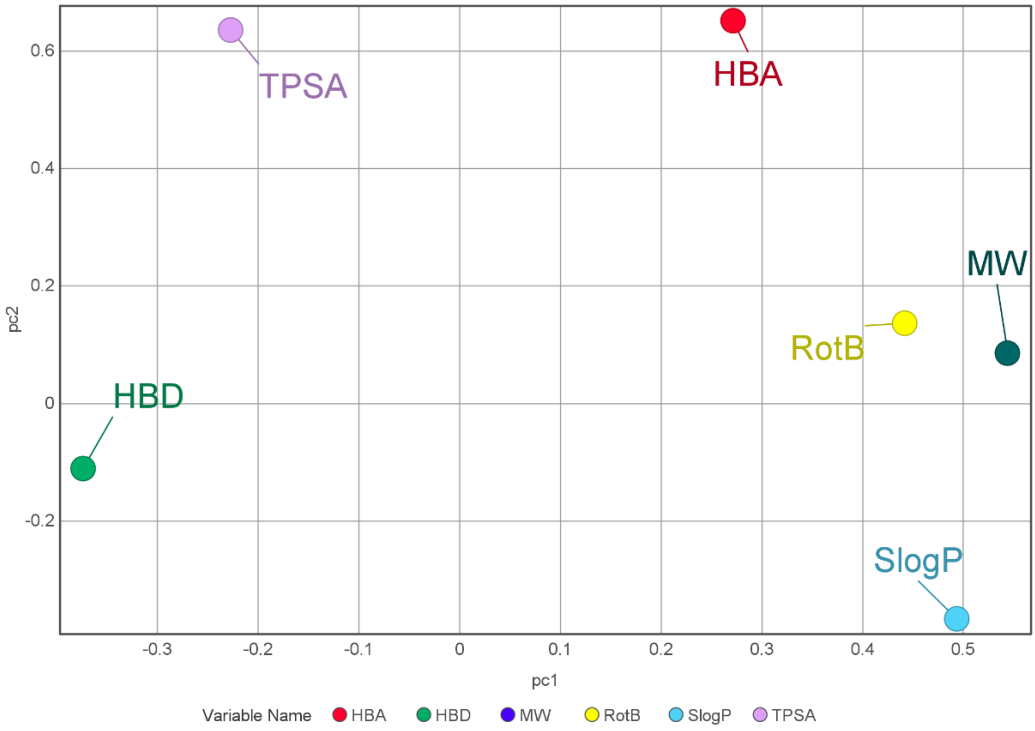


Figure 39: PC1 and PC2 loading plot for the physicochemical properties of the designed compounds.

The loading visual shows the results for principal components 1 and 2. Molecular weight, rotatable bond count and hydrogen bond acceptor count and SlogP have mostly positive load on principal component one, this component measures the property space description based on the correlation of these physicochemical properties. TPSA and HBA have a positive load on component two,

largely indicating its description of chemical space on the correlation of polar surface area and abundance of highly electronegative of atoms possessing polar bonds with hydrogen atoms.

Twenty-one compounds, violated either the PAINS, NIH, or Lipinski filters due to the presence of undesirable functional groups like azides and iso-thio-cyanates in the library. Consequently, only 50 compounds from the seventy-one met the filtering criteria. The Lipinski filters compounds based on the rule of five which is defined by four physicochemical parameters. This rule addresses the oral drug like properties of MW≤500; logP≤5; HBD≤5 and HBA≤10. The NIH and PAINS (Pan Assay Interference Compounds) filters identify potentially problematic or promiscuous compounds that may interfere with assay readouts or exhibit undesirable properties, it helps to identify compounds that may lead to false-positive or misleading results in biological assays (Baell & Holloway, 2010; Jadhav et al., 2010; Lipinski, 2004)

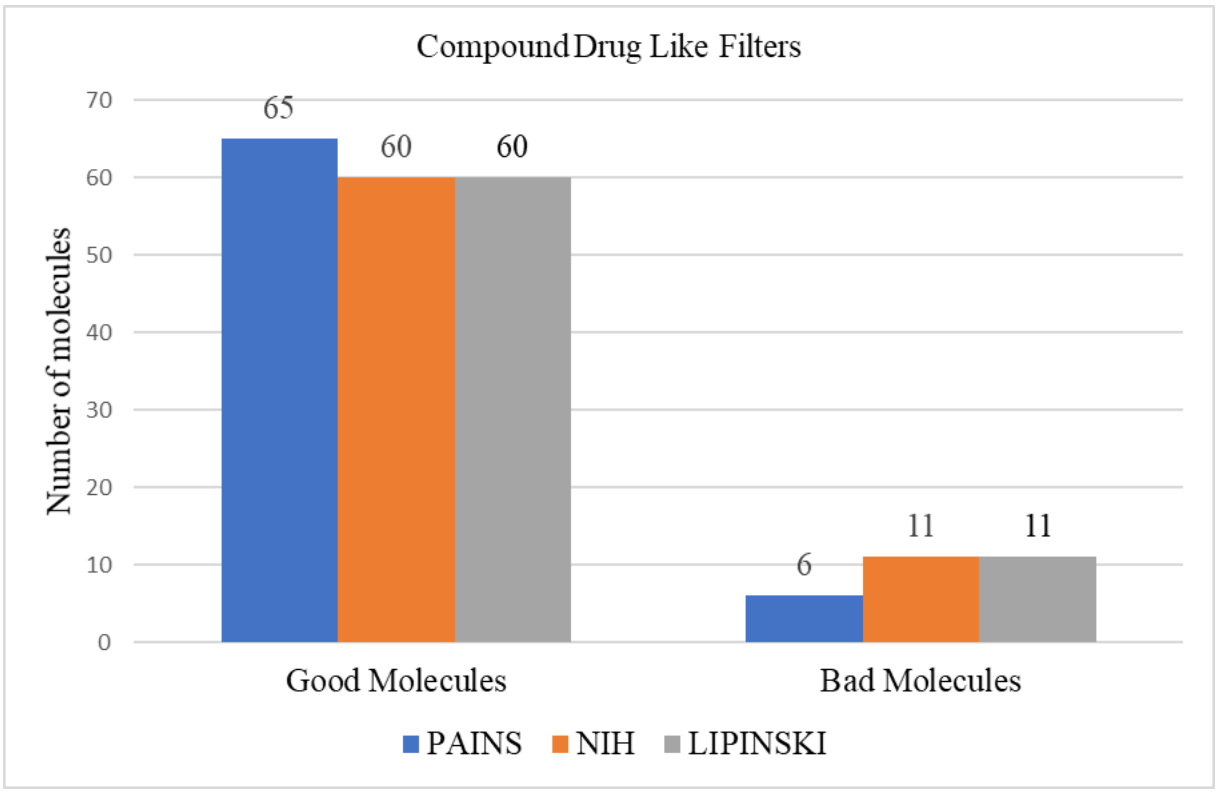


Figure 40: Drug-Like properties of designed compounds

Some common classes of PAINS alerts include Michael acceptors, which are compounds containing reactive electrophilic groups like α, β-unsaturated carbonyls or nitro groups. These groups can covalently bind to nucleophiles and interfere with assay results. Reactive functional

groups are compounds with chemically reactive groups such as thiols, hydrazines, or primary amines. These groups can react with assay components and cause false-positive or misleading results. Redox active compounds, such as quinones, phenols, or nitroso groups, can generate reactive oxygen species or interfere with redox processes in assays. Metal chelators are compounds containing specific chelating groups like dithiocarbamates or hydroxamic acids. These groups can complex with metal ions and interfere with metal-dependent assay systems.

Table 8: Compounds that fail Lipinski, PAINS and NIH filters

Compound	Lipinski Violation	PAINS	NIH
Gen_5_Cross_157058__3	HBD		×
Gen_5_Cross_239447__2	HBD		×
Gen_4_Cross_285367__1	HBD		×
Gen_1_Mutant_18_587834__2	HBD		Hydrazine
Gen_4_Cross_529873__3	HBD		×
Gen_5_Cross_502762__5	HBD		×
Gen_1_Mutant_75_26962__3	HBA		×
Gen_5_Cross_969434__5	HBA		×
Gen_4_Cross_671065__1	HBA		×
Gen_3_Cross_272672__3	HBA		×
Gen_4_Cross_29092__3	HBA		×
Gen_5_Mutant_8_570373__1	×	Azo	Azide
Gen_4_Mutant_6_664731__2	×	Azo	Azide
Gen_3_Mutant_9_489911__3	×	Azo	Azide, carbodiimide isocyanate
Gen_5_Cross_118324__1	×	Azo	Azide
Gen_4_Cross_225518__3	×	Azo	Azide, carbodiimide isocyanate
Gen_5_Mutant_6_956012__4	×	Azo	Azide
Gen_1_Mutant_13_20994__1	×	×	Azo, hydrazine
Gen_4_Mutant_10_672916__4	×	×	Carbodiimide, isothiocyanate
Gen_5_Mutant_10_644862__3	×	×	Carbodiimide, isothiocyanate
Gen_5_Mutant_11_883885__3	×	×	Carbodiimide, isocyanate

The effects of having a limited scaffold diversity can be seen in Table 8 where problematic functional groups are formed through mutation to introduce structural diversity in the compound library. Important physicochemical properties that should be strictly monitored during optimization of structures, logP and molecular weight were not violated by any of the molecules. Overall, the library was highly diverse structurally with good physicochemical properties.

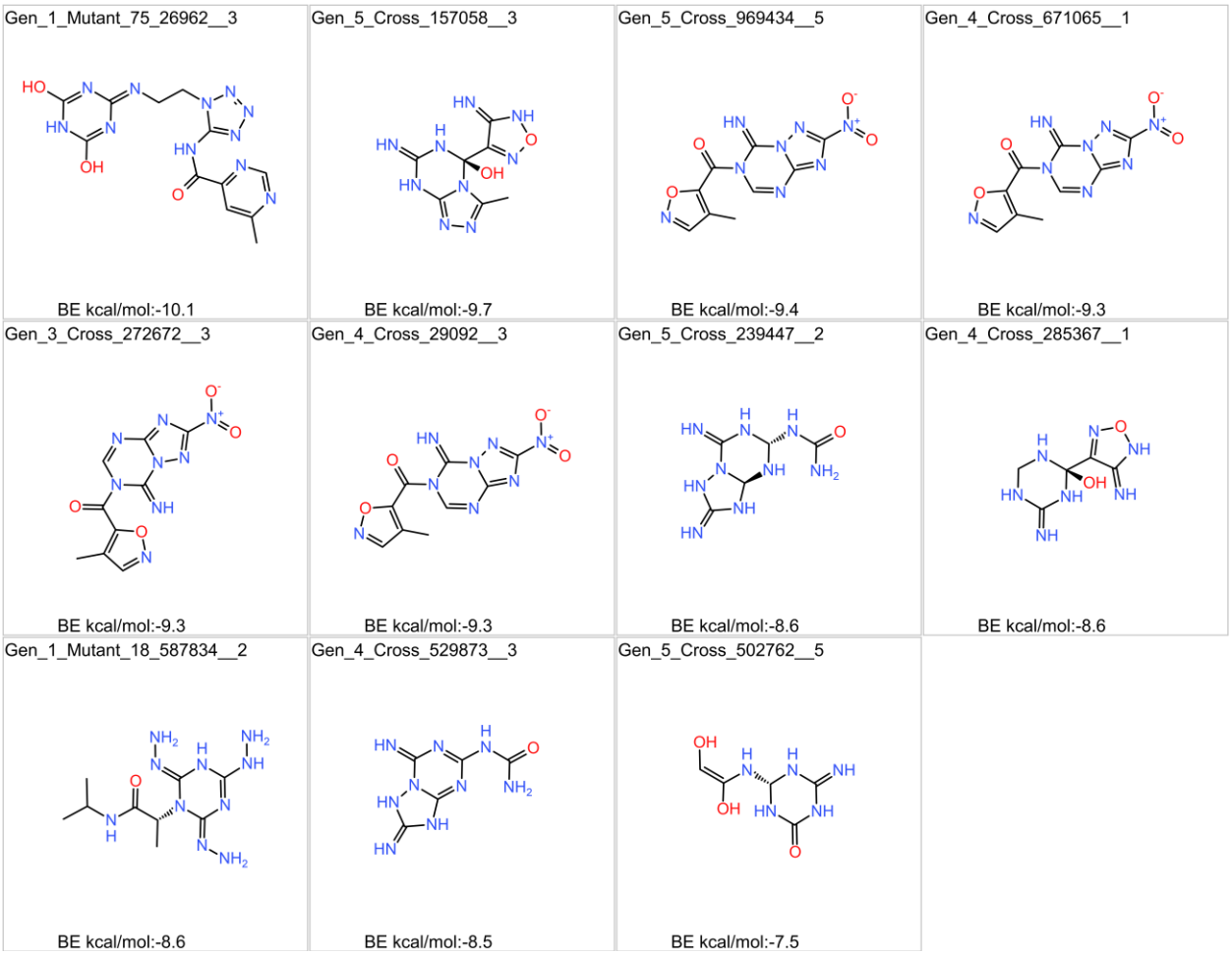


Figure 41: Compounds that fail Lipinski filter.

SwissADME was employed to profile the 50 compounds that made the selection, and a boiled egg analysis was carried out for the compounds. The boiled egg analysis was carried out for determining Blood Brain Barrier (BBB) permeability, passive absorption of molecules in the gastrointestinal tract (HIA), and the efflux of the compounds from the central nervous system (CNS) by P-glycoprotein (Daina et al., 2017; Daina & Zoete, 2016).

In the boiled egg analysis shown in Figure 42, the yellow region is the space occupied by compounds that permeate the blood brain barrier, the white region is the space occupied by compounds passively absorbed in the gastrointestinal tract (GIT). The Blue and red coloring for molecules is for compounds predicted to be effluated and not effluated by the P-glycoprotein from the nervous system, respectively. None of the tested compounds were predicted to permeate the BBB which acts as a protective barrier in the central nervous system, limiting the access of

substances into the brain. This is crucial to consider when designing drugs for neurological conditions. Out of the designed compounds, eleven were labeled blue, indicating that the P-glycoprotein would recognize them as substrates and actively transport them out of brain cells, leading to limited accumulation in the CNS and reduced neurological effects. On the other hand, the remaining 39 compounds were predicted not to be effluxed from the CNS since the P-glycoprotein would not recognize them as substrates. The compounds' oral administration effectiveness and bioavailability are largely determined by their passive absorption into the gastrointestinal tract (GIT). Among the compounds, five P-glycoprotein substrates and six non-P-glycoprotein substrates were predicted to be passively absorbed in the GIT.

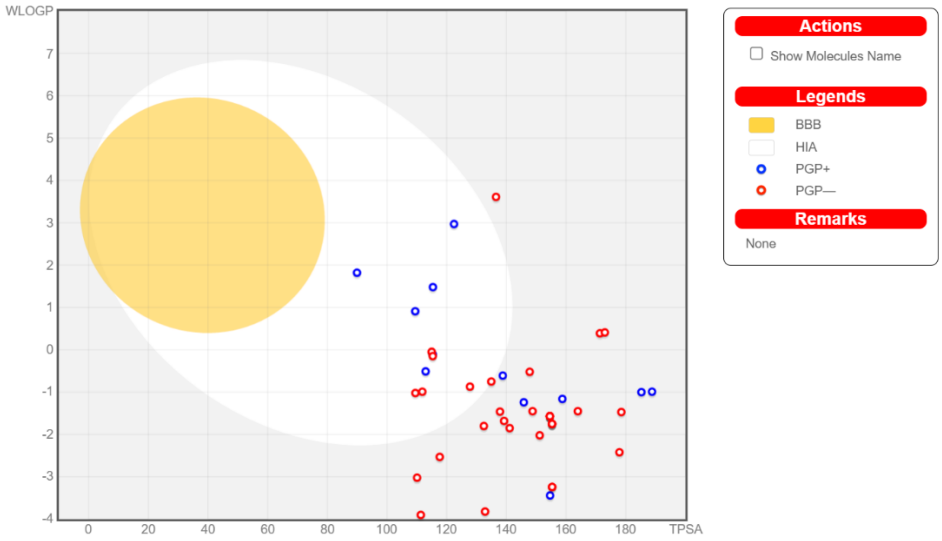


Figure 42: Boiled egg analysis for the 50 designed compounds (Daina et al., 2017).

Two of the four compounds with binding energies exceeding 10 kcal/mol were similar to one another in three dimensions. According to their scattering in 3D space, designed compounds have a high level of diversity overall (Figure 34).

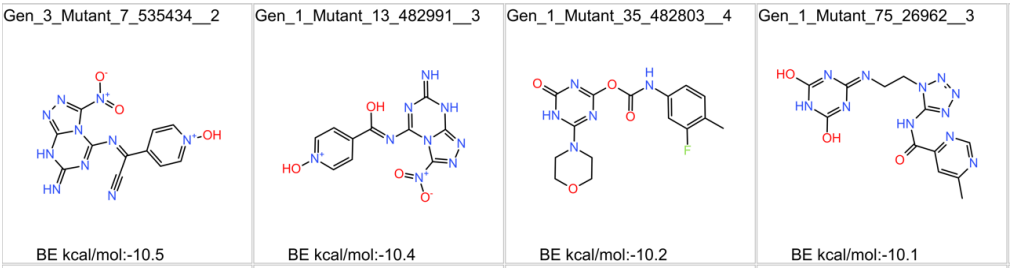


Figure 43: Top four compounds from *de novo* design

The first compound shown in Figure 44, Gen_3_Mutant_7_535434__2 with a binding affinity score of -10.5 kcal/mol had 5 hydrogen bonds with Thr9 HN, Thr10 HN, Gly199 HN, Asp203 OD2 and Ser338 HN. Two very strong hydrogen bonds were observed for Thr10 and Gly199 at distances of 1.8 Å and 1.9 Å respectively. Two other strong hydrogen bonds were observed for Ser338 and Asp203 at distances of 2.1 Å, 2.4 Å and 2.6 Å respectively, and a moderate strength hydrogen bond was observed for Thr9 at 2.6 Å. The strong hydrogen bond observed between the molecule and Gly199 is a key residue in the anchorage of ADP for the ribose section of the molecule.

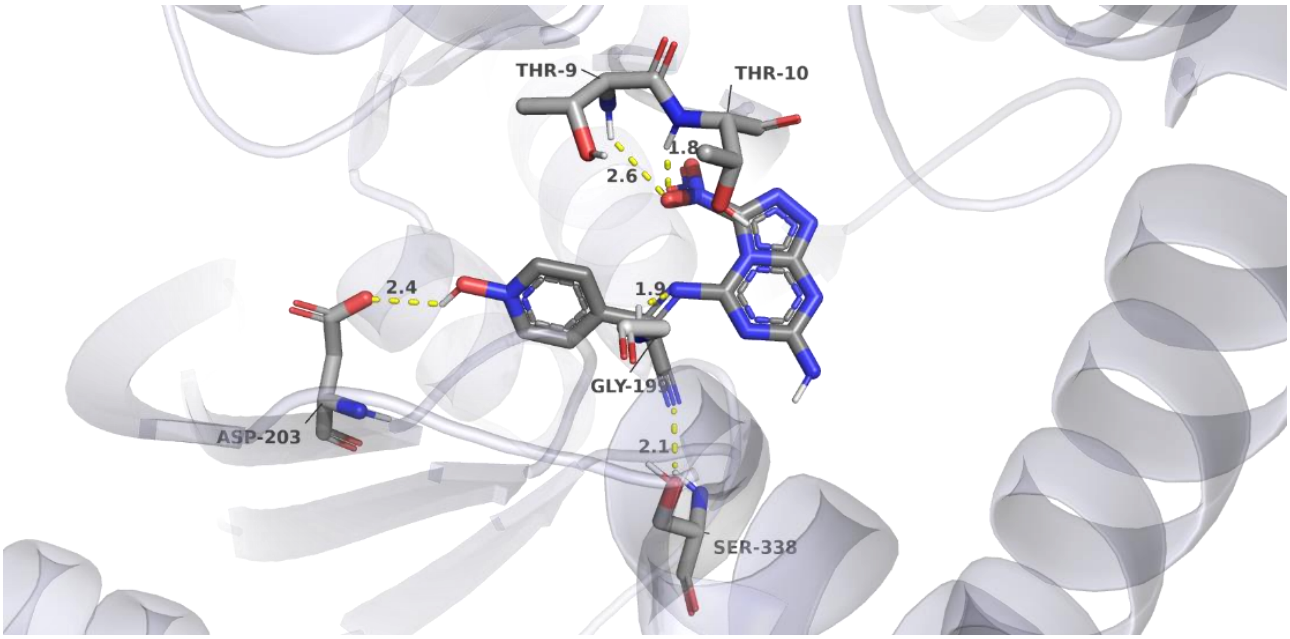


Figure 44: Gen_3_Mutant_7_535434__2 binding pose.

The second compound as shown in Figure 45, Gen_1_Mutant_13_482991__3 with a binding affinity of -10.4 kcal/mol had hydrogen bonds with Tyr11 O, Asp196 OD2 and Gly199 OD2. All four observed hydrogen bonds were strong, for Tyr11 and Asp196 the distances stood at 1.9 Å and 2.3A whilst Gly199 had two hydrogen bonds at a distance of 2.3 Å and 2.5 Å. Tyr11 is an important binding residue for ADP, being an important hydrogen bond anchor for both the phosphate groups. Gly199 also observed here is an important residue for anchoring the ribose on ADP.

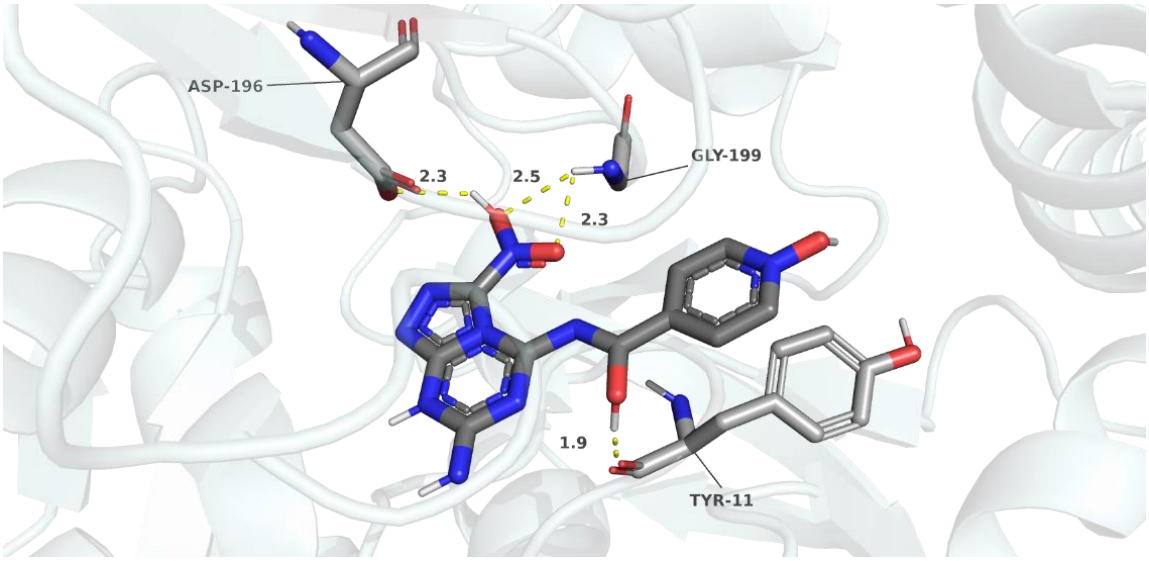


Figure 45: Gen_1_Mutant_13_482991__3 binding pose

The third compound as shown in Figure 46, Gen_1_Mutant_35_482803__4 with a binding affinity of -10.2 kcal/mol had hydrogen bonding with Asp196 OD2, Gly199 HN, Lys269 HZ1 and Arg270 HH. Two strong hydrogen bonds were observed for Asp196 and Gly199 at a distance of 2.0 Å and 2.1 Å respectively. Moderate strength hydrogen bonds at Lys269 and Arg270 were observed at distances of 2.7 Å for both, and lastly Asp196 had a weak hydrogen bond at a distance 3.3 Å. Residue Lys269 and Arg270 are important salt-bridge anchors for ADP, the ribose anchor for ADP residue Gly199 is also observed for this binding pose.

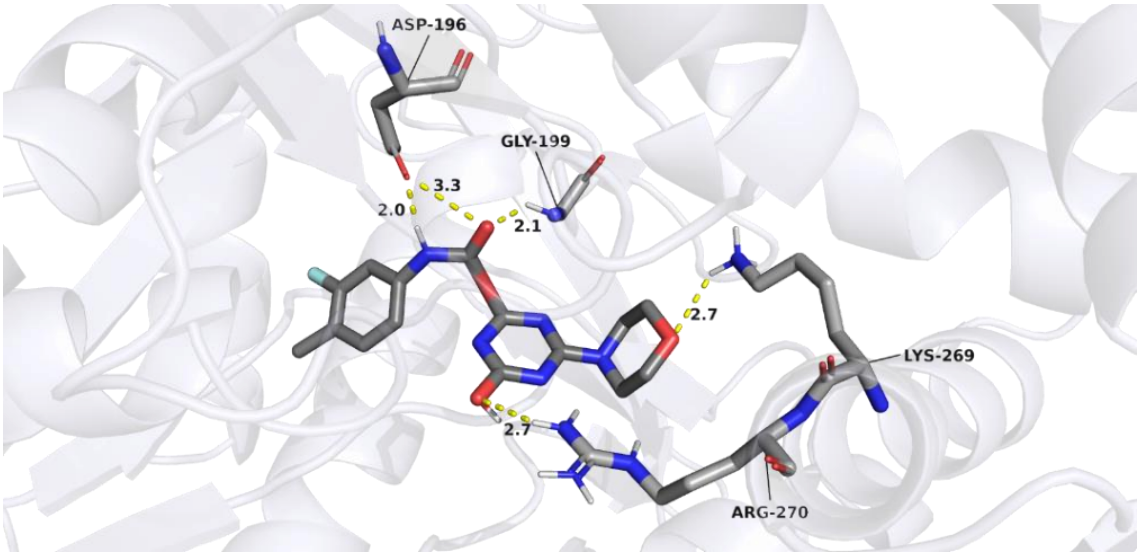


Figure 46: Gen_1_Mutant_35_482803__4 binding pose.

The fourth compound as shown in Figure 47, Gen_1_Mutant_75_26962__3 with a binding affinity of -10.1 kcal/mol had hydrogen bonds at Asp6 OD2, Leu197 O, Gly198 O, Gly199 O, Gly200 O and Ser338 OG. Two strong hydrogen bonds at Gly199 and Thr10 had distances of 1.3 Å and 2.2 Å respectively. Moderate hydrogen bonds were observed for Gly200, Asp6, Ser338, Leu197 and Gly198 at distances of 2.7 Å, 2.9 Å, 2.9 Å, 3.0 Å, and 3.1 Å respectively. One weak hydrogen bond was observed for Gly199 at a distance of 3.5 Å. The ribose anchor for ADP is once again observed with a strong hydrogen bond to the molecule, and a purine anchor Asp6 is also observed to participate in hydrogen bonding with moderate strength.

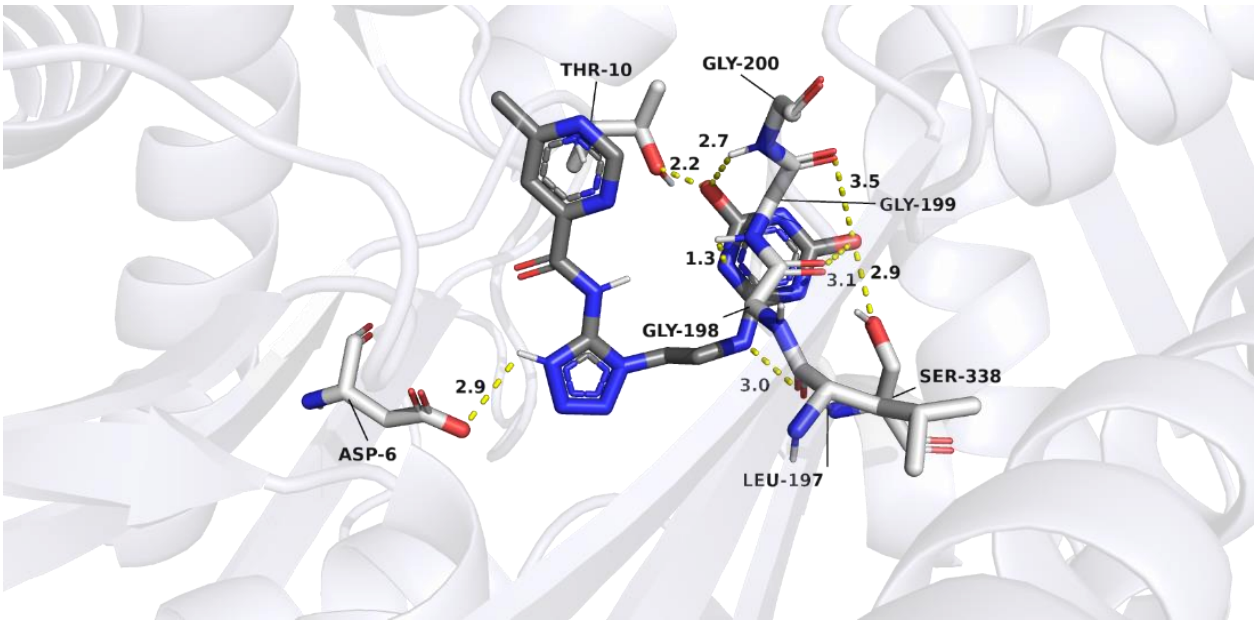


Figure 47: Gen_1_Mutant_75_26962__3 binding pose.

All the top four compounds had desirable binding poses, interacting with residues in the ADP binding site, however, Gen_1_Mutant_75_26962__3 (Figure 47) had a hydrogen bond donor violation for the Lipinski rule of five. Gen_3_Mutant_7_535434__2 (Figure 44) and Gen_1_Mutant_13_482991__3 (Figure 45) were highly similar structures separated by a hydroxyl and nitrile group.

4.3 Structural Refinement and Compound Selection

Compound selection was influenced primarily by the docking pose of ADP with the binding affinity of -8.76kcal/mol, the natural substrate to *Plasmodium falciparum* heat shock protein 70-1. The important residues that interact with one another in the ADP docking pose are depicted in

Figure 48, Table 9, and Table 10 below; these residues are thought to be crucial for competitive ADP inhibitors.

