AN IN-SILICO STUDY OF THE TYPE II NADH: QUINONE OXIDOREDUCTASE (NDH2). A NEW ANTI-MALARIA DRUG TARGET

a thesis submitted in fulfilment of the requirements for the degree

of

Doctor of Philosophy

in Bioinformatics

of

Rhodes University, South Africa

in the department of

Biochemistry and Microbiology

Faculty of Science

by

Bertha Cinthia Baye

ORCiD: 0000-0002-7822-9740

January 2022

Abstract

Malaria is caused by *Plasmodium* parasites, spread to people through the bites of infected female *Anopheles* mosquitoes. This study focuses on all 5 (*Plasmodium falciparum, Plasmodium knowlesi, Plasmodium malariae, Plasmodium ovale* and *Plasmodium vivax*) parasites that cause malaria in humans. Africa is a developing continent, and it is the most affected with an estimation of 90% of more than 400 000 malaria-related deaths reported by the World Health Organization (WHO) report in 2020, in which 61% of that number are children under the ages of five. Malaria resistance was initially observed in early 1986 and with the progression of time anti-malarial drug resistance has only increased. As a result, there is a need to study the malarial proteins mechanism of action and identify alternative treatment strategies for this disease. Type II NADH: quinone oxidoreductase (NDH2) is a monotopic protein that catalyses the electron transfer from NADH to quinone via FAD without a proton-pumping activity, and functions as an initial enzyme, either in addition to or as an alternative to proton-pumping NADH dehydrogenase (complex I) in the respiratory chain of bacteria, archaea, and fungal and plant mitochondrial. The structures for the *Plasmodium knowlesi*, *Plasmodium malariae*, *Plasmodium ovale* and *Plasmodium vivax* were modelled from the crystal structure of *Plasmodium falciparum* (5JWA). Compounds from the South African natural compounds database (SANCDB) were docked against both the NDH2 crystal structure and modelled structures. By performing *in silico* screening the study aimed to find potential compounds that might interrupt the electron transfer to quinone therefore disturbing the enzyme‟s function and thereby possibly eliminating the *plasmodium* parasite. CHARMM-GUI was used to create the membrane (since this work is with membrane-bound proteins) and to orient the protein on the membrane using OPM server guidelines, the interface produced GROMACS topology files that were used in molecular dynamics simulations. Molecular dynamics simulations were performed in the Centre for high performance computing (CHPC) cluster under the CHEM0802 project and the trajectories produced were further analysed. In this work not only were hit compounds from SANCDB identified, but also differences in behaviour across species and in the presence or absence of the membrane were described. This highlights the need to include the correct protein environment when studying these systems.

Declaration

I Bertha Cinthia Baye, declare that this is my own unaided work, except where duly acknowledged. It is being submitted for the degree of Doctor of Philosophy in Bioinformatics for the Faculty of Science at Rhodes University. It has not been submitted before for any degree for examination in any other University.

BERTHA CINTHIA BAYE

.....

Date: 23 August 2022

Acknowledgements

I would like to extend my sincere gratitude to my supervisor Prof Kevin Lobb and co-supervisor Dr Vuyani Moses. Dr Vuyani thank you for starting this journey with me and all your support and assistance. Prof Lobb thank you for finishing this journey with me, your valuable insights and support and presence when l needed you the most. I would like to thank Rhodes University for providing funding for part of this project.

I would like to thank my sister Vanessa Baye, for always being there for me and her love towards me and l would like to thank my family.

I also would like to acknowledge and thank my partner Kamangu, for the encouragement, the prayers and love towards me.

Thank you, my mother-in-law, for checking on me and support, thank you Br Nelson for always caring for me and treating me like family. Thank you, Lerato Moepeng for constantly checking on my writing progress and for the chats.

Thank you centre for high performance computing (CHPC), Cape town, for providing adequate computational time and resources, without you this work would not have been possible.

I would like to thank my RUBi family for the constructive discussions and encouragements.

Above all l would like to thank God Almighty for always being there for me and looking out for me even in my weakest moments, words cannot even explain my gratitude.

Dedication

~This thesis is dedicated to my son~

Mbae Ermias Maeti

Mummy loves you dearly

Table of contents

Table of Contents

List of figures

List of tables

List of equations

Abbreviations

2D	Two dimensional
3D	Three dimensional
ADMET	Absorption, distribution, metabolism, excretion, and toxicity
AIF-M1	Apoptosis inducing factor
CA/ Cα	carbon alpha
CSML	CHARMM Small Molecule Library
DCC	Dynamic cross correlation
FAD	Flavin Adenine Dinucleotide
HM	Homology modelling
MD	Molecular dynamics
NADH	nicotinamide adenine dinucleotide hydrogen
NDH2	Type II NADH dehydrogenases
PCA	Principal component analysis
PDB	Protein Data Bank
PMP	Peripheral membrane protein
*Pf*NDH2	*Plasmodium falciparum* Type II NADH dehydrogenases
*Pk*NDH2	*Plasmodium knowlesi* Type II NADH dehydrogenases
*Pm*NDH2	*Plasmodium malariae* Type II NADH dehydrogenases
*Po*NDH2	*Plasmodium ovale* Type II NADH dehydrogenases
*Pv*NDH2	*Plasmodium vivax* Type II NADH dehydrogenases
Rg	Radius of gyration
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
SANCDB	South African natural compound database

Amino acid	3-letters	1-letters
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Glyn	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Typtophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Talks and 3-minute thesis presentations

1. **Bertha Cinthia Baye** and Dr Vuyani Moses: *In silico* analysis of the Mg2+ ions coordinating Alternative NADH dehydrogenase enzyme of the *plasmodium* species. Rhodes University Center of postgraduate studies (CPGS) 2019 conference, (oral presentation)

2. **Bertha Cinthia Baye** and Dr Vuyani Moses: Natural compound inhibition of type II NADH:(NDH2) protein in *Plasmodium* species that causes malaria, Rhodes University Center of postgraduate studies (CPGS) 3-minute thesis, (https://www.youtube.com/watch?v=DbOlb6nEN34)

Thesis Overview

This thesis investigated the Type II NADH: dehydrogenase (NDH2) enzyme, utilising bioinformatics computational tools and techniques to study the structure and mechanical function of the protein. The aim was to identify hit compounds against the protein, factoring in the membrane as the protein is a peripheral membrane protein. This document consists of 5 chapters that cover and explain the methods adopted to reach the project aim.

CHAPTER 1

This chapter introduces malaria and its lifecycle and the parasites that spread it to human beings. From this point, targets in the malarial life cycle are presented with particular focus on the NDH2 protein and its human homologue, their sequence structures and function. The methods and packages used in this study are discussed, with reference to how these will be used with respect to the aims and objectives of the study.

CHAPTER 2

This chapter focuses on data retrieval and sequence study of the six protein structures. The 3-D construction of the other *plasmodium* species protein structures is also explored in this chapter.

CHAPTER 3

Drug discovery experiments were performed to identify hit compounds that could affect the functioning mechanism of the NDH2 protein. The protein ligand simulations were performed using GROMACS version 5.1.2 together with the CHARMM36 force filed in the absence of a membrane. Specific techniques were applied to study the MD trajectories.

CHAPTER 4

This chapter focuses on the MD simulation of the transmembrane protein. CHARMM-GUI was used to create the membrane and GROMACS topology files for simulation run and the OPM server was used to orient the protein on the membrane. The MD trajectories were visualised and analysed.

CHAPTER 5

This chapter summaries the results and findings from chapter 2-4 and proposed possible follow up experiments and future work are identified.

1

Literature review

1.1 BACKGROUND

Malaria is caused by a protozoan parasite, phylum Apicomplexa, genus *Plasmodium*, which is passed to human beings through the female mosquito anopheles‟ vector, which inoculates sporozoites into the human host (Cogswell, 1992; Lazarus et al., 2008). Malaria continues to be a fatal disease affecting large numbers of people living in regions with high temperatures, humidity, and rainfall. Africa is a developing continent, and it is the most affected with an estimation of 90% of more than 400 000 deaths reported by the World Health Organization (WHO) report in 2020, in which 61% of that number are children under the ages of five (Leffler et al., 2017; WHO, 2020).

Malaria epidemiology such as community prevalence, infection age-profile, disease syndromes and mortality, are affected by malaria transmission intensity, and this transmission intensity differs geographically (Gething et al., 2011). Malaria parasites infect the midgut, hemocoel, and salivary glands of the mosquito vector, which is then transmitted to the human being where it affects a variety of tissues, including the blood and liver of mammalian host (Matz et al., 2018). The parasite binds to the erythrocyte receptors invading the red blood cells. It is at this point that genetic variation may occur which may influence the progression of the infection (Cowman & Crabb, 2006; Leffler et al., 2017). The difference in the life cycle development of the parasite led to the conclusion that there must be more than one species of malaria parasite responsible for the difference in pattern of cyclical infection. There are a few plasmodial species that cause a significant health threat on humans; *Plasmodium falciparum* has an associated high mortality rate, *Plasmodium knowlesi* emerged as a local but important cause of the disease, predominantly zoonosis without evidence of primary human to human transmission, *Plasmodium malariae*, and *Plasmodium ovale* which both have fewer cases of significant death, and finally *Plasmodium vivax* is a major cause of illness around the globe with underestimated death cases (Ahmed & Cox-Singh, 2015; Cowman et al., 2016; Naing et al., 2014).

It is important to map the burden of malaria in different continents and countries, as this will give people a general sense of the effects of this disease so that they can take appropriate measures towards

controlling malaria. Global maps give a better visualisation of malaria problem and have long been used. Unfortunately, the most affected areas in the world, have the least well-developed health reporting systems and limited resources and knowledge to combat this disease (Snow et al., 1999). The environment plays a vital role in determining the distribution and diversity of the disease, as transmission is vector dependent. For example, the adult mosquito digests the blood meal faster and feeds more frequently in warmer climates and the parasite completes the extrinsic incubation in its anopheles vector in a shorter period of time leading to increase transmission intensity and infective mosquitoes (Minale & Alemu, 2018).

According to WHO‟s definition, malaria control is reducing the disease burden until it is no longer a public health problem.

Figure 1.1 Categorization of countries as malaria free, eliminating malaria, or controlling malaria adapted from (Feachem et al., 2010) with permission from authors and publisher

P. falciparum is an obligately sexual but facultatively outcrossing eukaryotic parasite; meiosis following a union of parasite gametes usually occurs in the midgut of the mosquito and mosquitoes that affect humans infected by a single genotype of the *plasmodium* result in self-fertilization of male and female gametes from the same genotype, instead of outcrossing between in related genotypes (Cerqueira et al., 2017). An unknown and probably variable proportion of invading sporozoites develop into dormant forms, in hepatocytes, the hypnozoites after infection by a mosquito (Cogswell, 1992). *Plasmodium* belongs to the phylum Apicomplexa, known to harbour a non-photosynthetic plastid like organelle of prokaryotic origin known as apicoplast (Foth et al., 2003). For the survival of the parasite, the organelle is indispensable and is the functional site for four major metabolic pathways (Saggu et al., 2016).

There are several emerging pathogens known to have cross-transmitted from the humans and the non-human host, wild primate populations are believed to have the potential to serve as origins and reservoirs of certain human pathogens ranging from viruses to helminths (Jongwutiwes et al., 2004). *P. knowlesi* is a malaria parasite of old-world monkeys thought to rarely affect humans until a report of many cases was submitted from the Kapit Division of Sarawak State, Malaysian Borneo (Cox-Singh et al., 2008). *Plasmodium knowlesi* is a highly pathogenic malaria parasite in humans, with a low pyrogenic threshold and a risk of severe disease that appears at least as high as that of *P. falciparum* in adults (Barber et al., 2013). There is a limit in accurate reporting of the true *knowlesi* malaria cases and its burden has likely been underestimated due to difficulties with microscopic diagnosis (Grigg et al., 2018). Death from *P. knowlesi* is associated primarily with respiratory distress; hypotension and acute kidney injury are common in older adults (Rajahram et al., 2016). Long-tailed macaque (Macaca fascicularis), pig-tailed macaque (Macaca nemestrina), and the banded-leaf monkey (Presbytis melalophos) are the parasite‟s natural reservoir hosts, and several mosquito species belong to Leucosphyrus group considered as vectors for *knowlesi* malaria, namely Anopheles hacker, Anopheles latens, Anopheles cracens, Anopheles balabacensis, Anopheles dirus and Anopheles introlatus (Amir et al., 2018a). Infection with *P. knowlesi* can rapidly progress into severe malaria that can be fatal due to its short asexual cycle of 24h, the treatment guideline developed is based on four core principles: early diagnosis and prompt effective treatment, rational use of anti-malarial drugs, combination therapy and appropriate weight-based dosing depending on the severity of the infection (Amir et al., 2018b).

Plasmodium malariae is considered a minor malaria parasite and its global disease burden is underappreciated with 4-24% cases across a wide geographical distribution. It may also modulate the clinical course and transmission of other species (Woodford et al., 2020). Production of immune complexes in the kidneys and the associated nephrotic syndrome are some of the infections associated with *P. malariae*; the essential lesions are a thickening of the glomerular basement membrane and endocapillary cell proliferation (Collins & Jeffery, 2007). This nephrotic syndrome can lead to progressive renal failure mainly in adolescents or young adults and is associated with high burden of anaemia (Yman et al., 2019). *P. malariae* is mostly found in Africa, tropical areas of Asia and Australasia, and this species has the ability to remain undiagnosed for long periods before multiplication and gametocytaemia occur, hindering malaria eradication efforts in endemic areas (Nabarro et al., 2018). It has a 72h intraerythrocytic intrahepatic development making it unique among the human-infective *plasmodium* species, it can persist in a human host for years to an entire lifetime (Rutledge, Marr, et al., 2017). According to (Rutledge, Böhme, et al., 2017), the *P. malariae* reference genome is 33.6Mb in size, has 6 540 genes and has an average guanine plus cytosine content of 24%.