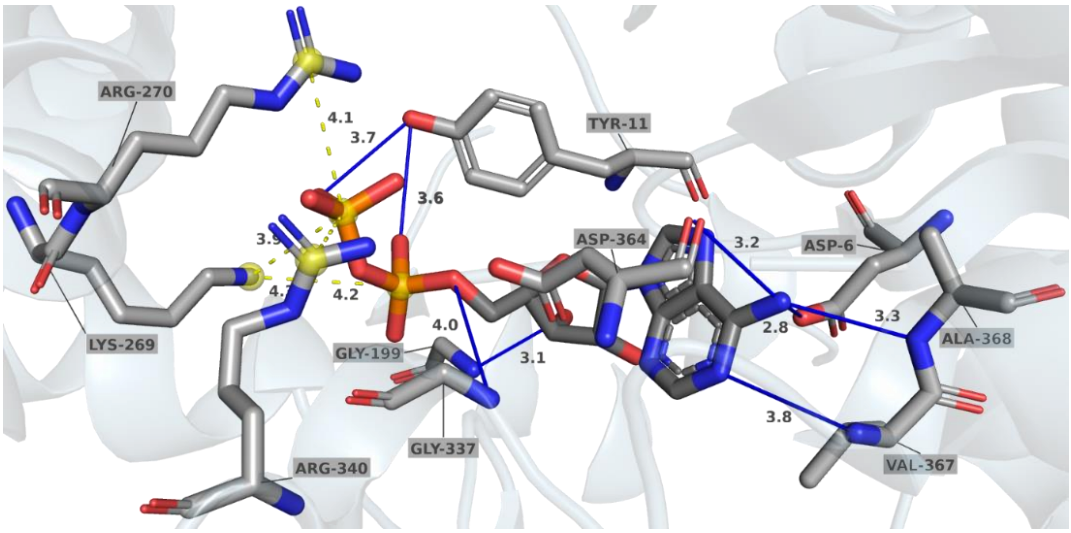


Figure 48: ADP binding pose

Table 9: ADP-*Pf*Hsp70-1 Hydrogen Bonds

Residue	Amino Acid	Distance Å	Atom	Protein Donor	Side Chain
6A	ASP	2.8	OD2	✓	✓
11A	TYR	3.6	OH	×	✓
11A	TYR	3.7	OH	✓	✓
199A	GLY	3.1	N	✓	×
337A	GLY	4	N	✓	×
364A	ASP	3.2	O	×	×
367A	VAL	3.8	N	✓	×
368A	ALA	3.3	N	✓	×

Table 10: ADP-*Pf*Hsp70-1 Salt Bridges

Residue	AA	Distance Å	Positive Protein?	Ligand Group
269A	LYS	3.90	✓	Phosphate
269A	LYS	4.30	✓	Phosphate

270A	ARG	4.08	✓	Phosphate
340A	ARG	4.20	✓	Phosphate

4.3.1 Selection of BN_01

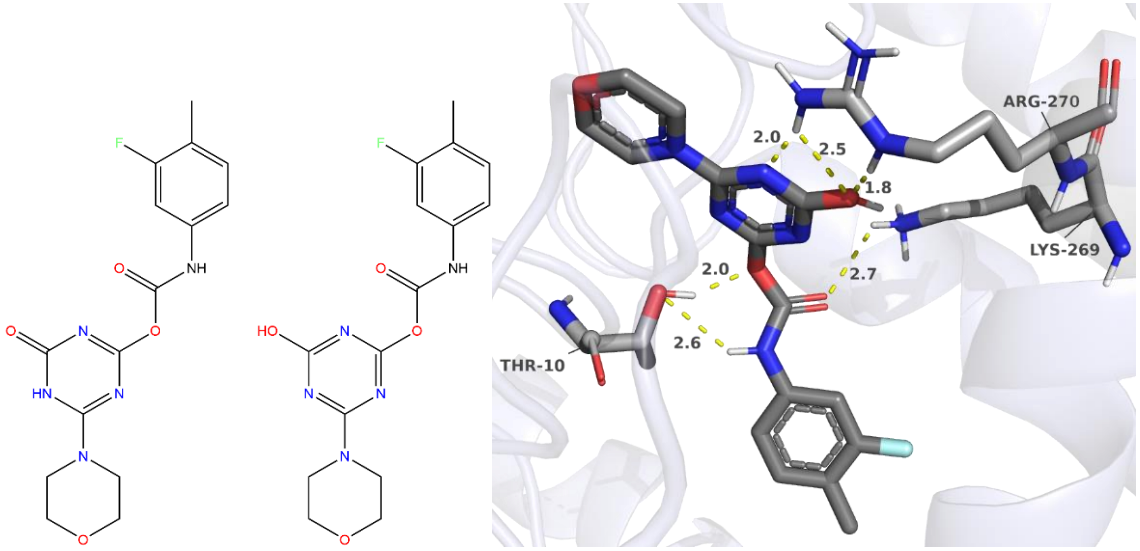


Figure 49: Structural refinements on BN_01 and the binding pose.

This was the first selected compound shown in Figure 49, the only modification on this compound was changing a double bond to a single bond, transforming from a carbonyl to an alcohol. This was in line with the availability of BN_01_TF for procurement. After structural modification the compound was reassessed via molecular docking and the resulting docking pose with a binding affinity of -9.5 kcal/mol had hydrogen bonding with Thr10 HG1 and OG1, Lys269 HZ1, Arg270 HE, and Arg270 HH. Four strong hydrogen bonds were present for Arg270 HE and Thr10 HG1 at distances of 1.8 Å, 2.0 Å, then two more for Arg270 HH at distances of 2.0 Å and 2.5 Å. Moderate-strength hydrogen bonds were observed for Thr1 OG1 and Lys269 at a distance of 2.7 Å and 2.7 Å. Arg270 and Lys269 were critical residues observed in the ADP binding pose, where the phosphate group was anchored.

The protein-ligand complex MDS starting from this binding position was used to determine the stability of the complex. RMSD calculations were performed independently for the protein-ligand complex, the backbone, and the ligand; the reference frame for superimposition was the initial frame along the protein backbone throughout these calculations.

The complex proved stable throughout 100 ns. This can be observed by a low RMSD fluctuation of below 2 nm for the protein-ligand complex, the protein backbone and the ligand captured independently (Figure 50). During the simulation, only minor conformational changes in the backbone are observed, the first after 30-40ns, the second after 60-70ns, and the third after 90ns.

Similar in-range conformational changes for the protein-ligand complex are observed around 40ns and 70ns, these conformational changes are independent of the ligand RMSD, and they were confirmed in the distribution histogram where 2 peaks show clear deviation from the original conformation by the protein-ligand complex (Figure 53). A coinciding RMSD movement for protein-ligand and ligand occurred just below 20ns, suggesting that the major conformational changes observed for the whole protein were independent of the ligand.

RMS calculations were carried out to clarify these observations, fluctuations for all residues overtime were plotted in the nanometer range (

Figure 51). These revealed a relatively stable residue fluctuation below 0.75 nm throughout the simulations for residues in the nucleotide and substrate binding domain. Major residue fluctuations were observed for residues present in the c-terminal of the protein (lid).

Mean square displacement (MSD) calculations were conducted to examine the ligand's motion within the binding site. The results presented in Figure 52, indicated that the ligand is firmly bound within the active site of the protein, with displacement values consistently below0.05 nm^2. In a study by Planes and coworkers to investigate the lateral movement of Endothelin 1 type A and Angiotensin II type 1 proteins, the mean square displacement was in the μm^2 range, setting the benchmark for ligand movement within the binding site (Planes et al., 2020).

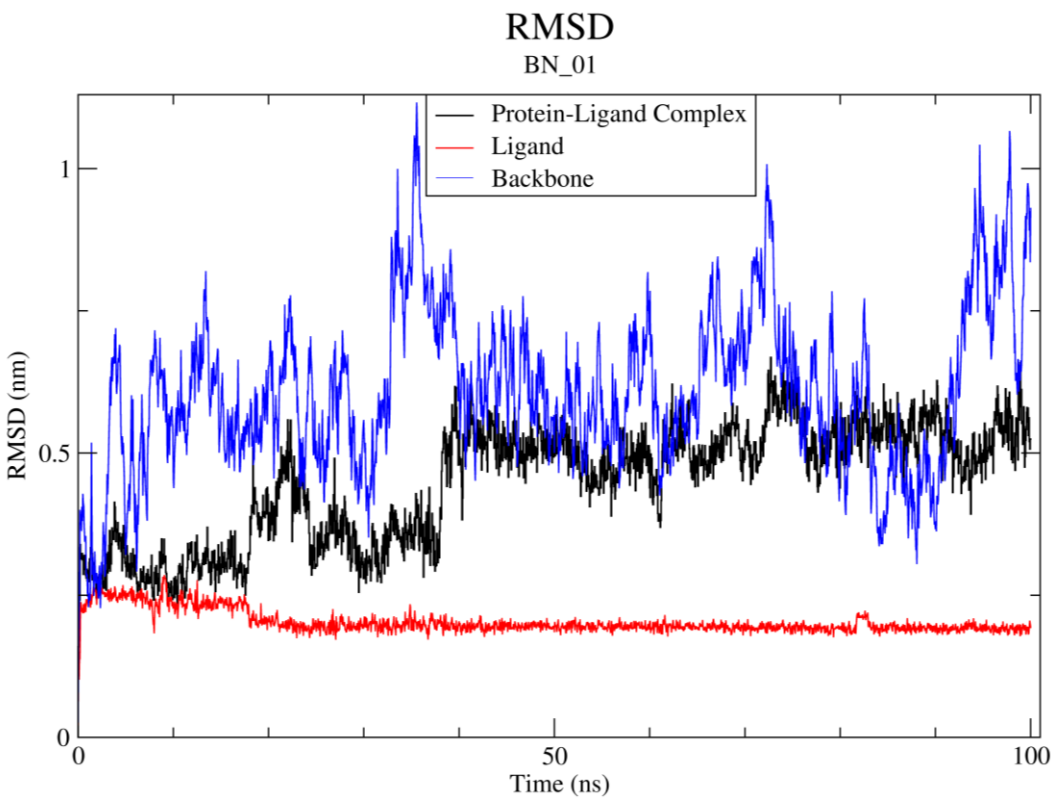


Figure 50: RMSD for protein-ligand complex, ligand and protein backbone for BN_01 MDS.

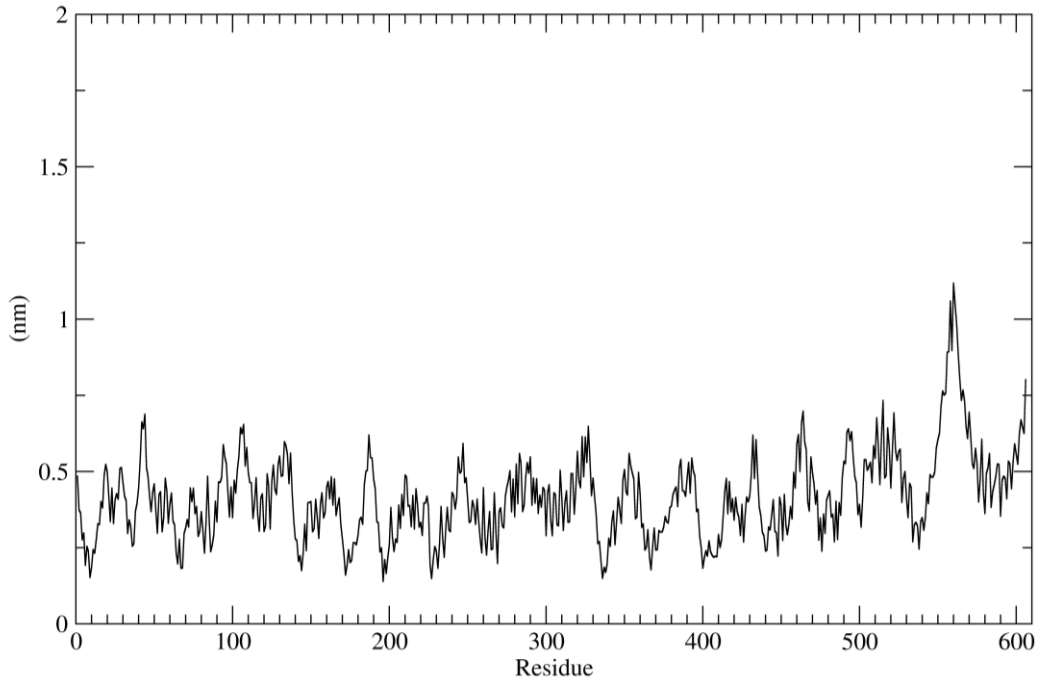


Figure 51: Residue RMS fluctuation for BN 01 MDS

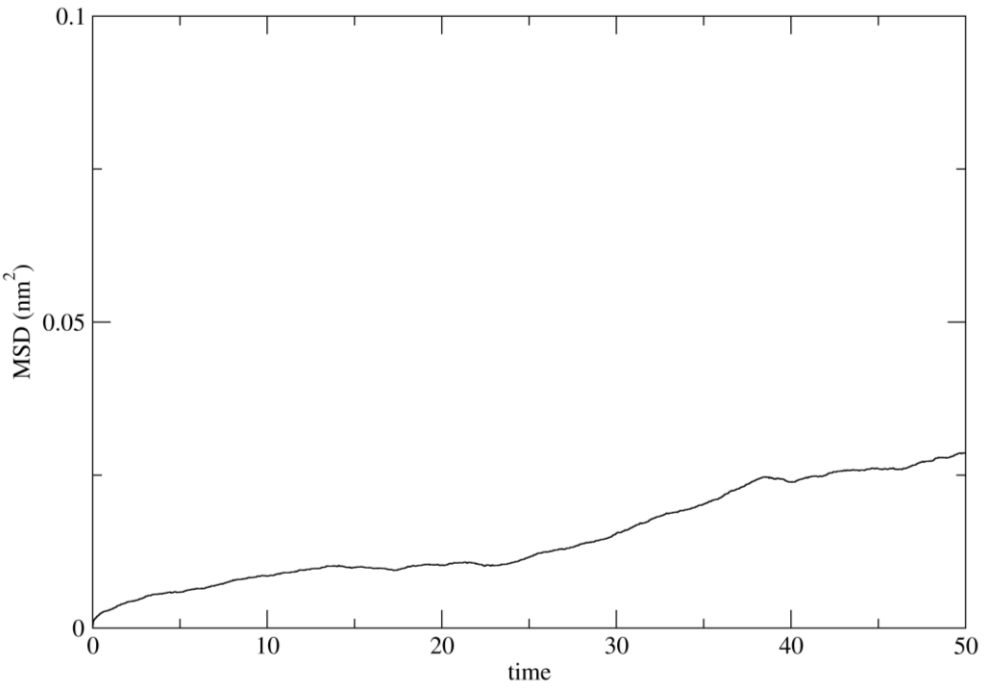


Figure 52: Mean Square Displacement of BN_01 MDS

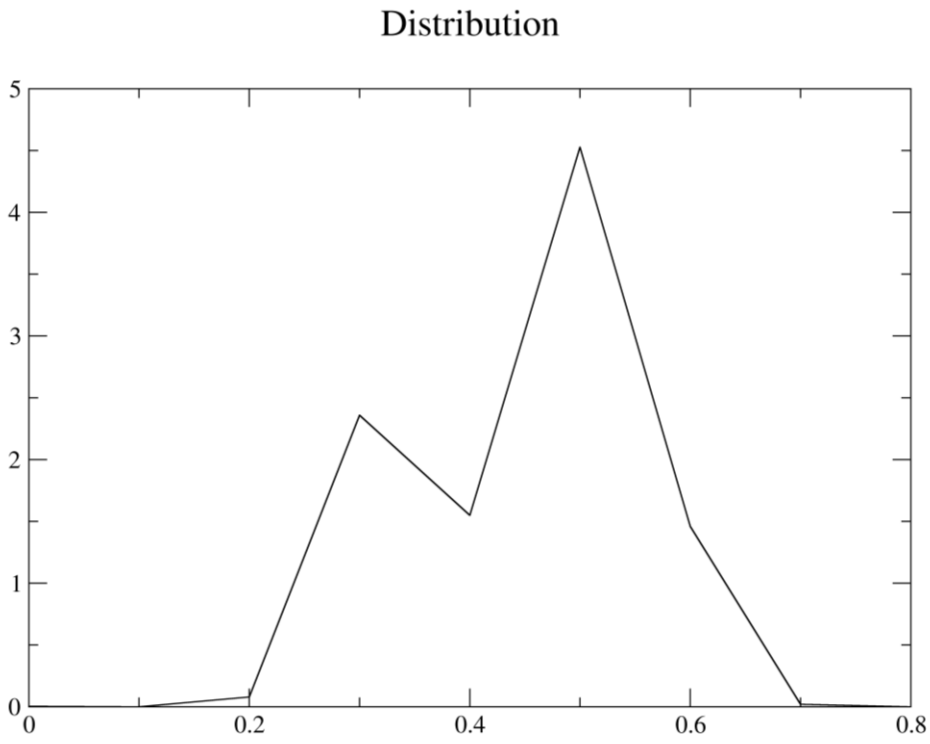


Figure 53: Protein backbone RMSD histogram distribution showing deviation from original conformation for BN_01 MDS

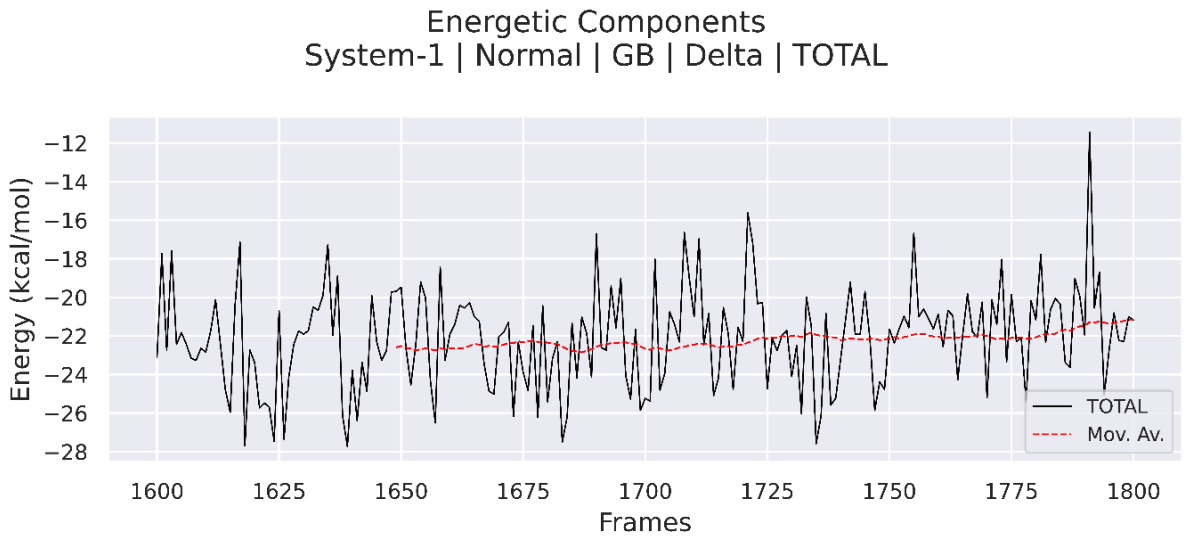


Figure 54: Total energy fluctuation for BN_01 MMGBSA

MMGBSA analysis for 200 frames, the point at which the protein-ligand complex was most stable revealed a Gibbs free binding energy of -27.73 kcal/mol being the lowest observed, and an average of -22 kcal/mol with huge contributions attributed to the residue Tyr11. Other contributing residues were Thr9, Ser12, Cys13, Arg270, Gly336, Gly337 Arg340 and Val367. Tyr11, Arg270, Gly337, Arg340 and Val367 are important residues for ADP binding participating in purine moiety and phosphate chain anchoring of the molecule. This compound possesses a similar binding pose to ADP with its interacting residues covering the same space as ADP, from the purine moiety to the phosphate groups. The residue Glu172 with an overall positive energy contribution to the protein-ligand complex is located 7.9 Å away from the ligand molecule, discounting any steric clashes between the two.

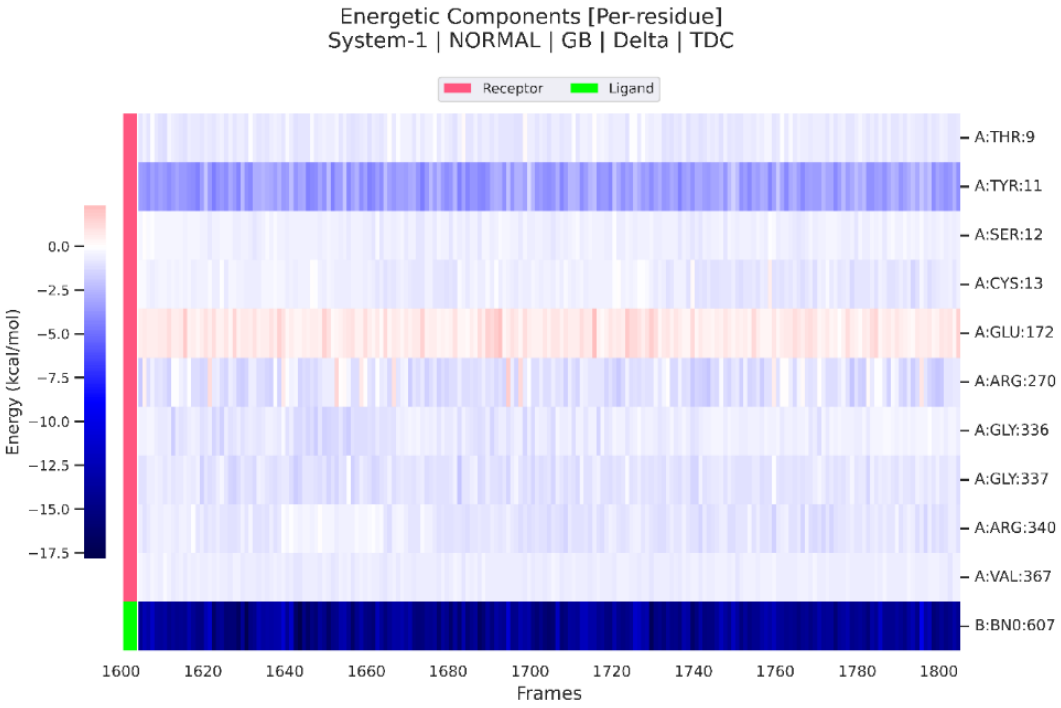


Figure 55: Stability of energetic contributions in binding for BN_01 MMGBSA

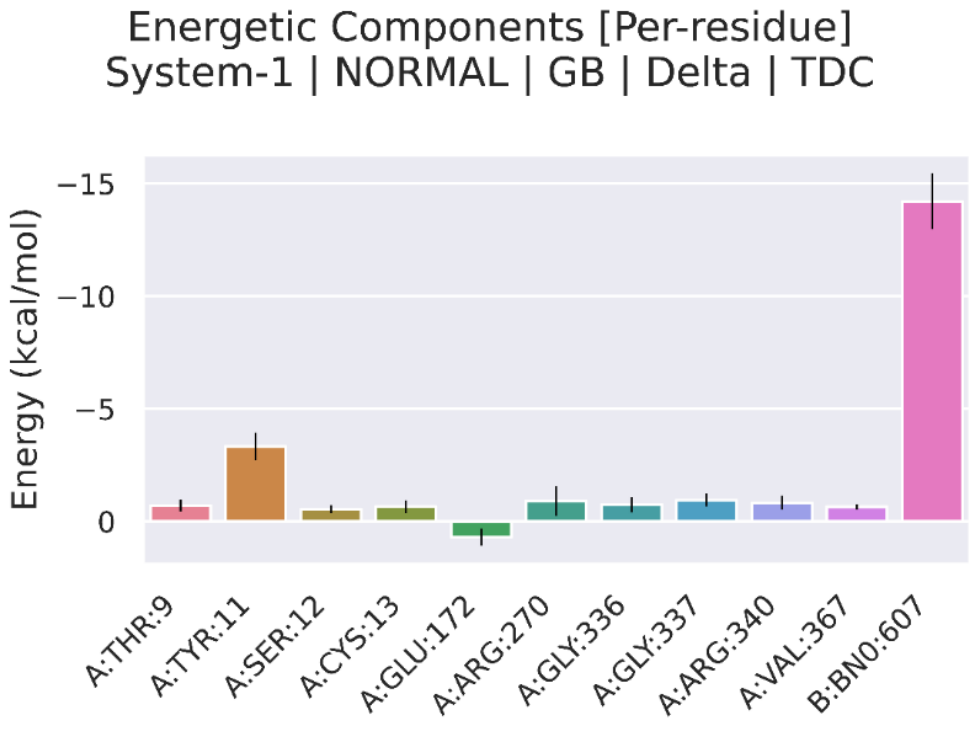


Figure 56: Residue energetic components for BN_01 MMGBSA

4.3.2 Selection of BN_02

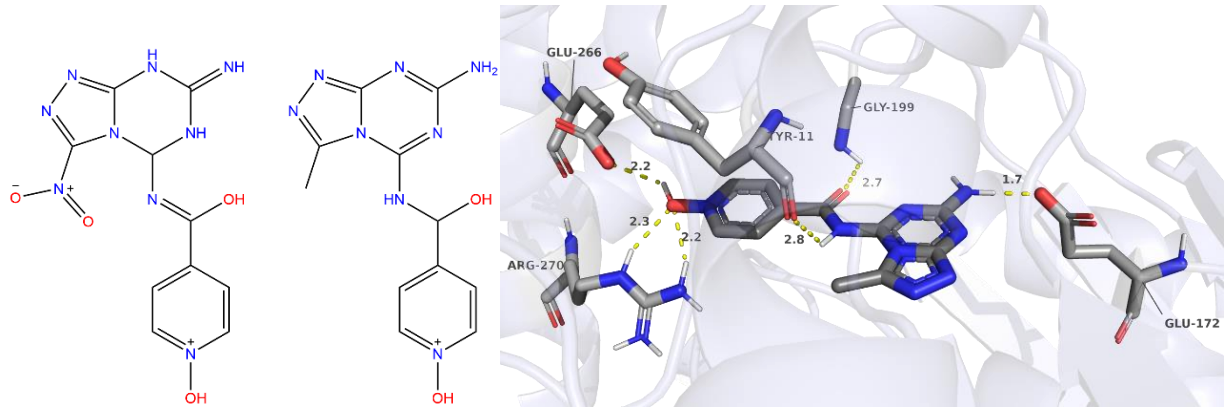


Figure 57: Structural refinements on BN_02 and the binding pose

This was the second selected compound shown in Figure 57, the iminol functional group was substituted with an amide functional group, the nitro group was substituted for a methyl group and the imine functional group was substituted for an amine. These substitutions were chosen based on the availability of the scaffolds for synthesis. The binding pose for BN_02 with a binding affinity of -8.7 kcal/mol revealed hydrogen bonding with residues Tyr11 O, Glu172 OE2, Gly199 HN, Glu172 OE2, Arg270 HH and Arg270 HE. Strong interactions were observed for Glu172, Glu266, and two for Arg270 at distances of 1.7 Å, 2.2 Å, 2.2 Å and 2.3 Å respectively. Moderate strength hydrogen bonds were observed for Gly199 and at distances of 2.7 Å and 2.8 Å respectively. Key residues in ADP binding were observed, these were Tyr11, Gly199 and Arg270.

Similar MDS was also carried out for BN_02 and BN_03 including all other post-simulation calculations. A similar trend for RMSD for protein-ligand complex and protein backbone is observed with conformational changes in-range for around 35ns and around 45ns. The conformational changes for the protein-ligand complex are within a 1 nm range proving the stability of the complex despite minor fluctuations in RMSD. The ligand was stable throughout the whole simulation with a minor RMSD change just below 20ns, this change coincided with a minor change for the protein backbone.

The RMS calculations for residues revealed a more pronounced residue fluctuation compared to BN_01, these account for the deviations from the original structure observed in Figure 61 of the distribution histogram where we observe two deviations from the original structure, the second deviation acquired stability momentarily before a return to the original structure. The stability of

the bound ligand can also be confirmed through the MSD graph (Figure 60) with values below 0.1 nm^2.

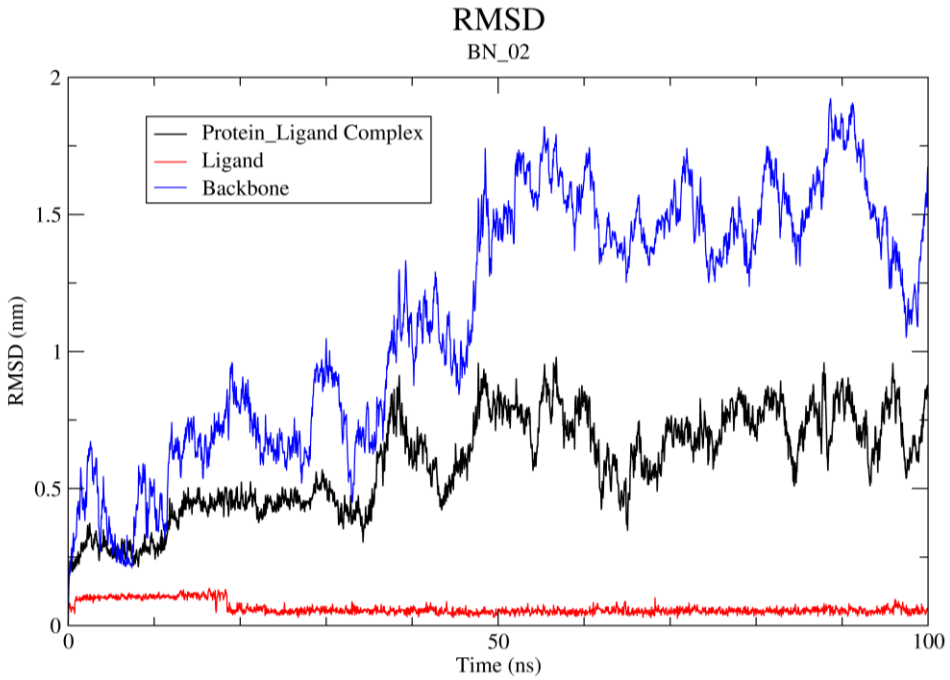


Figure 58: RMSD for protein-ligand complex, ligand and protein backbone for BN_02 MDS.

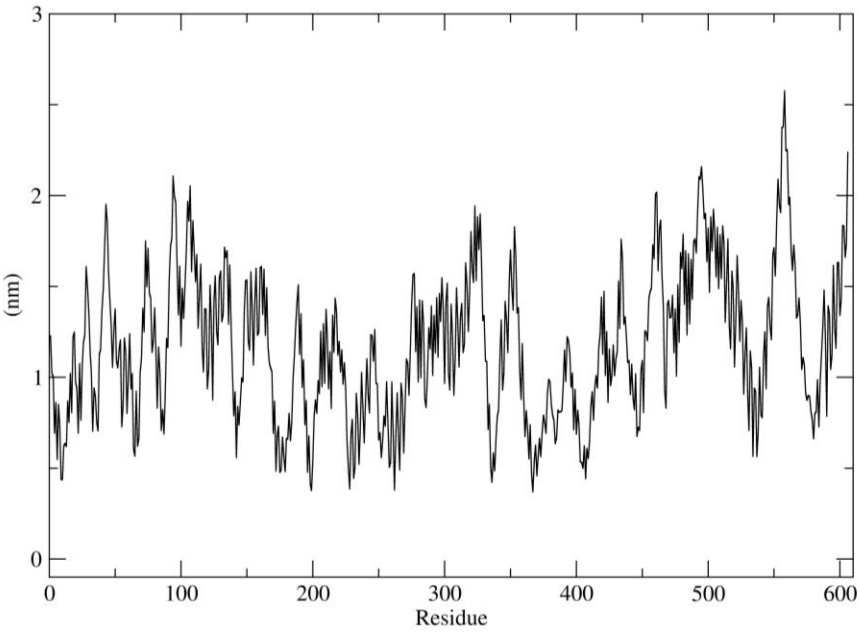


Figure 59: RMS fluctuation for residues for BN_02 MDS

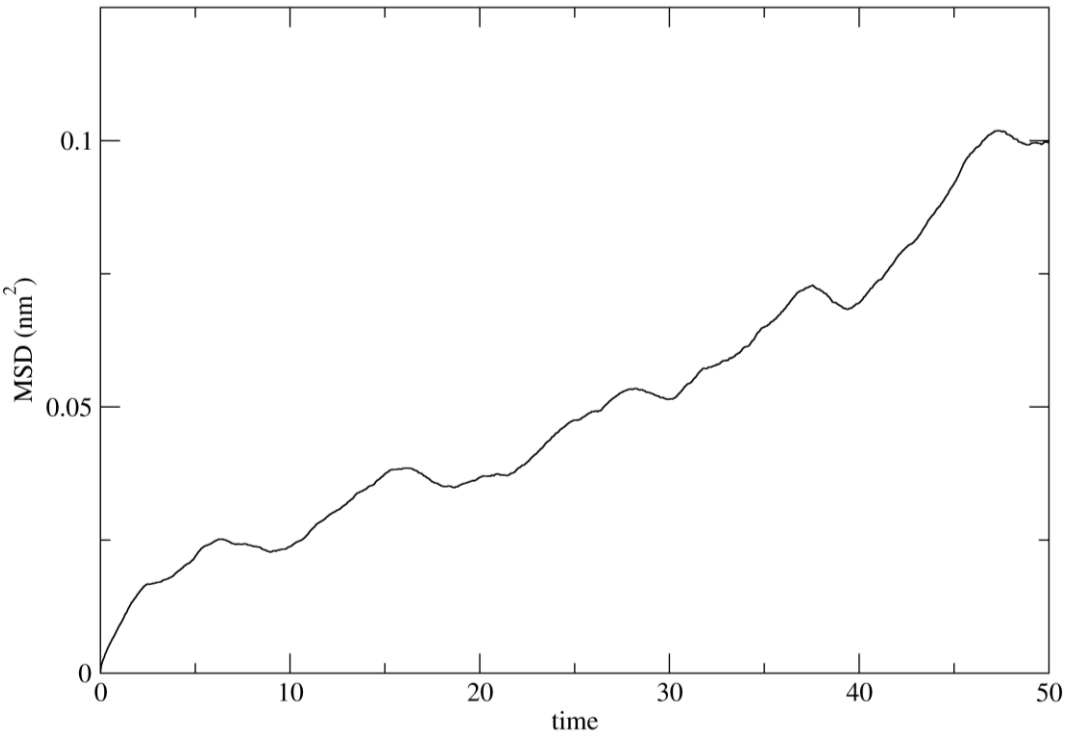


Figure 60: Mean Square Displacement for BN_02 MDS

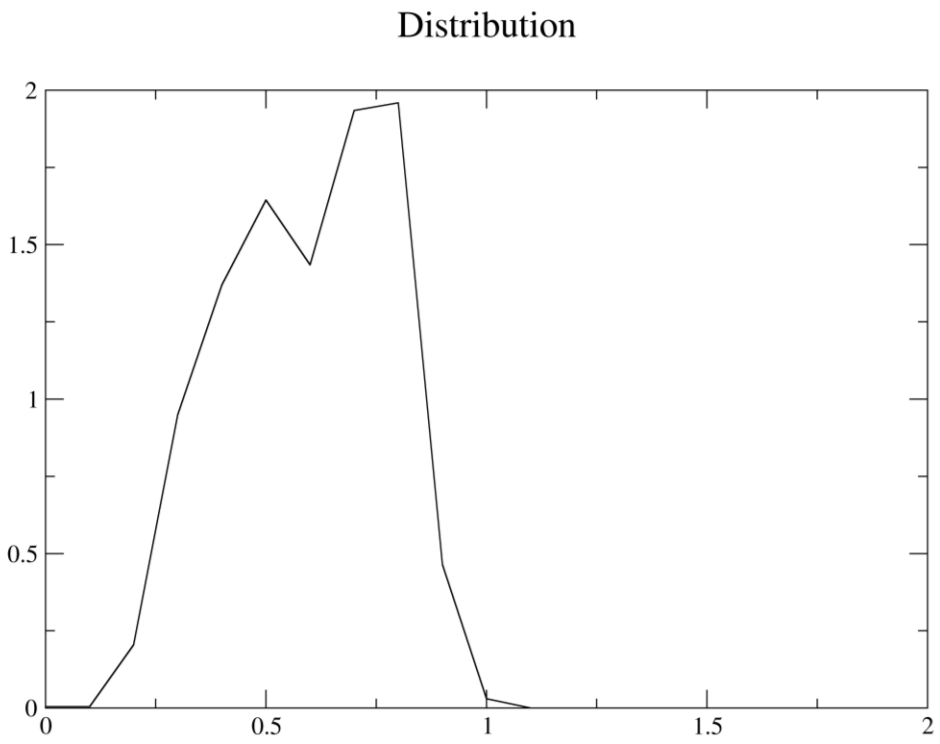


Figure 61: Protein backbone RMSD histogram distribution showing deviation from original conformation for BN_02 MDS

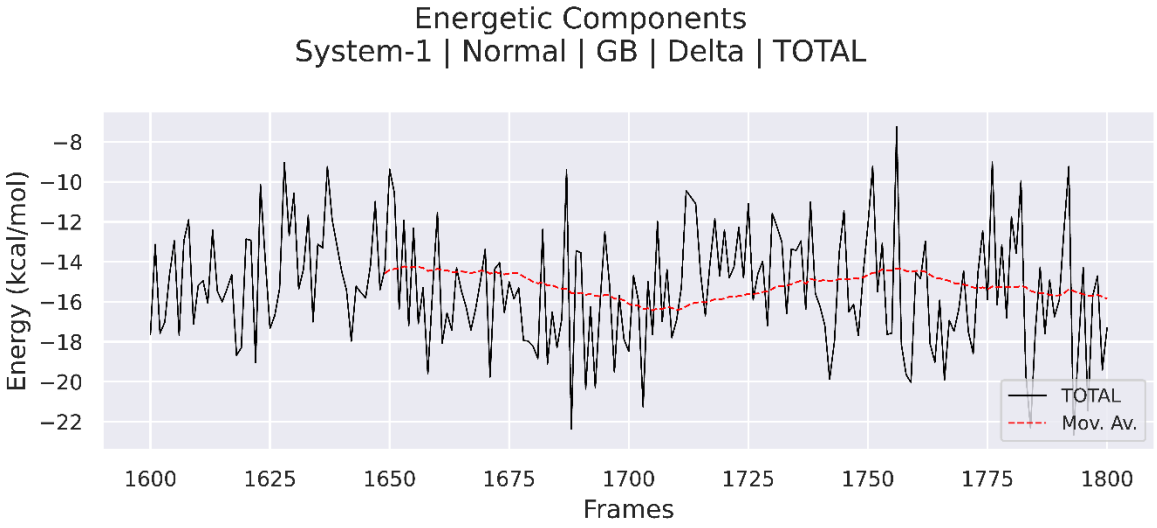


Figure 62: Total energy fluctuation for BN_02 MMGBSA

MMGBSA analysis for 200 frames, the point at which the protein-ligand complex was most stable, revealed a Gibbs free binding energy of -22.66 kcal/mol as shown in being the lowest observed

and an average ranging between 14 kcal/mol and 16 kcal/mol, with huge contributions also attributed to the residue Tyr11 with high stability throughout the analysis. Other contributing residues were Asp6, Thr9, Tyr11, Ser12, Cys13, Lys67, Glu266 and Val367. Asp6, Tyr11 and Val367 are important key residues in ADP binding, covering the ADP molecule from the purine moiety (Asp6 and Val367) to the phosphate groups (Tyr11). The Tyr11 hydrogen bond is observed to be the most stable throughout the simulation, an important residue for hydrogen bonding to the phosphate group of ADP and now BN_02.

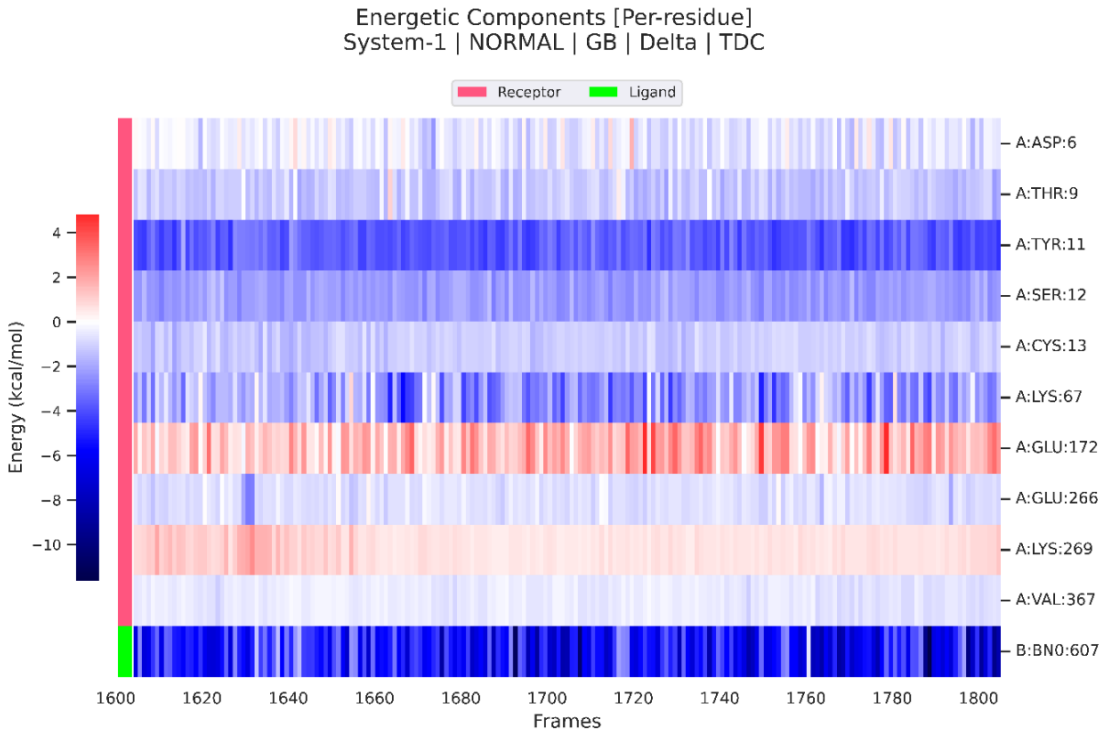


Figure 63: Stability of energetic contributions in binding for BN_02 MMGBSA.

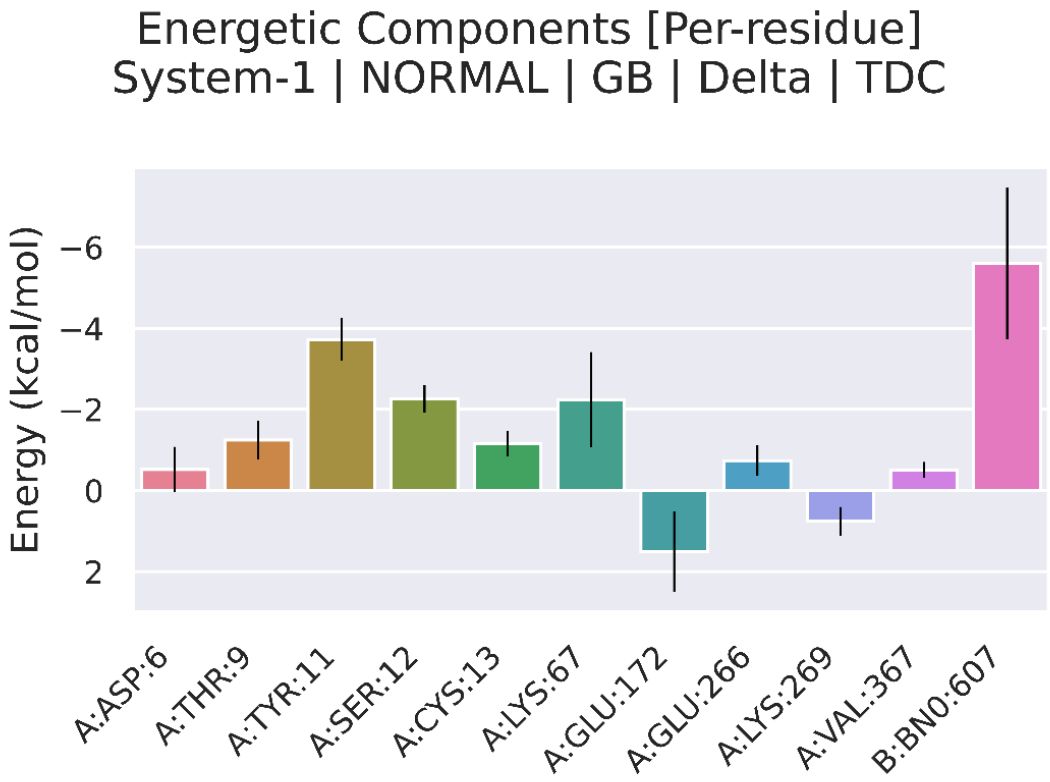


Figure 64: Residue energetic contributions for BN_02 MMGBSA

4.3.3 Selection of BN_03

The third compound BN_03 as shown in Figure 65, was derived from combining the triazine fragment from BN_01_TF with the acid fragment BN_02_TF to form a compound with an ester linkage. This compound had a binding affinity of -9.4 kcal/mol, and showed hydrogen bonding with the residues Lys269 HZ1, Arg270 HH, Asp364 O, Asp364 OD1 and Val367 HN. Residues Arg270 HH, Lys269 and Asp364 O and OD1 had strong hydrogen bonds with distances at 2.2 Å, 2.2 Å, 2.3 Å and 2.4 Å. Moderate-strength hydrogen bonds at a distance of 2.6 Å were observed for Val367 and Arg270 HH.

BN_03 has the same docking pose as ADP with interacting residues at Lys269 HZ1, Arg270 HH, Asp364 O, Asp364 OD1, and Val367 HN all being critical residues in ADP binding involved in anchoring the purine moiety and phosphate group.

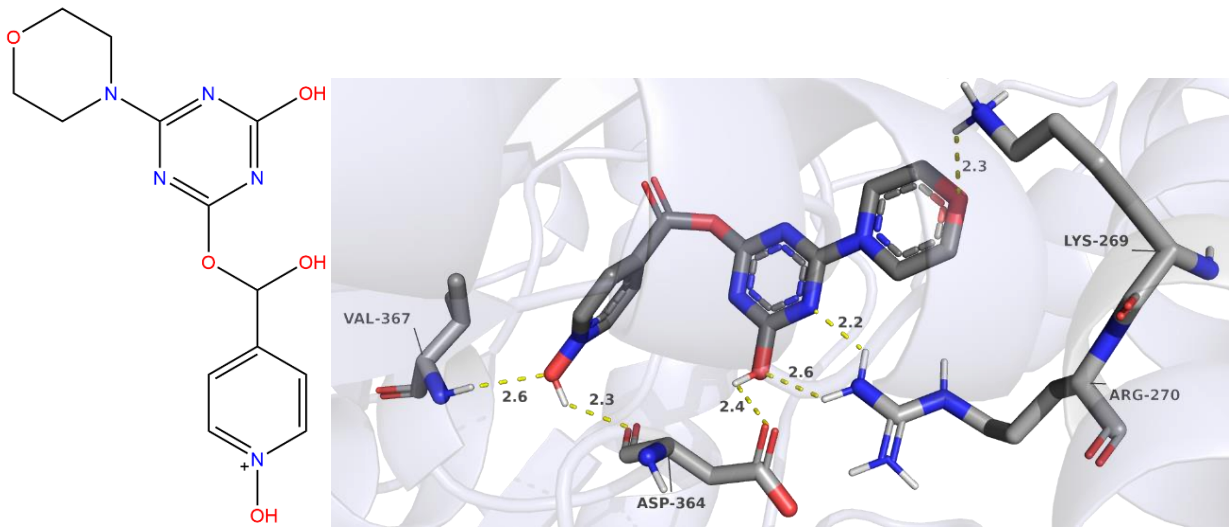


Figure 65: Binding pose of BN_03

A 100 ns MDS was run to determine the stability of the protein-ligand complex for *Pf*Hsp70-1 and BN_03. The resulting protein-ligand complex proved to be the most stable of the three, with a RMSD fluctuation below 0.4 (Figure 66) and a residue fluctuation below 0.5 (Figure 67). The protein-ligand pair; protein backbone and ligand all underwent a minor conformational change that coincided with each other at the region ranging from 40 ns to 60 ns, the order of this conformational change occurs for the protein-ligand complex and the ligand just below 45 ns and a following conformational change for the protein-ligand complex and for the protein backbone after 50 ns.

This change for the protein-ligand complex occurs independent of the ligand, suggesting only the protein continued changing of the conformation, and that the earlier conformation change was largely influenced by the side chains interacting with the ligand. This single conformational change can be observed through the RMSD histogram distribution, with one major deviation from the original structure clearly shown. The overall low RMSD from the complex suggests a tightly bound ligand with limited movement and this can also be confirmed by the mean square displacement of the ligand as shown in Figure 68, which sits below 0.005 nm^2.

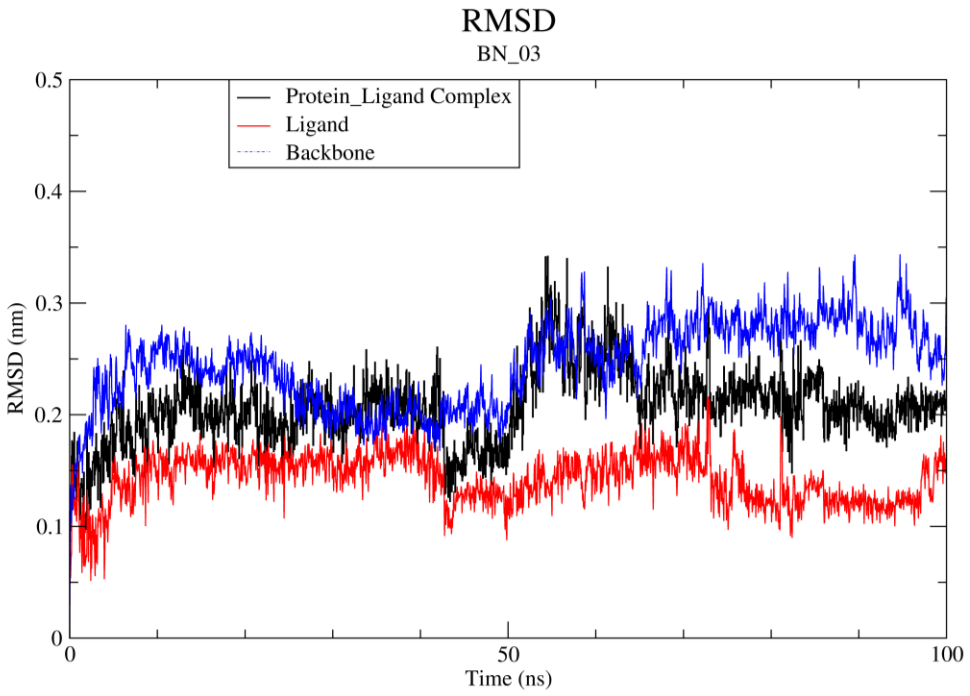


Figure 66: RMSD for protein-ligand complex, ligand and protein backbone for BN_03 MDS.

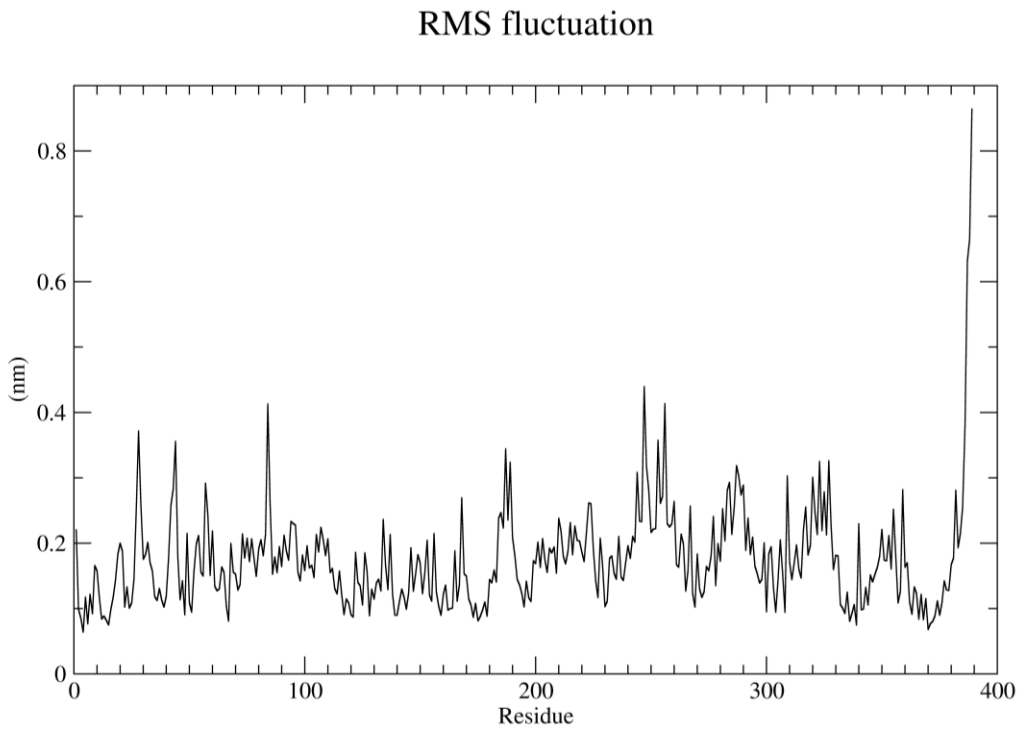


Figure 67: RMS fluctuation for residues for BN_03 MDS

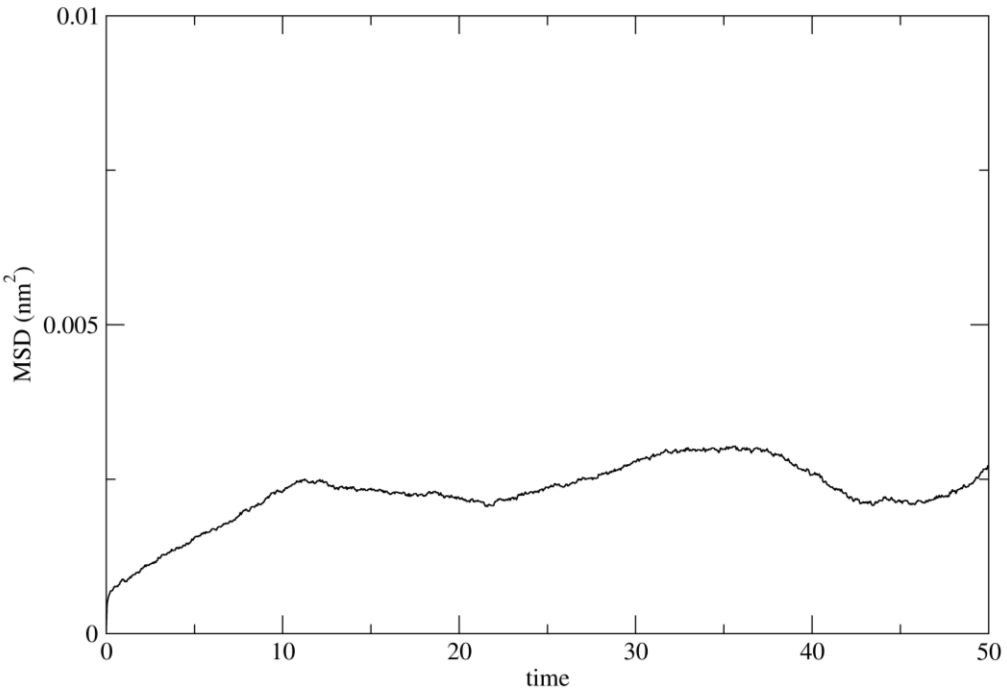


Figure 68: Mean Square Displacement for BN_03 MDS

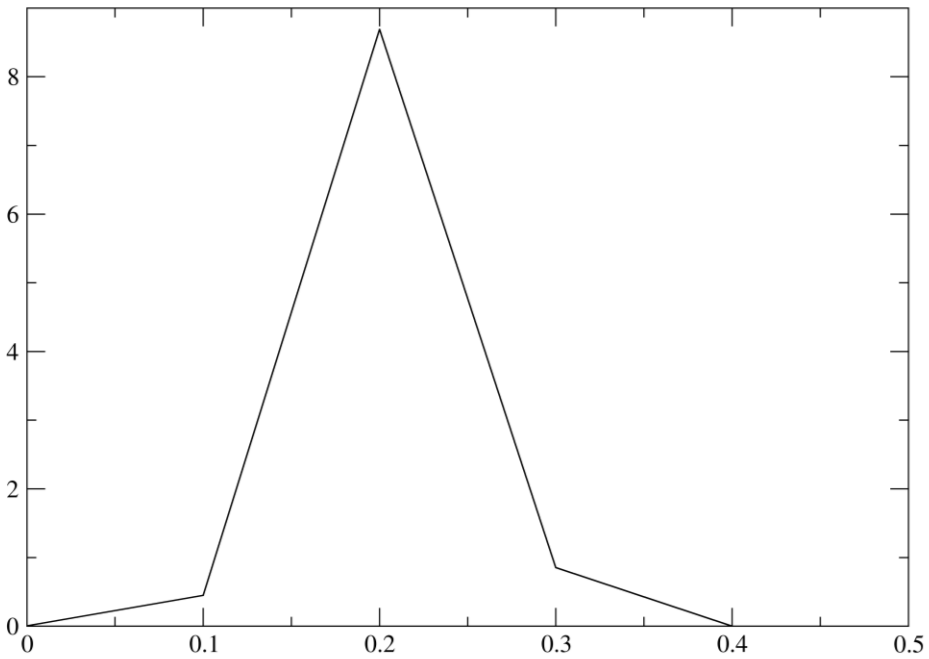


Figure 69: Protein backbone RMSD histogram distribution showing deviation from original conformation for BN_03 MDS

A stable 200 frames were put through MMGBSA analysis to obtain the lowest Gibbs free binding energy of -37.96 kcal/mol and an average between -27.5 to -31 kcal/mol (Figure 70), the lowest of the three with interactions at Tyr11, Ser12, Cys13, Thr33, Lys67, Gly198, Lys269, Arg270, Gly336 and Gly337 (Figure 71)

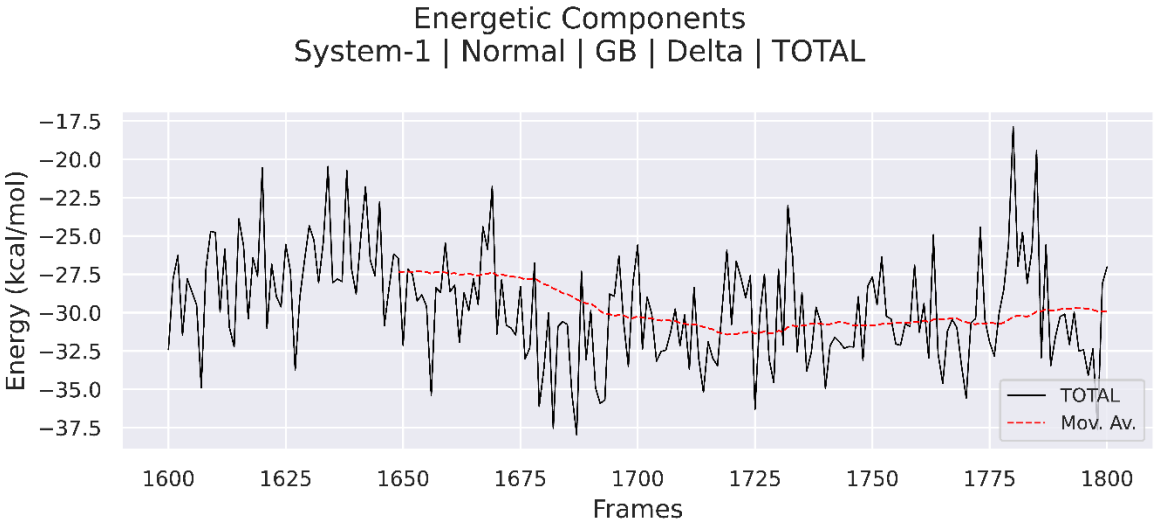


Figure 70: Total energy fluctuation for BN_03 MMGBSA

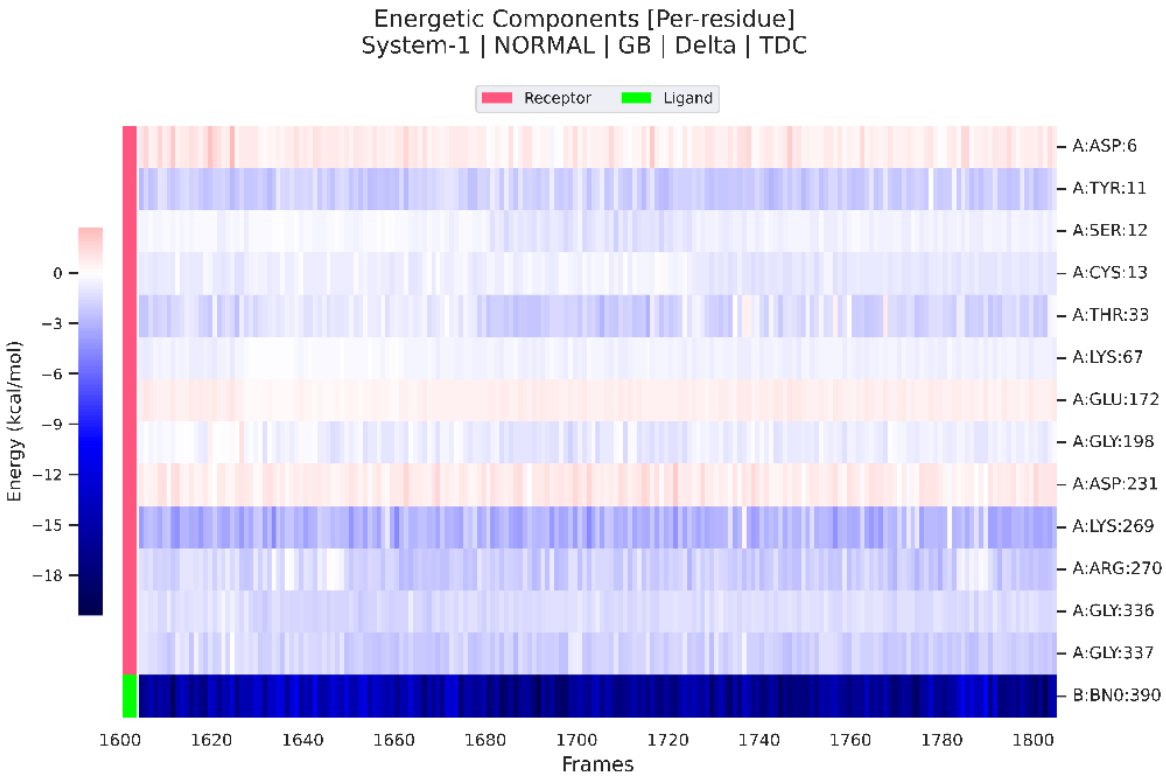


Figure 71: Stability of energetic contributions in binding for BN_03 MMGBSA.