In 1922, *Plasmodium ovale* was first reported as one of the five *plasmodium* species that cause malaria in human beings, it accounts for between 0.5-10.5% of all malaria cases and is geographically distributed in sub-Saharan Africa, Western Pacific, Timor, and Indonesia (Kotepui, Kotepui, et al., 2020). *P. ovale* demonstrates low parasitemia and has morphological similarities with *P. vivax* and mixed infections, leading to its infections being underestimated in comparison to other *plasmodium* species (Sofi et al., 2008). The presence of latent parasites (hypnozoites) in the liver long after first treatment with anti-malarial drugs can cause an infection relapse (Collins & Jeffery, 2005). This species can cause severe complications and death, complications including acute respiratory distress syndrome (ARDS), renal impairment, jaundice, hypotension, hyperbilirubinemia, pulmonary edema, shock, significant bleeding, and impaired consciousness (Wångdahl et al., 2019). The low species-specific parasitaemia and its short duration of patient infections might be the cause of rarity of *P. ovale* in published studies (Kotepui, Masangkay, et al., 2020). The *P. ovale* complex is formed by two sympatric species *Plasmodium ovale wallikeri* and *Plasmodium ovale curtisi,* it has a mild and often asymptomatic clinical course (Mitchell et al., 2021; Xia et al., 2020). The most sensitive method for detecting *P. ovale* and other malaria causing species is Polymerase chain reaction (PCR), even in cases of very low parasite density; the use of this has led to expansion in research involving *P. ovlae* and proves a far wider distribution in cases than previously anticipated (Mahittikorn et al., 2021).

Plasmodium vivax was once viewed as a benign infection but is now recognised as a cause of severe morbidity and mortality resulting in substantial negative effect on health worldwide; it is the most widespread species geographically and is the second largest contributor to clinical (symptomatic) malaria worldwide (Kevin Baird, 2013). For effective control and potential elimination, it is important to understand the spatiotemporal distribution and clinical burden of *P. vivax.* The species circulates in the peripheral blood at low but transmissible parasite densities, it can also form undetectable dormant liver stages (for example hypnozoites), which periodically awaken to case relapse infection and disease (Battle et al., 2019). *P. vivax* is endemic in Africa, although local surveillance for this parasite is rare (Twohig et al., 2019). 11.9 and 22 million *P. vivax* clinical cases were reported by the world Health Organisation (WHO) per year (Price et al., 2020). The risk of relapse of *Plasmodium vivax* ranges from 5% to 80% depending on the geographic location, primaquine therapy is the drug available to prevent relapse (Baird & Hoffman, 2004). It has been estimated that 38.8% of the population living at risk of *P. vivax* infection are unable to receive safe and effective primaquine therapy because of glucose-6-phosphate dehydrogenase (G6PD) deficiency and reduced CYP2D6 function (Cytochrome, 2019). The glucose-6-phosphate dehydrogenase (G6PD) deficiency is a genetic disorder that causes haemolysis of the red blood cells in response to certain medication or infections, the primaquine drug is a trigger leading to abdominal and back pains, and other symptoms (Parsanathan & Jain, 2020). Rapid diagnostic tests (RDTs) for *P. vivax* have reduced sensitivity in comparison to that used to diagnose *P. falciparum,* making the control and elimination of *P. vivax* more challenging.

This parasite usually circulates at low peripheral parasite densities, which is transmissible to the mosquito vector creating significant challenges for diagnosing infected individuals (Ding et al., 2017). Infected individuals often present with both asexual and sexual parasite stages in the peripheral circulation, resulting in efficient transmission before diagnosis and treatment (Douglas et al., 2013).

All the pathological symptoms of malaria are due to the asexual stage life cycle of the parasite in the host erythrocytes. The intra-erythrocytic parasite grows within the parasitophorous vacuole, divides to form new merozoites, which are released by rupture of the host cell and subsequently invade new erythrocytes. During this growth cycle, the parasite takes up and degrades a large amount of hemoglobin, an essential requirement for its growth, from the host cell. The uptake of host cytosol and hemoglobin occurs through cytostomal vesicles that traverse to the food vacuole (Lazarus et al., 2008; Thakur et al., 2015). Figure 1.2 below summarizes the malaria life cycle.

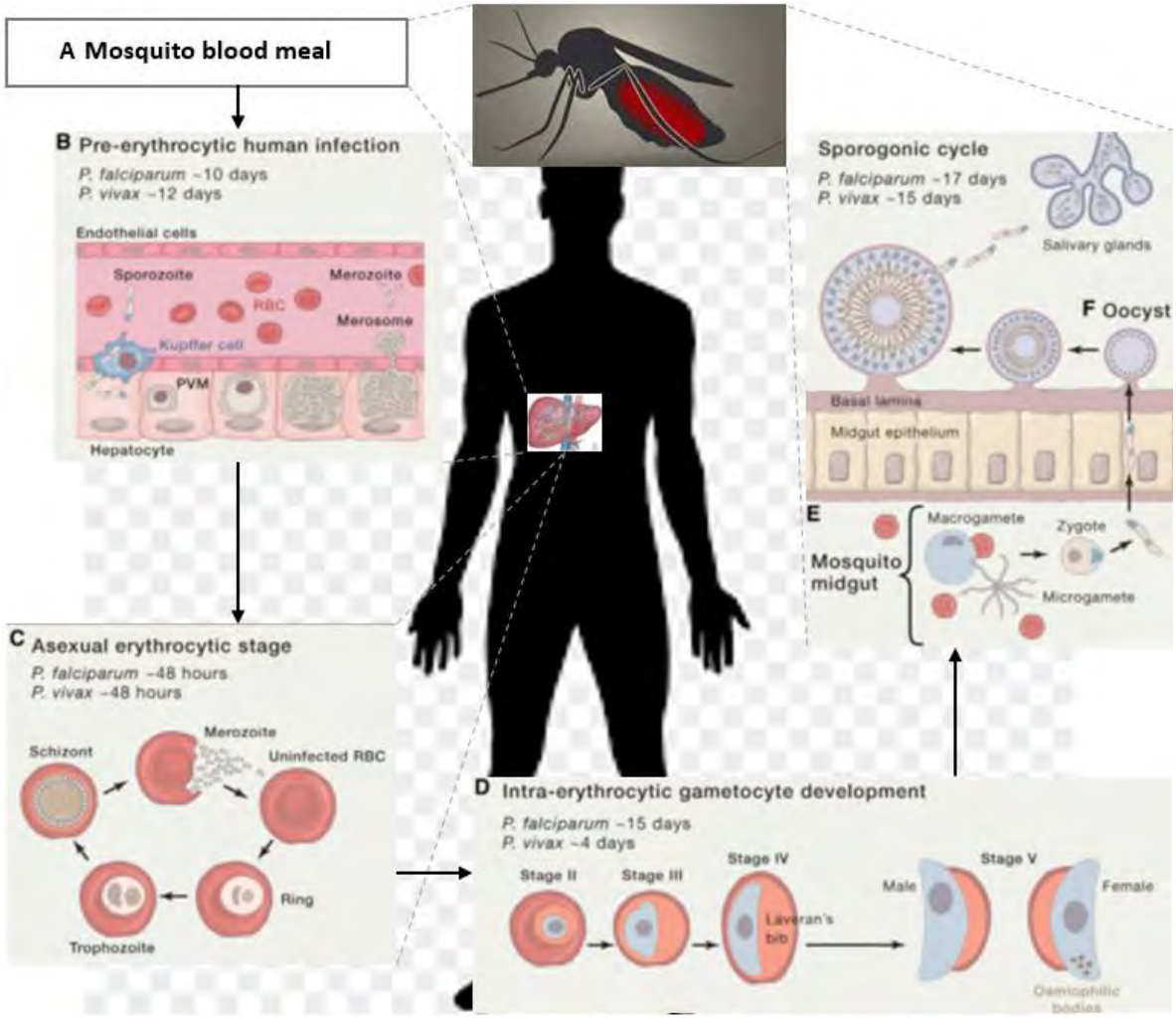

Figure 1.2 Malarial lifecycle figure adapted from, (Cowman et al., 2016) with permission from authors and publisher

Besides administration of highly effective anti-malarial drugs, there are other measures one can take to prevent malaria: measures such as, using insecticide treated nets (ITNs), indoor residual spraying (IRS) and improved diagnostics through the use of rapid diagnostic tests (RDTs) (Bhatt et al., 2015). A third of malaria endemic countries have a malaria elimination plan that is very distinct from regular malaria control strategies, which is the complete depletion of the parasite reservoir by targeting the gametocytes responsible for transmission and destroying silent hypnozoites that may relapse (Howes et al., 2012).

The world health organisation fears a major health crisis if resistance to the first-line drug Artemisinin (ART) anti-malarial drug spreads to India and Africa, as it is now evident in six countries of Southeast Asia (Tilley et al., 2016). The World Malaria Report discussed challenges met due to development and spread of mosquito resistance to insecticides, resistance to Artemisinin and its partner drugs, also funding shortfalls (Barber et al., 2017). There is a need of identifying new and different biochemical pathways in the parasite that can be used as pharmacological targets for new affordable and effective drugs (Siregar et al., 2015). The rise of resistance in *Plasmodium* depends on multiple factors, including the mutation rate of the parasite, the fitness costs associated with the resistance mutations, the overall parasite load, the strength of drug selection, and the treatment compliance (Petersen et al., 2011). The following figure 1.3 shows some of the current anti-malarial drugs on the market.

To describe and understand how a drug affects a specific target in a cell, producing a pharmacological effect in the body, the drug‟s mechanism of action (MoA) can be studied, this can produce information about the safety of the drug and help identify the correct dose of the drug (Swinney, 2011). Other researchers state that target identification of a new drug and elucidating the mechanism of action is important early in drug discovery process, before initiating human clinical trials (Davis, 2020). In the pre-molecular era, drug MoA was explored against whole tissues and rarely on isolated proteins; however now almost all drug discovery begins with activity of molecules on a molecular target (Gregori-Puigjané et al., 2012). Toxic monomeric α-Hematin (ferriprotoporphyrin IX) is released as the by-product when haemoglobin is catabolised by *plasmodium*, this product catalyses reactive oxygen species (ROS) production deposited on and damaging cell membranes (Pillat et al., 2020).

For uncomplicated malaria in Africa Amodiaquine is used as the first or second line of treatment and is predicted to have similar mode of action as that of chloroquine. The drug is a more pharmacoeconomical member of the chloroquine hybrids and has a related structure as chloroquine and is children friendly as it has a palatable taste (Aliyu et al., 2021). Prophylactic use of Amodiaquine has been associated with fatal cases of agranulocytosis and hepatitis (Adjei et al., 2008).

The artemisinin-based combination therapies (ACTs) are the first-line medicines for the treatment of malaria worldwide (Windle et al., 2020a). Artemisinin is a sesquiterpene lactone compound derived from the sweet wormwood plant (*Artemisia annua* L) and has a unique chemical structure (J. Wang et al., 2019). For the treatment and management of uncomplicated *P. falciparum* malaria the artemisinin-based combination therapies (ACTs) are the mainstay drugs and the mechanism of action for artemisinin-based combination therapies is unclear (O'Neill et al., 2010). There are possibilities that have been suggested which involve the degradation of the endoperoxide bridge in a heme-dependent process to form carbon centred radicals which then alkylate multiple targets including heme and proteins (Bergman et al., 2019). Other researchers suggest the alkylation of heme to be the trigger and target of artemisinin drug, also the drug influences reactive oxygen species (ROS) leading to the death of the parasite (N. Ma et al., 2020). Despite the artemisinin-based combination therapies (ACTs) being safe drugs, it should be noted that there is a possibility of adverse effects such as hepatitis and delayed haemolytic anaemia (Pousibet-Puerto et al., 2016).

The mechanism of action of Atovaquone involves the blocking of the mitochondrial electron transport chain at complex III of the respiratory chain of *plasmodium,* inducing the collapse of the mitochondrial membrane potential blocking energy supply to the parasite (Srivastava et al., 1997). It is a competitive inhibitor of ubiquinol binding that inhibits the complex III (cytochrome bc 1) of the electron transport chain leading to the loss of the membrane potential and parasite death (Egwu et al., 2021). The drug has potent sub-nanomolar activity against the erythrocytic *P. falciparum* in vitro and has activity against the liver stage and erythrocytic parasites (Bakshi et al., 2018). Atovaquone is combined with prodrug proguanil as it has insufficient efficacy in malaria treatment due to recrudescence and resistance (Verdaguer et al., 2021). Atovaquone causes few side effects and is generally well tolerated, however it should be noted that one might experience mild rash, fever, vomiting, diarrhoea, abdominal pain and headache (Nixon et al., 2013).

Chloroquine interacts with membrane stability and alter signalling pathways and transcriptional activity, resulting in inhibition of cytokine production and modulation of certain co-stimulatory molecules, it also interferes with lysosomal activity and autophagy (Schrezenmeier & Dörner, 2020a). Chloroquine inhibits the heme polymerase in malarial trophozoites, preventing the conversion of heme to hemazoin leading to accumulation of toxic heme into the parasite eventually killing it (A. H. Kumar, 2020). This drug is one of the safest and cheapest drugs of all time. Chloroquine can be protonated in the acidic environments of the low pH organelles within the cell and is a weak diprotic base (pK_a = 10.1), it exhibits its main anti-malarial activity in the asexual stages (Pillat et al., 2020). The drug accumulates inside the food vacuole and interferes with the detoxification of heme a haemoglobin catabolism product (Hyde, 2002). Hydroxychloroquine is the derivative of chloroquine (Carrière et al., 2020). The mode of action is probably context-dependent on the inflammatory conditions or the affected tissues or organs, as with various therapeutic interventions of the immune system

(Schrezenmeier & Dörner, 2020b). Chloroquine induce a dysfunction of the lysosomal enzymes, leading to the impairment of intracellular degradation processes in conjunction with the accumulation of pathological metabolic products (glycogen and phospholipids) (Tönnesmann et al., 2012).
The lumefantrine mechanism of action has not been fully interpreted and has been implicated in detoxification of products of heme degradation (Santos et al., 2021). The drug is a weak base aryl amino alcohol and highly lipophilic, it is responsible for elimination of residual parasites as it is a blood schizonticide (Resende et al., 2019). The hemozoin formation in the digestive vacuole of the erythrocytic-stage malaria parasite can be inhibited by the drug and the hemozoin serves as a crystalline repository for the sequestration and detoxification of ferri-protoporphyrin heme molecules released by hemoglobin digestion (Windle et al., 2020b). According to (Assefa et al., 2010), there is need for further study on blood concentration of lumefantrine and pharmaco-vigilance on the drug toxicity as it causes mouth ulcer in children.