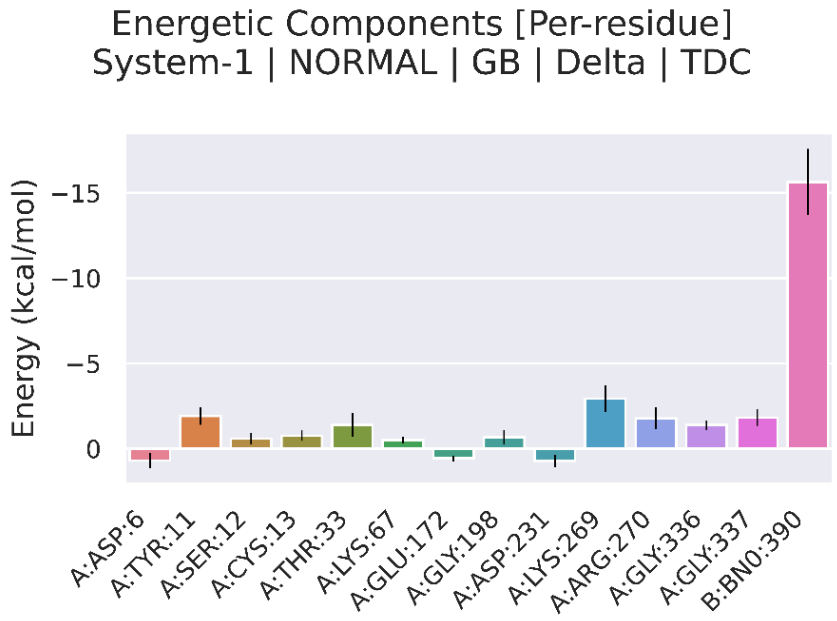


Figure 72: Residue energetic contributions for BN_03 MMGBSA

Key residues in ADP binding covering the phosphate group Tyr11, Lys269, Arg270, and Gly337, were observed for BN_03 with huge energy contributions and considerable stability. Two additional residues with considerable stability and huge energy contributions were observed to be Thr33 and Gly336.

The diversity of these three compound BN_01, BN_02 and BN_03 was examined with PCA based on their physicochemical properties, logP, TPSA, Mw, RotB, HBA and HBD.

Table 11: Physicochemical properties for the three synthesis candidates with SwissADME

Molecule name	**logP**	**TPSA**	**MW g/mol**	**Rotatable Bonds**	**HBD**	**HBA**
BN01	1.54	109.70	349.32	5	2	8
BN02	-1.53	135.20	287.26	3	3	6
BN03	-0.89	121.78	320.10	4	2	8

Table 12: PCA for 3 synthesized compounds

Property	PC1	PC2	PC3
MW	0.44	-0.01	0.88
SLogP	0.41	-0.40	-0.31
HBA	-0.37	0.58	0.08
HBD	-0.39	-0.50	0.13
TPSA	-0.44	0.01	0.21
RotB	0.39	0.50	-0.21
Proportion / %	85.79	14.21	0

PC1, with an 85.79% share, is the leading principal component, capturing the majority of the variance in the dataset. The properties with the greatest impact on PC1 are MW, SlogP, and TPSA. Positive MW and SLogP loadings on PC1 imply a link between higher molecular weight and

lipophilicity. TPSA has a negative loading, it shows it has an inverse connection with PC1. PC2 captures a lesser share of the variance, with a value of 14.21%. The characteristics HBA and RotB have the most influence on it. The positive loading for HBA implies a positive connection with PC2, implying that substances with more hydrogen bond acceptors have higher PC2 values. The positive RotB loading implies that compounds having a greater number of rotatable bonds will have higher PC2 values.

Both BN_01 and BN_02 occupy the same chemical space along the third principal component, indicating these two compounds share a chemical space in terms of the third principal component as shown in Figure 73. This suggests that they may share specific physicochemical attributes and structural traits that contribute to their location along PC3. In contrast to BN_01 and BN_03 along PC3, BN_02 occupies a unique region in the chemical space. It implies that BN_02 has distinct physicochemical and structural qualities that set it apart from the other two chemicals.

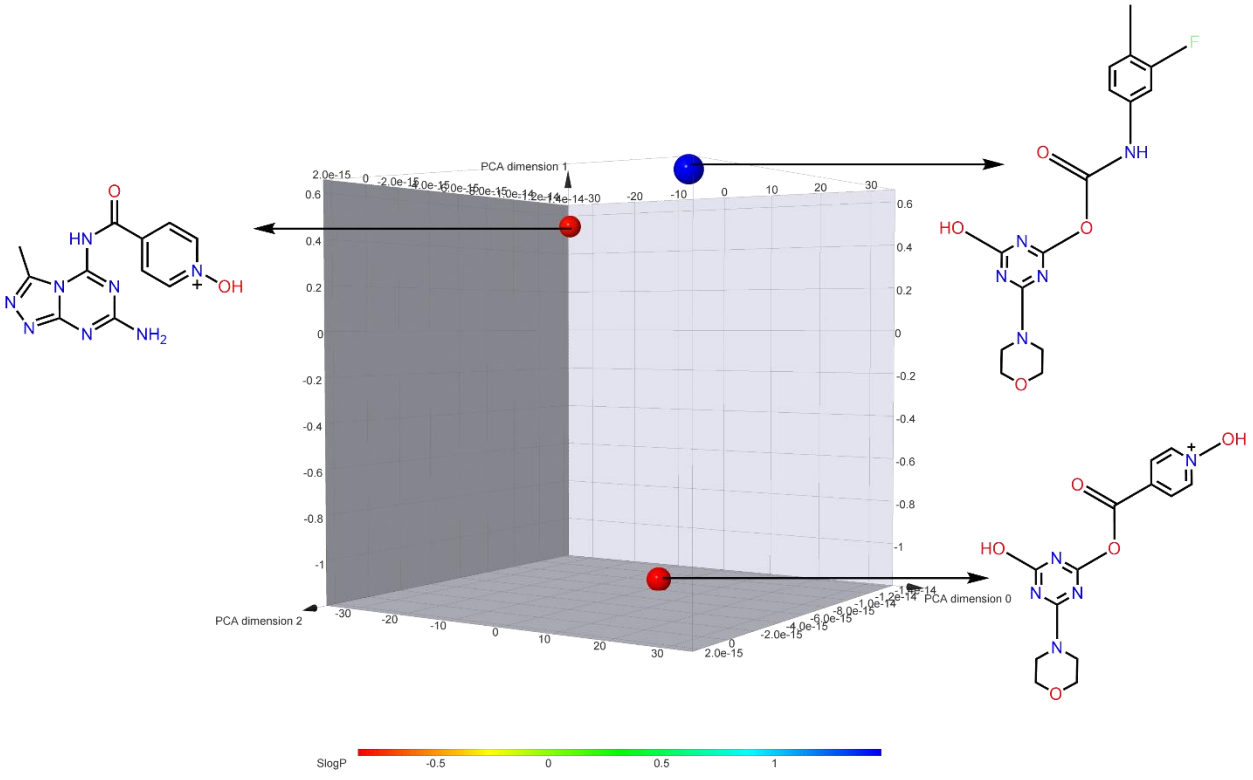


Figure 73: Principal Component Analysis for synthesis candidates.

From the physicochemical profiling with SwissADME shown in Figure 74, a boiled egg analysis was carried out for determining Blood Brain Barrier (BBB) permeability, passive absorption of molecules in the gastrointestinal tract (HIA), and the efflux of the compounds from the central nervous system (CNS) by P-glycoprotein (Daina et al., 2017; Daina & Zoete, 2016). The three compounds were predicted to not cross the BBB, also predicted to be effluated by the P-glycoprotein from the CNS. BN_01 and BN_03 were also predicted to be passively absorbed in the GIT whilst BN_02 is not passively absorbed.

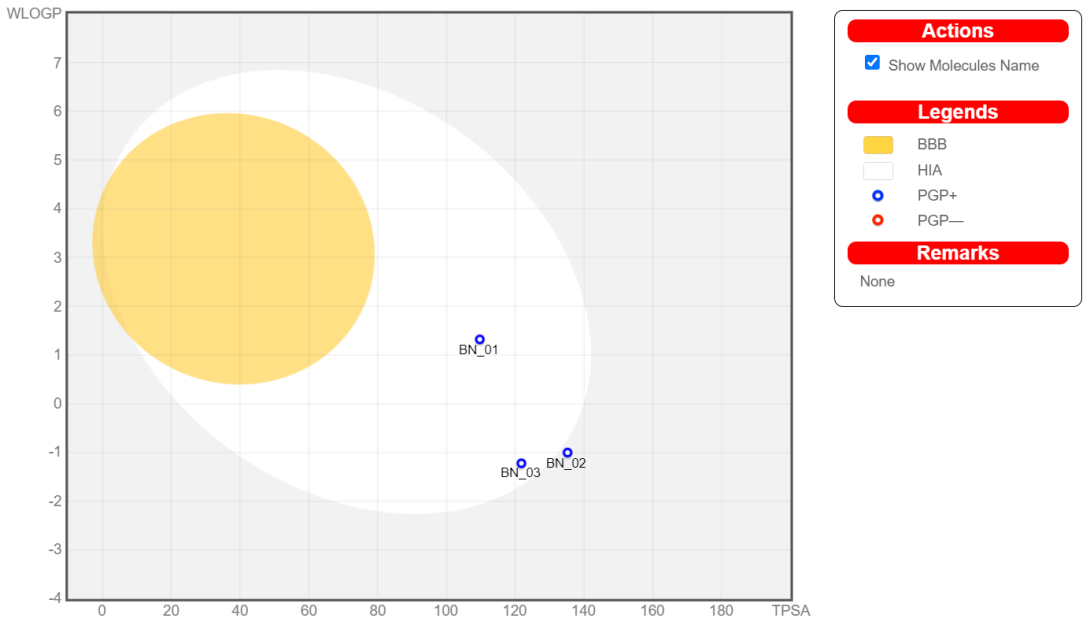


Figure 74: Boiled egg analysis for BN_01, BN_02 and BN_03 (Daina et al., 2017).

Table 13: Melting point, percentage yield and general appearance of the compound.

Compound	Melting Point ^{o}C	% Yield	Appearance
BN_01	280	77	Long needle like transparent crystals
BN_02	220	94.8	Transparent crystals
BN_03	249	96	Fine needle like transparent crystals

CHAPTER 5 DISCUSSION

5.1 Comparative Modeling

Modeling the nucleotide binding domain of *Pf*Hsp70-1, to which ATP/ADP competitive inhibitors shall be developed in, specifically the ATP/ADP binding pocket, was the goal of the comparative modeling step. The nucleotide binding domain of *Pf*Hsp70-1 was successfully modelled, underscoring the accuracy and precision attained. The positive assessment criteria, which show strong structural quality, give reason to believe that the protein model is reliable. This accomplishment is crucial because it creates a strong basis for other research processes like the design and production of inhibitors. *Pf*Hsp70-1's 3D structure was modelled using the homologous protein *Pf*Hsp70-x, which covered the nucleotide binding domain, and *Escherichia coli* K12 heat shock protein 70, which covered the substrate binding domain. The structure resembled native proteins according to the DOPE Z-score of -10.6 in Figure 22a, which indicates that the structure had a strong resolution and matched structures from X-ray investigations of 3D structures, models with a score of -1.0 and less are thought to mimic proteins in their original state. (Hatherley et al., 2016). It has been previously demonstrated that the DOPE Z-score has more accuracy compared to ModPipe-Comb, DFIRE, and Rosetta (Kryshtafovych et al., 2021; Martí-Renom et al., 2000; M. Shen & Sali, 2006). A Schistosomiasis hydroxycinnamoyl CoA-quinate enzyme model modeled with Modeller9v5 was published recently in 2019 with a DOPE Z-Score of -1.596 (Kaushik & Saini, 2019). A study on the repurposing of FDA-approved drugs for the inhibition of Anterior Gradient 2 homodimer saw the modelling of five models for the protein with the Z-Scores ranging from 0.366 to 2.265 (Ullah et al., 2022). All these publications speak to the range of Z-scores acceptable for modelled proteins, a value obtained for *Pf*Hsp70-1 indicated the high native nature of the protein. It has been reported that the section of the protein that corresponds to the substrate binding domain is an N-terminus covering amino acids at position 1 to position 374 (Misra & Ramachandran, 2009).

The findings of the modeled protein exhibited high local model quality, except for a few amino acids outside the nucleotide binding region that did not quite match the knowledge-based assumptions a negative value is generally considered favorable. However, this had little impact as the key domain for this project was the nucleotide binding domain, where the ADP-competitive inhibitors were to be designed and developed. The local model quality score, often based on knowledge-based energy potentials, evaluated the quality of the model at the local level, focusing

on individual residues and regions. The linker region of the protein was identified as the section that needed the most refinement, with a positive energy. A negative knowledge-based energy score shows that the modeled region or interaction has a lower energy than what is expected on average. It suggests that the modeled region or interaction is energetically favorable and consistent with the patterns observed in the known protein structures used to derive the energy score. The goal of comparative modeling was to obtain protein structures and interactions that mimic the characteristics of native or biologically relevant states. The score suggests that the modeled region or interaction agrees with the energetics observed in natural proteins, increasing the confidence in the accuracy and quality of the model.

The results from the 3D-1D (Figure 24) analysis confirmed that protein-structural clashes for the nucleotide binding domain were not significant to affect the functionality of the protein in the nucleotide binding domain. As a requirement, residues should possess a positive 3D-1D score for sustained protein function and 80% of the residues should possess a score equal to or above 0.2 as an indicator of coherence between the primary sequence and the overall three-dimensional arrangement of the entire protein molecule. For the whole model, 80% of the residues had a score equal to or above 0.2 with all the residues in the nucleotide binding domain possessing positive scores throughout. The negative scores were observed in the region between 375 and 392, which corresponds to the linker fragment that links the nucleotide binding domain and the substrate binding domain. The extent of this can be visualized in the local model quality visualization (Figure 23) with an extended red region for the linker. A positive 3D-1D score indicates good agreement between the model's structure and the expected structural properties based on the sequence, these results suggest the folding pattern and secondary structure elements are consistent with the primary sequence of the protein.

According to Ramachandran, which is still the most widely used method in computational biology to evaluate the validity of models, at least 80% of the residues must be present in the most favored region (Ramachandran & Sasisekharan, 1968). This study saw 92.1% of residues present in the most favored region and a grand total of 99.8% of residues were in the allowed region and 0.2% of residues were in the disallowed region. The residues of concern were Pro419 and Lys555, shown in Figure 75. Proline is known for its unique rigid cyclic structure with distinct conformational constraints therefore, this residue often occupies regions of the Ramachandran plot different from

other amino acids, particularly disallowed regions. The presence of proline in the disallowed region at a non-critical domain of the protein did not warrant further structural refinement. Lysine has a flexible side chain, its presence in the disallowed region most likely indicated conformational strain and side chain orientation inaccuracy, however, its position in the chain is at a non-critical region, and it too did not warrant further structural refinement.

These two residues, with disallowed dihedral angles for the backbone conformation in 3D space, had their location on the protein modelled with the substrate binding domain modeled after *E. coli K12 heat* shock protein 70 which had a 48% sequence similarity to *Pf*Hsp70-1. The problematic regions shown in red (Figure 76) do not affect the ATP activity of the protein, hence the *Pf*Hsp70-1 structure generated through comparative modeling was a good 3D model resembling that of native protein structures.

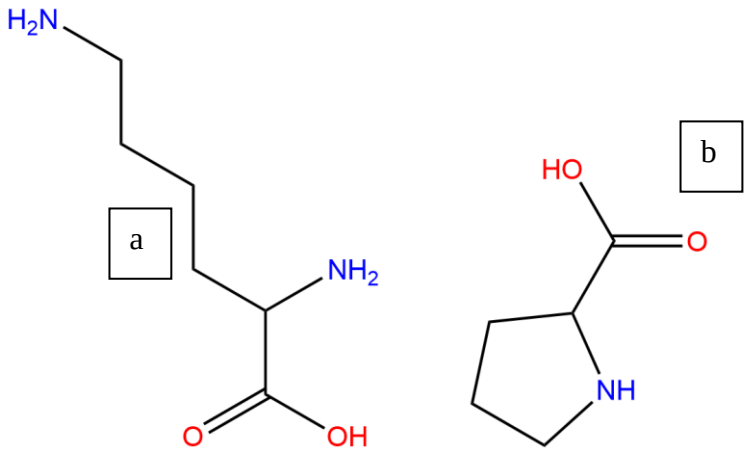


Figure 75: (a) Lysine and (b) Proline

The four metrics, namely DOPE Z-score, 3D-1D score, local model quality (knowledge-based energy), and Ramachandran plot, are complementary and provide different aspects of information about the quality and accuracy of a protein structure. The DOPE Z-score and 3D-1D both assessed the overall agreement between a protein structure and known structural information. The DOPE Z-score indicates a high-quality model with low energy and the 3D-1D score is positive and high, it suggests that the model has a favorable overall quality and accurately represents the expected structural properties based on the sequence. The local model quality, often represented as a knowledge-based energy score, assesses the quality of specific regions or interactions within the

protein structure. It evaluates how well these regions conform to the energetics observed in known protein structures. The local model quality score is mostly negative for the important region, the NBD, with favorable energy. It reinforces the notion that the NBD has been accurately modelled and is consistent with natural protein structures. The Ramachandran plot provides information about the backbone conformation of amino acid residues in the protein structure, 99.8% of the residues fell within the allowed regions of the plot, thus, it evaluates the agreement between the observed backbone dihedral angles and the expected conformational space

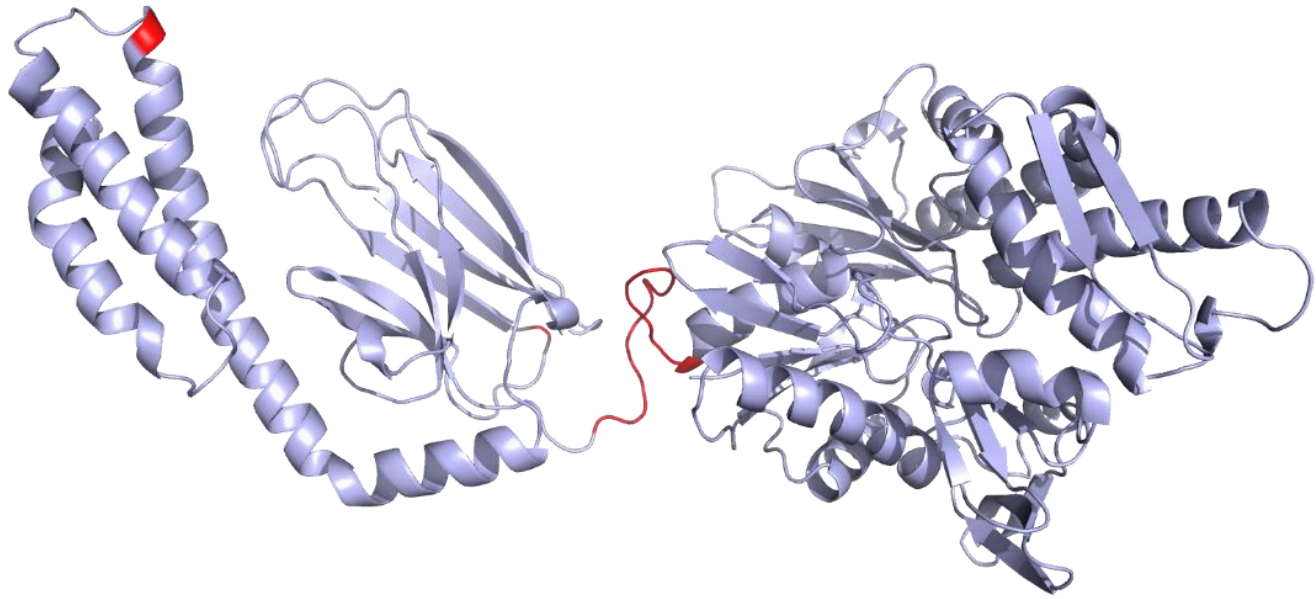


Figure 76: *Pf*Hsp70-1 with unsatisfied residues in red

5.2 *De Novo* Design

The goal of this objective was to design 1,3,5-triazine-2-amine derivatives as potential anti-malarial compounds targeting *Pf*Hsp70-1. To achieve this objective, 71 compounds were initially designed. Subsequently, through a rigorous selection process, 50 compounds were chosen for further analysis.

The diversity of the designed compounds was ensured by compound diversity in the seed library, to which, when analyzed in two dimensions, the 2D Tanimoto distance matrix scatter plot based on the Morgan fingerprints for the seed compounds had an R^2 value of 0.0004 indicating the high

dissimilarity of the 1,3,5-triazine compounds. A multivariate analysis of 6 physicochemical properties based on the principal component analysis used to explore the distribution and scattering of the seed library in 3D space focused on six physicochemical properties MW; LogP; HBD and HBA; TPSA; and RotB, which are key parameters in drug design (Shultz, 2019). They directly influence the drug's behavior and interactions within the biological system. SlogP, lipophilicity is a critical property because it determines the drug's ability to cross biological barriers, such as cell membranes, and reach its target site. Optimal lipophilicity ensures sufficient drug uptake, distribution, and penetration into tissues. The polar surface area of a drug molecule provides insights into its water solubility and permeability. Optimal TPSA ensures a balance between hydrophilicity and lipophilicity, allowing the drug to efficiently navigate biological barriers and interact with its target. Molecular weight influences a drug's physicochemical properties, including solubility, bioavailability, and metabolic stability. It is important to optimize the molecular weight to maintain a balance between efficient drug delivery, target interaction, and elimination from the body. HBD and HBA relate to a drug's ability to form hydrogen bonds, which are essential for molecular recognition and binding to target proteins. The presence of suitable donors and acceptors facilitates specific interactions with the target and enhances the drug's affinity and potency. number of rotatable bonds affects a drug's conformational flexibility and shape. Too many rotatable bonds can lead to excessive conformational freedom, making it challenging to specifically interact with the target. Optimization of RotB helps in improving drug-like properties and reducing the complexity of compound libraries (Caron et al., 2021; Stegemann et al., 2023). The compounds designed in the study averaged acceptable rule of five values for physicochemical properties as shown in Table 6. The two values of concern were LogP at -0.97 and TPSA at 143.50 A^2, these speak into the hydrophilic nature of these compounds, that might present challenges with membrane permeability. Compounds that pass through the membrane lipid bilayer actively are lipophilic, hence these designed compounds with hydrophilic nature would require active transport mechanisms, such as transport proteins or channels (Seddon et al., 2009; Venable et al., 2019).

This PCA method reveals a clear scattering of all compounds in 3D space without specific clustering for any group of compounds. This also indicates the high diversity present within the seed library. The most important physicochemical properties to observe for growing fragments in drug discovery are molecular weight and logP which are supposed to be kept below 300 g/mol and 3 respectively. Analysis of the averages for each of the six properties reveals that, with the

exception of the hydrogen bond acceptor count, all properties were within the limits defined by the rule of 3. According to this rule, the molecular weight of a fragment should be less than 300 Da, the cLogP should be equal to or less than 3, and the number of hydrogen bond donors and acceptors should not exceed 3. The property that exceeded this rule was the number of hydrogen bond acceptors, with an average count of 7. This was an expected outcome for the compounds as the curation of the seed library focused on the functional groups required for Click Chemistry reactions and Robust Reactions supplied by Durrant and Hartenfeller (Durrant & McCammon, 2012; Hartenfeller et al., 2011). These reactions are highly modular with low sensitivity to water, oxygen and solvents hence, and are ideal for building a library of synthetically accessibility compounds *insilco* (Hein et al., 2008; Hou et al., 2012; Smedley et al., 2022). One-Pot synthesis techniques with simple post-synthesis cleanups were the main focus of this project, aiming at simplified synthesis procedures adaptable to constrained laboratory resources. Seventy-one compounds were designed from this step to create a library of 1,3,5-triazine-2-amine derivatives which were further analyzed for drug-like properties.

The first analysis performed was the Tanimoto Distance Matrix analysis based on the Morgan fingerprints. This analysis measures the structural similarity between compounds by comparing their fingerprints. It plays a crucial role in assessing the chemical diversity and similarity within the compound library. By analyzing the Tanimoto distances, we gain insights into the structural variations and uniqueness of the designed compounds. This test helps in identifying diverse compounds with distinct structural features, which is essential for finding potential lead compounds. The result of the diversity in the seed library yielded a highly diverse library with a 2D Tanimoto distance matrix scatter plot based on the Morgan fingerprints R^2 of 0.0001 indicating a lack of correlation in the chemical space of the compounds (Ying et al., 2021).

Next, the analysis of physicochemical properties was conducted. This test involved evaluating properties such as MW, TPSA, SlogP, RotB, HBA, and HBD. These properties provide valuable information about the drug-like nature and pharmacokinetic profile of the compounds. Assessing these physicochemical properties helps in identifying compounds that adhere to Lipinski's Rule of Five, which predicts favorable drug-like characteristics (Leeson & Empfield, 2010; Meanwell, 2011). The significance of this analysis lies in its ability to prioritize compounds with desirable

physicochemical properties for further development. The output from this analysis was used in the principal component analysis of the physicochemical properties of the designed compounds.

PCA is a statistical technique that helps in exploring the chemical space occupied by the compounds and identifying patterns within the dataset. By examining the distribution of compounds along the principal components, we gain insights into the diversity and clustering of the compound library. This analysis aids in identifying structurally distinct compounds and assessing their potential as lead compounds. The PCA analysis contributes significantly to understanding the chemical space and diversity of the designed compounds, this was the complementary test to the Tanimoto Distance Matrix analysis, linking structural features to the physicochemical properties observed within the dataset (Akella & DeCaprio, 2010; O' Connor et al., 2012; Xue et al., 1999). This was significant, given the limited diversity of the scaffolds of the designed compound which was 1,3,5-triazine-2-amine, however the results yielded diversity within the designed compounds. Except for the top two ranking compounds, which could be distinguished by the iminol and nitrile functional groups, this analysis showed a scattering of compounds along the three dimensions in the absence of any specific clustering for a particular class of substances. Averaging all six physicochemical properties for the generated library revealed the general average of the library does not violate the Lipinski rule of five which stipulates molecular weight of a compound to be <500 Da, the cLogP ≤5, the number of HBD ≤5 and the number of HBA ≤10. Of note, molecular weight and the partition coefficient average for the library were 256.56 g/mol and -0.97, which puts them within range for good lead molecules for structural optimization and QSAR studies which recommends molecular weight to be below 300 g/mol and the partition coefficient to also be below 3 (Lipinski, 2004).

From the 71 compounds designed from the seed library, 21 compounds failed the Lipinski, PAINS and NIH filters due to hydrogen bond donors and acceptors violations for Lipinski, azo functional groups for PAINS and Azide, hydrazine, carbodiimide, isocyanate, gte4_basic_N and isothiocyanate for NIH filters.

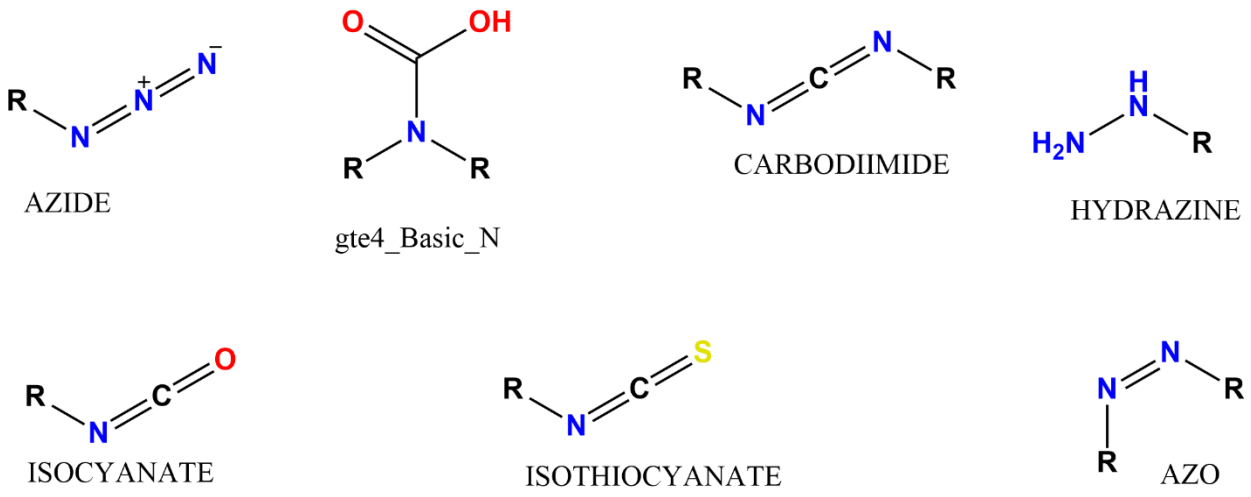


Figure 77: Structural Alerts

None of the compounds in the library violated Lipinski's molecular weight or logP values despite including these undesired functional groups; as a result, the library is quite diverse and has a lot of room for optimization.

Last, the SwissADME analysis for Blood-Brain Barrier (BBB) permeability was performed via the boiled egg experiment (Daina et al., 2017). The BBB is a physiological barrier that limits the passage of substances from the bloodstream into the brain. Assessing the BBB permeability of the compounds is crucial, as it provides insights into their ability to cross this barrier and reach the intended target in the central nervous system. This analysis helps in evaluating the potential of the compounds to act on malaria parasites within the brain. By considering BBB permeability, we gain insights into the suitability of the compounds as anti-malarial drugs, none of the 50 compounds were predicted to penetrate the blood brain barrier, which in an important character required for anti-malarial drugs, most malaria patients do not present with cerebral malaria. It has been reported that around 1% of children infected with *P. falciparum* progress to cerebral malaria (Luzolo & Ngoyi, 2019b)

Because compounds were developed within the substrate binding domain active site, this implied a compound must outperform ADP, the natural substrate, and have a similar binding pose to ADP in order to be a competitive inhibitor (Botha et al., 2011; Persky et al., 2020). The binding pose of ADP becomes important in choosing molecules for optimization. The binding pose of ADP can be split into three sections, focusing on the residues and their interacting atoms on the ADP. ADP is

composed of a purine moiety, a ribose moiety and a phosphate group. Clustering the interacting residues according to ADP molecule sections resulted in three sets of interacting residues: purine anchors, ribose anchors, and phosphate anchors. Asp6, Ala368, Val367 and Asp364 are the anchor residues for the purine moiety, Gly199 is the ribose anchor residue in the active site. The phosphate anchors are Tyr11, Lys269, Arg270, Gly337 and Arg340. These residues are critical in understanding the relationship between a prospective inhibitor's binding posture and the possibility of inhibition of the protein, as well as its competitiveness.

The series of tests conducted in this study played a vital role in addressing the research objective of designing 1,3,5-triazine-2-amine derivatives as potential anti-malarial compounds. The Tanimoto Distance Matrix analysis and PCA were complementary approaches that provided valuable insights into the chemical space occupied by the compounds. The Tanimoto analysis assessed the structural similarity and diversity, while PCA explored the distribution and clustering of compounds based on their physicochemical properties. The integration of the physicochemical property analysis within PCA allowed for a comprehensive understanding of the relationship between chemical structure and properties. This information proved useful for identifying lead-like compounds and understanding their potential as anti-malarial agents.

Evaluating physicochemical properties played a crucial role in the PCA analysis, as it provided a quantitative basis for assessing the chemical space and identifying structurally distinct compounds. By considering properties such as molecular weight, polar surface area, partition coefficient, and hydrogen bonding capacity, the PCA analysis gained deeper insights into the diversity and clustering of the compound library. This integration of physicochemical property analysis with PCA enhanced the selection and prioritization of compounds with desirable drug-like characteristics.

Additionally, the application of the Lipinski, NIH, and PAINS filtering provided an additional layer of selection, ensuring that compounds adhered to Lipinski's Rule of Five and are void of PAINS, which predicts favorable drug-like properties. This filtering step eliminated compounds with unfavorable characteristics and focused the research on those with higher potential for success in subsequent stages of drug discovery.

Furthermore, the Swiss ADME analysis for blood-brain barrier (BBB) permeability assessment was crucial in evaluating the compounds' ability to cross the BBB and reach their target in the

central nervous system. This analysis considered the compounds' physicochemical properties related to BBB penetration, providing valuable information on their suitability as anti-malarial drugs that can act on malaria parasites within the brain.

By integrating Tanimoto analysis, PCA, physicochemical property analysis, Lipinski, NIH and PANS filtering, and Swiss ADME assessment, this research identified 50 promising 1,3,5-triazine-2-amine derivatives for further development as effective anti-malarial drugs. These analytical tools collectively provided comprehensive insights into the chemical space, diversity, drug-likeness, and potential BBB permeability of the compounds. Such insights were essential for making informed decisions and advancing the discovery of target-specific anti-malarial compounds, specifically BN_01, BN_02, and BN_03.

5.3 Molecular Dynamics and Synthesis

The synthesis of the designed compounds was a crucial step in the research process, aiming to validate their potential as anti-malarial agents. Several steps were undertaken to evaluate the stability, binding interactions, and overall behavior of the compounds within the protein target.

Initially, binding pose analysis was performed to identify compounds that interacted with critical residues involved in binding. By comparing the binding poses of the compounds with the reference ADP binding pose (purine anchors, ribose anchors, and phosphate anchors) two compounds with favorable binding interactions and high VINA binding affinity score were selected for further analysis. This step ensured that the synthesized compounds could engage with the target protein in a manner similar to known ligands.

The top two compounds (BN_01 and BN_02) selected from the binding pose analysis including a third compound (BN_03) from the recombination of BN_01 and BN_02 were subjected to protein-ligand molecular dynamics simulations for a length of 100ns. These simulations provided insights into the stability and conformational dynamics of the protein-ligand complex. Various parameters were evaluated, including RMSD (Root Mean Square Deviation), which quantifies the deviation of the complex structure from the reference structure. The RMS fluctuation analysis assessed the local fluctuations and flexibility of residues within the protein. Additionally, the Mean Square Displacement (MSD) of the ligand from the protein was calculated, providing information on the diffusive behavior and mobility of the ligand within the complex. The RMSD histogram analysis further revealed the distribution of structural deviations, allowing the evaluation of conformational

variability of the system. A trend for the three-synthesis candidate compounds; BN_01, BN_02, and BN_03, according to MSD, shows BN_03 has the highest immobility, followed by BN_01 and the least immobilized compound as BN_02. This trend is observed in the average Gibbs free energies which were -30 kcal/mol, -22 kcal/mol, and -15 kcal/mol for BN_03, BN_01, and BN_02 respectively, this clearly shows the immobilized compounds are tightly bound via hydrogen bonding also observed with MMGBSA. All the three compounds had stable complexes with their protein counterparts, the trend observed for MSD and Gibbs free energy also extends to protein-ligand rmsd which was below 0.4 nm, 0.6 nm, and 1.0 nm for BN_03, BN_01 and BN_02 respectively, indicating ligand stability within the protein plays a huge role in the binding energies observed.

The RMSD histogram for BN_03 has a single peak, this suggests that the protein-ligand complex has a well-defined, stable conformation and binding pose that it frequently visits during the simulation. This stable state represents a favorable binding mode with strong interactions between the ligand and the protein. The absence of additional peaks in the histogram indicates that the system is not undergoing significant large-scale conformational changes or transitions between distinct states. The ligand is confined to a specific conformational space around the equilibrium structure, all this crucial in understanding the high binding affinity of BN_03 among the three compounds.

BN_01 and BN_02 show the presence of two merged peaks indicating that the protein-ligand complex is exploring two distinct conformations. These different states represent alternative stable conformational sub-states of the system, the system is transitioning between these two states relatively frequently. BN_02 had a second blunt peak with a small plateau indicating the presence of a second conformational state that is less stable and less frequently sampled by the system. The plateau suggests that the complex spends some time transitioning between the two states before reaching the more stable state. For both BN_01 and BN_02 the two peaks might correspond to different ligand conformational changes in the protein that are induced by the ligand as discussed earlier from observation of the ligand RMSD analysis.

For RMSD of BN_03, the observation of a conformational change in the ligand and the protein-ligand complex at around 40 ns indicates that the ligand itself may have experienced a transition or rearrangement within the binding site. The protein-ligand complex's altered conformation raises

the possibility that the interaction between the ligand and protein may have impacted the protein's conformation. The protein-ligand change likely speaks to the flexibility and rearrangements of side chains upon induced fit binding of BN_03. Furthermore, the significant RMSD change in the backbone and protein-ligand complex at around 50 ns indicates that the protein has undergone a notable conformational transition or flexibility change. This could imply that the protein experienced a conformational rearrangement or a transition to a different stable state during the simulation. The backbone RMSD change specifically suggests that certain regions or secondary structure elements of the protein may have shifted or undergone structural adjustments, all this has been alluded to in the RMSD histogram observation.

For BN_01 the RMSD change at around 18 ns for the protein-ligand and ligand indicates that both the protein-ligand complex and the ligand itself underwent a conformational transition or rearrangement suggesting that the ligand might have explored different binding poses or conformational states within the binding site, potentially sampling different interactions with the protein. The significant RMSD change in the backbone at around 30 ns suggests that the protein experienced a flexibility shift at this time point. This change in the protein backbone structure may be influenced by ligand binding. The RMSD changes at around 40 ns for both the protein-ligand complex and the backbone indicate that the protein's conformation and its interaction with the ligand are undergoing significant changes. This time point might represent another important transition or rearrangement event for the protein-ligand complex, potentially associated with the ligand's stabilization or reorientation within the binding site. The RMSD changes in the protein backbone at 30 ns and the protein-ligand complex at 40 ns suggest that the protein exhibits conformational flexibility, adapting to changes induced by ligand binding.

For BN_02 the RMSD change at around 25 ns for the protein-ligand complex and the backbone suggests that both the protein-ligand complex and the protein backbone undergo structural adjustments. This might indicate initial relaxation or equilibration of the protein-ligand complex as it settles into a stable binding mode. The RMSD change for the ligand at around 20 ns, accompanied by a minor RMSD change for the backbone, suggests that the ligand itself experienced a conformational rearrangement while potentially influencing the protein's conformation. This ligand-induced effect on the backbone is indicative of induced-fit binding. The significant RMSD changes in the backbone and the protein-ligand complex at 35 ns and 45 ns

suggest additional conformational transitions. These time points might represent subsequent conformational rearrangements or structural adaptations in response to ligand. The RMSD changes at multiple time points (20 ns, 25 ns, 35 ns, and 45 ns) indicate that the protein-ligand complex samples multiple conformational states during the simulation. These states might correspond to different ligand binding modes, these can be corroborated in the MSD for BN_02, which had the highest mobility of the three compounds.

The molecular dynamics study indicates a trend for the three synthesized compounds. BN_03 was tightly held in the active site, this is further corroborated by a much higher binding affinity from MMGBSA.

A study to investigate the inhibition potential of dihydroorate dehydrogenase in small cell lung cancer reported MSD values below 0.2 as speaking to the stability of the complex and retention of the ligand within the active site (Zhi et al., 2021). Overall, the three compounds were sufficiently stable with enough confidence on their immobility. This objective was a proof of concept for the feasibility of synthesis of the designed compounds which lays a foundation for the future work of this project. Each compound underwent multiple synthesis runs in triplicate, and the melting point tests showed consistent results, indicating purity of synthesized compounds which were obtained via slow crystallization processes. This achievement marks an exciting stepping stone for future work, signaling the feasibility and reproducibility of the compound synthesis. It sets the stage for further exploration and experimentation, such as additional biological assays, structure-activity relationship studies, and potential preclinical testing.

5.4 Conclusion

This work has significantly contributed to the field of anti-malarial drug discovery and design cutting across through subfields of computational biology, molecular dynamics and chemical synthesis. The research outputs in this study include:

First, a 3D homology structure of *Pf*Hsp70-1was successfully generated. This structural model provides a valuable tool for understanding the protein's architecture and for further studies on its function and interaction with potential inhibitors. The 3D structure serves as a foundation for rational drug design and facilitates structure-based drug discovery approaches.

Second, a set of fifty lead-like 1,3,5-triazine-2-amine derivatives were designed specifically for the inhibition of *Pf*Hsp70-1. These compounds were strategically designed using computational structure-based methods to target the nucleotide binding domain, a critical region of the protein involved in its function as an ATPase in protein folding. The design process considered drug-likeness criteria, physicochemical properties, and structural considerations, resulting in a diverse and promising library of compounds with potential anti-malarial activity. The compounds can be described as lead-like owing to their physicochemical properties of molecular weight and logP below 300 Da and 3 respectively, providing ample room for optimization to lead candidates.

Last, three compounds, namely BN_01, BN_02, and BN_03 shown in Figure 78, Figure 79, and Figure 80 respectively, were synthesized with successful isolation of crystals, which speaks to the validation of the design process aiming at modular structures through click chemistry and robust reactions. These compounds represent the culmination of the synthetic efforts and demonstrate the practical feasibility of producing the designed lead-like derivatives. The synthesis of these compounds validates the synthetic routes and provides tangible evidence of the potential for further optimization and development of novel anti-malarial drugs targeting *Pf*Hsp70-1.

Collectively, these research outputs lay a solid foundation for the discovery and development of new anti-malarial agents. The 3D homology structure of *Pf*Hsp70-1 provides insights into its molecular architecture, while the designed 1,3,5-triazine-2-amine derivatives offer a diverse set of compounds with potential inhibitory activity. The successful synthesis of three compounds confirms the feasibility of the design and paves the way for future optimization and evaluation. Overall, this dissertation contributes to the ongoing efforts to combat malaria and underscores the importance of structure-based design in the development of effective anti-malarial drugs.

5.5 Recommendations

In the future, leveraging the robust 3D structure of the *Pf*Hsp70-1, along with the availability of 50 synthetically accessible 1,3,5-triazine-2-amine derivatives and the successful synthesis of three compounds obtained as crystalline compound, an extensive series of investigations must be carried out. These include comprehensive characterizations of synthesized compounds using techniques like carbon and proton NMR, and high-resolution mass spectrometry to gain detailed insights into the nature and structure synthesized compounds. Additionally, enzymatic biological assays will have to be performed, encompassing a range of techniques such as Surface Plasmon Resonance

(SPR), Malate Dehydrogenase (MDH), and fluorescence assays. These assays are capable of determining the association and dissociation constants of the compounds to *Pf*Hsp70-1, the abrogation of the protein's function upon binding of the compounds, and assess enzyme activity respectively. Furthermore, the potential for whole cell assays has to be explored in providing a more comprehensive understanding of the compounds' anti-malarial activity on non-resistant strains and resistant strains to currently administered antimalarial compounds that are artemisinin and aryl amino alcohol based.

Moving forward, the assessment will also encompass absorption, distribution, metabolism, excretion, and toxicity tests, which will shed light on the compounds' pharmacokinetic, pharmacodynamic, and toxicity profiles. By undertaking this systematic approach, the research has a potential bridge the gap between molecular design and potential anti-malarial drug candidates, paving the way for the development of novel therapeutic interventions in the fight against malaria.

6.0 REFERENCES

Acharya, C., Coop, A., Polli, J. E., & Mackerell, A. D. J. (2011). Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. *Current Computer-Aided Drug Design*, *7*(1), 10–22. https://doi.org/10.2174/157340911793743547

Acharya, P., Kumar, R., & Tatu, U. (2007). Chaperoning a cellular upheaval in malaria: Heat shock proteins in Plasmodium falciparum. *Molecular and Biochemical Parasitology*, *153*(2), 85–94. https://doi.org/10.1016/j.molbiopara.2007.01.009

Acun, B., Hardy, D. J., Kale, L. V., Li, K., Phillips, J. C., & Stone, J. E. (2018). Scalable molecular dynamics with NAMD on the Summit system. *IBM Journal of Research and Development*, *62*(6). https://doi.org/10.1147/JRD.2018.2888986

Agamah, F. E., Mazandu, G. K., Hassan, R., Bope, C. D., Thomford, N. E., Ghansah, A., & Chimusa, E. R. (2019). Computational/in silico methods in drug target and lead prediction. *Briefings in Bioinformatics*, *00*(May), 1–13. https://doi.org/10.1093/bib/bbz103

Al Khaja, K. A. J., & Sequeira, R. P. (2021). Drug treatment and prevention of malaria in pregnancy: a critical review of the guidelines. *Malaria Journal*, *20*(1), 1–13. https://doi.org/10.1186/s12936-020-03565-2

Alam, A., Goyal, M., Iqbal, M. S., Pal, C., Dey, S., Bindu, S., Maity, P., & Bandyopadhyay, U. (2009). Novel antimalarial drug targets: Hope for new antimalarial drugs. In *Expert Review of Clinical Pharmacology* (Vol. 2, Issue 5, pp. 469–489). https://doi.org/10.1586/ecp.09.28

Anwar-Mohamed, A., Barakat, K. H., Bhat, R., Noskov, S. Y., Tyrrell, D. L., Tuszynski, J. A., & Houghton, M. (2014). A human ether-á-go-go-related (hERG) ion channel atomistic model generated by long supercomputer molecular dynamics simulations and its use in predicting drug cardiotoxicity. *Toxicology Letters*, *230*(3), 382–392. https://doi.org/10.1016/J.TOXLET.2014.08.007

Arabi, A. A. (2017). Routes to drug design via bioisosterism of carboxyl and sulfonamide groups. *Future Medicinal Chemistry*, *9*(18), 2167–2180. https://doi.org/10.4155/fmc-2017-0136

Ashley, E. A., Pyae Phyo, A., & Woodrow, C. J. (2018). Malaria. *The Lancet*, *391*(10130), 1608–1621. https://doi.org/10.1016/S0140-6736(18)30324-6

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*(7), 2719–2740. https://doi.org/10.1021/jm901137j

Bahar, A. A., Liu, Z., Garafalo, M., Kallenbach, N., & Ren, D. (2015). Controlling Persister and Biofilm Cells of Gram-Negative Bacteria with a New 1,3,5-Triazine Derivative. *Pharmaceuticals 2015, Vol. 8, Pages 696-710*, *8*(4), 696–710. https://doi.org/10.3390/PH8040696

Bain, A. D., Hughes, D. W., Anand, C. K., Nie, Z., & Robertson, V. J. (2010). Problems, artifacts and solutions in the INADEQUATE NMR experiment. *Magnetic Resonance in Chemistry*, *48*(8), 630–641. https://doi.org/10.1002/MRC.2639

Banumathy, G., Singh, V., Pavithra, S. R., & Tatu, U. (2003). Heat shock protein 90 function is essential for Plasmodium falciparum growth in human erythrocytes. *Journal of Biological Chemistry*, *278*(20), 18336–18345. https://doi.org/10.1074/jbc.M211309200

Barakat, K., & Tuszynski, J. (2011). Relaxed complex scheme suggests novel inhibitors for the lyase activity of DNA polymerase beta. *Journal of Molecular Graphics and Modelling*, *29*(5), 702–716. https://doi.org/10.1016/J.JMGM.2010.12.003

Barrett, M. P., Kyle, D. E., Sibley, L. D., Radke, J. B., & Tarleton, R. L. (2019). Protozoan persister-like cells and drug treatment failure. *Nature Reviews Microbiology*, *17*(10), 607–620. https://doi.org/10.1038/s41579-019-0238-x

Beck-Johnson, L. M., Nelson, W. A., Paaijmans, K. P., Read, A. F., Thomas, M. B., & Bjørnstad, O. N. (2013). The effect of temperature on Anopheles mosquito population dynamics and the potential for malaria transmission. *PLoS ONE*, *8*(11), e79276. https://doi.org/10.1371/journal.pone.0079276

Beckstein, O., & Sansom, M. S. P. (2005). A Hydrophobic Gate in an Ion Channel: The Closed State of the Nicotinic Acetylcholine Receptor. *Physical Biology*, *3*(2), 147–159. https://doi.org/10.1088/1478-3975/3/2/007

Beeson, J. G., Drew, D. R., Boyle, M. J., Feng, G., Fowkes, F. J. I., & Richards, J. S. (2016). Merozoite surface proteins in red blood cell invasion, immunity and vaccines against malaria. *FEMS Microbiology Reviews*, *40*(3), 343–372. https://doi.org/10.1093/femsre/fuw001

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977). The protein data bank: A computer-based archival file for macromolecular structures. *Journal of Molecular Biology*, *112*(3), 535–542. https://doi.org/10.1016/S0022-2836(77)80200-3

Bertalan, É., Lešnik, S., Bren, U., & Bondar, A. N. (2020). Protein-water hydrogen-bond networks of G protein-coupled receptors: Graph-based analyses of static structures and molecular dynamics. *Journal of Structural Biology*, *212*(3), 107634. https://doi.org/10.1016/J.JSB.2020.107634

Berthold, M. R., Cebron, N., Dill, F., Di Fatta, G., Gabriel, T. R., Georg, F., Meinl, T., Ohl, P., Sieb, C., & Wiswedel, B. (2006). KNIME: The konstanz information miner. *4th International Industrial Simulation Conference 2006, ISC 2006*, *11*(1), 58–61. https://doi.org/10.1145/1656274.1656280

Bitencourt-Ferreira, G., & de Azevedo, W. F. (2018). Development of a machine-learning model to predict Gibbs free energy of binding for protein-ligand complexes. *Biophysical Chemistry*, *240*, 63–69. https://doi.org/10.1016/J.BPC.2018.05.010

Blanshard, A., & Hine, P. (2021). Atovaquone-proguanil for treating uncomplicated Plasmodium falciparum malaria. *Cochrane Database of Systematic Reviews*, *2021*(1). https://doi.org/10.1002/14651858.CD004529.PUB3/MEDIA/CDSR/CD004529/IMAGE_N/NCD004529-CMP-017.15.SVG

Bon, M., Bilsland, A., Bower, J., & McAulay, K. (2022). Fragment-based drug discovery—the importance of high-quality molecule libraries. *Molecular Oncology*, *16*(21), 3761–3777. https://doi.org/10.1002/1878-0261.13277

Boratyn, G. M., Thierry-Mieg, J., Thierry-Mieg, D., Busby, B., & Madden, T. L. (2019). Magic-BLAST, an accurate RNA-seq aligner for long and short reads. *BMC Bioinformatics*, *20*(1), 1–19. https://doi.org/10.1186/S12859-019-2996-X/FIGURES/11

Borhani, D. W., & Shaw, D. E. (2011). The future of molecular dynamics simulations in drug discovery. *Journal of Computer-Aided Molecular Design 2011 26:1*, *26*(1), 15–26. https://doi.org/10.1007/S10822-011-9517-Y

Borrel, A., Regad, L., Xhaard, H., Petitjean, M., & Camproux, A. C. (2015). PockDrug: A model for predicting pocket druggability that overcomes pocket estimation uncertainties. *Journal of Chemical Information and Modeling*, *55*(4), 882–895. https://doi.org/10.1021/CI5006004/SUPPL_FILE/CI5006004_SI_009.XLS

Bosman, A., & Mendis, K. N. (2007). A major transition in malaria treatment: The adoption and deployment of artemisinin-based combination therapies. *American Journal of Tropical Medicine and Hygiene*, *77*(SUPPL. 6), 193–197. https://doi.org/10.4269/ajtmh.2007.77.193

Botha, M., Pesce, E. R., & Blatch, G. L. (2007). The Hsp40 proteins of Plasmodium falciparum and other apicomplexa: Regulating chaperone power in the parasite and the host. *International Journal of Biochemistry and Cell Biology*, *39*(10), 1781–1803. https://doi.org/10.1016/j.biocel.2007.02.011

Bowie, J. U., Lüthy, R., & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science (New York, N.Y.)*, *253*(5016), 164–170. https://doi.org/10.1126/SCIENCE.1853201

Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., … Karplus, M. (2009). CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry*, *30*(10), 1545–1614. https://doi.org/10.1002/JCC.21287

Brown, S. P., & Muchmore, S. W. (2007). Rapid estimation of relative protein-ligand binding affinities using a high-throughput version of MM-PBSA. *Journal of Chemical Information*

and Modeling, *47*(4), 1493–1503. https://doi.org/10.1021/CI700041J/ASSET/IMAGES/MEDIUM/CI700041JN00001.GIF

Brown, S. P., & Muchmore, S. W. (2009). Large-scale application of high-throughput molecular mechanics with Poisson-Boltzmann surface area for routine physics-based scoring of protein-ligand complexes. *Journal of Medicinal Chemistry*, *52*(10), 3159–3165. https://doi.org/10.1021/JM801444X/SUPPL_FILE/JM801444X_SI_007.PDB

Burg, J. S., Ingram, J. R., Venkatakrishnan, A. J., Jude, K. M., Dukkipati, A., Feinberg, E. N., Angelini, A., Waghray, D., Dror, R. O., Ploegh, H. L., & Garcia, K. C. (2015). Structural basis for chemokine recognition and activation of a viral G protein-coupled receptor. *Science*, *347*(6226), 1113–1117. https://doi.org/10.1126/SCIENCE.AAA5026/SUPPL_FILE/BURG.SM.PDF

Cascioferro, S., Parrino, B., Spanò, V., Carbone, A., Montalbano, A., Barraja, P., Diana, P., & Cirrincione, G. (2017). 1,3,5-Triazines: A promising scaffold for anticancer drugs development. *European Journal of Medicinal Chemistry*, *142*, 523–549. https://doi.org/10.1016/j.ejmech.2017.09.035

Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., & Woods, R. J. (2005). The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, *26*(16), 1668–1688. https://doi.org/10.1002/JCC.20290

CDC. (2018). *Malaria: Lifecycle*. Centers for Disease Control and Prevention. https://www.cdc.gov/malaria/about/biology/index.html

Chakravarty, D., McElfresh, G. W., Kundrotas, P. J., & Vakser, I. A. (2020). How to choose templates for modeling of protein complexes: Insights from benchmarking template-based docking. *Proteins: Structure, Function, and Bioinformatics*, *88*(8), 1070–1081. https://doi.org/10.1002/PROT.25875

Chen, I. J., & Foloppe, N. (2010). Drug-like bioactive structures and conformational coverage with the ligprep/confgen suite: Comparison to programs MOE and catalyst. *Journal of Chemical*

Information and Modeling, *50*(5), 822–839. https://doi.org/10.1021/CI100026X/SUPPL_FILE/CI100026X_SI_001.PDF

Chen, Y., & Shoichet, B. K. (2009). Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nature Chemical Biology*, *5*(5), 358–364. https://doi.org/10.1038/nchembio.155

Clore, G. M., & Gronenborn, A. M. (1998). Determining the structures of large proteins and protein complexes by NMR. *Trends in Biotechnology*, *16*(1), 22–34. https://doi.org/10.1016/S0167-7799(97)01135-9

Cohen, P., & Alessi, D. R. (2013). Kinase drug discovery - What's next in the field? *ACS Chemical Biology*, *8*(1), 96–104. https://doi.org/10.1021/CB300610S/ASSET/IMAGES/MEDIUM/CB-2012-00610S_0006.GIF

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science : A Publication of the Protein Society*, *2*(9), 1511–1519. https://doi.org/10.1002/PRO.5560020916

Counihan, N. A., Modak, J. K., & de Koning-Ward, T. F. (2021). How Malaria Parasites Acquire Nutrients From Their Host. In *Frontiers in Cell and Developmental Biology* (Vol. 9). Frontiers Media S.A. https://doi.org/10.3389/fcell.2021.649184

Cox, F. E. (2010). History of the discovery of the malaria parasites and their vectors. *Parasites and Vectors*, *3*(1), 1–9. https://doi.org/10.1186/1756-3305-3-5

Cozzetto, D., & Tramontano, A. (2005). Relationship between multiple sequence alignments and quality of protein comparative models. *Proteins: Structure, Function, and Bioinformatics*, *58*(1), 151–157. https://doi.org/10.1002/PROT.20284

Cui, L., & Su, X. Z. (2009a). Discovery, mechanisms of action and combination therapy of artemisinin. *Expert Review of Anti-Infective Therapy*, *7*(8), 999–1013. https://doi.org/10.1586/ERI.09.68

Cui, L., & Su, X. Z. (2009b). Discovery, mechanisms of action and combination therapy of artemisinin. *Expert Review of Anti-Infective Therapy*, *7*(8), 999–1013. https://doi.org/10.1586/ERI.09.68

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), 1–13. https://doi.org/10.1038/srep42717

Daina, A., & Zoete, V. (2016). A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules. *ChemMedChem*, *11*(11), 1117–1121. https://doi.org/10.1002/cmdc.201600182

Daniyan, M. O., Boshoff, A., Prinsloo, E., Pesce, E. R., & Blatch, G. L. (2016). The Malarial Exported PFA0660w Is an Hsp40 Co-Chaperone of PfHsp70-x. *PLoS ONE*, *11*(2). https://doi.org/10.1371/journal.pone.0148517

Daniyan, M. O., Przyborski, J. M., & Shonhai, A. (2019). Partners in mischief: Functional networks of heat shock proteins of Plasmodium falciparum and their influence on parasite virulence. *Biomolecules*, *9*(7). https://doi.org/10.3390/biom9070295

Dao, F., Djonor, S. K., Ayin, C. T. M., Adu, G. A., Sarfo, B., Nortey, P., Akuffo, K. O., & Danso-Appiah, A. (2021). Burden of malaria in children under five and caregivers' health-seeking behaviour for malaria-related symptoms in artisanal mining communities in Ghana. *Parasites and Vectors*, *14*(1), 1–11. https://doi.org/10.1186/S13071-021-04919-8/TABLES/2

Dasgupta, R., Pradhan, A. K., & Ghosh, S. (2020). First report of an insilico study of Luciferase in Mycobacterium sp. EPa45. *BioRxiv*, 2020.10.30.362764. https://doi.org/10.1101/2020.10.30.362764

Daugelaite, J., O' Driscoll, A., & Sleator, R. D. (2013). An Overview of Multiple Sequence Alignments and Cloud Computing in Bioinformatics. *ISRN Biomathematics*, *2013*, 1–14. https://doi.org/10.1155/2013/615630

Dey, F., & Caflisch, A. (2008). Fragment-based de Novo Ligand design by multiobjective evolutionary optimization. *Journal of Chemical Information and Modeling*, *48*(3), 679–690. https://doi.org/10.1021/CI700424B/SUPPL_FILE/CI700424B-FILE021.PDF

Docampo-Álvarez, B., Gómez-González, V., Montes-Campos, H., Otero-Mato, J. M., Méndez-Morales, T., Cabeza, O., Gallego, L. J., Lynden-Bell, R. M., Ivaništšev, V. B., Fedorov, M. V., & Varela, L. M. (2016). Molecular dynamics simulation of the behaviour of water in nano-confined ionic liquid–water mixtures. *Journal of Physics: Condensed Matter*, *28*(46), 464001. https://doi.org/10.1088/0953-8984/28/46/464001

Dondorp, A. M., Yeung, S., White, L., Nguon, C., Day, N. P. J., Socheat, D., & Von Seidlein, L. (2010). Artemisinin resistance: Current status and scenarios for containment. *Nature Reviews Microbiology*, *8*(4), 272–280. https://doi.org/10.1038/nrmicro2331

Dorn, M., E Silva, M. B., Buriol, L. S., & Lamb, L. C. (2014). Three-dimensional protein structure prediction: Methods and computational strategies. *Computational Biology and Chemistry*, *53*(PB), 251–276. https://doi.org/10.1016/J.COMPBIOLCHEM.2014.10.001

Duffy, P. E., & Mutabingwa, T. K. (2006). Artemisinin combination therapies. *Lancet*, *367*(9528), 2037–2039. https://doi.org/10.1016/S0140-6736(06)68900-9

Durrant, J. D., Amaro, R. E., & McCammon, J. A. (2009). AutoGrow: A Novel Algorithm for Protein Inhibitor Design. *Chemical Biology & Drug Design*, *73*(2), 168–178. https://doi.org/10.1111/J.1747-0285.2008.00761.X

Durrant, J. D., De Oliveira, C. A. F., & McCammon, J. A. (2011). POVME: An algorithm for measuring binding-pocket volumes. *Journal of Molecular Graphics and Modelling*, *29*(5), 773–776. https://doi.org/10.1016/J.JMGM.2010.10.007

Durrant, J. D., Lindert, S., & McCammon, J. A. (2013). AutoGrow 3.0: An improved algorithm for chemically tractable, semi-automated protein inhibitor design. *Journal of Molecular Graphics and Modelling*, *44*, 104–112. https://doi.org/10.1016/J.JMGM.2013.05.006

Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biology*, *9*(1), 1–9. https://doi.org/10.1186/1741-7007-9-71/FIGURES/4

Durrant, J. D., & McCammon, J. A. (2012). Autoclickchem: Click chemistry in silico. *PLoS Computational Biology*, *8*(3). https://doi.org/10.1371/journal.pcbi.1002397

Dutta, S., & Berman, H. M. (2005). Large macromolecular complexes in the Protein Data Bank: a status report. *Structure (London, England : 1993)*, *13*(3), 381–388. https://doi.org/10.1016/J.STR.2005.01.008

Eastman, P., Friedrichs, M. S., Chodera, J. D., Radmer, R. J., Bruns, C. M., Ku, J. P., Beauchamp, K. A., Lane, T. J., Wang, L. P., Shukla, D., Tye, T., Houston, M., Stich, T., Klein, C., Shirts, M. R., & Pande, V. S. (2013). OpenMM 4: A reusable, extensible, hardware independent library for high performance molecular simulation. *Journal of Chemical Theory and Computation*, *9*(1), 461–469. https://doi.org/10.1021/CT300857J/ASSET/IMAGES/MEDIUM/CT-2012-00857J_0008.GIF

Egwu, C. O., Augereau, J. M., Reybier, K., & Benoit-Vical, F. (2021). Reactive oxygen species as the brainbox in malaria treatment. In *Antioxidants* (Vol. 10, Issue 12). MDPI. https://doi.org/10.3390/antiox10121872

El-Faham, A., & Albericio, F. (2011). Peptide coupling reagents, more than a letter soup. *Chemical Reviews*, *111*(11), 6557–6602. https://doi.org/10.1021/cr100048w

Ensign, D. L., & Pande, V. S. (2009). The Fip35 WW Domain Folds with Structural and Mechanistic Heterogeneity in Molecular Dynamics Simulations. *Biophysical Journal*, *96*(8), L53–L55. https://doi.org/10.1016/J.BPJ.2009.01.024

Fairhurst, R. M., & Dondorp, A. M. (2016). Artemisinin-Resistant Plasmodium falciparum Malaria . *Microbiology Spectrum*, *4*(3), 409–429. https://doi.org/10.1128/microbiolspec.ei10-0013-2016

Fitch, C. D. (2004). Ferriprotoporphyrin IX, phospholipids, and the antimalarial actions of quinoline drugs. In *Life Sciences* (Vol. 74, Issue 16, pp. 1957–1972). Elsevier Inc. https://doi.org/10.1016/j.lfs.2003.10.003

Fitch, C. D., Dutta, P., Kanjananggulpan, P., & Chevli, R. (1984). Ferriprotoporphyrin IX: a mediator of the antimalarial action of oxidants and 4-aminoquinoline drugs. *Progress in Clinical and Biological Research*, *155*, 119–130. https://api.semanticscholar.org/CorpusID:40561581

Florens, L., Washburn, M., Muster, N., Wolters, D., Gardner, M., Anthony, R., Haynes, D., Moch, K., Sacci, J., Witney, A., Grainger, N., Holder, A., Wu, Y., Yates, J., & Carucci, D. (2002). A proteomic view of the malaria parasite life cycle. *Proceedings 50th ASMS Conference on Mass Spectrometry and Allied Topics*, *419*(6906), 59–60.