The Mefloquine anti-malarial drug is a nuclear factor kappa B (NF-kB) inhibitor that blocks the activation of IkBα kinase, leading to reduction of IkBα degradation, decrease of p65 phosphorylation and suppressed expression of NF-kB target genes in colorectal cancer (CRC) cells (Xu et al., 2018). The mechanism of action of mefloquine as an anti-malarial drug includes the lysosomal/vacuole inhibitor, a disruptor of the mitochondrial proton motive force, a protein translation inhibitor, and a purine nucleoside phosphorylase (PNP) inhibitor (Montoya et al., 2019). As a quinoline-based compound, the drug modulates several cellular phenomena, such as alteration of membrane potential, induction of oxidative stress, imbalance of ion homeostasis, disruption of metabolism, failure of organelle function, leading to cell cycle arrest and programmed cell death (Ghosh et al., 2021). Mefloquine (a 4-quinolinemethanol synthetic quinoline) is recommended as a prophylactic for malaria endemic areas and has long retention in the human body and high efficacy (Martins et al., 2021). Mefloquine package insert mentions psychiatric symptoms including abnormal dreams, anxiety, paranoia, agitation, confusion, memory impairment and hallucination (Cameron Ritchie et al., 2013).
There is no direct evidence that exists of piperaquine (PQ) mechanism of action at the molecular level as few investigations of its clinical pharmacokinetics has been conducted (Sacchi et al., 2019). This drug is an essential component of the mainstay artemisinin-based therapies used for the treatment of malaria globally. The H NMR spectroscopy was used to investigate the interaction of the drug with heme, finding the formation of compound-heme complexes as the mechanism of action (R. Ma et al., 2019). Mild headache, listlessness, nausea and dizziness were some of the side effects listed in the first human studies of piperaquine (Bouth Denis et al., 2002).
Primaquine is effective against all exoerythrocytic forms of the parasite used in conjunction with other anti-malarial treatments for *P. vivax* and *P. ovale*, it is the available transmission-blocking drug, displaying a marked activity against gametocytes of all human malaria species (Vale et al., 2009).

This drug is not effective against endo-erythrocytic forms of plasmodia and is co-administered with blood-schizontocides (Baird & Hoffman, 2004). Primaquine prevents primary parasitemia by destroying the *plasmodium* parasite in the liver before reaching the bloodstream and cause disease and is therefore used as the primary prophylactic (Hill et al., 2006). Primaquine prevents relapse in *P. vivax* and *P. ovale*, however it is not given to patients with glucose-6-phosphate dehydrogenase (G6PD) deficiency due to haemolytic toxicity (Ashley et al., 2014). The primary targets of primaquine are the ROS-labile Fe-S groups and yeast respiratory growth (Lalève et al., 2016). Primaquine causes dose-limiting side effects such as haemolytic anemia and methemoglobinemia mainly in patients with glucose-6-phosphate dehydrogenase (G6PD) deficiency (Ganesan et al., 2009).

The hepatic microsomal enzyme cytochrome P450 (CYP) 3A4 metabolises the quinine drug, the phenotypic variability in debrisoquine metabolism depends on CYP2D6 enzyme system and does not influence oral elimination of the drug (Krishna & White, 1996). Structure-activity studies show that the conformation around atoms C-8 and C-9 of the cinchona alkaloids, particularly the direction of the aliphatic N-H and (9C)0-H bonds relative to each other, are crucial to antimalarial activity (Karle et al., 1992). The drug is isolated from the bark of several species of Cinchona and Remijia trees a stalwart of anti-malarial chemotherapy and chemoprophylaxis (G. Z. Yang et al., 2019). Quinine and its derivatives act on a variety of ion channels including several types of potassium channels, members of family of ligand gated ion channels such as the 5-HT3-type of serotonin receptor and nicotinic acetylcholine receptors (nAChR) (Gisselmann et al., 2018). The Quinine drug is toxic to the auditory system by commonly inducing hearing loss and tinnitus, presumably due to its ototoxic effects on disruption of cochlear hair cells and blockade of ion channels of neurons in the auditory system (Zou et al., 2018). It evokes different physiological consequences by interacting with the striatum neurons and sympathetic neurons and modulates neuronal physiological by interacting with intrinsic ion channels or synaptic targets (Hirasawa et al., 2000). According to (Punihaole et al., 2018), the ability of quinine to bind to DNA and potentially inhibit transcription and translation has been considered as the mode of action for its anti-malarial activity.

Toxic effects of most of these drugs, offset the robust usage in humans, hence new anti-malarial drug targets are considered. The physical abnormalities in the nervous system, cardiovascular system, neuromuscular junctions, and the liver caused as side effects by some of these administered drugs is a cause of concern for the continued use by the public (Mulenga-cilundika et al., 2020; Tse et al., 2019). Due to these side effects, it is necessary to find more effective and safe drugs that can be used to eradicate malaria, in this study we will look at potential hit compounds on the new anti-malarial drug target type II NADH: quinone oxidoreductase (NDH2) protein.

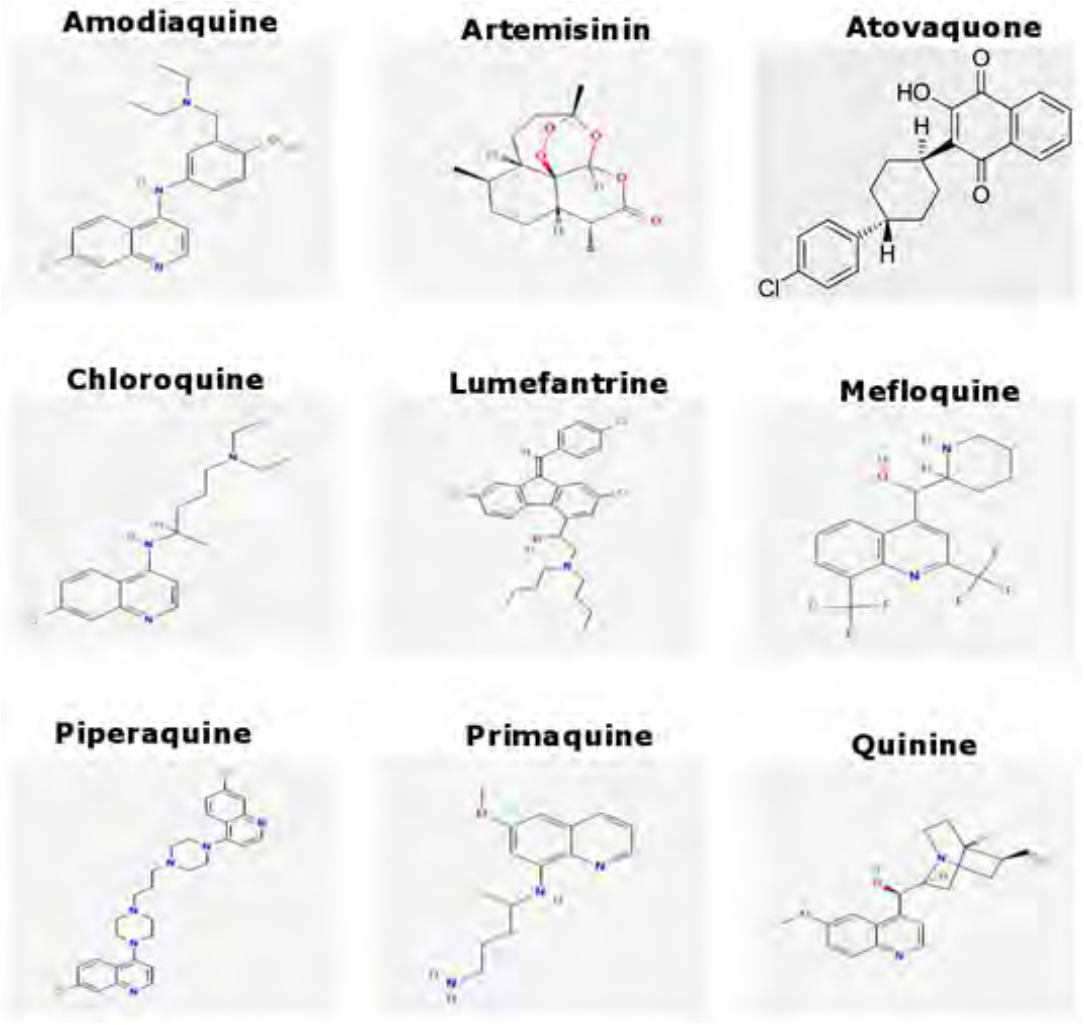

Figure 1.3 A list of few anti-malarial drugs currently on the market their name and structure

1.1.1 Type II NADH: quinone oxidoreductase (NDH2) and Apoptosis-inducing factor 1 (AIF-M1)

Due to the increase in drug resistance, it has been of importance to develop new and effective drugs that can be used to eradicate this disease. Type II NADH: quinone oxidoreductase (NDH2) is a 40-70kDa single subunit monotopic membrane protein that catalyses the electron transfer from NADH to quinone via FAD or FMN without a proton-pumping activity (Nakatani et al., 2017). It functions as an initial enzyme, either in addition to or as an alternative to proton-pumping NADH dehydrogenase (complex I) in the respiratory chain of bacteria, archaea, and fungal and plant mitochondria (Yamashita et al., 2018). NDH2 is a homodimer, peripheral membrane protein containing chain A and chain H with a FAD cofactor and four magnesium ions on each chain and the PDB ID is 5JWA (Y. Yang et al., 2017).

NDH2 is distinct from two other functionally related NDHs, the proton pumping type NDH (NDH-I/complex I) and the sodium pumping type NDH (NQR) (Nakatani et al., 2020).

The complete malaria genome project (Gardner et al., 1999) showed that *P. falciparum* mitochondria lacks the NADH: dehydrogenase which is present in most mammalian mitochondria and it contains the alternative complex I (type II NADH: dehydrogenase) (Biagini et al., 2006). The Type II NADH: quinone oxidoreductase (NDH2) enzymes have emerged as potential drug targets against *Plasmodium falciparum*, the enzyme carries out the NADH: quinone oxidoreductase activity in the electron transport chain (Iwata et al., 2012). Type II NADH dehydrogenases do not pump protons across the membrane in contrast to the type of NADH dehydrogenases present in humans, contributing to the membrane electrical potential ($\Delta\psi$) (Feng et al., 2012).

One important property of NDH2 enzymes is that they are membrane-bound, and, in particular, are monotopically bound to the cell membrane (Lencina et al., 2019). NDH2 is non-proton pumping and bacteria, for example, prefer to use NDH2 instead of the proton-pumping NDH-1 in which both enzyme complexes are found in the genome, because there is higher rate of ATP synthesis at a lower energetic efficiency of the respiratory chain when NDH2 mediate NADH oxidation (Heikal et al., 2014). NDH2 does not transfer protons across the membrane due to its structural and catalytic simplicity and does not contribute to the generation of proton motive force, which might be advantageous for some organisms in the maintenance of the NAD+/NADH redox balance and generation of ATP because its catalytic function cannot be compromised by proton motive force back-pressure (Blaza et al., 2017).

Membrane receptor proteins are responsible for signal transduction across membranes, for communication between cells and between a cell and its surrounding environment (Mori et al., 2016). Transmembrane proteins play a role in energy production, regulation, and metabolism in living cells; this might be a contributing factor why half of present-day drugs have some effect on transmembrane proteins (Kozma et al., 2013). About 25% of the human genome might code transmembrane proteins (Fagerberg et al., 2010). Substrate transports and signal transductions are key functions of membrane proteins. Atomic structural information of membrane transporters and receptors contribute greatly to the understanding of the biological phenomena; however, structure determination using X-ray, NMR and cryo-EM techniques is challenging compared to soluble proteins (Kozma et al., 2013; White, 2004). There are a few identified NDH2 inhibiters with moderate activities such as, Iodonium derivatives, flavones, quinolones, phenothiazines, and nanaomycin A and polymyxin B (Biagini et al., 2006; Fang & Beattie, 2002). A well-validated, specific NDH2 inhibitor would provide a novel tool to study NDH2 function as well as enable preclinical pharmacological validation of NDH2 as a target for TB therapy, *plasmodium* malaria and other micro-organisms that causes disease in human beings (Harbut et al., 2018.).

NDH2 catalyses the transfer of electrons from NADH into the mycobacterial respiratory pathway and has been proposed to be targeted by several early-stage inhibitors (Murugesan et al., 2018). NDH2 contains two Rossmann folds which are responsible for binding NADH and housing the co-factor FAD (noncovalently), and which are central to NADH oxidation. These domains are followed by a C-terminal membrane-anchoring domain in which the quinone binding site (Q-site) is localised, allowing electron transfer from the reduced FAD to the acceptor quinone pooled in the membrane (Bridges et al., 2018). This is understood in the context that NDH2 is a peripheral membrane protein that oxidizes NADH and reduces quinones; a central feature of the respiratory chain in many species except mammals (Blaza et al., 2017). The interface between the protein monomers is extensive, with a buried surface area of *ca* 1.920Å^2 per monomer, which is ~9% of the solvent-accessible surface of each monomer and suggests that the homodimer is physiologically relevant (Iwata et al., 2012).