Fu, Y., Zhao, J., & Chen, Z. (2018). Insights into the Molecular Mechanisms of Protein-Ligand Interactions by Molecular Docking and Molecular Dynamics Simulation: A Case of Oligopeptide Binding Protein. *Computational and Mathematical Methods in Medicine*, *2018*. https://doi.org/10.1155/2018/3502514

Gahtori, P., Ghosh, S. K., Parida, P., Prakash, A., Gogoi, K., Bhat, H. R., & Singh, U. P. (2012). Antimalarial evaluation and docking studies of hybrid phenylthiazolyl-1,3,5-triazine derivatives: A novel and potential antifolate lead for Pf-DHFR-TS inhibition. *Experimental Parasitology*, *130*(3), 292–299. https://doi.org/10.1016/j.exppara.2011.12.014

Gaillard, T. (2018). Evaluation of AutoDock and AutoDock Vina on the CASF-2013 Benchmark. *Journal of Chemical Information and Modeling*, *58*(8), 1697–1706. https://doi.org/10.1021/ACS.JCIM.8B00312/SUPPL_FILE/CI8B00312_SI_002.ZIP

Ganesan, A., Coote, M. L., & Barakat, K. (2017). Molecular dynamics-driven drug discovery: leaping forward with confidence. *Drug Discovery Today*, *22*(2), 249–269. https://doi.org/10.1016/J.DRUDIS.2016.11.001

Garman, E. F. (2014). Developments in X-ray Crystallographic Structure Determination of Biological Macromolecules. *Science*, *343*(6175), 1102–1108. https://doi.org/10.1126/SCIENCE.1247829

Giordano, D., Biancaniello, C., Argenio, M. A., & Facchiano, A. (2022). Drug Design by Pharmacophore and Virtual Screening Approach. *Pharmaceuticals*, *15*(5). https://doi.org/10.3390/ph15050646

Golenser, J., Waknine, J. H., Krugliak, M., Hunt, N. H., & Grau, G. E. (2006). Current perspectives on the mechanism of action of artemisinins. In *International Journal for Parasitology* (Vol. 36, Issue 14, pp. 1427–1441). https://doi.org/10.1016/j.ijpara.2006.07.011

Gong, J., Zheng, H., Wu, Z., Chen, T., & Zhao, X. (2009). Genome shuffling: Progress and applications for phenotype improvement. *Biotechnology Advances*, *27*(6), 996–1005. https://doi.org/10.1016/J.BIOTECHADV.2009.05.016

González-Domínguez, J. (2021). Fast and Accurate Multiple Sequence Alignment with MSAProbs-MPI. *Methods in Molecular Biology*, *2231*, 39–47. https://doi.org/10.1007/978-1-0716-1036-7_3/COVER

Grant, B. J., Lukman, S., Hocker, H. J., Sayyah, J., Brown, J. H., McCammon, J. A., & Gorfe, A. A. (2011). Novel Allosteric Sites on Ras for Lead Generation. *PLOS ONE*, *6*(10), e25711. https://doi.org/10.1371/JOURNAL.PONE.0025711

Gribbon, P., & Andreas, S. (2005). High-throughput drug discovery: What can we expect from HTS? *Drug Discovery Today*, *1*(10), 17–22. https://doi.org/10.1016/S1359-6446(04)03275-1

Guedes, I. A., Pereira, F. S. S., & Dardenne, L. E. (2018). Empirical scoring functions for structure-based virtual screening: Applications, critical aspects, and challenges. *Frontiers in Pharmacology*, *9*(SEP), 1089. https://doi.org/10.3389/FPHAR.2018.01089/BIBTEX

Gunjan, S., Singh, S. K., Sharma, T., Dwivedi, H., Chauhan, B. S., Imran Siddiqi, M., & Tripathi, R. (2016). Mefloquine induces ROS mediated programmed cell death in malaria parasite: Plasmodium. *Apoptosis*, *21*(9), 955–964. https://doi.org/10.1007/s10495-016-1265-y

Haldar, K., Bhattacharjee, S., & Safeukui, I. (2018). Drug resistance in Plasmodium. *Nature Reviews Microbiology*, *16*(3), 156–170. https://doi.org/10.1038/nrmicro.2017.161

Hansson, T., Oostenbrink, C., & Van Gunsteren, W. F. (2002). Molecular dynamics simulations. *Current Opinion in Structural Biology*, *12*(2), 190–196. https://doi.org/10.1016/S0959-440X(02)00308-1

Hartenfeller, M., Eberle, M., Meier, P., Nieto-Oberhuber, C., Altmann, K. H., Schneider, G., Jacoby, E., & Renner, S. (2011). A collection of robust organic synthesis reactions for in silico molecule design. *Journal of Chemical Information and Modeling*, *51*(12), 3093–3098. https://doi.org/10.1021/ci200379p

Hatherley, R., Brown, D. K., Glenister, M., & Bishop, Ö. T. (2016). PRIMO: An interactive homology modeling pipeline. *PLoS ONE*, *11*(11), 1–20. https://doi.org/10.1371/journal.pone.0166698

Hoffer, L., Muller, C., Roche, P., & Morelli, X. (2018). Chemistry-driven Hit-to-lead Optimization Guided by Structure-based Approaches. *Molecular Informatics*, *37*(9). https://doi.org/10.1002/minf.201800059

Hogeweg, P. (2011). The roots of bioinformatics in theoretical biology. *PLoS Computational Biology*, *7*(3). https://doi.org/10.1371/journal.pcbi.1002021

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

Hollingsworth, S. A., & Karplus, P. A. (2010). A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomolecular Concepts*, *1*(3–4), 271–283. https://doi.org/10.1515/BMC.2010.022/MACHINEREADABLECITATION/RIS

Hospital, A., Goñi, J. R., Orozco, M., & Gelpí, J. L. (2015). Molecular dynamics simulations: advances and applications. *Advances and Applications in Bioinformatics and Chemistry : AABC*, *8*(1), 37. https://doi.org/10.2147/AABC.S70333

Hussein, H. A., Borrel, A., Geneix, C., Petitjean, M., Regad, L., & Camproux, A. C. (2015). PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. *Nucleic Acids Research*, *43*(W1), W436–W442. https://doi.org/10.1093/NAR/GKV462

Idro, R., Marsh, K., John, C. C., & Newton, C. R. J. (2010). Cerebral Malaria; Mechanisms Of Brain Injury And Strategies For Improved Neuro-Cognitive Outcome. *Pediatric Research*, *68*(4), 267. https://doi.org/10.1203/PDR.0B013E3181EEE738

Jadhav, A., Ferreira, R. S., Klumpp, C., Mott, B. T., Austin, C. P., Inglese, J., Thomas, C. J., Maloney, D. J., Shoichet, B. K., & Simeonov, A. (2010). Quantitative analyses of aggregation, autofluorescence, and reactivity artifacts in a screen for inhibitors of a thiol protease. *Journal of Medicinal Chemistry*, *53*(1), 37–51. https://doi.org/10.1021/jm901070c

Jaghoori, M. M., Bleijlevens, B., & Olabarriaga, S. D. (2016). 1001 Ways to run AutoDock Vina for virtual screening. *Journal of Computer-Aided Molecular Design*, *30*(3), 237–249. https://doi.org/10.1007/S10822-016-9900-9/TABLES/3

Jämsä-Jounela, S. L. (2007). Future trends in process automation. *Annual Reviews in Control*, *31*(2), 211–220. https://doi.org/10.1016/J.ARCONTROL.2007.08.003

Janson, G., Grottesi, A., Pietrosanto, M., Ausiello, G., Guarguaglini, G., & Paiardini, A. (2019). Revisiting the "satisfaction of spatial restraints" approach of MODELLER for protein homology modeling. *PLOS Computational Biology*, *15*(12), e1007219. https://doi.org/10.1371/JOURNAL.PCBI.1007219

Jaradat, D. M. M. (2018). Thirteen decades of peptide synthesis: key developments in solid phase peptide synthesis and amide bond formation utilized in peptide ligation. *Amino Acids*, *50*(1), 39–68. https://doi.org/10.1007/s00726-017-2516-0

Jong, K., Grisanti, L., & Hassanali, A. (2017). Hydrogen Bond Networks and Hydrophobic Effects in the Amyloid β30-35 Chain in Water: A Molecular Dynamics Study. *Journal of Chemical Information and Modeling*, *57*(7), 1548–1562. https://doi.org/10.1021/ACS.JCIM.7B00085/SUPPL_FILE/CI7B00085_SI_001.PDF

K. Gopalakrishnan, G. Sowmiya, S. S. Sheik, & K. Sekar. (2007). Ramachandran Plot on The Web (2.0). *Protein & Peptide Letters*, *14*(7), 669–671. https://doi.org/10.2174/092986607781483912

Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, *22*(12), 2577–2637. https://doi.org/10.1002/BIP.360221211

Kalidas, Y., & Chandra, N. (2008). PocketDepth: A new depth based algorithm for identification of ligand binding sites in proteins. *Journal of Structural Biology*, *161*(1), 31–42. https://doi.org/10.1016/J.JSB.2007.09.005

Kandt, C., Ash, W. L., & Peter Tieleman, D. (2007). Setting up and running molecular dynamics simulations of membrane proteins. *Methods*, *41*(4), 475–488. https://doi.org/10.1016/J.YMETH.2006.08.006

Karplus, M., & Petsko, G. A. (1990). Molecular dynamics simulations in biology. *Nature 1990 347:6294*, *347*(6294), 631–639. https://doi.org/10.1038/347631a0

Kato, T., Makino, F., Nakane, T., Terahara, N., Kaneko, T., Shimizu, Y., Motoki, S., Ishikawa, I., Yonekura, K., & Namba, K. (2019). The 1.54 Å resolution structure of apoferritin by CRYOARM300 with Cold-FEG. *Microsc. Microanal.*, *25*(S2), 998–999. https://doi.org/10.1017/s1431927619005725

Kaushik, P., & Saini, D. K. (2019). Sequence Analysis and Homology Modelling of SmHQT Protein, a Key Player in Chlorogenic Acid Pathway of Eggplant. *BioRxiv*, 599282.

Kinyanjui, S. M., Mberu, E. K., Winstanley, P. A., Jacobus, D. P., & Watkins, W. M. (1999). The antimalarial triazine WR99210 and the prodrug PS-15: Folate reversal of in vitro activity against Plasmodium falciparum and a non-antifolate mode of action of the prodrug. *American Journal of Tropical Medicine and Hygiene*, *60*(6), 943–947. https://doi.org/10.4269/ajtmh.1999.60.943

Klenke, B., Barrett, M. P., Brun, R., & Gilbert, I. H. (2003). Antiplasmodial activity of a series of 1,3,5-triazine-substituted polyamines. *Journal of Antimicrobial Chemotherapy*, *52*(2), 290–293. https://doi.org/10.1093/JAC/DKG307

Klenke, B., Stewart, M., Barrett, M. P., Brun, R., & Gilbert, I. H. (2001). Synthesis and biological evaluation of s-triazine substituted polyamines as potential new anti-Trypanosomal drugs. *Journal of Medicinal Chemistry*, *44*(21), 3440–3452. https://doi.org/10.1021/jm010854+

Ko, J., Park, H., Heo, L., & Seok, C. (2012). GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Research*, *40*(W1), W294–W297. https://doi.org/10.1093/NAR/GKS493

Konopka, J. K., Task, D., Afify, A., Raji, J., Deibel, K., Maguire, S., Lawrence, R., & Potter, C. J. (2021). Olfaction in Anopheles mosquitoes. *Chemical Senses*, *46*, 1–24. https://doi.org/10.1093/CHEMSE/BJAB021

Kralj, S., Jukič, M., & Bren, U. (2023). Molecular Filters in Medicinal Chemistry. In *Encyclopedia* (Vol. 3, Issue 2, pp. 501–511). https://doi.org/10.3390/encyclopedia3020035

Kravats, A. N., Hoskins, J. R., Reidy, M., Johnson, J. L., Doyle, S. M., Genest, O., Masison, D. C., & Wickner, S. (2018). Functional and physical interaction between yeast Hsp90 and Hsp70. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(10), E2210–E2219. https://doi.org/10.1073/pnas.1719969115

Krieger, E., Nabuurs, S. B., & Vriend, G. (2003). Homology Modeling. *Structural Bioinformatics*, *857*, 507–508. https://doi.org/10.1007/978-1-61779-588-6

Krivov, S. V. (2011). The free energy landscape analysis of protein (FIP35) folding dynamics. *Journal of Physical Chemistry B*, *115*(42), 12315–12324. https://doi.org/10.1021/JP208585R/ASSET/IMAGES/MEDIUM/JP-2011-08585R_0009.GIF

Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., & Moult, J. (2021). Critical assessment of methods of protein structure prediction (CASP)—Round XIV. *Proteins: Structure, Function and Bioinformatics*, *89*(12), 1607–1617. https://doi.org/10.1002/prot.26237

Kundrotas, P. J., Zhu, Z., Janin, J., & Vakser, I. A. (2012). Templates are available to model nearly all complexes of structurally characterized proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(24), 9438–9441. https://doi.org/10.1073/PNAS.1200678109

Kyaw, M. P., Nyunt, M. H., Chit, K., Aye, M. M., Aye, K. H., Aye, M. M., Lindegardh, N., Tarning, J., Imwong, M., Jacob, C. G., Rasmussen, C., Perin, J., Ringwald, P., & Nyunt, M. M. (2013). Reduced susceptibility of Plasmodium falciparum to artesunate in southern Myanmar. *PloS One*, *8*(3), e57689–e57689. https://doi.org/10.1371/journal.pone.0057689

Lam, S. D., Das, S., Sillitoe, I., & Orengo, C. (2017). An overview of comparative modelling and resources dedicated to large-scale modelling of genome sequences. *Urn:Issn:2059-7983*, *73*(8), 628–640. https://doi.org/10.1107/S2059798317008920

Laskowski, R. A., MacArthur, M. W., Moss, D. S., Thornton, J. M., & IUCr. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Urn:Issn:0021-8898*, *26*(2), 283–291. https://doi.org/10.1107/S0021889892009944

Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR 1996 8:4*, *8*(4), 477–486. https://doi.org/10.1007/BF00228148

Legraverend, M., & Grierson, D. S. (2006). The purines: Potent and versatile small molecule inhibitors and modulators of key biological targets. *Bioorganic & Medicinal Chemistry*, *14*(12), 3987–4006. https://doi.org/10.1016/J.BMC.2005.12.060

Levitt, M. (1992). Accurate modeling of protein conformation by automatic segment matching. *Journal of Molecular Biology*, *226*(2), 507–533. https://doi.org/10.1016/0022-2836(92)90964-L

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

Lilburn, T. G., Cai, H., Gu, J., Zhou, Z., & Wang, Y. (2014). Exploring systems affected by the heat shock response in Plasmodium falciparum via protein association networks. *International Journal of Computational Biology and Drug Design*, *7*(4), 369–383. https://doi.org/10.1504/IJCBDD.2014.066554

Lim, F. P. L., & Dolzhenko, A. V. (2014). 1,3,5-Triazine-based analogues of purine: From isosteres to privileged scaffolds in medicinal chemistry. *European Journal of Medicinal Chemistry*, *85*, 371–390. https://doi.org/10.1016/J.EJMECH.2014.07.112

Lipinski, C. A. (2004). Lead- and drug-like compounds: The rule-of-five revolution. *Drug Discovery Today: Technologies*, *1*(4), 337–341. https://doi.org/10.1016/j.ddtec.2004.11.007

Liu, K., & Houry, W. A. (2014). Heat shock proteins of malaria. In A. Shonhai & G. L. Blatch (Eds.), *Heat Shock Proteins of Malaria* (Vol. 9789400774, pp. 1–223). Springer. https://doi.org/10.1007/978-94-007-7438-4

Liu, Q., Liang, C., & Zhou, L. (2019). Structural and functional analysis of the Hsp70/Hsp40 chaperone system. In *Protein Science*. https://doi.org/10.1002/pro.3725

Liu, T., W. Tang, G., & Capriotti, E. (2011). Comparative modeling: the state of the art and protein drug target structure prediction. *Combinatorial Chemistry & High Throughput Screening*, *14*(6), 532–547. https://doi.org/10.2174/138620711795767811

Lüthy, R., Bowie, J. U., & Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature*, *356*(6364), 83–85. https://doi.org/10.1038/356083A0

Luzolo, A. L., & Ngoyi, D. M. (2019a). Cerebral malaria. *Brain Research Bulletin*, *145*, 53–58. https://doi.org/10.1016/j.brainresbull.2019.01.010

Luzolo, A. L., & Ngoyi, D. M. (2019b). Cerebral malaria. *Brain Research Bulletin*, *145*, 53–58. https://doi.org/10.1016/j.brainresbull.2019.01.010

Mabate, B., Zininga, T., Ramatsui, L., Makumire, S., Achilonu, I., Dirr, H. W., & Shonhai, A. (2018). Structural and biochemical characterization of Plasmodium falciparum Hsp70-x reveals functional versatility of its C-terminal EEVN motif. *Proteins: Structure, Function, and Bioinformatics*, *86*(11), 1189–1201. https://doi.org/10.1002/PROT.25600

Maier, A. G., Matuschewski, K., Zhang, M., & Rug, M. (2019). Plasmodium falciparum. *Trends in Parasitology*, *35*(6), 481–482. https://doi.org/10.1016/j.pt.2018.11.010

Majeux, N., Scarsi, M., Apostolakis, J., Ehrhardt, C., & Caflisch, A. (1999). Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins: Structure, Function and Genetics*, *37*(1), 88–105. https://doi.org/10.1002/(SICI)1097-0134(19991001)37:1<88::AID-PROT9>3.0.CO;2-O

Makenga, G., Menon, S., Baraka, V., Minja, D. T. R., Nakato, S., Delgado-Ratto, C., Francis, F., Lusingu, J. P. A., & Van Geertruyden, J. P. (2020). Prevalence of malaria parasitaemia in school-aged children and pregnant women in endemic settings of sub-Saharan Africa: A systematic review and meta-analysis. *Parasite Epidemiology and Control*, *11*, e00188. https://doi.org/10.1016/J.PAREPI.2020.E00188

Manohar, S., Khan, S. I., & Rawat, D. S. (2011). Synthesis of 4-aminoquinoline-1,2,3-triazole and 4-aminoquinoline-1,2,3-triazole-1,3,5-triazine Hybrids as Potential Antimalarial Agents. *Chemical Biology & Drug Design*, *78*(1), 124–136. https://doi.org/10.1111/J.1747-0285.2011.01115.X

Manyika, J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P., & Dewhurst, M. (2017). Harnessing Automation for a Future that Works. *McKinsey Global Institute*, *January*, 1–4.

Marius Clore, G., & Iwahara, J. (2009). Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chemical Reviews*, *109*(9), 4108–4139. https://doi.org/10.1021/CR900033P/ASSET/CR900033P.FP.PNG_V03

Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., & Šali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Biomolecular Structure*, *29*, 291–325. https://doi.org/10.1146/annurev.biophys.29.1.291

Matz, J. M. (2022). Plasmodium's bottomless pit: properties and functions of the malaria parasite's digestive vacuole. *Trends in Parasitology*, *38*(7), 525–543. https://doi.org/https://doi.org/10.1016/j.pt.2022.02.010

Mayr, L. M., & Bojanic, D. (2009). Novel trends in high-throughput screening. *Current Opinion in Pharmacology*, *9*(5), 580–588. https://doi.org/10.1016/J.COPH.2009.08.004

Mbengue, A., Bhattacharjee, S., Pandharkar, T., Liu, H., Estiu, G., Stahelin, R. V., Rizk, S. S., Njimoh, D. L., Ryan, Y., Chotivanich, K., Nguon, C., Ghorbal, M., Lopez-Rubio, J. J., Pfrender, M., Emrich, S., Mohandas, N., Dondorp, A. M., Wiest, O., & Haldar, K. (2015). A molecular mechanism of artemisinin resistance in Plasmodium falciparum malaria. *Nature*, *520*(7549), 683–687. https://doi.org/10.1038/nature14412

McGreevy, R., Teo, I., Singharoy, A., & Schulten, K. (2016). Advances in the molecular dynamics flexible fitting method for cryo-EM modeling. *Methods*, *100*, 50–60. https://doi.org/10.1016/J.YMETH.2016.01.009

Mehlhoff, J. D., Stearns, F. W., Rohm, D., Wang, B., Tsou, E. Y., Dutta, N., Hsiao, M. H., Gonzalez, C. E., Rubin, A. F., & Ostermeier, M. (2020). Collateral fitness effects of mutations. *Proceedings of the National Academy of Sciences*, *117*(21), 11597–11607. https://doi.org/10.1073/PNAS.1918680117

Ménard, R. (2000). The journey of the malaria sporozoite through its hosts: Two parasite proteins lead the way. *Microbes and Infection*, *2*(6), 633–642. https://doi.org/10.1016/S1286-4579(00)00362-2

Miller, B. R., McGee, T. D., Swails, J. M., Homeyer, N., Gohlke, H., & Roitberg, A. E. (2012). MMPBSA.py: An efficient program for end-state free energy calculations. *Journal of Chemical Theory and Computation*, *8*(9), 3314–3321. https://doi.org/10.1021/ct300418h

Miller, R. J. D. (2014). Femtosecond Crystallography with Ultrabright Electrons and X-rays: Capturing Chemistry in Action. *Science*, *343*(6175), 1108–1116. https://doi.org/10.1126/SCIENCE.1248488

Misra, G., & Ramachandran, R. (2009). Hsp70-1 from Plasmodium falciparum: Protein stability, domain analysis and chaperone activity. *Biophysical Chemistry*, *142*(1–3), 55–64. https://doi.org/10.1016/j.bpc.2009.03.006

Mita, T., & Tanabe, K. (2012). Evolution of Plasmodium falciparum drug resistance: Implications for the development and containment of artemisinin resistance. *Japanese Journal of Infectious Diseases*, *65*(6), 465–475. https://doi.org/10.7883/yoken.65.465

Moffett, K., Konteatis, Z., Nguyen, D., Shetty, R., Ludington, J., Fujimoto, T., Lee, K. J., Chai, X., Namboodiri, H., Karpusas, M., Dorsey, B., Guarnieri, F., Bukhtiyarova, M., Springman, E., & Michelotti, E. (2011). Discovery of a novel class of non-ATP site DFG-out state p38 inhibitors utilizing computationally assisted virtual fragment-based drug design (vFBDD). *Bioorganic & Medicinal Chemistry Letters*, *21*(23), 7155–7165. https://doi.org/10.1016/J.BMCL.2011.09.078

Mohandas, N., & An, X. (2012). Malaria and human red blood cells. *Medical Microbiology and Immunology*, *201*(4), 593–598. https://doi.org/10.1007/s00430-012-0272-z

Mok, S., Ashley, E. A., Ferreira, P. E., Zhu, L., Lin, Z., Yeo, T., Chotivanich, K., Imwong, M., Pukrittayakamee, S., Dhorda, M., Nguon, C., Lim, P., Amaratunga, C., Suon, S., Hien, T. T., Htut, Y., Faiz, M. A., Onyamboko, M. A., Mayxay, M., … Bozdech, Z. (2015). Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. *Science*, *347*(6220), 431–435. https://doi.org/10.1126/science.1260403

Moore, C. M., Wang, J., Lin, Q., Ferreira, P., Avery, M. A., Elokely, K., Staines, H. M., & Krishna, S. (2022). Selective Inhibition of Plasmodium falciparum ATPase 6 by Artemisinins and Identification of New Classes of Inhibitors after Expression in Yeast. *Antimicrobial Agents and Chemotherapy*, *66*(5). https://doi.org/10.1128/aac.02079-21

Mori, T., & Saito, S. (2015). Dynamic heterogeneity in the folding/unfolding transitions of FiP35. *The Journal of Chemical Physics*, *142*(13), 135101. https://doi.org/10.1063/1.4916641

Mugumbate, G., Abrahams, K. A., Cox, J. A. G., Papadatos, G., Van Westen, G., Lelièvre, J., Calus, S. T., Loman, N. J., Ballell, L., Barros, D., Overington, J. P., & Besra, G. S. (2015). Mycobacterial dihydrofolate reductase inhibitors identified using chemogenomic methods and in vitro validation. *PLoS ONE*, *10*(3), 1–17. https://doi.org/10.1371/journal.pone.0121492

Müller, I. B., & Hyde, J. E. (2010). Antimalarial drugs: Modes of action and mechanisms of parasite resistance. *Future Microbiology*, *5*(12), 1857–1873. https://doi.org/10.2217/fmb.10.136

Müntener, T., Böhm, R., Atz, K., Häussinger, D., & Hiller, S. (2020). NMR pseudocontact shifts in a symmetric protein homotrimer. *Journal of Biomolecular Nmr*, *74*(8), 413. https://doi.org/10.1007/S10858-020-00329-7

Muralidharan, V., Oksman, A., Pal, P., Lindquist, S., & Goldberg, D. E. (2012). Plasmodium falciparumheat shock protein 110 stabilizes the asparagine repeat-rich parasite proteome during malarial fevers. *Nature Communications*, *3*, 1310. https://doi.org/10.1038/ncomms2306

Nagasundaram, N., Doss, G. P. C., Chakraborty, C., Karthick, V., Kumar, D. T., Balaji, V., Siva, R., Lu, A., Ge, Z., & Zhu, H. (2016). Mechanism of artemisinin resistance for malaria PfATP6 L263 mutations and discovering potential antimalarials: An integrated computational approach. *Scientific Reports*, *6*. https://doi.org/10.1038/srep30106

Nandakumar, D. N., Nagaraj, V. A., Vathsala, P. G., Rangarajan, P., & Padmanaban, G. (2006). Curcumin-artemisinin combination therapy for malaria. *Antimicrobial Agents and Chemotherapy*, *50*(5), 1859–1860. https://doi.org/10.1128/AAC.50.5.1859-1860.2006

Ni, S., Yuan, Y., Huang, J., Mao, X., Lv, M., Zhu, J., Shen, X., Pei, J., Lai, L., Jiang, H., & Li, J. (2009). Discovering potent small molecule inhibitors of cyclophilin A using de novo drug design approach. *Journal of Medicinal Chemistry*, *52*(17), 5295–5298. https://doi.org/10.1021/JM9008295/SUPPL_FILE/JM9008295_SI_001.PDF

Njunge, J., Ludewig, M. H., Boshoff, A., Pesce, E.-R., & Blatch, G. L. (2013). Hsp70s and J Proteins of Plasmodium Parasites Infecting Rodents and Primates: Structure, Function, Clinical Relevance, and Drug Targets. *Current Pharmaceutical Design*, *19*(3), 387–403. https://doi.org/10.2174/138161213804143734

Njunge, J. M., Mandal, P., Przyborski, J. M., Boshoff, A., Pesce, E. R., & Blatch, G. L. (2015). PFB0595w is a Plasmodium falciparum J protein that co-localizes with PfHsp70-1 and can stimulate its in vitro ATP hydrolysis activity. *The International Journal of Biochemistry & Cell Biology*, *62*, 47–53. https://doi.org/10.1016/J.BIOCEL.2015.02.008

Notredame, C. (2007). Recent Evolutions of Multiple Sequence Alignment Algorithms. *PLOS Computational Biology*, *3*(8), e123. https://doi.org/10.1371/JOURNAL.PCBI.0030123

Ouji, M., Augereau, J.-M., Paloque, L., & Benoit-Vical, F. (2018). Plasmodium falciparum resistance to artemisinin-based combination therapies: A sword of Damocles in the path toward malaria elimination. *Parasite*, *25*, 24. https://doi.org/10.1051/parasite/2018021

Pal, S., Kumar, V., Kundu, B., Bhattacharya, D., Preethy, N., Reddy, M. P., & Talukdar, A. (2019). Ligand-based Pharmacophore Modeling, Virtual Screening and Molecular Docking Studies for Discovery of Potential Topoisomerase I Inhibitors. *Computational and Structural Biotechnology Journal*, *17*, 291–310. https://doi.org/10.1016/j.csbj.2019.02.006

Paloque, L., Ramadani, A. P., Mercereau-Puijalon, O., Augereau, J. M., & Benoit-Vical, F. (2016). Plasmodium falciparum: multifaceted resistance to artemisinins. *Malaria Journal*, *15*(1), 1–12. https://doi.org/10.1186/s12936-016-1206-9

Pan, B. Bin, Yang, F., Ye, Y., Wu, Q., Li, C., Huber, T., & Su, X. C. (2016). 3D structure determination of a protein in living cells using paramagnetic NMR spectroscopy. *Chemical Communications*, *52*(67), 10237–10240. https://doi.org/10.1039/C6CC05490K

Papathoti, N. K., Kiddeejing, D., Daddam, J. R., Le Thanh, T., & Buensanteai, N. (2020). Identification of Protease Inhibition Mechanism by Iturin A against Agriculture Cutworm (Spodoptera litura) by Homology Modeling and Molecular Dynamics. *The Open Bioinformatics Journal*, *13*(1), 119–128. https://doi.org/10.2174/1875036202013010119

Pavithra, S. R., Kumar, R., & Tatu, U. (2007a). Systems Analysis of Chaperone Networks in the Malarial Parasite Plasmodium falciparum. *PLOS Computational Biology*, *3*(9), e168. https://doi.org/10.1371/JOURNAL.PCBI.0030168

Pavithra, S. R., Kumar, R., & Tatu, U. (2007b). Systems analysis of chaperone networks in the malarial parasite Plasmodium falciparum. *PLoS Computational Biology*, *3*(9), 1701–1715. https://doi.org/10.1371/journal.pcbi.0030168

Pesce, E. R., & Blatch, G. L. (2014). Plasmodial Hsp40 and Hsp70 chaperones: Current and future perspectives. In *Parasitology* (Vol. 141, Issue 9, pp. 1167–1176). Cambridge University Press. https://doi.org/10.1017/S003118201300228X

Pesce, E.-R., Cockburn, I. L., Goble, J. L., Stephens, L. L., & Blatch, G. L. (2010). Malaria heat shock proteins: drug targets that chaperone other drug targets. Infectious disorders drug targets. *Infectious Disorders - Drug Targets*, *10*(3), 147–157. https://doi.org/10.2174/187152610791163417

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., & Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, *26*(16), 1781–1802. https://doi.org/10.1002/JCC.20289

Pitman, M. R., & Menz, R. I. (2006). Methods for Protein Homology Modelling. *Applied Mycology and Biotechnology*, *6*(C), 37–59. https://doi.org/10.1016/S1874-5334(06)80005-5

Planes, N., Vanderheyden, P. P. M. L., Gratton, E., & Caballero-George, C. (2020). Image mean square displacement to study the lateral mobility of Angiotensin II type 1 and Endothelin 1 type A receptors on living cells. *Microscopy Research and Technique*, *83*(4), 381–392. https://doi.org/10.1002/jemt.23425

Przyborski, J. M., Diehl, M., & Blatch, G. L. (2015). Plasmodial HSP70s are functionally adapted to the malaria parasite life cycle. *Frontiers in Molecular Biosciences*, *2*(JUN), 1–7. https://doi.org/10.3389/fmolb.2015.00034

Quiliano, M., Mendoza, A., Fong, K. Y., Pabón, A., Goldfarb, N. E., Fabing, I., Vettorazzi, A., López de Cerain, A., Dunn, B. M., Garavito, G., Wright, D. W., Deharo, E., Pérez-Silanes, S., Aldana, I., & Galiano, S. (2016). Exploring the scope of new arylamino alcohol derivatives: Synthesis, antimalarial evaluation, toxicological studies, and target exploration. *International Journal for Parasitology: Drugs and Drug Resistance*, *6*(3), 184–198. https://doi.org/10.1016/j.ijpddr.2016.09.004

Rabaan, A. A., Al-Ahmed, S. H., Haque, S., Sah, R., Tiwari, R., Malik, Y. S., Dhama, K., Yatoo, M. I., Bonilla-Aldana, D. K., & Rodriguez-Morales, A. J. (2020). SARS-CoV-2, SARS-CoV, and MERS-CoV: A comparative overview. *Infezioni in Medicina*, *28*(2), 174–184.

Rabal, O., Urbano-Cuadrado, M., & Oyarzabal, J. (2011). Computational medicinal chemistry in fragment-based drug discovery: What, how and when. *Future Medicinal Chemistry*, *3*(1), 95–134. https://doi.org/10.4155/fmc.10.277

Radfar, A., Diez, A., & Bautista, J. M. (2008). Chloroquine mediates specific proteome oxidative damage across the erythrocytic cycle of resistant Plasmodium falciparum. *Free Radical Biology and Medicine*, *44*(12), 2034–2042. https://doi.org/10.1016/j.freeradbiomed.2008.03.010

Ramachandran, G. N., & Sasisekharan, V. (1968). Conformation of Polypeptides and Proteins. *Advances in Protein Chemistry*, *23*(C), 283–437. https://doi.org/10.1016/S0065-3233(08)60402-7

Rayan, A. (2009). New tips for structure prediction by comparative modeling. *Bioinformation*, *3*(6), 263. https://doi.org/10.6026/97320630003263

Reddy, B. P., Khadgapathi, Podili., & Reddy, G. K. (2012). *Raltegravir* (Patent US20120270888A1). FDA.

Sahu, S., Ghosh, S. K., Gahtori, P., Pratap Singh, U., Bhattacharyya, D. R., & Bhat, H. R. (2019). In silico ADMET study, docking, synthesis and antimalarial evaluation of thiazole-1,3,5-

triazine derivatives as Pf-DHFR inhibitor. *Pharmacological Reports*, *71*(5), 762–767. https://doi.org/10.1016/j.pharep.2019.04.006

Sahu, S., Ghosh, S. K., Kalita, J. M., Ginjupalli, M. C., & K, K. R. (2020). Discovery of potential 1,3,5-Triazine compounds against strains of Plasmodium falciparum using supervised machine learning models. *European Journal of Pharmaceutical Sciences*, *144*, 105208. https://doi.org/10.1016/j.ejps.2019.105208

Sahu, U., Sahoo, P. K., Kar, S. K., Mohapatra, B. N., & Ranjit, M. (2013). Association of TNF level with production of circulating cellular microparticles during clinical manifestation of human cerebral malaria. *Human Immunology*, *74*(6), 713–721. https://doi.org/10.1016/j.humimm.2013.02.006

Saifi, A. (2014). Antimalarial drugs: Mode of action and status of resistance. *African Journal of Pharmacy and Pharmacology*, *7*(5), 148–156. https://doi.org/10.5897/ajppx12.015

Salmaso, V., & Moro, S. (2018a). Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview. *Frontiers in Pharmacology*, *9*(AUG), 923. https://doi.org/10.3389/FPHAR.2018.00923/BIBTEX

Salmaso, V., & Moro, S. (2018b). Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview. *Frontiers in Pharmacology*, *9*(AUG), 923. https://doi.org/10.3389/FPHAR.2018.00923/BIBTEX

Sánchez, R., & Šali, A. (1997). Advances in comparative protein-structure modelling. *Current Opinion in Structural Biology*, *7*(2), 206–214. https://doi.org/10.1016/S0959-440X(97)80027-9

Sánchez-Linares, I., Pérez-Sánchez, H., Cecilia, J. M., & García, J. M. (2012). High-Throughput parallel blind Virtual Screening using BINDSURF. *BMC Bioinformatics*, *13*(SUPPL 1). https://doi.org/10.1186/1471-2105-13-S14-S13

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*(8), 1879–1881. https://doi.org/10.1021/JM0341913/ASSET/IMAGES/MEDIUM/JM0341913N00001.GIF

Schulze, J., Kwiatkowski, M., Borner, J., Schlüter, H., Bruchhaus, I., Burmester, T., Spielmann, T., & Pick, C. (2015). The Plasmodium falciparum exportome contains non-canonical PEXEL/HT proteins. *Molecular Microbiology*, *97*(2), 301–314. https://doi.org/10.1111/mmi.13024

Seraphim, T. V, Chakafana, G., Shonhai, A., & Houry, W. A. (2019). Plasmodium falciparum R2TP complex: driver of parasite Hsp90 function. *Biophysical Reviews*, *11*(6), 1007–1015. https://doi.org/10.1007/s12551-019-00605-3

Shahinas, D., Folefoc, A., & Pillai, D. R. (2013). Targeting plasmodium falciparum Hsp90: Towards reversing antimalarial resistance. *Pathogens*, *2*(1), 33–54. https://doi.org/10.3390/pathogens2010033

Shahinas, D., Folefoc, A., Taldone, T., Chiosis, G., Crandall, I., & Pillai, D. R. (2013). A Purine Analog Synergizes with Chloroquine (CQ) by Targeting Plasmodium falciparum Hsp90 (PfHsp90). *PLoS ONE*, *8*(9). https://doi.org/10.1371/journal.pone.0075446

Shen, M., & Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Science*, *15*(11), 2507–2524. https://doi.org/10.1110/ps.062416606

Shen, Q., Xiong, B., Zheng, M., Luo, X., Luo, C., Liu, X., Du, Y., Li, J., Zhu, W., Shen, J., & Jiang, H. (2011). Knowledge-based scoring functions in drug design: 2. can the knowledge base be enriched? *Journal of Chemical Information and Modeling*, *51*(2), 386–397. https://doi.org/10.1021/CI100343J/SUPPL_FILE/CI100343J_SI_003.TXT

Sheng, C., & Zhang, W. (2013). Fragment informatics and computational fragment-based drug design: an overview and update. *Medicinal Research Reviews*, *33*(3), 554–598. https://doi.org/10.1002/MED.21255

Shetty, A. K., & Steele, R. W. (1999). Plasmodium falciparum malaria. *The Journal of the Louisiana State Medical Society : Official Organ of the Louisiana State Medical Society*, *151*(1), 29–32. https://doi.org/10.4269/ajtmh.1971.20.655

Shibeshi, M. A., Kifle, Z. D., & Atnafie, S. A. (2020). Antimalarial drug resistance and novel targets for antimalarial drug discovery. *Infection and Drug Resistance*, *13*, 4047–4060. https://doi.org/10.2147/IDR.S279433

Shonhai, A. (2010). Plasmodial heat shock proteins: targets for chemotherapy. *FEMS Immunology & Medical Microbiology*, *58*(1), 61–74. https://doi.org/10.1111/J.1574-695X.2009.00639.X

Shonhai, A., Boshoff, A., & Blatch, G. L. (2007). The structural and functional diversity of Hsp70 proteins from Plasmodium falciparum . *Protein Science*, *16*(9), 1803–1818. https://doi.org/10.1110/ps.072918107

Smith, G. D., Bedrov, D., & Borodin, O. (2000). Molecular Dynamics Simulation Study of Hydrogen Bonding in Aqueous Poly(Ethylene Oxide) Solutions. *Physical Review Letters*, *85*(26), 5583. https://doi.org/10.1103/PhysRevLett.85.5583

Stank, A., Kokh, D. B., Horn, M., Sizikova, E., Neil, R., Panecka, J., Richter, S., & Wade, R. C. (2017a). TRAPP webserver: predicting protein binding site flexibility and detecting transient binding pockets. *Nucleic Acids Research*, *45*(W1), W325–W330. https://doi.org/10.1093/NAR/GKX277

Stank, A., Kokh, D. B., Horn, M., Sizikova, E., Neil, R., Panecka, J., Richter, S., & Wade, R. C. (2017b). TRAPP webserver: Predicting protein binding site flexibility and detecting transient binding pockets. *Nucleic Acids Research*, *45*(W1), W325–W330. https://doi.org/10.1093/nar/gkx277

Sutcliffe, M. J., Haneef, I., Carney, D., & Blundell, T. L. (1987). Knowledge based modelling of homologous proteins, part I: three-dimensional frameworks derived from the simultaneous superposition of multiple structures. *Protein Engineering, Design and Selection*, *1*(5), 377–384. https://doi.org/10.1093/PROTEIN/1.5.377

Tahir ul Qamar, M., Alqahtani, S. M., Alamri, M. A., & Chen, L. L. (2020). Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants. *Journal of Pharmaceutical Analysis*, *10*(4), 313–319. https://doi.org/10.1016/J.JPHA.2020.03.009

Talapko, J., Škrlec, I., Alebić, T., Jukić, M., & Včev, A. (2019). Malaria: The past and the present. *Microorganisms*, *7*(6), 179. https://doi.org/10.3390/microorganisms7060179

Thapar, M. M., Gupta, S., Spindler, C., Wernsdorfer, W. H., & Bj6rkman, A. (2003). *Pharmacodynamic interactions among atovaquone, proguanil and cycloguanil against Plasmodium falciparum in vitro*.

Tilocca, B., Soggiu, A., Sanguinetti, M., Musella, V., Britti, D., Bonizzi, L., Urbani, A., & Roncada, P. (2020). Comparative computational analysis of SARS-CoV-2 nucleocapsid protein epitopes in taxonomically related coronaviruses. *Microbes and Infection*, *22*(4–5), 188–194. https://doi.org/10.1016/J.MICINF.2020.04.002

Tjong, E., Dimri, M., & Mohiuddin, S. S. (2023). *Biochemistry, Tetrahydrofolate*. StatPearls Publishing, Treasure Island (FL). http://europepmc.org/books/NBK539712

Toure, O. A., Landry, T. N., Assi, S. B., Kone, A. A., Gbessi, E. A., Ako, B. A., Coulibaly, B., Kone, B., Ouattara, O., Beourou, S., Koffi, A., Remoue, F., & Rogier, C. (2018). Malaria parasite clearance from patients following artemisinin-based combination therapy in Côte d’Ivoire. *Infection and Drug Resistance*, *11*, 2031–2038. https://doi.org/10.2147/IDR.S167518

Townsend, J. P., Hassler, H. B., Wang, Z., Miura, S., Singh, J., Kumar, S., Ruddle, N. H., Galvani, A. P., & Dornburg, A. (2021). The durability of immunity against reinfection by SARS-CoV-2: a comparative evolutionary study. *The Lancet Microbe*, *2*(12), e666–e675. https://doi.org/10.1016/S2666-5247(21)00219-6

Trape, J. F. (2001). The public health impact of chloroquine resistance in Africa. *American Journal of Tropical Medicine and Hygiene*, *64*(1-2 SUPPL.), 12–17. https://doi.org/10.4269/ajtmh.2001.64.12

Ullah, S., Khan, S. U., Khan, A., Junaid, M., Rafiq, H., Htar, T. T., Zhao, Y., Shah, S. A. A., & Wadood, A. (2022). Prospect of Anterior Gradient 2 homodimer inhibition via repurposing FDA-approved drugs using structure-based virtual screening. *Molecular Diversity*, *26*(3), 1399–1409. https://doi.org/10.1007/s11030-021-10263-x

Valdés-Tresanco, M. S., Valdés-Tresanco, M. E., Valiente, P. A., & Moreno, E. (2021). Gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. *Journal of Chemical Theory and Computation*, *17*(10), 6281–6291. https://doi.org/10.1021/acs.jctc.1c00645

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, *26*(16), 1701–1718. https://doi.org/10.1002/JCC.20291

van Eijk, A. M., Hill, J., Noor, A. M., Snow, R. W., & ter Kuile, F. O. (2015). Prevalence of malaria infection in pregnant women compared with children for tracking malaria transmission in sub-Saharan Africa: A systematic review and meta-analysis. *The Lancet Global Health*, *3*(10), e617–e628. https://doi.org/10.1016/S2214-109X(15)00049-2

Venable, R. M., Krämer, A., & Pastor, R. W. (2019). Molecular Dynamics Simulations of Membrane Permeability. *Chemical Reviews*, *119*(9), 5954–5997. https://doi.org/10.1021/ACS.CHEMREV.8B00486/ASSET/IMAGES/MEDIUM/CR-2018-00486Z_0032.GIF

Vilaivan, T., Saesaengseerung, N., Jarprung, D., Kamchonwongpaisan, S., Sirawaraporn, W., & Yuthavong, Y. (2003). Synthesis of Solution-Phase Combinatorial Library of 4,6-Diamino-1,2-dihydro-1,3,5-triazine and Identification of New Leads Against A16V+S108T Mutant Dihydrofolate Reductase of Plasmodium falciparum. *Bioorganic & Medicinal Chemistry*, *11*(2), 217–224. https://doi.org/10.1016/S0968-0896(02)00344-9

Wagner, J. R., Sørensen, J., Hensley, N., Wong, C., Zhu, C., Perison, T., & Amaro, R. E. (2017). POVME 3.0: Software for Mapping Binding Pocket Flexibility. *Journal of Chemical Theory and Computation*, *13*(9), 4584–4592. https://doi.org/10.1021/ACS.JCTC.7B00500/SUPPL_FILE/CT7B00500_SI_002.PDF

Wallner, B., & Elofsson, A. (2005). All are not equal: A benchmark of different homology modeling programs. *Protein Science*, *14*(5), 1315–1327. https://doi.org/10.1110/PS.041253405

Wang, S., Liu, D., Ding, M., Du, Z., Zhong, Y., Song, T., Zhu, J., & Zhao, R. (2021). SE-OnionNet: A Convolution Neural Network for Protein–Ligand Binding Affinity Prediction. *Frontiers in Genetics*, *11*, 1805. https://doi.org/10.3389/FGENE.2020.607824/BIBTEX

Wang, W. J., Huang, Q., Zou, J., Li, L. L., & Yang, S. Y. (2015). TS-Chemscore, a Target-Specific Scoring Function, Significantly Improves the Performance of Scoring in Virtual Screening. *Chemical Biology & Drug Design*, *86*(1), 1–8. https://doi.org/10.1111/CBDD.12470

Wang, X., Song, K., Li, L., & Chen, L. (2018). Structure-Based Drug Design Strategies and Challenges. *Current Topics in Medicinal Chemistry*, *18*(12), 998–1006. https://doi.org/10.2174/1568026618666180813152921

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. https://doi.org/10.1093/NAR/GKY427

Watkins, W. M., & Sixsmith, D. G. (1984). The activity of proguanil and its metabolites, cycloguanil and p-chlorophenylbiguanide, against Plasmodium falciparum in vitro. *Annals of Tropical Medicine and Parasitology*, *78*(3), 273–278. https://doi.org/10.1080/00034983.1984.11811816

Webb, B., & Sali, A. (2016). Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics*, *54*(1), 5.6.1-5.6.37. https://doi.org/10.1002/CPBI.3

Webb-Robertson, B. J. M., McCue, L. A., & Lawrence, C. E. (2008). Measuring Global Credibility with Application to Local Sequence Alignment. *PLOS Computational Biology*, *4*(5), e1000077. https://doi.org/10.1371/JOURNAL.PCBI.1000077

Welsch, M. E., Snyder, S. A., & Stockwell, B. R. (2010). Privileged scaffolds for library design and drug discovery. *Current Opinion in Chemical Biology*, *14*(3), 347–361. https://doi.org/10.1016/j.cbpa.2010.02.018

WHO. (2019a). *World malaria report,November,20* (WHO, Ed.; Issue November,30). https://www.who.int/publications-detail/world-malaria-report-2019

WHO. (2019b). *World malaria report,November,20* (WHO, Ed.; Issue November,30).

WHO. (2020). World Malaria Report: 20 years of global progress and challenges. In WHO (Ed.), *World Health Organization*. https://www.who.int/publications/i/item/9789240015791

WHO. (2021). Word Malaria Report 2021. In WHO (Ed.), *Word Malaria report Geneva: World Health Organization. (2021). Licence: CC*. https://www.who.int/publications/i/item/9789240040496

WHO. (2022). *World malaria report 2022*. https://www.who.int/teams/global-malaria-programme

WHO. (2023). *World malaria report 2022*. https://www.wipo.int/amc/en/mediation/

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, *35*(suppl_2), W407–W410. https://doi.org/10.1093/NAR/GKM290

Yadav, R., Choudhury, C., Kumar, Y., & Bhatia, A. (2020). Virtual repurposing of ursodeoxycholate and chenodeoxycholate as lead candidates against SARS-Cov2-Envelope protein: A molecular dynamics investigation. *Journal of Biomolecular Structure and Dynamics*. https://doi.org/10.1080/07391102.2020.1868339/SUPPL_FILE/TBSD_A_1868339_SM2155.PDF

Ye, J., McGinnis, S., & Madden, T. L. (2006). BLAST: improvements for better sequence analysis. *Nucleic Acids Research*, *34*(suppl_2), W6–W9. https://doi.org/10.1093/NAR/GKL164

Yip, K. M., Fischer, N., Paknia, E., Chari, A., & Stark, H. (2020). Atomic-resolution protein structure determination by cryo-EM. *Nature 2020 587:7832*, *587*(7832), 157–161. https://doi.org/10.1038/s41586-020-2833-4

Yoganathan, S., & Miller, S. J. (2013). N-methylimidazole-catalyzed synthesis of carbamates from hydroxamic acids via the lossen rearrangement. *Organic Letters*, *15*(3), 602–605. https://doi.org/10.1021/OL303424B/SUPPL_FILE/OL303424B_SI_001.PDF

Yu, H. Y., Ziegelhoffer, T., Osipiuk, J., Ciesielski, S. J., Baranowski, M., Zhou, M., Joachimiak, A., & Craig, E. A. (2015). Roles of Intramolecular and Intermolecular Interactions in Functional Regulation of the Hsp70 J-protein Co-Chaperone Sis1. *Journal of Molecular Biology*, *427*(7), 1632–1643. https://doi.org/10.1016/J.JMB.2015.02.007

Yuan, Y., Pei, J., & Lai, L. (2011). LigBuilder 2: A practical de novo drug design approach. *Journal of Chemical Information and Modeling*, *51*(5), 1083–1091. https://doi.org/10.1021/CI100350U/SUPPL_FILE/CI100350U_SI_001.PDF

Yuan, Y., Pei, J., & Lai, L. (2020). LigBuilder V3: A Multi-Target de novo Drug Design Approach. *Frontiers in Chemistry*, *8*, 142. https://doi.org/10.3389/FCHEM.2020.00142/BIBTEX

Yui, K., & Inoue, S. I. (2021). Host-pathogen interaction in the tissue environment during Plasmodium blood-stage infection. *Parasite Immunology*, *43*(2), e12763. https://doi.org/10.1111/PIM.12763

Yuthavong, Y. (2013). Antifolate Drugs. In M. Hommel & P. G. Kremsner (Eds.), *Encyclopedia of Malaria* (pp. 1–12). Springer New York. https://doi.org/10.1007/978-1-4614-8757-9_2-1

Zarou, M. M., Vazquez, A., & Vignir Helgason, G. (2021). Folate metabolism: a re-emerging therapeutic target in haematological cancers. In *Leukemia* (Vol. 35, Issue 6, pp. 1539–1551). Springer Nature. https://doi.org/10.1038/s41375-021-01189-2

Zhang, Q. C., Petrey, D., Norel, R., & Honig, B. H. (2010). Protein interface conservation across structure space. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(24), 10896–10901. https://doi.org/10.1073/PNAS.1005894107

Zhang, Q., Ma, C., Oberli, A., Zinz, A., Engels, S., & Przyborski, J. M. (2017). Proteomic analysis of exported chaperone/co-chaperone complexes of P. falciparum reveals an array of complex protein-protein interactions. *Scientific Reports*, *7*(August 2016), 1–15. https://doi.org/10.1038/srep42188

Zhao, Q., Capelli, R., Carloni, P., Lüscher, B., Li, J., & Rossetti, G. (2021). Enhanced Sampling Approach to the Induced-Fit Docking Problem in Protein-Ligand Binding: The Case of Mono-ADP-Ribosylation Hydrolase Inhibitors. *Journal of Chemical Theory and Computation*, *17*(12), 7899–7911. https://doi.org/10.1021/ACS.JCTC.1C00649/ASSET/IMAGES/MEDIUM/CT1C00649_M003.GIF

Zhi, H. Y., Zhao, L., Lee, C. C., & Chen, C. Y. C. (2021). A novel graph neural network methodology to investigate dihydroorotate dehydrogenase inhibitors in small cell lung cancer. *Biomolecules*, *11*(3). https://doi.org/10.3390/biom11030477

Zhou, A. Q., O'Hern, C. S., & Regan, L. (2011). Revisiting the Ramachandran plot from a new angle. *Protein Science*, *20*(7), 1166–1171. https://doi.org/10.1002/PRO.644

Zhu, H., & Pisabarro, M. T. (2011). MSPocket: an orientation-independent algorithm for the detection of ligand binding pockets. *Bioinformatics*, *27*(3), 351–358. https://doi.org/10.1093/BIOINFORMATICS/BTQ672

Zininga, T. (2015). *Characterization of heat shock protein 70-z (PfHsp70-z) from plasmodium falciparium*. *November*.

Zininga, T., Achilonu, I., Hoppe, H., Prinsloo, E., Dirr, H. W., & Shonhai, A. (2015). Overexpression, purification and characterisation of the Plasmodium falciparum Hsp70-z (PfHsp70-z) protein. *PLoS ONE*, *10*(6), 1–13. https://doi.org/10.1371/journal.pone.0129445

Zininga, T., Anokwuru, C., Sigidi, M., Tshisikhawe, M., Ramaite, I., Traoré, A., Hoppe, H., Shonhai, A., & Potgieter, N. (2017). Extracts Obtained from Pterocarpus angolensis DC and Ziziphus mucronata Exhibit Antiplasmodial Activity and Inhibit Heat Shock Protein 70 (Hsp70) Function. *Molecules*, *22*(8), 1224. https://doi.org/10.3390/molecules22081224

Zininga, T., Ramatsui, L., Makhado, P. B., Makumire, S., Achilinou, I., Hoppe, H., Dirr, H., & Shonhai, A. (2017). (−)-Epigallocatechin-3-gallate inhibits the chaperone activity of Plasmodium falciparum Hsp70 chaperones and abrogates their association with functional partners. *Molecules*, *22*(12). https://doi.org/10.3390/molecules22122139

Zininga, T., & Shonhai, A. (2014). Are heat shock proteins druggable candidates? *American Journal of Biochemistry and Biotechnology*, *10*(4), 209–210. https://doi.org/10.3844/ajbbsp.2014.209.210

7.0 APPENDIX

FASTA sequence for *Pf*Hsp70-1

>PF3D7_0818900 | Plasmodium falciparum 3D7 | heat shock protein 70 | genomic | Pf3D7_08_v3 forward | (geneStart+0 to geneEnd+0) | length=4431

TGTGTTAGGTAAAAATACTATATTAATATAAAAATATAAATAACCTTAAGAAAATATACA

ATTATTTTTTTTCATTTCATTTAATGATGCTATTAAAAGCGGAATATTTTACAAAAAATA

AAAATATGTTATATATATATAAATATATATATATATATATATAATATATTTTATATGTAA

GTAAAATTTCATGTATAAAGAAATATAGTATTTTTTCTTATTTTATTTTATATATATAAA

TCGAGAAAATTAAAAAATGCTTTTAAATAATAATACAAATTATTTTGTATATAATAAAAT

AATATATTATATATATTATTTATATATATACGCAAAATTATGAAAATCCACATAATAAAA

ATATATATATATAAATATATATAATTTATATTATATATATATAATAATTTAGGCCCCGTG

CCCTTATTTTGATTTTTTTTTTTTTTTGGAATTTATAAACTTGTGATAAAAATATATATA

ATAATATTTTATATATATAATTTTATATAAAATATAATAATAAATGTATTATATATTATA

TATATATATATATATATATATATATATATATATTATATAAACTATATATTTTAAAAATATAT

ATTAATAAATATTGTATAGATACCTATAGAAACATGCAATTGAATATATTTCTTGTAAAA

TTTACATAGAGTACTTATATTTTTTTCTGCAGGGTAGGAAAAAAAAAAAACAGAAGGGGC

GAAATGGCCCCGCATTTTTGGGATTTTTATTTGGATTAAAAAAAATAAGAAATAACGAAT

TTTAGTCCCTTGAACTTGCAAAGAAAAAAAAACTTATAAAAGTAGGATAAAAAATTTTTT

TTTTTTTTTTATAAAAAATATATGAATGTAGCATTTATAATATTTTTTTATACAAAATAT

AAATTATTCGCATAAATATCTGGTGAAATACAAACAAGAATAAATTCAACAATATTTATA

TATTTTTTATTTATTATATATATATAATATTTTTTTATGATACTACTATTTATATATATA

TATATATATTTATATGTATTATATATCTCGTTTATAACAAATTGAAAAATATAAAATATA

TTTATATTCAAATAAAAGTAAAAATAAAGTAATTATATTATATATATTATTAATACTTAT

TATATATATATTATATAACATAATATTTTAATGGTTTATTTATAATAATATATACTATAT

ATGTTATATATAAATAAATACCATAGTTCTTATATATATATATTTTTTATATATTGTAAT

ATTATAAAAAAATTCGTAAAATATAATTTTTTTTATTTTATTATATATTTAACATTAGTA

ATTATTTATATTATTTATATTTATATACATAATATAATATATGTAAAAAAAATAAAATAA

AATAAAATAAAATAATATCTTATACAAATATAATAATAAATTATTATATATAAATAAAAT

ATATATAATATATTTACAAAATAAATAAAAAATAGTAAATATTTATATAATATTAATAAA

TAAATATATTCTATATATTTTTTATAAAGATAAATAAAATAATAATAATATATTATAAAA

TAATATATTTATAAATATATATTTTTTATATATTTCCAAAAAAAAAAAAAAAAGATAAAAT

ATTTTCGTATCTTAAGAATAATTTTTCTTATATTAATTATATTATATTATATTATAATTA

TTTAAAAATATATTATTTGTAAAAGTTTTATAATTATATAATAAAATAGTATAAATATAA

TATATTATTATATATATATATAATATATAAATATATTTATATTTATTATTATTATTATTATA

TATATATATATATATACAAAAGAATTAAGAAAAAATGGCTAGTGCAAAAGGTTCAAAACC

AAATTTACCAGAATCCAATATCGCTATTGGAATTGATTTAGGTACTACTTATTCTTGTGT

TGGTGTATGGAGAAATGAAAATGTAGATATTATTGCTAATGACCAAGGTAATAGAACAAC

CCCATCTTATGTTGCTTTCACCGATACCGAAAGATTAATTGGAGATGCTGCTAAAAACCA

AGTAGCTAGGAATCCAGAAAATACAGTATTTGATGCTAAGAGATTAATTGGTAGAAAATT

TACAGAATCATCAGTACAAAGTGATATGAAACATTGGCCATTCACTGTTAAATCAGGTGT

TGATGAGAAACCAATGATTGAAGTTACCTATCAAGGAGAAAAGAAATTATTCCATCCAGA

AGAAATTTCTTCTATGGTATTACAAAAAATGAAAGAAAATGCTGAAGCATTTTTAGGAAA

ATCTATAAAGAATGCTGTCATTACCGTTCCAGCTTATTTTAACGATTCACAAAGACAAGC

TACTAAAGATGCTGGTACAATTGCAGGATTAAATGTTATGAGAATTATTAATGAACCTAC

TGCAGCTGCTATTGCATATGGTTTACACAAAAAAGGAAAAGGTGAAAAGAACATTTTAAT

TTTCGACTTAGGAGGAGGTACATTTGATGTATCATTATTAACTATTGAAGATGGTATTTT

TGAAGTAAAAGCTACTGCTGGTGATACTCATTTAGGTGGTGAAGATTTCGATAACAGATT

AGTAAATTTCTGTGTTGAAGATTTCAAAAGAAAAAACAGAGGTAAAGATTTATCAAAAAA

TAGTAGAGCCTTAAGAAGATTAAGAACACAATGTGAAAGAGCAAAACGTACTTTATCATC

ATCTACACAAGCTACAATTGAAATAGATTCCTTATTTGAAGGTATTGATTACAGTGTTAC

TGTAAGTAGAGCAAGATTTGAAGAATTATGTATCGACTATTTCCGTGATACTTTAATTCC

AGTAGAAAAAGTTTTAAAAGATGCTATGATGGATAAAAAAAGTGTACATGAAGTTGTTTT

AGTTGGTGGTTCTACAAGAATTCCAAAAATCCAAACTTTAATAAAAGAATTCTTTAATGG

TAAAGAAGCATGCAGATCAATTAACCCAGATGAAGCTGTTGCATATGGTGCAGCTGTACA

AGCAGCCATTTTATCTGGTGACCAATCAAATGCTGTCCAAGATTTATTATTATTAGATGT

TTGCTCCTTATCATTAGGTTTAGAAACTGCTGGTGGTGTTATGACCAAATTAATTGAAAG

AAACACAACCATACCTGCTAAAAAGAGTCAAATCTTTACTACTTATGCTGATAACCAACC

AGGTGTCTTAATTCAAGTATATGAAGGTGAAAGAGCCTTAACCAAAGATAACAATTTATT

AGGAAAATTTCACTTAGATGGTATTCCACCTGCACCAAGAAAGGTACCACAAATCGAAGT

TACATTCGATATCGATGCTAACGGTATCTTAAACGTTACGGCTGTAGAAAAATCCACTGG

TAAACAAAACCATATTACAATTACCAACGACAAAGGAAGATTATCTCAAGATGAAATTGA

TCGTATGGTTAATGATGCTGAAAAATACAAAGCAGAAGATGAAGAAAACAGAAAAAGAAT

CGAAGCAAGAAACAGCCTTGAAAATTACTGCTATGGAGTTAAAAGCTCATTAGAAGACCA

AAAAATTAAAGAAAAATTACAACCAGCTGAAATTGAAACATGTATGAAAACTATTACAAC

CATACTTGAATGGTTAGAAAAAAACCAACTTGCTGGAAAAGATGAATATGAAGCCAAACA

AAAAGAAGCAGAATCGGTTTGTGCTCCAATTATGTCTAAAATCTATCAAGATGCTGCTGG

TGCAGCCGGTGGTATGCCAGGAGGTATGCCCGGTGGAATGCCCGGTGGAATGCCAGGTGG

TATGAATTTCCCAGGAGGTATGCCCGGAGCAGGAATGCCAGGAAATGCCCCAGCTGGAAG

TGGACCAACAGTTGAAGAAGTTGATTAAACTAATAAAAAAAAAAAACATTAAACAGGACA

AATATAAAAATATATATATTATAAAAATATATATATATATATATATTTTTTTTTTTTT

TTTTACATTTTTGTAAATTAATATATATTGAATACATATATGCCAAACAAAAAAGAAAAA

TATATCTACATATATTTATATATATATATATATATATATATATATTCATGTGTAATAATA

TATACAATATTTATACATATAAAAATATTTTTTATATTTTCGGTTATATAAATAAACAAG

CTATTTATGGATCACACAAATATAATTATATATTTGTATATGTATTTTTATATATATATA

TATATATATGTATGTGTAATTATTTATTTTTTTTTTCCTCATAAAATTATAATAAACTTT

TGTACTGCTTAGAAATTAAAAAATTAATAAATTAATAAATCATTAAAGGTTTTTATTTAA

TTTGAAAAATAAGTTTTTTTTTAAAGAATATACATCCATACAAAATGTTTTTAAAAATAT

ATGTATTATTAGGCATGAAATATATATATTATAAATATATATATATATATA

Table 14: Curated Seed Compounds

Smiles	ID
Nc1nc(N)nc(NC2CC2)n1	ZINC000000001239
Cc1cc2nc(N)nc(N)n2n1	ZINC000000001242
CCOc1nc(N)nc(NC)n1	ZINC000000066140
Cc1nc(N)nc(N(C)C)n1	ZINC000000076646
COc1nc(C)nc(N)n1	ZINC000000193078
CN(C)c1nc2n(c(=O)n1)CCS2	ZINC000000267522
CN(C)c1nc2n(c(=O)n1)CCN2C	ZINC000000267525
CN(C)c1nc2n(c(=O)n1)CCO2	ZINC000000270497
COc1nc(N)nc(N(C)C)n1	ZINC000000295277
CN(C)c1ncnc(N)n1	ZINC000000394205
CC(C)C(=O)c1nc(N)nc(N)n1	ZINC000000394880
CN(C)c1nc(N)n(CCO)c(=O)n1	ZINC000000494371
Nc1nc(N)nc(Cl)n1	ZINC000000896288
Nc1nc(N)nc(N)n1	ZINC000000897751
COc1nc(N)nc(OC)n1	ZINC000000967443
Nc1nc(N)nc(C(=O)[O-])n1	ZINC000001263542
COc1nc(NC#N)nc(N(C)C)n1	ZINC000001423972
Nc1[nH]c(=O)nc(N2CCOCC2)n1	ZINC000001427973
Nc1ncnc(N)n1	ZINC000001445475