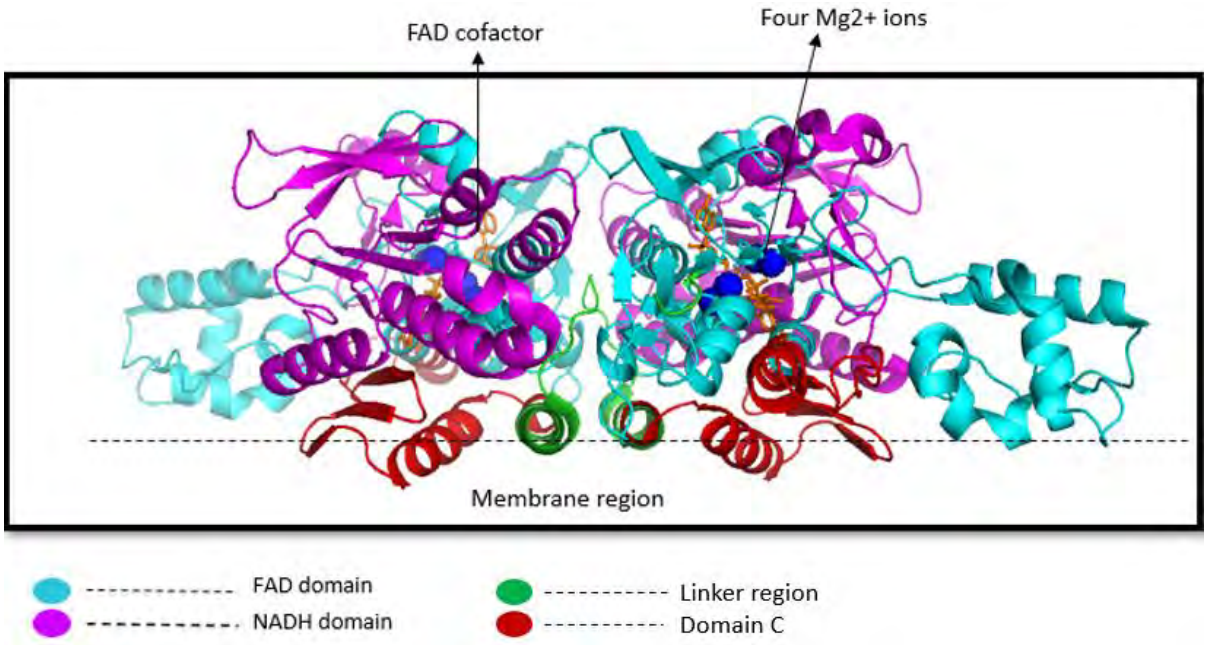

Figure 1.4 The homodimer 5JWA NDH2 *Plasmodium falciparum* crystal structure, showing its different domains, chain A and H, the FAD cofactor in orange and the four magnesium ions represented as blue spheres ion each chain

Two protein sequences are considered homologous if they descend from a common ancestor (Ponting & Russell, 2000). They tend to have similar structure, depending on the degree of divergence, similar functions, cellular localisation, or ligands (in our case cofactors) (Klim et al., 2018; Mills et al., 2018). The apoptosis-inducing factor (AIF) is a mitochondrial flavoprotein oxidoreductase that contributes to cell death, programmes, and participates in the assembly of the respiratory chain (Romero-Tamayo et al., 2021; Sahu et al., 2021). AIF-M1 is considered a flavoprotein as it contains, as an electron acceptor flavin adenine dinucleotide (FAD) an electron carrier like NAD in its action (Hammerstad & Hersleth, 2021; Lienhart et al., 2013). Deficiency of AIF causes mitochondrial dysfunction causing

muscle atrophy and neurodegeneration in model organisms and humans (Bano & Prehn, 2018).According to (Elguindy & Nakamaru-ogiso, 2015), Apoptosis-inducing factor 1 (AIF-M1) is an NDH2 enzyme which also functions as NDH2. They also showed that the human protein has NADH-quinone oxidoreductase activity as such this protein was used as the human homologue in this study. AIF-M1 shown in figure 1.5 below is a mitochondrial FAD-containing monomer, NADH-dependent oxidoreductase that plays a role in oxidative phosphorylation (Delettre et al., 2006; Miramar et al., 2001; Sevrioukova, 2011). AIF binds NAD(P)H and forms a long-lived charge-transfer complex (CTC) between its FAD cofactor and the NAD(P)H nicotinamide moiety and therefore is modulated (Brosey et al., 2016). AIF and NDH2 possess FAD and NADH molecules at similar binding regions, supporting that AIF may have a common functional ancestor with NDH2 (Trisolini et al., 2019). AIF has limited access to other cellular membranes and is constrained to the mitochondria, it can lead to severe mitochondrial diseases when mutated. According to (Martin Vabulas, 2021), the AIF-M1 is anchored in the inner membrane of mitochondria exposing its C-terminal part to the intermembrane space. The protein is proteolytically processed to be released from the membrane and subsequently translocated to the cytosol and the nuclei upon apoptosis induction. A substantial fraction approximately 30% of the protein residues at the outer mitochondrial membrane on the cytosolic side and might be sufficient to cause cell death even without an input from the intramitochondrial pool (S.-W. Yu et al., 2009).

AIF-M1 resides in the mitochondria and plays a role in mitochondrial regulation after hypoxia-ischemia not only through the loss of the mitochondria‟s energy-producing function but also through the protein release into the cytosol and subsequent translocation into the nucleus where it induces apoptosis through chromatin condensation and large-scale DNA fragmentation in a caspase-independent manner (Rodriguez et al., 2020). It is involved in cellular bioenergetics, assisting in stabilization and maintenance of the electron transport chain (ETC), the protein‟s mutation in humans has been associated with impaired ETC function resulting in mitochondrial encephalomyopathy (Kadam et al., 2020). The residue arginine (R201) found in the FAD binding pocket confers conformational stability to the flavoprotein. The protein can induce nuclear apoptosis in the absence of caspases by triggering chromatin condensation, its advantage is that its effects are not inhibited by pharmacological caspase inhibitors in the cytoplasm (H. Wang et al., 2020). Apoptosis inducing factor has a vital role in maintenance of mitochondrial morphology and energy metabolism (Letkovska et al., 2021).

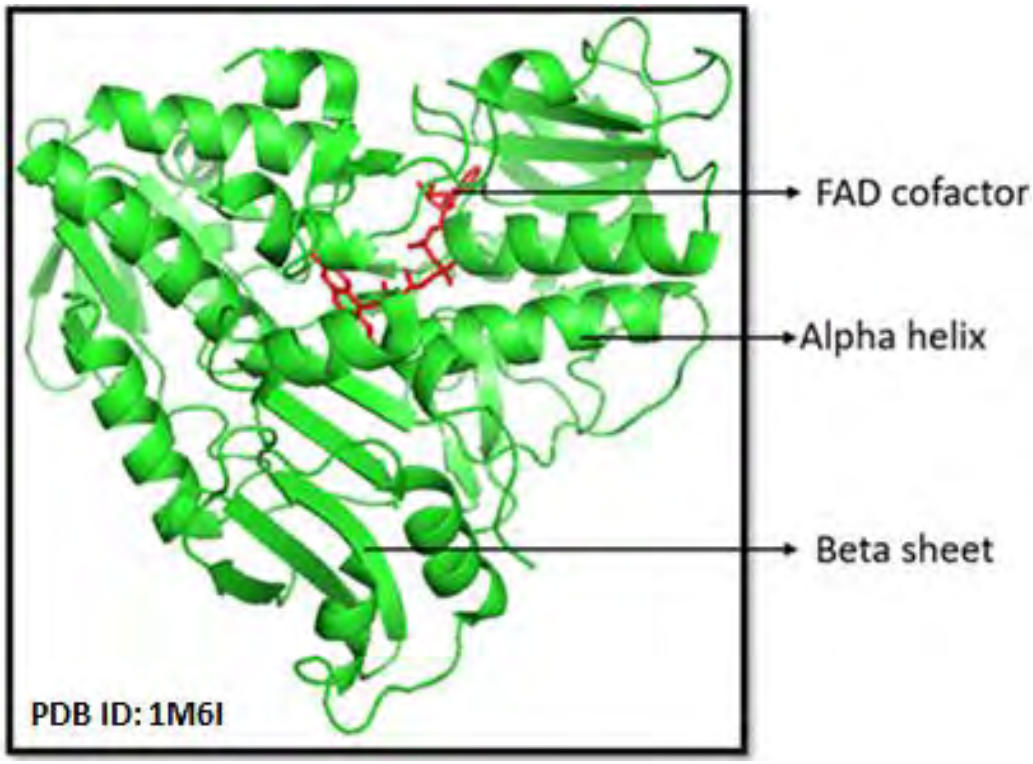

Figure 1.5 The crystal structure of AIF-M1 protein and the FAD cofactor shown in red

1.1.2 Proteins and membrane proteins

Protein structures are 3-dimensional (3-D) arrangements of atoms from amino acid chain molecule derived from a combinations of 20 amino acids which build the protein structures. Alpha helices and beta sheets as shown in figure 1.5 above, are the stable folding patterns that make a secondary protein structure, and they adopt these 3-D structures due to the hydrogen bonding between amino groups and carboxyl groups in neighbouring regions of the protein chain. When the protein backbone adopts a right handed helical conformation with 3.6 residues per turn and a set of hydrogen bonds formed between the main chain carbonyl (CO) of the i^{th} residue and the main chain NH of the $(i+4)^{th}$ residue, an alpha (α) helix is formed (Sun et al., 2004). A beta (β) sheet consists of at least two β-strands, each approaching an extended backbone conformation with dihedral angles confined to the region where the φ torsion angle is between −60° and −180°, and the ψ torsion angle is between 30° and 180° (Sun et al., 2004).

The function and structure of a protein will vary widely as determined by the arrangement of amino acids in a chain. There are components that can affect the function or conformation of the protein, for example temperature, pressure, or pH value. The final folded structure or conformation of a protein is usually the one in which the free energy is minimised. The cell unit expresses the protein with specific functionality, topology and different proteins will react differently to change in environment and this will affect their longevity and function (Consortium, 2019).

Membrane proteins position themselves in a lipid bilayer and contain stretches of residues that are exposed to the hydrophobic environment at the core of the membrane; these transmembrane segments usually have one of two structure types: α-helices or β-strands (Kelm et al., 2010; Smith et al., 2004).

The structure and function differ from protein to protein, some are embedded completely and others partially in membranes, whereas some interact with other proteins or are standalones. The proteins under investigation in this study are membrane proteins (specifically peripheral membrane proteins) and studying membranes is important in the context of the effect of protein environment on the dynamics and behavior of the protein. There are many functions served by the biological membranes such as acting as permeability barriers, energy transduction in respiratory and photosynthetic systems, transporting information between the exterior and interior phases, they can be excitable, they can also give a cell its individuality and finally but not least they support catalytic functions, hence they are important drug targets (Engel & Gaub, 2008). Membrane proteins have a large variety of fundamental physiological functions. Different mechanisms control the formation of membrane proteins and their functionality, including post-translational modifications such as proteolytic ectodomain shedding which is a form of limited proteolysis and thus an irreversible post-translational modification (Lichtenthaler et al., 2018). There are more than 350 unique membrane protein structures solved by X-ray crystallography, available at Protein Data Bank (I. Moraes et al., 2014). Depending on the nature of their interaction with the membrane, these proteins can be classified into two main categories: integral and peripheral membrane proteins. Peripheral membrane proteins are highly unique amphipathic proteins that interact with the membrane indirectly, using electrostatic or hydrophobic interactions, or directly using hydrophobic tails or GPI-anchors (Boes et al., 2021). Not many membrane proteins have been solved due to certain challenges such as extraction of the membrane proteins from the native lipid environment while maintaining structural and functional integrity. Structural characterization has conventionally been achieved using detergents, which are amphipathic molecules that keep membrane proteins in solution by forming micelles around the hydrophobic domains, however this fails to mimic the complexity of the native membrane environment (Autzen et al., 2019).

Each membrane is made up of a phospholipid double layer and houses proteins or protein complexes providing specific communication channels between the cell and its environment. Cell membranes are composed of hundreds of different lipid types that are asymmetrically distributed between the two surfaces, with many membrane proteins embedded covering an estimated membrane area as large as 30% at a lipid/protein ratio of about 50-100 (Corradi et al., 2018). The native host environment of all membrane proteins is highly dynamic and heterogeneous and this constitutes a bottleneck for their direct structural analysis and poses a considerable challenge for the preparation of proteins for many biophysical methods (Hoi et al., 2021; Urner et al., 2020).

Peripheral membrane proteins (PMP) are located directly on the membrane surface and adhere temporarily to the biological membrane with which they are associated, membrane proteins comprise soluble domains that extend into the membrane (Fisette et al., 2016). Little is known about peripheral membrane protein mechanism of function, as they only bind transiently to a membrane, making it difficult to calculate binding conformations as they can be found in states attached to the membrane as well as completely solvated in water (Corey et al., 2020). PMP are attractive propositions in drug search for diseases such as tuberculosis, cancer and parasitic infections, however their highly amphipathic nature and their dependencies on lipid interaction limits structure and function investigation and ability to be drug design target (Boes et al., 2021).

In terms of computational work on membrane proteins, it is important to orient the protein correctly on the membrane before minimization or performing any MD simulation, Orientations of Proteins in Membranes (OPM) database can be used to find the correct orientation of the protein. This server or database provides spatial positions of membrane protein structures with respect to the lipid bilayer, the implicit solvation model of the lipid bilayer is used to calculate the position of the proteins and the results are verified against experimental studies of spatial arrangement of transmembrane and peripheral proteins in the membranes (Kutateladze & Overduin, 2001; A. L. Lomize et al., 2011; M. A. Lomize et al., 2012). How this database was used in this study will be discussed in chapter 4 of the document. MD simulations may be used to analyse the protein-membrane interaction (A. G. Lee, 2018). There are different types of lipids which include dimyristoylphosphatidylcholine (DMPC), dilauroylphosphatidylcholine (DLPC), 1-palmitoyl-2-oleoyl-snglycero-3-phosphocholine (POPC), 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC), dipalmitoylphosphatidylethanolamine (DPPE), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE), and 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) (Jo et al., 2009).

1.2 Methodology adopted and used in this study

As this project aimed to identify hit compounds that could be used as anti-malarial drugs to eradicate the disease without harming or with less side effects to humans, bioinformatics techniques and procedures were applied and followed to achieve this goal. The figure 1.6 below shows the techniques that were used together with their associated tools using personal computers and servers; we will further explain how these techniques were applied together with their relevance. NCBI, RCSB and UniProtKB webservers were used to retrieve data. MAFFT, MUSCLE and PROMALS3D tools were used for multiple sequence alignment study and MEME was used to calculate the protein motifs. The MEGA tool was used to calculate and build phylogenetic trees. Homology modelling was performed in MODELLER whereas, AutoDock Vina was used for docking studies employing the SANCDB compounds. Molecular dynamics simulations was used to study of NDH2 free from the membrane using GROMACS, and with a membrane bound to the proteins CHARMM-GUI was used. MDM-task

was used to calculate the Principal Component Analysis (PCA) and MD-task was used to calculate the dynamic cross correlation of the proteins.

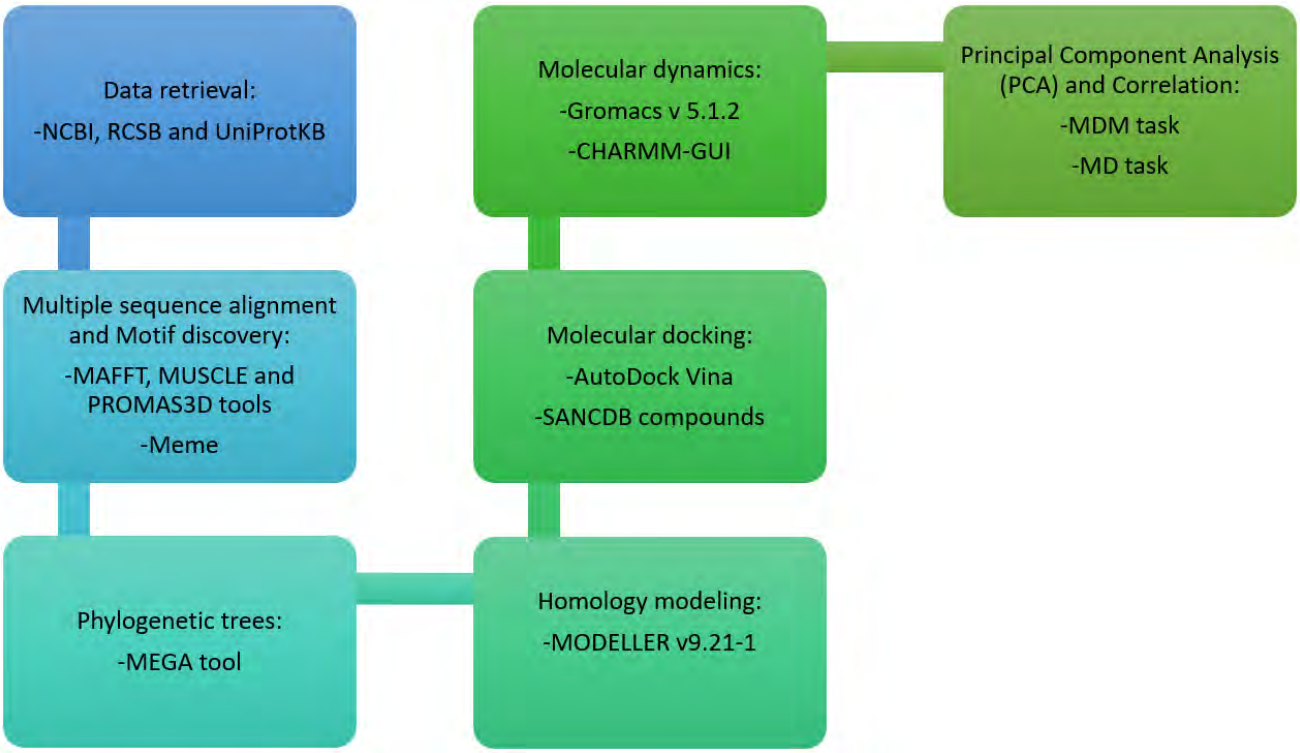

Figure 1.6 Methodology applied in the study, together with the tools and webservers used

1.2.1 3-D protein structure prediction

In this study we focused on five *plasmodium* species that causes malaria in humans, *P. falciparum* had a crystal structure with (PDB ID: 5JWA) solved in the databases and in the absence of the crystal structures for the other *plasmodium* it was necessary to generate appropriate models. Homology modelling is a structure prediction method dependent on the observation that the structural conformation of a protein is more conserved than its amino acid sequence (Cavasotto & Phatak, 2009). Crystal structures present in the databases can be used as templates for predicting the 3-D structure of target sequences and the sequence identity between the template and target must be higher than 30% to be considered reliable (Dalal & Atri, 2014). Template selection is very important as it impacts the quality of the calculated homology models (Muhammed & Aki-Yalcin, 2019). Inspecting the probability values, E-values, identities, secondary structure scores and alignment to identify the best templates is necessary. The best template is generally represented by high scores with similar length to the query, few gaps in the pairwise query template alignment (Fiser, n.d.). Alignment algorithms are used in the multiple sequence alignment between the target and the template, together with other relevant homologs to obtain more accurate results. There are many tools for personal computers and webservers that can be used to perform homology modelling, ranging from automated webservers, downloadable programs, interactive webservers, and standalone programs mainly in

Fortran and Python (Vyas et al., 2012). Homology modelling (HM) is an important molecular modelling technique to model 3-D structures of scientific importance in drug discovery, as determining the atomic and molecular structure of a crystal using Nuclear magnetic resonance spectroscopy (NMR), cryo-electronic microscope (cryoEM) and X-ray crystallography which can be a long and costly process (Er et al., 2018; Nwanochie & Uversky, 2019; Vénien-Bryan et al., 2017). HM method allows one to use single or multiple templates depending on the need and aim of the study.

MODELLER (Šali & Blundell, 1993), HHpred (Söding et al., 2005), SWISS-MODEL (Waterhouse et al., 2018) and PRotein Interactive MOdeling (PRIMO) (Hatherley et al., 2016) are some of the tools and programs used in 3-D structure prediction. Many of these programs are reported to be reliable, fast, and accurate, however it is recommended to validate the quality of the stereo-chemical properties of the generated models are assessed using external tools such as PROSA (Wiederstein & Sippl, 2007), VERIFY3D (Eisenberg, 1997), ERRAT, PROCHECK (Laskowski et al., 1993) and AMOEBA (Laury et al., 2015). Figure 1.7 below summaries the comparative modelling process. Some of these tools used to validate the models were also adopted in this study and will be used and explained in the following chapter.

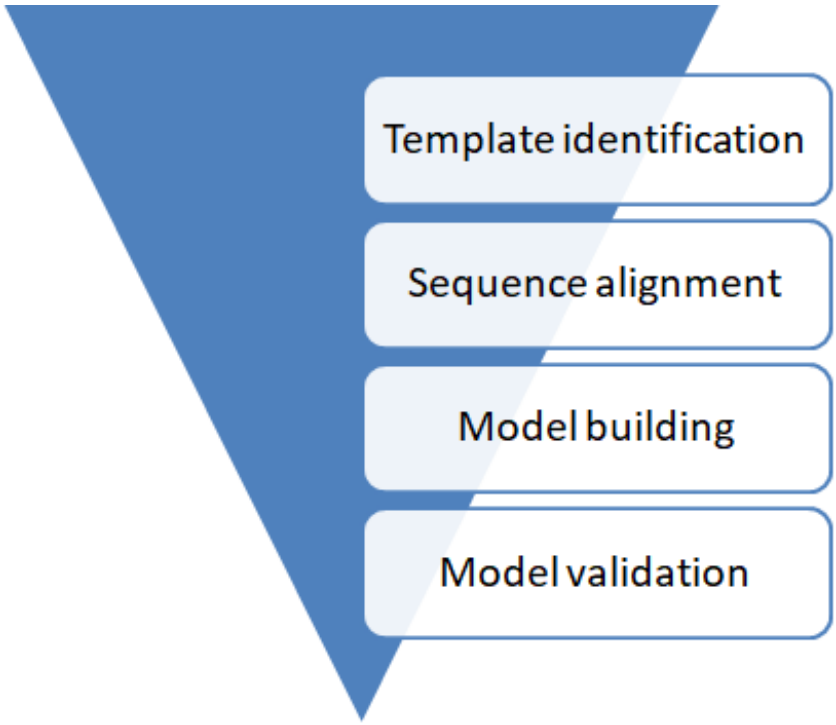

Figure 1.7 Flow chart of the homology modeling process

1.2.2 High throughput virtual screening

The completion of the human genome project has led to the availability of a vast sequence database and expansion of sequential and structural bioinformatics. Bioinformatics techniques are used to model and analyse protein targets, virtual screening can be performed with large compound libraries to identify hit and effective compounds that can be used to combat disease at a cost effective rate (Brown & Tastan Bishop, 2017; F. Moraes & Góes, 2016). There are three important steps involved in modern drug discovery which include identification of targets related to the disease discovery, novel lead discovery and optimisation, and drug development using target or ligand (Liu et al., 2018).

The behaviour of small molecules in the binding site of a target protein is explored using the molecular docking method. For this, protein structures may be obtained from X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy, which are the experimental procedures, together with homology modelled structures for determining structures that have not yet been solved. Docking predicts the orientation of the ligand bound to the protein receptor using shape and electrostatic interactions to quantify the interaction based on scoring sampling and scoring function algorithms. Formation of hydrogen bonds and van der Waals interactions also play a vital role (Pagadala et al., 2017). The docking score approximates the sum of all the interactions, representing the binding free energy values. There are more than 60 different docking tools and programs that have been developed for academic and commercial use; examples include DOCK (Venkatachalam et al., 2003), AutoDock (Österberg et al., 2002), FlexX (Rarey et al., 1996), Surflex (Jain, 2003), GOLD (Jones et al., 1997), ICM, Glide (Friesner et al., 2004), Cdocker, LigandFit (Venkatachalam et al., 2003), AutoDock Vina (Trott & Olson, 2009) and rDock (Ruiz-Carmona et al., 2014).

Emil Fischer in 1894 explained the „lock-key" model to interpret the binding of a ligand to a protein (Fischer, 1894). The docking algorithm finds the optimal binding mode of the ligand to the active site of the protein target, identifying drugs that can bind to the protein target more strongly than the natural substrate, drugs that could potentially alter or prevent the biochemical reaction that the target molecule catalyses (Thomsen & Christensen, 2006). The conformation/orientation sampling (pose generation) and a scoring function that associates a score to each predicted pose are the two stages of molecular docking (Huang & Zou, 2010). Scoring functions play the role of a pose selector, used to discriminate putative correct binding modes and binders from non-binders in the pool of poses generated by the sampling engine. There are three types of scoring functions: force-field based scoring functions; empirical scoring functions; and knowledge-based scoring functions (Salmaso & Moro, 2018).

Virtual screening (VS), identifies the potential drug leads for a given protein target by searching a large ligand database (Li et al., 2019). Virtual screening speeds up the discovery process to reduce the number of candidates to be tested experimentally, however does not replace *in vitro* and *in vivo* assays

(Neves et al., 2018). VS can be categorised in two groups: ligand-based methods and receptor-based methods (Gimeno et al., 2019). Ligand based virtual screening uses features of molecules derived from chemical structures of known binders to score new molecules from a chemical library (Cournia et al., 2020). Structure-based virtual screening attempts to predict the best interaction between ligands against a molecular target to form a complex, the ligands are ranked according to their affinity to the target and the most promising compounds are shown at the top of the list (Maia et al., 2020).

To evaluate the binding ability of arbitrary poses at the respective protein target scoring functions are used, the functions select potential active candidates from large chemical libraries and are used to rank different compound orientations based on their binding scores (Aggarwal & Koes, 2020).

The use of a crystal structure is recommended to validate the molecular docking technique by redocking the protein‟s co-factor or a known ligand; the pose of the redocked system should match the experimental crystal structure pose (C et al., 2020). As large libraries of drugs are used in virtual screening using docking, one must narrow down the ligands to the specific ones that are considered hit compounds (Issa et al., 2019). Dibenziodolium chloride (DPI), diphenyliodonium chloride (IDP), 1-hydroxy-2-dodecyl-4 (1H) and quinolone (HDQ) are some of the previously characterised NDH2 inhibitors, however they do not inhibit *Pf*NDH2 activity (Dong et al., 2009). There are many NDH2 inhibitors that have been identified, however the lack of information regarding their mode of action and associated inhibitor-bound NDH2 structure has impeded rational drug development (Petri et al., 2018).

1.2.3 Molecular dynamics simulation

To study the progression of molecular systems at temperature and physiological conditions, molecular dynamics simulations are applied. During molecular dynamics, trajectories are generated by numerically solving the Newton‟s equations of motion, and the atoms interact during the simulation showing the proper dynamic evolution of the system. For MD, temperature, pH and pressure and residue mutations can be adjusted in separate simulations to demonstrate the effect of these environmental changes on structural variation; as such it can show events such as misfolding or change in conformation of the protein throughout simulation (Campos et al., 2010).

Highly optimized software packages on HPC resources are used to perform MD simulations, trajectories produced are generally large in size ranging from gigabytes to terabytes making it difficult to impractical, to convert trajectories into a range of different formats, therefore the trajectories are produced as a frame by frame description of the motion of particles in the simulation (Richard et al., 2019). Using long simulation times and realistic boundary conditions, together with increasing computational power, simulations can now be performed on even larger systems, and complex systems such as membrane protein structures can also now be successfully and accurately simulated (Gordiz et al., 2015; Hansson et al., 2002). Due to the complexity of the simulation, large systems

require significant computing resources to capture behaviour of the system in full atomic detail (Dreher et al., 2013; Hollingsworth & Dror, 2018). There are fundamental challenges due to the high computational cost of evaluating forces between all particles combined with integrating over short time steps (~2 fs). The structural motions of the atoms are obtained by solving the classical Newton equations for the systems, over time forces acting on the atoms are computed, depending on the particle position and total potential energy of the system (V_{tot}) (Liguori et al., 2020).

The steps taken in MD simulations are shown in figure 1.8 below in which step 1 involves solvation of the system, providing a force field that is reliable for the type of simulation (Jiang et al., 2018; Schuetz et al., 2019). Energy minimisation in step 2 is used to release steric clashes using classic algorithms such as steepest descent and conjugate gradient methods; in step 3 a periodic boundary condition is applied, the ensemble to sample from is chosen and proper thermostat and barostat is set. Before the production run periodic boundaries on all sides are activated and initial simulations in canonical ensemble (with a constant number of particles, volume, and temperature (NVT)) and isothermal-isobaric ensemble (with a constant number of particles, pressure, and volume (NPT)), with harmonic position restraints on all heavy atoms are conducted to equilibrate the systems. Step 4 involves equilibration of the system and, in simulations with membranes, relaxes the membrane, protein and water with position restraints on selected molecules to minimize perturbation. Step 5 produces the trajectory or analysis, that may be checked for convergent results and in step 6 more detailed analysis, such as root mean square deviations (RMSD), root mean square fluctuation (RMSF), radius of gyration, dynamics cross correlations (DCC) and Principal component analysis (PCA) are performed for full analysis of the MD results (Lemkul, 2019; Sheik Amamuddy et al., 2021a).

1.2.3.1 Analysis for the MD simulations

The root mean square deviation (RMSD) is the measure of average distance between atoms of superimposed proteins and are represented in angstrom (Å), the equation below is used to calculate root mean square deviation (RMSD) where the averaging is performed over the n pairs of equivalent atoms and d_i is the distance between the two atoms in the i-th pair (Kufareva & Abagyan, 2012).

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n} d_i^2}$$

Equation 1.1 The equation used to calculate the RMSD

The root mean square fluctuation (RMSF) measures the average deviation of a residue over time from a reference position, it analyses the structure portions of structure that are fluctuating from their mean structure the most (Benson & Daggett, 2012). In the root mean square fluctuation (RMSF) equation

below T is the time over which root mean square fluctuation (RMSF) is averaged, r_i^{ref} is the reference position of particle *i* and the time averaged position of the same *i* particle will be the reference position.

$$RMSF_i = \left[\frac{1}{T}\sum_{t_j=1}^{T}\left|\Gamma_i(t_j) - \Gamma i^{\Gamma ef}\right|^2\right]^{\frac{1}{2}}$$

Equation 1.2 The RMSF equation

The radius of gyration (R_g) measures the compactness of a protein structure and the size of the protein molecules (Arnittali et al., 2019). The equation below shows m_i as the mass of atom *i* and r_i the position of *i* with respect to the center of mass of the molecule.

$$R_g = \left(\frac{\Sigma_i|\Gamma_i|^2 m_i}{\Sigma_i m_i}\right)^{1/2}$$

Equation 1.3 The equation used to calculate the Radius of gyration

Principal component analysis (PCA) calculates the relationship between statistically meaningful conformations sampled during the trajectory, the dataset variables are narrowed down to "principal components" preserving still the differences between the data (Martínez, 2015). Principal components are new variables that are constructed as linear combinations or mixtures of the initial variables.