CN(C)c1nc(N(C)C)[nH]c(=O)n1	ZINC000001446067
CN(C)c1nc(N)nc(N(C)C)n1	ZINC000001446096
Cc1nc(N)nc(N)n1	ZINC000001555389
Nc1nc(N)[nH]c(=S)n1	ZINC000001586434
Nc1nc(N)nc(N2CCOCC2)n1	ZINC000001603965
NCc1nc(N)nc(N)n1	ZINC000001607506
CN(C)c1nc(NN)nc(N(C)C)n1	ZINC000001657295
NC(=O)CCc1nc(N)nc(N)n1	ZINC000001661526
CN(C)c1nc(N)nc(N)n1	ZINC000001673367
Nc1nc(N)nc(CC(=O)[O-])n1	ZINC000001674446
O=c1nc2n(c(=O)[nH]1)CCN2	ZINC000001682354
Nc1ncncn1	ZINC000001685262
CNc1nc(NC)nc(NC)n1	ZINC000001688106
COc1nc(N)nc(N)n1	ZINC000001696268
Nc1ncnc(N2CCOCC2)n1	ZINC000001701973
Nc1nc(N)nc(NCO)n1	ZINC000001845500
CCN(CC)c1nc(N)nc(N)n1	ZINC000001869706
Nc1nc(N)nc(N2CCCC2)n1	ZINC000001869707
CC(=O)Nc1nc(N)nc(N)n1	ZINC000001869711
CCNc1nc(N)nc(OC)n1	ZINC000002011037
Cc1ncnc(N)n1	ZINC000002012340

CC(C)Oc1nc(N)nc(N)n1	ZINC000002014311
CNc1nc(C)nc(N)n1	ZINC000002016226
CCNc1ncnc(N)n1	ZINC000002034775
COc1ncnc(N)n1	ZINC000002039878
Nc1nc(N)n2cnnc2n1	ZINC000002169000
Nc1nc(N)nc(c2nonc2N)n1	ZINC000002241178
Cc1nnc2nc(N)nc(N)n12	ZINC000002308063
c1ncnc(N2CCNCC2)n1	ZINC000002540621
CNc1nc(OC)nc(N(C)C)n1	ZINC000002831342
Nc1nc(N)n2nc(N)nc2n1	ZINC000003174241
Cn1c(N)nc(=N)nc1N	ZINC000004411342
NNc1nc(NN)nc(NN)n1	ZINC000004530618
Nc1[n+]([O-])c(N)[n+]([O-])c(N)[n+]1[O-]	ZINC000004548826
Nc1nc(N)n2c([N+](=O)[O-])nnc2n1	ZINC000004558145
Nc1nc(N)nc(C(F)F)n1	ZINC000004566807
COc1nc(NN)nc2nnnn12	ZINC000004580226
COc1nc(N)nc(C(=O)[O-])n1	ZINC000004785861
COc1nc(N2CCCC2)[nH]c(=O)n1	ZINC000005129957
Nc1nc(N)nc(N2CC2)n1	ZINC000005134150
Nc1nc(NCO)nc(NCO)n1	ZINC000005178906
COc1nc(N(C)C)[nH]c(=O)n1	ZINC000005799912

Nc1nc(N)nc(CCO)n1	ZINC000006261688
NC(=O)Cc1nc(N)nc(N)n1	ZINC000006261690
Nc1nc(N)n(N)c(=S)n1	ZINC000006494845
N=C(N)Nc1nc(N)nc(N)n1	ZINC000006523906
Nc1ncnc2ncnn12	ZINC000008618478
Nc1nc[nH]c(=O)n1	ZINC000012358729
CN(C)c1nc(N)nc(CCN)n1	ZINC000012879670
Nc1[nH]c(=O)[nH]c(=O)n1	ZINC000013529033
O=c1[nH]c(NC2CC2)nc(=O)[nH]1	ZINC000013529039
OCCNc1nc(O)nc(O)n1	ZINC000013529042
CCNc1[nH]c(=O)[nH]c(=O)n1	ZINC000013529045
CC(C)Nc1nc(N)nc(O)n1	ZINC000013542453
CCNc1nc(N)[nH]c(=O)n1	ZINC000013542454
Cc1[nH]c(=O)nc(N)n1	ZINC000014989103
Nc1nc(N)nc(C2CCNCC2)n1	ZINC000015442151
O=c1[nH]c(N2CCOCC2)nc(=O)[nH]1	ZINC000016138192
Nc1nc(N)n2c(N)nnc2n1	ZINC000016889989
Nc1nc(C2CC2)[nH]c(=O)n1	ZINC000017820339
Nc1nc(N)[nH]c(=O)n1	ZINC000017968769
CN(C)c1[nH]c(=O)[nH]c(=O)n1	ZINC000018194808
CSc1nc2ncnc(N)n2n1	ZINC000019093141

Nc1[nH]c(=O)[nH]c(=S)n1	ZINC000019819667
Cc1nc2ncnc(N)n2n1	ZINC000020153788
CCc1nc2ncnc(N)n2n1	ZINC000020213761
NNc1nc(N)[nH]c(=O)n1	ZINC000022126362
NNc1ncnc(N)n1	ZINC000022126365
N#CNc1nc(N)nc(N)n1	ZINC000032191040
NC(=O)Nc1nc(N)nc(N)n1	ZINC000034295699
COc1nc(C)nc(NC(=O)N)n1	ZINC000034515534
Nc1ncnc(SCC(=O)[O-])n1	ZINC000034927032
NC(=O)CSc1ncnc(N)n1	ZINC000034927033
Nc1nc(N)nc(CO)n1	ZINC000036761540
Nc1nc(N)nc(NCCO)n1	ZINC000038363127
Nc1nc(O)n2ccnc2n1	ZINC000039095354
Nc1nc(O)nc2ncnn12	ZINC000040175507
CNc1nc(N)[nH]c(=O)n1	ZINC000044787417
CN(C)c1nc(N)nc(CN)n1	ZINC000054957011
O=c1[nH]c(N2CCNCC2)nc(=O)[nH]1	ZINC000205220209

Table 15: Designed Compounds

Structure	Molecule Name	Vina Score kcal/mol
N#C/C(=N\c1nc(=N)[n-]c2nnc([N+](=O)[O-])n12)c1cc[n+](O)cc1	Gen_3_Mutant_7_535434__2	-10.5
N=c1nc(/N=C(\O)c2cc[n+](O)cc2)n2c([N+](=O)[O-])nnc2[n-]1	Gen_1_Mutant_13_482991__3	-10.4
Cc1ccc(NC(=O)Oc2nc(N3CCOCC3)[nH]c(=O)n2)cc1F	Gen_1_Mutant_35_482803__4	-10.2
Cc1cc(C(=O)Nc2[nH+]nnn2CCN=c2nc([O-])[n-]c([O-])n2)ncn1	Gen_1_Mutant_75_26962__3	-10.1
Cc1ccc([N-]C(=O)O[C@H]2[N-]C(=O)NC(=[NH2+])N2)cc1F	Gen_2_Cross_646659__4	-9.8
Cc1nnc2n1[C@@](O)(c1no[nH]c1=N)NC(=[NH2+])N2	Gen_5_Cross_157058__3	-9.7
CC(C)c1cccc(NC(=O)OCCn2c(N)[nH+]c(N(C)C)nc2=O)c1	Gen_1_Mutant_35_728042__2	-9.7
CO[C@@H]1N(C(=O)C[C@@H]2CCC(=O)C2)/C(=N/N)N[C@@H]2[NH2+]NNN21	Gen_1_Mutant_13_20994__1	-9.6
CCc1nn2c(=N)n(C(=O)c3oncc3C)c(=O)[nH]c2[nH+]1	Gen_4_Cross_407229__1	-9.6
N=c1[nH]c(C2CC[NH2+]CC2)[nH+]c(=N)n1C(=O)[N-]c1ccc(Cl)c(C(F)(F)F)c1	Gen_1_Mutant_31_461033__3	-9.6
Cc1c[nH+]oc1C(=O)N1CNc2nc([N+](=O)[O-])nn2C1=N	Gen_5_Cross_134568__5	-9.5

Cc1nnc2[n-]c(=N)nc(/N=C(\O)N3CCOC[C@H]3C)n12	Gen_3_Mutant_31_505985__1	-9.5
Cc1c[nH+]oc1C(=O)n1cnc2nc([N+](=O)[O-])nn2c1=N	Gen_5_Cross_969434__5	-9.4
Cc1cnoc1C(=O)n1cnc2nc([N+](=O)[O-])nn2c1=N	Gen_4_Cross_671065__1	-9.3
Cc1cnoc1C(=O)n1c[nH+]c2[nH+]c([N+](=O)[O-])nn2c1=N	Gen_3_Cross_272672__3	-9.3
Cc1cnoc1C(=O)n1c[nH+]c2[nH+]c([N+](=O)[O-])[nH+]n2c1=N	Gen_4_Cross_29092__3	-9.3
Cc1nnc2[n-]c(=N)nc(N(C([NH-])=O)c3nccn(C)c3=O)n12	Gen_5_Mutant_87_658992__1	-9.2
N#CCC(=O)N1C(=[NH2+])N[C@@H](N)N[C@@H]1CC(=O)[O-]	Gen_4_Cross_692010__1	-9.1
CCc1nc2ncn(C(=O)c3o[nH+]cc3C)c(=N)n2[nH+]1	Gen_2_Cross_80190__3	-9.1
CCc1nc2[nH+]cn(C(=O)c3oncc3C)c(=N)n2[nH+]1	Gen_1_Mutant_23_772577__2	-9.1
N#C[C@H](F)C(=O)n1c(=O)[nH]c2nc([N+](=O)[O-])nn2c1=N	Gen_5_Cross_576088__1	-9
N=c1[nH]c2nnc([N+](=O)[O-])n2c(N)[nH+]1	ZINC000004558145__2	-9
Cc1[nH+]nc2[n-]c(=N)nc([NH2+]C(N)=O)n12	Gen_4_Cross_57267__1	-9
N=c1[nH]c(=N)n2c([N+](=O)[O-])nnc2[nH]1	ZINC000004558145__1	-8.9

N=c1[n-]c(NC(=O)OCCOCC(F)(F)F)n/c(=N/C([NH-])=O)[nH]1	Gen_1_Mutant_12_415455__5	-8.9
N=c1nc([N-]C(=O)CO)[n-]c2nnc([N+](=O)[O-])n12	Gen_4_Mutant_23_361508__3	-8.9
N#CCC(=O)N1C(=[NH2+])N[C@@H]([NH3+])[NH2+][C@@H]1CC(=O)[O-]	Gen_2_Mutant_26_713002__1	-8.9
CO[C@@H]1[N@H+]2NC=N[C@H]2[N-]C(=O)[N@H+]1C(=O)[C@@H](F)C#N	Gen_4_Mutant_21_168407__1	-8.9
N=c1[n-]c(N=C=S)nc2nnc([N+](=O)[O-])n12	Gen_4_Mutant_10_672916__4	-8.9
C=C(C#N)[NH2+]/C(S)=[N+]1\CCN2C(=O)[N-]C(=O)N[C@@H]21	Gen_5_Cross_343011__2	-8.8
[N-]=[N+]=Nc1nc2n[nH+]c([N+](=O)[O-])n2c(=N)[n-]1	Gen_5_Mutant_8_570373__1	-8.8
N=c1nc(N)n2c([N+](=O)[O-])nnc2[n-]1	Gen_5_Cross_226315__5	-8.8
N=c1nc2[nH][nH+]c([N+](=O)[O-])n2c(=N)[nH]1	Gen_3_Cross_358542__4	-8.8
CCN(C(=O)Nc1cccc(SC(F)(F)F)c1)c1nc(O)nc(=O)[n-]1	Gen_1_Mutant_31_975975__3	-8.8
N=c1[n-]c(=N)n2c([N+](=O)[O-])nnc2[nH]1	Gen_4_Cross_790404__1	-8.8
C=C(C#N)[NH2+]/C(S)=[N+]1\CCN2C(=O)NC(=O)N[C@@H]21	Gen_4_Cross_533200__1	-8.8
CN(C)[C@@]1(N2CCNCC2)[N-]C(O)=N/C(=[NH+]\C#N)N1	Gen_2_Cross_393282__5	-8.8

C=C(N=[N+]=[N-])Nc1nc(=N)[nH+]c2cc(C)[nH]n12	Gen_4_Mutant_6_664731__2	-8.8
Cc1[nH+]nc2[n-]c(=N)[nH]/c(=N\C([NH-])=O)n12	Gen_5_Cross_113416__3	-8.7
C/C(O)=N/c1nc(=N)[n-]c2n[nH+]c(C)n12	Gen_5_Cross_145555__4	-8.7
N=c1nc(N)n2c([N+](=O)[O-])nnc2[nH]1	Gen_5_Cross_197997__1	-8.7
[N-]=[N+]=Nc1n/c(=N\C(=[NH2+])N=C=O)[nH]c(=N)[nH]1	Gen_3_Mutant_9_489911__3	-8.7
N=c1[nH]c(=N)n2c([N+](=O)[O-])[nH+]nc2[nH]1	Gen_4_Cross_199626__3	-8.7
CC(=O)Nc1nc(=N)[n-]c2n[nH+]c(C)n12	Gen_4_Cross_556970__1	-8.7
N=c1nc(N)n2c([N+](=O)[O-])nnc2[n-]1	Gen_4_Cross_421361__3	-8.7
CC[NH2+][C@@H](C)[C@@H](C)C(=O)n1c(N)nc(NC(C)=O)[nH+]c1=N	Gen_1_Mutant_13_143186__1	-8.7
N#CCc1ccc(N2C(=[NH2+])N[C@@H](CC(=O)[O-])N[C@H]2[NH3+])cn1	Gen_1_Mutant_18_783381__2	-8.7
C[C@@H](F)COC(=O)[N-]c1nc(=N)[nH+]c2[nH]c(=N)[n-]n12	Gen_1_Mutant_12_156572__1	-8.7
C=C(N=[N+]=[N-])/[NH+]=c1\[nH]c(=N)[nH+]c2cc(C)[nH]n12	Gen_5_Cross_118324__1	-8.6
CCc1nc2n(n1)C(=[NH2+])N(C(=O)c1oncc1C)CN2	Gen_5_Cross_905278__3	-8.6

NC(=O)N[C@H]1NC(=[NH2+])N2NC(=[NH2+])N[C@@H]2N1	Gen_5_Cross_239447__2	-8.6
N=c1[n-]n2c(=N)[nH]/c(=N/C(=O)N=C=S)[nH+]c2[nH]1	Gen_5_Mutant_10_644862__3	-8.6
Nc1nc(OC(=O)[C@@H]2CSCC2=O)nc(N2CCOCC2)n1	Gen_1_Mutant_14_417776__3	-8.6
N=c1[n-]onc1[C@]1(O)NC(=[NH2+])NC[NH2+]1	Gen_4_Cross_285367__1	-8.6
CC(C)[N-]C(=O)[C@@H](C)n1/c(=N/N)nc([NH2+]N)[nH]/c1=N\N	Gen_1_Mutant_18_587834__2	-8.6
[N-]=[N+]=Nc1nc(=N)[nH]/c(=N\C(=[NH2+])N=C=O)[nH]1	Gen_4_Cross_225518__3	-8.6
[NH2+]=C1N[C@@H](OC(=O)[C@@H]2CSCC2=O)N[C@H]([NH3+])N1	Gen_2_Cross_341531__1	-8.6
CC(=O)NC(=S)N1CCn2c([O-])nc(=O)nc21	Gen_3_Cross_52076__4	-8.6
N=c1[n-]c2nnc([N+](=O)[O-])n2c(=N)[nH]1	Gen_5_Cross_998726__2	-8.5
N=c1nc2[nH]nc([N+](=O)[O-])n2c([O-])[nH+]1	Gen_5_Cross_375402__4	-8.5
C=C(C#N)N/C(S)=[N+]1\CCN2C(=[NH2+])N=C(O)N[C@@H]21	Gen_5_Cross_686463__4	-8.5
N=c1[n-]n2c(=N)nc(NC(N)=O)nc2[nH]1	Gen_4_Cross_529873__3	-8.5
CC(C)Nc1nc(O)n/c(=N/c2c[n-]c(=S)c(N)n2)[n-]1	Gen_1_Mutant_87_206816__1	-8.5

N=c1[nH]c(N)[nH+]c2n[nH+]c([N+](=O)[O-])n12	Gen_3_Cross_813651__5	-8.5
N=c1[n-]c(N)nc2nnc([N+](=O)[O-])n12	Gen_2_Cross_321836__5	-8.5
C=C(C)C[C@H](N)C(=O)/N=c1\nc[nH]c(N2CCOCC2)n1	Gen_1_Mutant_13_957976__4	-8.5
COc1ccc(Nc2nc(=N)nc(N3CCCC3)[nH]2)c(F)c1	Gen_1_Mutant_87_650199__1	-8.5
N=c1[n-]c2nnc([N+](=O)[O-])n2c(=N)[nH]1	Gen_1_Cross_143883__1	-8.5
N=c1[nH]c(N)[nH+]/c(=N\C(=[NH2+])N=C=O)[nH]1	Gen_5_Mutant_11_883885__3	-8.4
N=c1[nH]c(N)nc2n[nH+]c([N+](=O)[O-])n12	Gen_5_Cross_41646__3	-8.4
CC(=O)/N=C(\S)N1CCn2c(N=[N+]=[N-])nc(=O)[nH+]c21	Gen_5_Mutant_6_956012__4	-8.2
CC(=O)NC(=S)N1CCn2c(O)nc(=N)[nH+]c21	Gen_5_Cross_872536__1	-8.1
CC(=O)N/C(S)=[N+]1/CCn2c(NCO)nc(=O)[nH+]c21	Gen_5_Cross_274454__5	-7.9
C[NH+](C)[C@H]1NC(=O)N(/C(O)=C/C#N)[C@@H](N)[NH2+]1	Gen_5_Cross_820333__5	-7.7
[NH2+]=C1NC(=O)N[C@H]([N-]/C(O)=C\O)N1	Gen_5_Cross_502762__5	-7.5

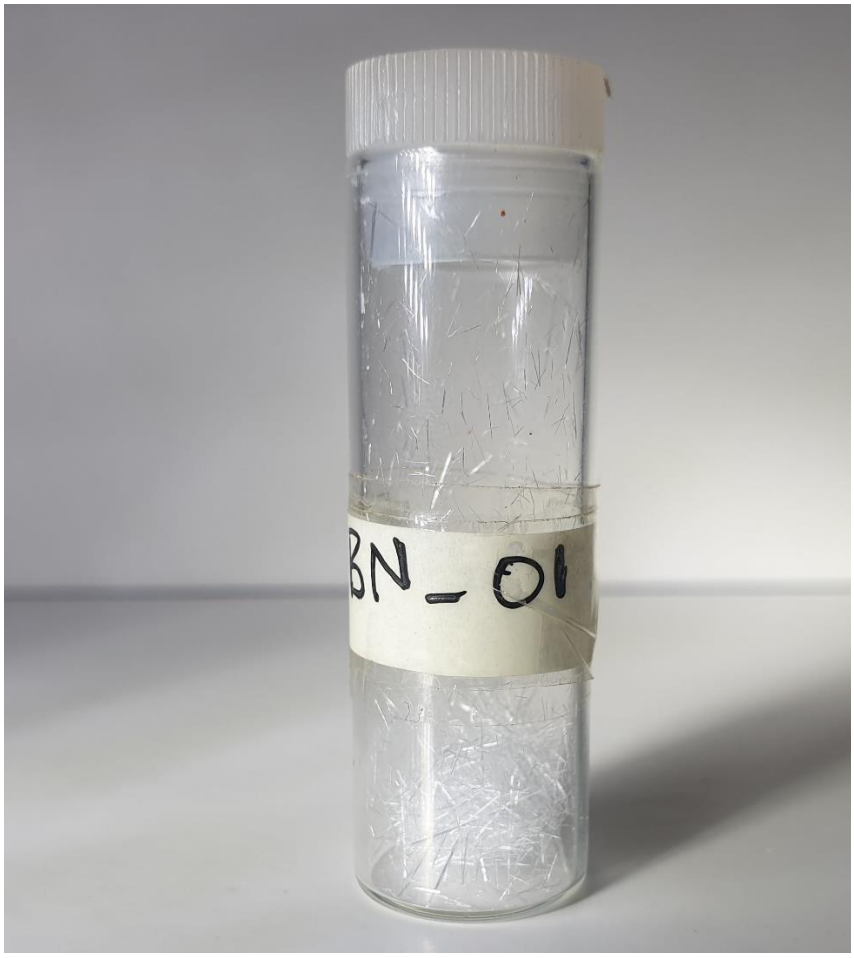


Figure 78: Picture of BN_01 crystals.

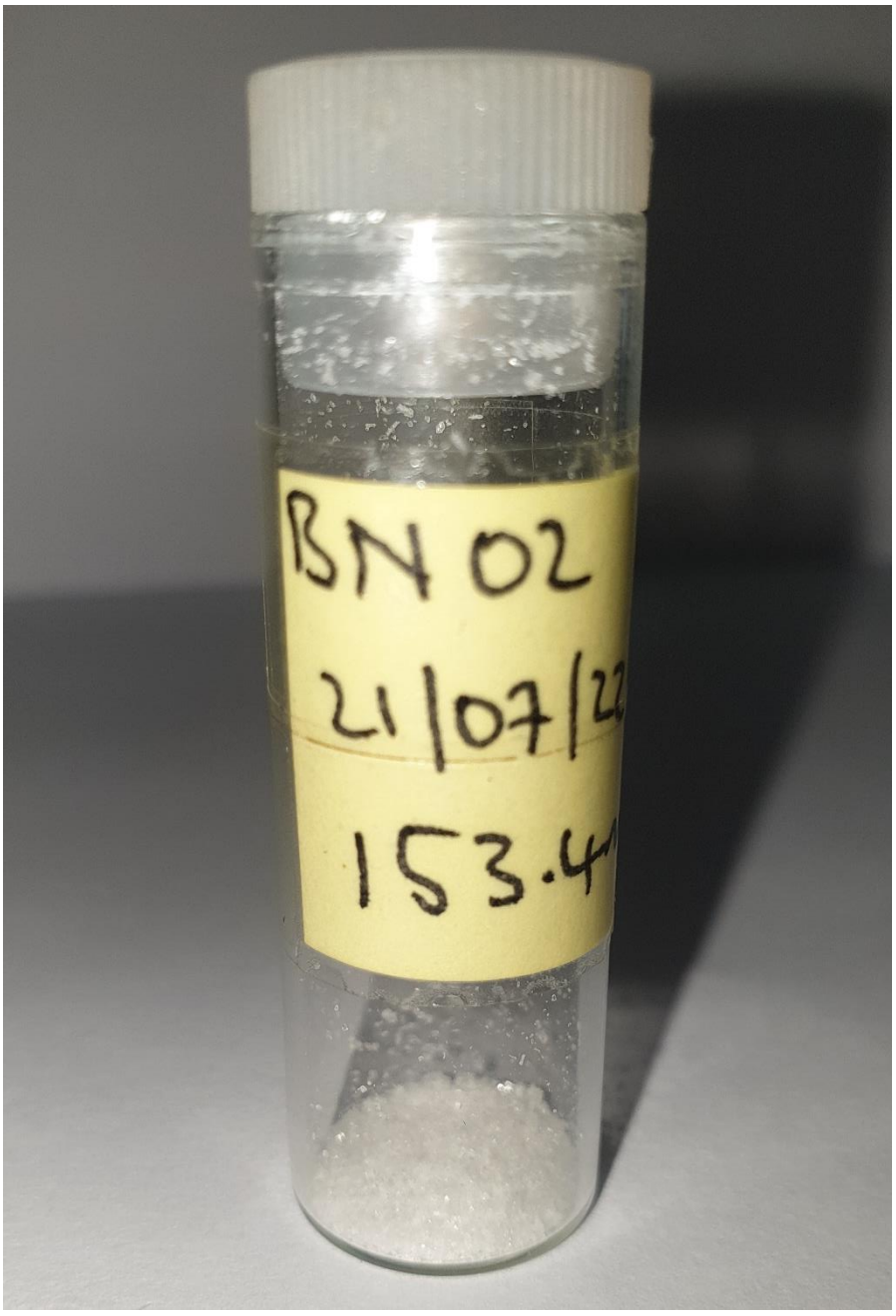


Figure 79: Picture of BN_02 crystals.

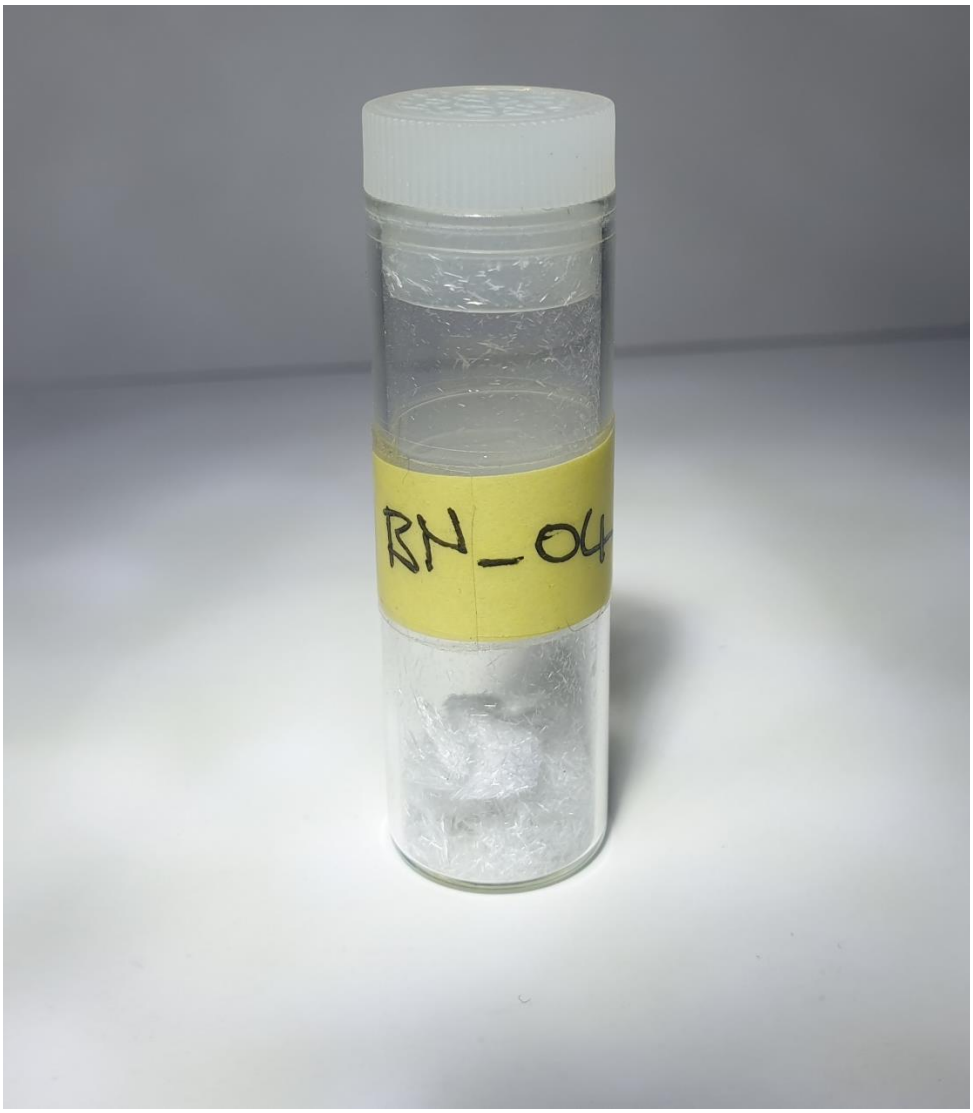


Figure 80: Picture of BN_03 crystals.