Dynamic cross correlation (DCC) calculates the degree to which residues move together and this can be represented in an NxN heatmap, where N is the number of Cα atoms of the system each element corresponds to the dynamic cross-correlation between each i,j atom (Brown et al., 2017a). The nature of a protein is dynamic and this change in conformation can be inspected using DCC, regions of high correlation reflect motions in the same direction and anti-correlated motions show atoms moving in opposite direction (H. Yu & Dalby, 2020). In the equation below $r_i(t)$ denotes the vector of the *i*th atom's coordinates as a function of time t (Kasahara et al., 2014).

$$DCC(ij) = \frac{\left\langle \Delta\Gamma_i(t) \cdot \Delta\Gamma_j(t)\right\rangle_t}{\sqrt{\left\langle \left\|\Delta_{\Gamma_i}(t)\right\|^2\right\rangle_t}\sqrt{\left\langle \left\|\Delta_{\Gamma_j}(t)\right\|^2\right\rangle_t}}$$

Equation 1.4 The Dynamic cross correlation equation

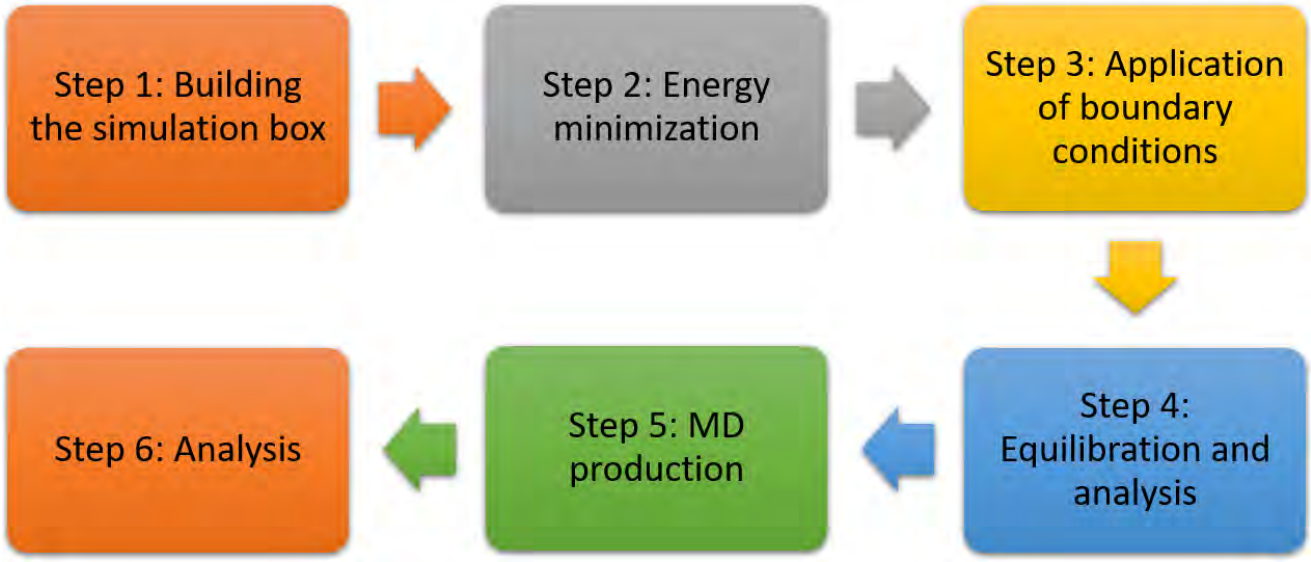

Figure 1.8 The workflow of molecular dynamics simulations

There are a number of programs and software used to perform MD simulation programs like NAMD (Phillips et al., 2020), GROMACS (Berendsen et al., 1995), AMBER (Salomon-Ferrer et al., 2013), CHARMM (Lamoureux & Roux, 2003), OpenMM (Eastman et al., 2017) and LAMMPS (Thompson et al., 2022). Visualisation programs such as VMD (Humphrey et al., 1996), CHIMERA (Pettersen et al., 2004) or PyMOL (Yuan et al., 2016) can be used to visually inspect or make measurements on the MD simulations. MD simulations hold great promise in characterising the structural and dynamical aspects of protein-ligand interactions important in the functioning of a membrane protein; atomistic simulations allow for temporal resolutions of about 1fs and spatial resolutions of sub-angstrom level (Muller et al., 2019).

1.3 Project motivation and knowledge gap

Due to the emergence of drug resistance in the parasite against current anti-malarial drugs there is a need to develop new and effective drugs that can be used to eradicate malaria. Type II NADH: quinone oxidoreductase (NDH2) a protein that catalyses the electron transfer electrons from NADH to quinone through the FAD cofactor has been proven to be a new drug target.

As malaria is still a cause of concern around the world and claiming many lives each year especially in children under the ages of 5, there is a need for more studies and drug discoveries to conquer this disease. Further, studying the 5 *plasmodium* species that causes malaria in human beings is necessary as it allows one to compare the mechanism of action across these species. The inclusion of a membrane in this study is to explore and understand the purpose of the membrane towards the functioning of the NDH2 protein. There have not been many NDH2 studies where the membrane has been included in simulation. To understand the biological function of a protein it is vital to gather the knowledge of the atomic motions and their collective correlated character (Karplus & Ichiye, 1995).

Little is known about how membrane proteins function and how their structure is defined by amino acid sequences. We seek to understand the functional mechanism of NDH2 and possibly identify compounds that can be used to interrupt its function in electron transfer.

1.4 Research aims and objectives

Bioinformatics tools and techniques will be used to study the NDH2 protein in the context of drug discovery. We aim to identify lead compounds, explore the protein-ligand interactions, and compare the *plasmodium* species that causes malaria in human beings. The aims of the project include:

1. Sequence analysis of the *plasmodium* species, the human homologue and 11 other species, using Multiple sequence analysis, motif analysis, phylogenetic analysis, and sequence analysis
2. 3-D structural construction of NDH2 models for the four *plasmodium* species that causes malaria in human beings, where crystal structures have not yet been solved.
3. Perform high throughput virtual screening on all six species against the 623 SANCDB database and select hit compounds for further analysis.
4. Calculate the physical movements of the atoms using Molecular Dynamics simulations, first without a membrane and secondly with a membrane as NDH2 is a peripheral membrane protein.
5. Other techniques such as PCA and DCC will be used in analysing the MD trajectories to investigate the protein-ligand interactions.

2

Sequence analysis

Chapter overview

Proteins perform most of the work in living cells; they control cell division, metabolism, and the flow of materials and information in and out of the cells (Frauenfelder & McMahon, 1998). It is essential for one to study how proteins function to understand how cells work (Zhang et al., 2013). In Chapter 1 we have addressed that type II NADH: quinone oxidoreductase catalyses the electron transfer from NADH (electron donor) to a quinone (electron acceptor). However, this information is not complete in the understanding of the functioning of this protein in an organism. In this study investigations have been taken with added aim to understand the context in which the biochemical activity is used and to know the role that the protein has on the growth or development of the organism. The study starts with sequence analysis, which was performed in which the FASTA sequence of the NDH2 was compared with sequences of other species and of the human homologue. Multiple sequence alignment (MSA), motif analysis, phylogenetic analysis and sequence identity were the bioinformatics methods used to study the sequences. The MSA results allowed us to compare between the NDH2 and AIF-M1 protein so to find similarity and homologous between the two proteins and phylogenetic analysis was also conducted to assess the shared evolutionary origins and genetic diversity in the sequences.

2.1 INTRODUCTION

2.1.1 Data retrieval

The first step in solving and studying protein as primary sequence, this data retrieval or sequence search and sequence similarity is widely accepted as the best marker for substantiating homologous relationships (Gabler et al., 2020). The name of the protein or its sequence can be used as a query in sequence databases, and the advantage of these retrieval systems is that they also add additional important information in related databases. The NCBI website, for example, uses some measure of similarity between sequences to distinguish biologically significant relationships from chance similarities. BLAST, a tool in NCBI, finds regions of similarity between biological sequences and

compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. BLAST uses a heuristic method to identify homologous sequences. This method finds short matches initially between two sequences, the algorithm then finds all common words between the query sequence and the hit sequences (Vej & Telephone, 2007). BLAST programs use the statistical methods of Karlin and Altschul and were designed for fast database searching without affecting the sensitivity for distantly related sequences (Altschup et al., 1990). BLAST (Ladunga, 2017; Zhuang et al., 2012), HMMER (Potter et al., 2018; Prakash et al., 2017), HHblits (Remmert et al., 2012), HHpred (Söding, 2005; Steinegger et al., 2019) protein sequence and domain databases SCOPe (Fox et al., 2014), ECOD (Cheng et al., 2014; Schaeffer et al., 2018), Pfam (Coggill et al., 2008; El-Gebali et al., 2019), RefSeq (O'Leary et al., 2016), UniProt (Pundir et al., 2016), National Center for Biotechnology Information Web Resources (NCBI) (Agarwala et al., 2018; Gibney & Baxevanis, 2011; M. Yang et al., 2020) are some of the high quality and reliable sequence search engines.

2.1.2 Multiple sequence alignment and motif analysis

Multiple Sequence Alignment (MSA) is generally the alignment of three or more biological sequences (protein or nucleic acid) of similar length. This process forms the basis of a wide ranges of biological data analysis, including the description of the relationship between protein sequences assumed to have a common ancestry, descendance and protein structural and functional similarity (Rozewicki et al., 2019). For comparison sake and accuracy, in our study three multiple sequence alignment tools MAFFT, MUSCLE and PROMALS3D were used to study the sequences of interest using default parameters.

MSA using fast Fourier Transform (MAFFT) is a method in which the initial alignment is constructed by the method and refined by the iterative refinement method. In the method a rough distance between every pair of input sequences is rapidly calculated based on the number of 6-tuples shared by the two sequences and iterative refinement optimizes the weighted sum-of-pairs (WSP) score, using an approximate group-to-group alignment algorithm and the tree-dependent restricted partitioning technique (Edgar et al., 2004; Gotoh, 1995; Hirosawa et al., 1995; Katoh et al., 2002, 2005). The latest version of MAFFT incorporates pairwise alignment information into the objective function, using the following different algorithms: the G-INS-i method which incorporates global pairwise alignment information; the H-INS-i algorithm which incorporates local pairwise information from the fasta34 program in FASTA and; finally, the F-INS-i option which uses the fasta34 program without the Smith–Waterman optimisation (Pearson & Lipman, 1988; Reigosa et al., 2001).

Multiple Sequence Comparison by Log-Expectation (MUSCLE) uses two distance measures for a pair of sequences - a *k*-mer distance which is a contiguous subsequence of length *k*-tuple (for an unaligned

pair) and the Kimura distance correction for multiple substitutions at a single site (for an aligned pair) (Edgar et al., 2004). The first step in MUCSLE is to rapidly generate a rough draft of the alignment using a very crude guide tree, and the tools also implement the log expectation (LE) score to align profiles during the progressive alignment; this has been shown to outperform other scoring functions in homology searches (Wallace et al., 2005). Refining the rough draft by generating a more accurate guide tree, based on the initial alignment is the second stage in the process, generating a second progressive alignment. New pairwise profile alignments are calculated only for those subtrees that have changed relative to the initial tree to increase speed, and to further improve the alignment quality an optional tree-based iteration step has been factored in (Hirosawa et al., 1995).

Profile Multiple Alignment with local Structures and 3-D constrains (PROMALS3D) is a tool for protein MSA construction enhanced with additional evolutionary and structural information from database searches (Pei & Grishin, 2014). This method integrates advanced alignment techniques such as probabilistic consistency of profile–profile comparisons, and additional information from database homologs and predicted secondary structures to derive consistency-based alignments (Pei et al., 2008). PROMALS3D uses the scoring function of weighted sum-of-pairs of BLOSUM62 to align similar sequences (Henikoff & Henikoff, 1992). The output is a consensus alignment enriched with sequence and structural information about input proteins and their homologs. The PROMALS3D webserver allows input of sequences and structures; it extracts sequences from input structures combining them with other input sequences to form the final input sequence set.

Proteins are polymers from the combinations of twenty amino acids linearly connected and are complete biological molecules when in their stable conformation. The chemical interactions and physicochemical properties of amino acid residues determine the conformation of proteins and form a relationship between protein sequences and structures (Kim et al., 2011). Understanding the close relationship between protein sequences and has been one of the primary interests in bioinformatics research. A motif is an amino acid sequence pattern that is widespread and contributes to the biological function of the sequence in which it resides, biological significance such as, binding sites and conserved domains (Grant et al., 2011). For example, motif discovery plays an important role in identification of Transcription Factor Binding Sites (TFBSs) that help in learning the mechanisms for regulation of gene expression (Hashim et al., 2019). Sequence motifs can predict other protein‟s structural or functional behaviours and discovering sequence motifs is a key task to comprehend the connection of sequences with their structures. Biological sequence motifs are short, with a fixed length, and include sequence patterns that may represent important structural or functional features in nucleic acid and protein sequences such as transcription binding sites, splice junctions, active sites, or interaction interfaces (Mohamed et al., 2016). To discover the protein sequence motifs a MEME algorithm, from the Multiple Expectation Maximisation for Motif Elicitation MEME suite was used. The MEME suite

was used as it is freely available for academic use and is a powerful tool for studying sequence motifs in proteins, DNA and RNA.

2.1.4 Sequence identity and phylogenetic analysis

The Maximum Likelihood (ML) statistical method was used to infer evolutionary relationships while calculating trees for the top three models for each gap deletion option for each protein families. The Molecular Evolutionary Genetics Analysis (MEGA) is a software that provides a biologist centric, integrated suite of tools for statistical analyses of DNA and protein sequence data from an evolutionary standpoint, was used to calculate the phylogenetic trees (Tamura et al., 2011). In this package, for nucleotide substitutions, the GTR and five nested models are available, whereas six models with and without empirical frequencies (+F) have been programmed for the amino acid substitutions (Tamura et al., 2011). MEGA5 provides the goodness of-fit of the substitution models with and without assuming the existence of evolutionary rate variation among sites, which is modelled by a discrete Gamma distribution (+G) (Yang, 1994) and/or an allowance for the presence of invariant sites (+I) (Fitch & Margoliash, 1967; Shoemaker & Fitch, 1989). The Bayesian information criterion (BIC, Schwarz 1978) measures the goodness-of-fit of each model to the data and is corrected by the Akaike information criterion (AICc), (Hurvich & Tsai, 1989; Posada & Buckley, 2004).

2.1.5 Structure prediction (3-D)

Structure prediction is the prediction of the three-dimensional structure of a protein from its amino acid sequence, predicting its folding, and its secondary, tertiary and quaternary structure from its primary structure (Mattea et al., 2018). Experimental and computational methods can be used for protein structure prediction, in which the experimental methods include X-ray crystallography. NMR spectroscopy and cryo-EM, and the computational methods include homology modelling, fold recognition and Ab initio modelling (Kuhlman & Bradley, 2019).Comparative or homology modelling constructs a useful 3-D model for a protein that is related to a known protein structure, if the experimentally determined structure is not yet available (Bradley et al., 2005; Eswar et al., 2014; Fiser, 2004; Misura et al., 2006; Petrey & Honig, 2005). This computational modelling method is used to cover the gap between number of known protein sequences and Protein data bank (PDB) entries, by complementing the experimental structure determination (Studer et al., 2021). Comparative modelling interpolates structural information from homologous structural information from homologous structures providing protein models that are of sufficient high quality and accuracy to guide structure-based research. There are four main steps involved in comparative modelling: (1) fold assignment which identifies similarity between the target and at least one known template structure, (2) alignment of the target sequence and the template or templates, (3) building a model based on the alignment with the chosen template or templates, (4) predicting model errors (Eswar et al., 2014). There are several computer programs and Webservers that automate the comparative modelling process, and, in this study, MODELLER was used. This program automatically calculates a model

containing all non-hydrogen atoms, within minutes on a reasonable processor and with no user intervention (Eswar et al., 2006). MODELLER can calculate additional auxiliary tasks such as fold assignment (Eswar et al., 2003), alignment of two protein sequences or their profile (Marti-Renom, 2004), multiple alignment of protein sequences (Madhusudhan et al., 2006), phylogenetic tree calculation and *de novo* modelling of loops in the protein structure, apart from model building (Watanabe et al., 2004).

MODELLER generates protein 3-D structures by satisfying spatial restraints imposed by the sequence alignment with the template structure and applying the terms of the CHARMM-22 force-field. From this a 3-D protein model is obtained by optimising the molecular probability density function while simultaneously minimising input restraint violations (Bernard R. Brooks et al., 1983; Evers et al., 2003). Several homology models are generated in this step, to guarantee sufficient conformational sampling of each active-site residue.

The following external evaluating tools were used for structure quality assessment of the selected model, Verify3D, PROCHECK and Protein Structure Analysis (PROSA) webservers (Eisenberg, 1997; Laskowski et al., 1993; Wiederstein & Sippl, 2007). Verify3D compares the 3-D protein model to its amino acid sequence to check the accuracy of the model by using a 3-D profile computed from atomic coordinates of the structure, 3-D profiles of the correct protein structures match their own with high scores (Eisenberg, 1997). This program enables one to locate parts of the protein that are likely to have the correct conformation or to look for misfolded regions through assessment of structures on the residue level (Von Grotthuss et al., 2003).

The PROCHECK program provides a detailed check on the stereochemistry of the protein structure; the output is comprised of several plots in postscript format and a comprehensive residue by residue listing giving an overall quality of the structure as compared with well refined structures of the same resolution and highlights regions that may need further investigation (Laskowski et al., 1993). PROCHECK is a suite of separate Fortran and C programs that run successively via a shell script. Hydrogen and atoms with zero occupancy are not included in the analyses and only the highest-occupancy conformations are retained where atoms are found in alternate conformations.

The ProSA program (Protein Structure Analysis) is a tool widely used to check 3-D models of protein structures for potential errors, its range of application includes error recognition in experimentally determined structures, theoretical models, and protein engineering (Wiederstein & Sippl, 2007).

Besides the fact that each of the tools used in the construction of a model, template selection, alignment, model building, and refinement have their own internal measures of quality, ultimately, the most meaningful criterion for the quality of a model is its conformational energy (Petrey & Honig, 2005).

2.2 METHODOLOGY

2.2.1 Data retrieval

The *Plasmodium falciparum* NDH2 sequence with primary accession number (Q8I302) was retrieved from Universal Protein Resource Knowledgebase (UniProtKB) database and the crystal structure (PDB ID: 5JWA) was retrieved from RCSB Protein Data Bank (Consortium, 2019; Rose et al., 2011). The P*f*NDH2 protein is expressed by the PF3D7_0915000 gene. The known crystal structure (ID:5JWA) was used as the query sequence to retrieve the sequences of the other four *plasmodium* species that cause malaria in human beings using the Basic Local Alignment Search Tool (BLAST) from the National Center for Biotechnology Information (NCBI) (Altschup et al., 1990) (http://www.ncbi.nlm.nih.gov/). The protein blast (blastp) algorithm was used, the Protein Data Bank proteins (PDB) database was selected, and the FASTA sequence obtained from RCSB was pasted into the webserver query search box.

The query search had a percentage identity of 100%, an E-value of 0.0, query coverage of 100% and the length of the protein was 521 amino acids. The X-ray diffraction method was used to extract the protein, it had a resolution of 2.16 Å, an R-value free of 0.228, R-value work of 0.188 and R-value observed of 0.190. With alternative NADH dehydrogenase (NDH2) being absent in human beings, it was necessary to retrieve a human homologue as mentioned in literature (Elguindy & Nakamaru-ogiso, 2015). For this homologue the protein sequence of the Apoptosis-inducing factor 1 (AIF-M1) with an accession number of (O95831) was retrieved from UNiProtKB and the crystal structure (PDB ID: 1M6I) was retrieved from RCSB Protein Data Bank (Consortium, 2019; Rose et al., 2011). The crystal structure of apoptosis inducing factor (AIF-M1) under the classification of oxidoreductase enzyme was extracted using the X-ray diffraction method, it had a resolution of 1.80 Å, R-value free of 0.241 and R-value work of 0.225.

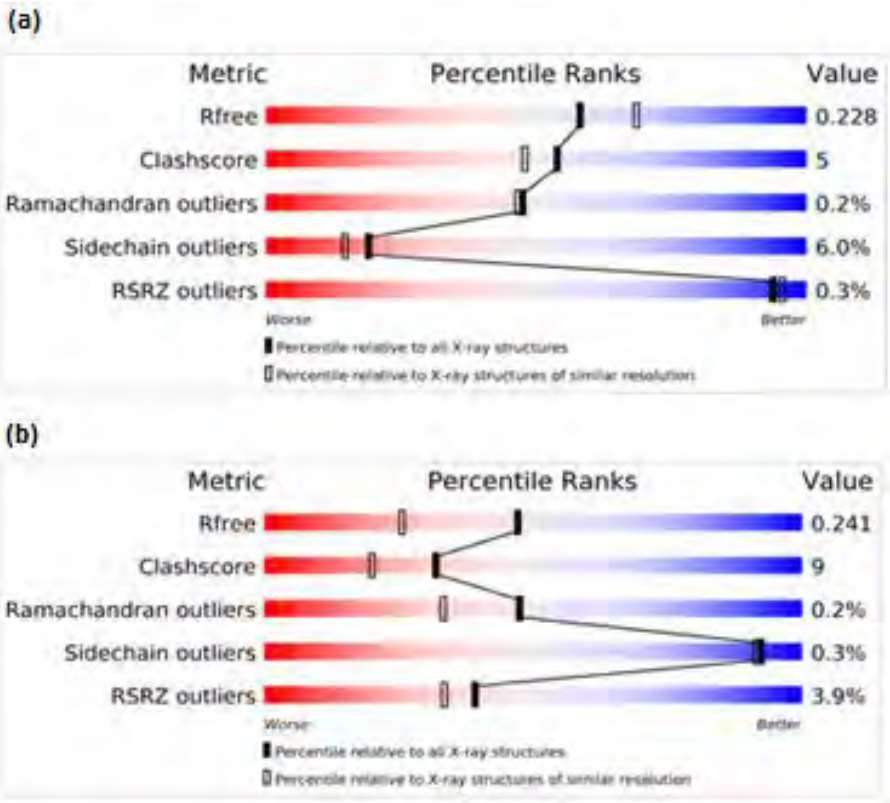

Figure 2.1 (a) The X-ray crystal quality of *Pf*NDH2 (PDB ID: 5JWA), (b) Apoptosis inducing factor (PDB ID: 1M6I)

2.2.2 Multiple sequence alignment

A multiple sequence alignment calculation of proteins was performed in which the sequences of the *Plasmodium*, human being, bacteria, and fungi were included. Three multiple alignment tools were used to perform the calculations for comparison and accuracy reasons, and a total of 17 sequences were included in the multiple sequence alignment. National Center for Biotechnology Information (NCBI) and Universal Protein Resource Knowledgebase (UNiProtKB) was used to retrieve the sequence of the bacteria and fungi (http://www.ncbi.nlm.nih.gov/) and (https://www.uniprot.org/uniprotkb/). The FASTA format file, NCBI or UNiProtKB accession number of the 17 sequences for NDH2 which included *Plasmodium falciparum* (5JWA), *Plasmodium knowlesi* (OTN67496), *Plasmodium malariae* (SBT70925), *Plasmodium ovale* (SCQ16138), *Plasmodium vivax* (SGX76671), *Homo sapiens* (1M6I), *Mycobacterium tuberculosis* (L7N5D1), *Escherichia coli* (P00393), *Corallococcus corraloides* (QAT82924), *Staphylococus aeureus* (Q2FZV7), *Neurospora crassa* (Q7S1W8), *Pseudomonas aeruginosa* (AZP62247), *Saccharomyces cerevisiae* (S288C NP_010198), *Yarrowia lipolytica* (VBB82463), *Aspergilluus thermomutatus* (RHZ55328), *Aspergillus turcosus* (RLL97513) and *Caldalkalibacillus thermarum* (F5L3B8), were uploaded as an input to the multiple alignment tool webservers (https://www.ebi.ac.uk/Tools/msa/mafft/, https://www.ebi.ac.uk/Tools/msa/muscle/ and http://prodata.swmed.edu/promals3d/). These species cause disease and health problems in humans

and contain the alternative NADH: ubiquinone oxidoreductases that catalyses the same redox reaction as respiratory chain complex 1 (Kerscher, 2000). The output files from the multiple alignment webservers were visualised, analysed, and edited using Jalview version 2.10 software (Alzohairy, 2014; Hirosawa et al., 1995).

2.2.3 Motif discovery

MEME version 4.11.2 was used to identify significant patterns in the data sequences that may represent important structural or functional features (Bailey et al., 2009). An in-house motif analysis python script was run using a cluster (YODA) to perform motif calculations, the script contained the parameters of the run in which 50 motifs with a motif width of 6-20 residues were run for each class. The input file was the FASTA file consisting of the 17 sequences mentioned earlier. The first 30 motifs were considered significant for further analysis.

An in-house MATLAB script was used to create a heatmap, to show the conserved motifs in the NDH2 sequences. To identify overlapping motifs the Motif Alignment and Search Tool (MAST) was used; this tool takes a group of motifs as the query and compares it to each sequence in the database of sequences (Bailey et al., 1998). An alignment free method was applied, so to find all motifs that satisfies the input constraints (Aneja et al., 2016). MEME gives two output text files the mast.txt and meme.txt, in which the mast.txt output consist of three representations of the results, excerpts of which are shown in the three figure panels. The E-value score of the overall match of the motifs in the input, followed by the relative locations of significant matches of the motifs in the sequences and finally a detailed picture of the motif matches, showing the exact location and *p*-value score of each motif match aligned above the target sequence (Bailey et al., 2009). The meme.txt file was used as the input file to create a heatmap showing the motifs.

The mast.txt file advised that motifs that had correlations above 0.60 should be removed from query because this could cause some combined *p*-values and E-values to be underestimated and reduce accuracy (Bailey et al., 2009). The E-value of a sequence is the expected number of sequences in a random database of the same size that would match the motifs as well as the sequence does and is equal to the combined *p*-value of the sequence times the number of the sequences in the database. The combined sequence of a sequence measures the strength of the match of the sequence to all the motifs and is calculated by: (1) finding the score of the single best match of each motif to the sequence, (2) calculating the sequence *p*-value of each score, (3) forming the product of the *p*-values and (4) taking the *p*-value of the product. Each of the 17 sequences selected as best motifs for further analysis had an E-value of less than 10 and a *p*-value of 0.0001.

2.2.4 Sequence identity and phylogenetic analysis

A separate in-house python script was used to calculate sequence identity of the 17 sequences. Phylogenetic tree calculations were carried out to study evolutionary relationships using the Molecular Evo-

lutionary Genetic Analysis (MEGA) vs7.0 tool (S. Kumar et al., 2016). Three gap deletion options 90%, 95% and 100% were used to calculate the models, and the best three evolutionary models for each deletion option were selected based on the lowest Bayesian information criterion (BIC) scores. The Maximum Likelihood (ML) statistical method was used to infer evolutionary relationship while calculating trees for the top three models for each gap deletion option for each protein families (Tamura et al., 2011). A strong branch swap filter and 1000 bootstrap replicates were used for each tree calculation. The trees were then compared to the bootstrap consensus trees to ensure that branching patterns were accurate and the best model and gap deletion for each case was, chosen.

2.2.5 Construction of *Plasmodium* species (3-D)

The 3-D structures of NDH2 for *P. knowlesi*, *P. malariae*, *P. ovale* and *P. vivax* which do not currently have crystal structures in the PDB were constructed by homology modelling using MODELLER version 9.21 (Sali, 2013). The *Pf*NDH2 with accession number (ID:5JWA) was used as the template in structural modelling, the models were homodimers and included the FAD cofactor and four magnesium ions (Mg^{2+}).

Before MODELLER automatically calculates a model containing all non-hydrogen atoms the target sequence was converted to a PIR file format, which is a readable MODELLER format. Appropriately, we named the files (name of *plasmodium* species-5JWA.ali) which MODELLER used to read and write sequences and alignments. The first line in the PIR formatted sequence consisted of >P1; (name of the protein), the second line, had ten fields separated by colons, containing details about the structure and the rest of the file contained the sequence of the target sequence and specific *plasmodium* species with an asterisk (*) marking its end. The chains were separated by a backslash, and the two FAD cofactors and four magnesium ions (Mg^{2+}) were included in the modelling file by adding eleven backslashes followed by a full stop after every slash (/./. /…) just before the asterisk (*). The amino acids in the sequence were represented using a standard uppercase single letter.

A python script named model2.py was used to perform the actual building of the models, the script loaded standard Modeller classes and the auto-model class, following which it requested a verbose output, creating a new MODELLER environment to build the model. The (*plasmodium* species-5JWA.ali) was used as the input file. 100 models were built at a very slow refinement; the top model which had the lowest z-DOPE (Discrete Optimized Protein Energy) score was selected for further analysis and validation. A text file was produced with the z-DOPE score of the template and all the models created. DOPE is an atomic distance-dependent statistical potential calculated from a sample of native protein structures; it is grounded entirely in probability theory (Shen & Sali, 2006).

Internal evaluations using MODELLER were considered before using external evaluation or validating tools, the z-DOPE score was used to evaluate the fold of the selected model and there were no errors or restraint violations in the log file (evaluate_model.log) produced after running the python

script evaluate_model.py (Eswar et al., 2014). The following external evaluating tools were used for structure quality assessment of the selected model, Verify3D, PROCHECK and Protein Structure Analysis (PROSA) webservers (Eisenberg, 1997; Laskowski et al., 1993; Wiederstein & Sippl, 2007).

2.3 RESULTS AND DISCUSSION

This chapter aimed to study the protein sequences of the *Plasmodium* species (NDH2), together with the Apoptosis inducing factor (AIF-M1) homologue, using MSA, phylogenetic analysis and motif analysis. The 3-D structure determines the function of the protein; thus, the determination of a protein‟s structure is a vital step in understanding how the protein achieves its function. Experimental structure determination can be long and difficult, and errors can occur, the determined structure may adopt a non-physiological fold due to non-physiological constraints imposed by the crystal in the case of a X-ray crystallography (Kalman & Ben-Tal, 2010). Therefore, it is of importance to calculate the accuracy of the models created.

2.3.1 Data retrieval and Sequence alignment

The *Pf*NDH2 protein structure with accession number (ID:5JWA), was coded for by the gene PF3D7_0915000, it was extracted using the X-ray method, with a resolution of 2.16 Å, two chains (A and H) and a length of 494 amino acids starting from position 39-533. The Blastp output contained the local alignment, statistics of the score, sequence length and a hit score with number of identical residues.

Multiple sequence alignment identifies common features between species and identify residues that serve an important function, this technique is the foundation for a range of computational methods including the prediction of protein secondary structure and solvent accessibility, functional sites, active sites, and interaction sites.

There is no perfect multiple alignment tool and hence it is important to use a variety of available tools and compare the results that best suit your situation and obtain the best alignment possible (Waterhouse et al., 2009). The results obtained in MAFFT were visualised in the Jalview tool as shown in figure 2.2 below, the conserved regions were highlighted, in which the blue showed hydrophobic regions (A, I, L, M, F, W, V), orange depicted the glycines (G), green showed the polar regions (Q), cyan showed the aromatic regions (H, Y), pink showed the cysteines groups (C), magenta depicted (D) the negative charges and yellow showed the prolines (P).

Figure 2.3 below shows the alignment results from MAFFT, MUSCLE and PROMALS3D webservers visualised in Jalview software. MAFFT was selected as the tool that had the best alignment, as it had fewer gaps.

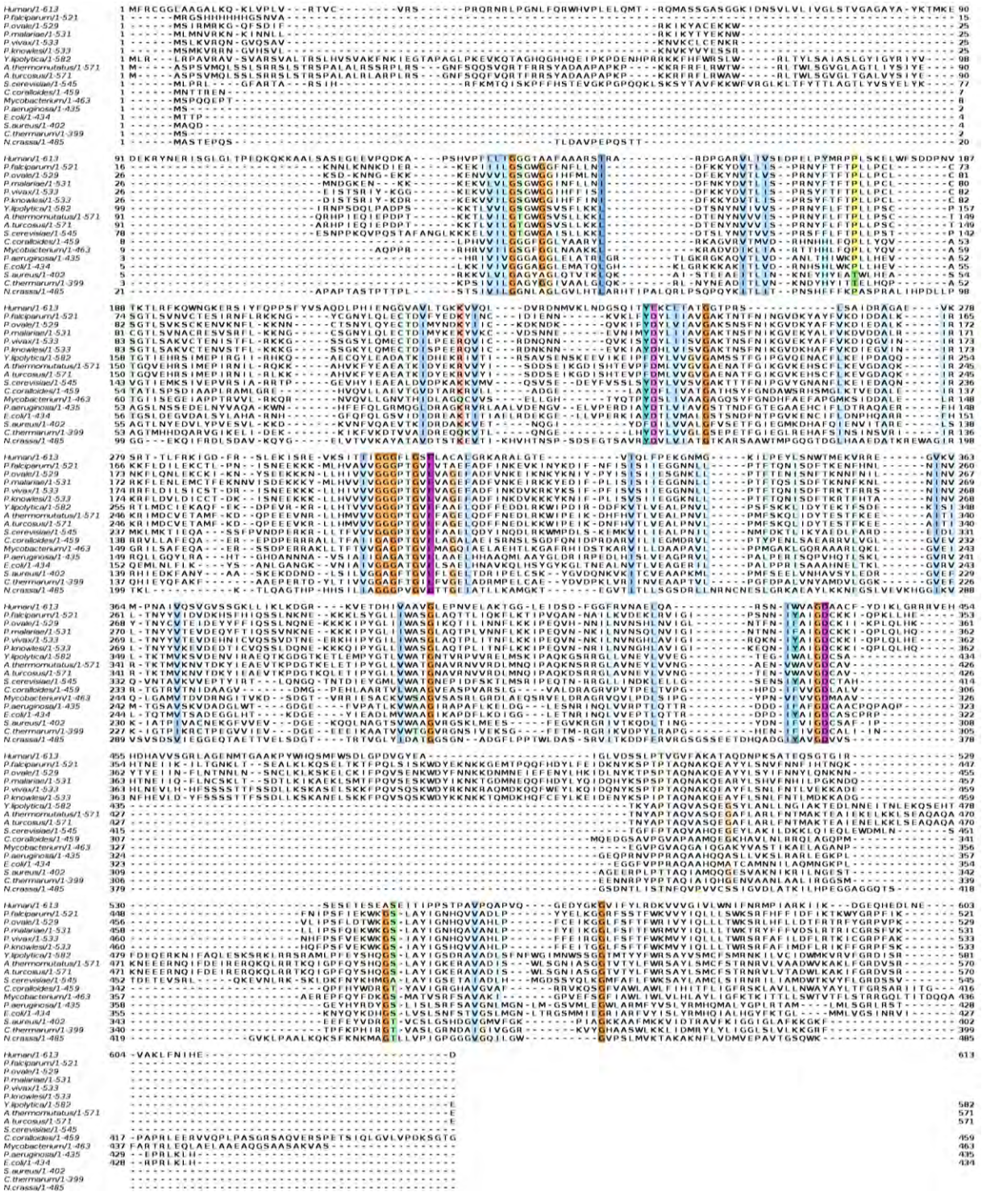

Figure 2.2 MAFFT multiple sequence alignment output visualized in Jalview version 2.10 software

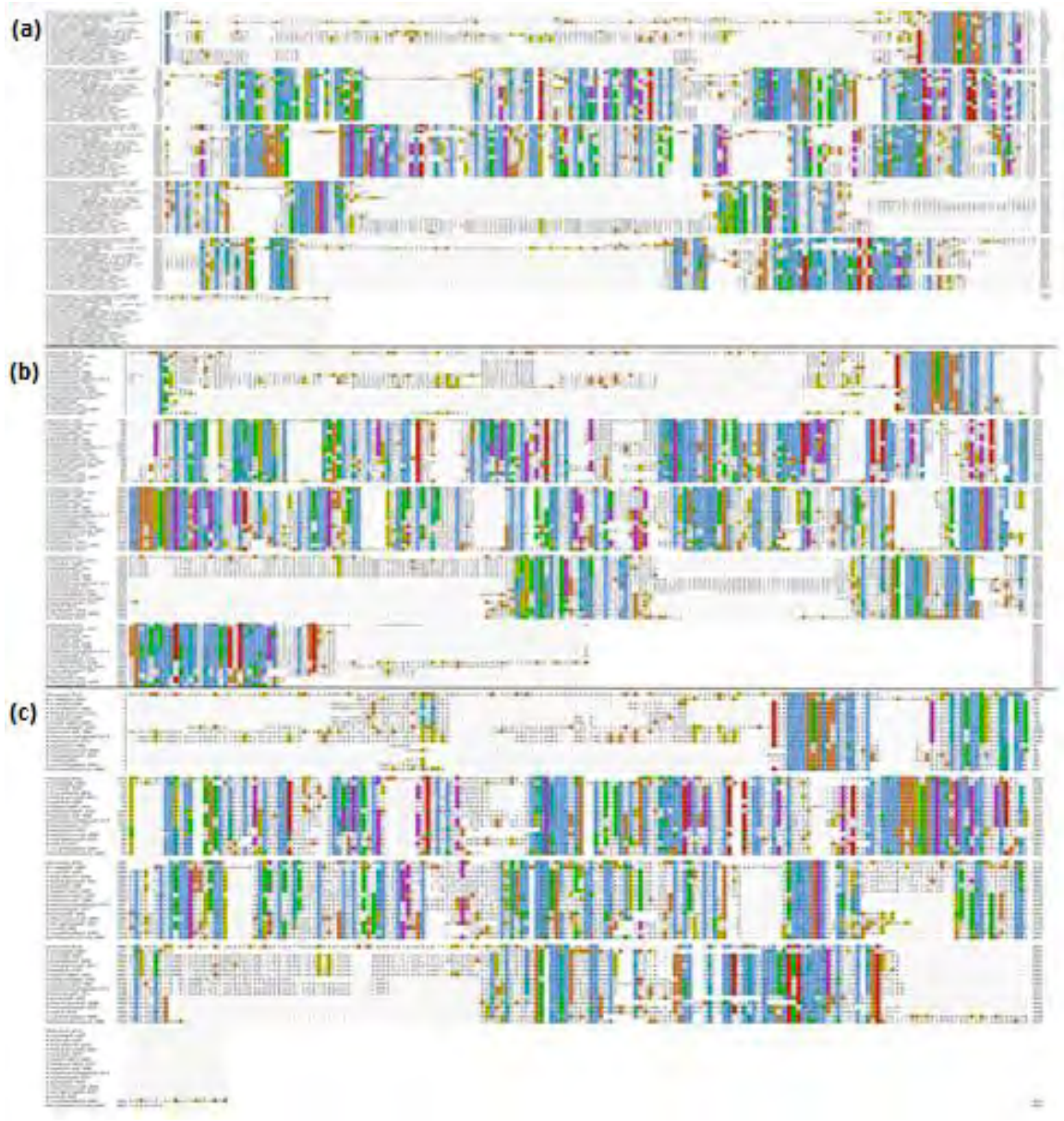

Figure 2.3 The multiple sequence alignment output results from (a) MAFFT, (b) MUSCLE and (c) PROMALS3D visualized in Jalview version 2.10 software

2.3.2 **Motif discovery and phylogenetic analysis**

Motif analysis for the 17 sequences of NDH2 from different organisms was performed. An in-house python script was used to create a heat map shown below of the first 30 motifs. The motifs that were prominent in all sequences were mapped on the structures as shown in the figure 2.4 below. The MEME suite numbers the motifs and 0.0 on the legend depicts that there is no motif conservation in the proteins and 1.0 value indicates that there is 100% motif conservation in the protein sequences. The naming on top of the heat map shows the names of the organisms and the species group, they fall under. Motif 1, 9 and 10 showed 100% conservation in all sequences. Motif 2 had 95% preservation

in all sequences excluding the human. Motif 3-6, 8 and 11 had 90% preservation in all sequences except the human and *N. crassa* fungi. Motif 12 had 75% motif conservation in most sequences except *C. thermanum, P. aeruginosa* and *N. crassa.* The motifs that had 100% conservation in all protein sequences were mapped on the *plasmodium* NDH2 and human AIF-M1 structures.

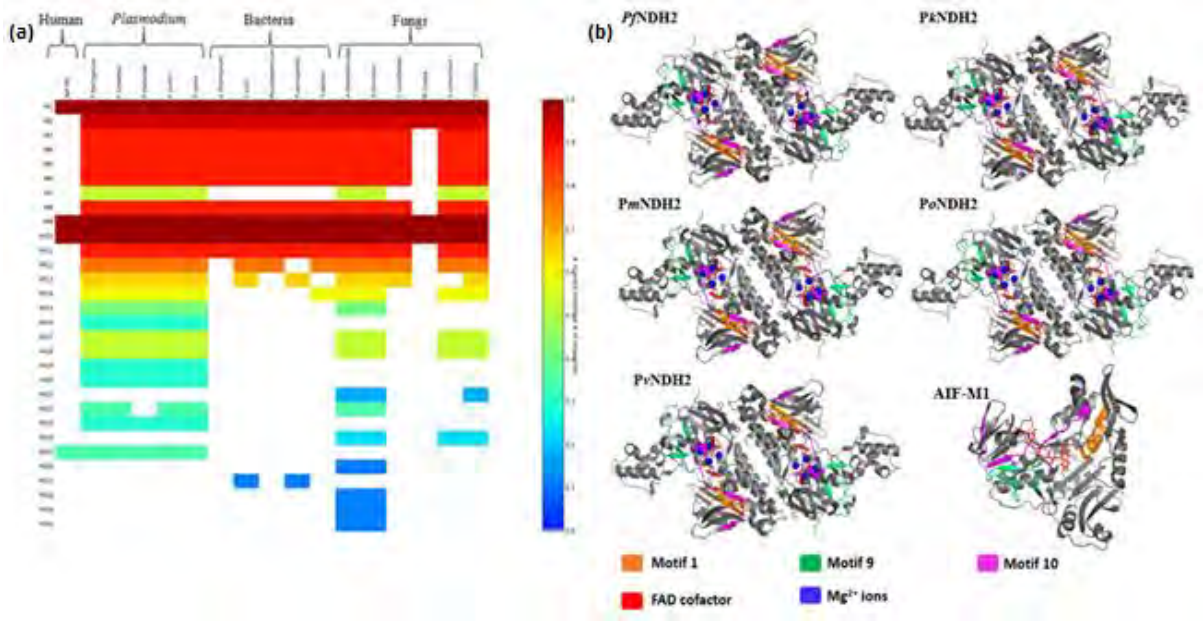

Figure 2.4 (a) A heatmap showing the first 30 motifs calculated by MEME, (b) the common motifs mapped on the *plasmodium* and human structures

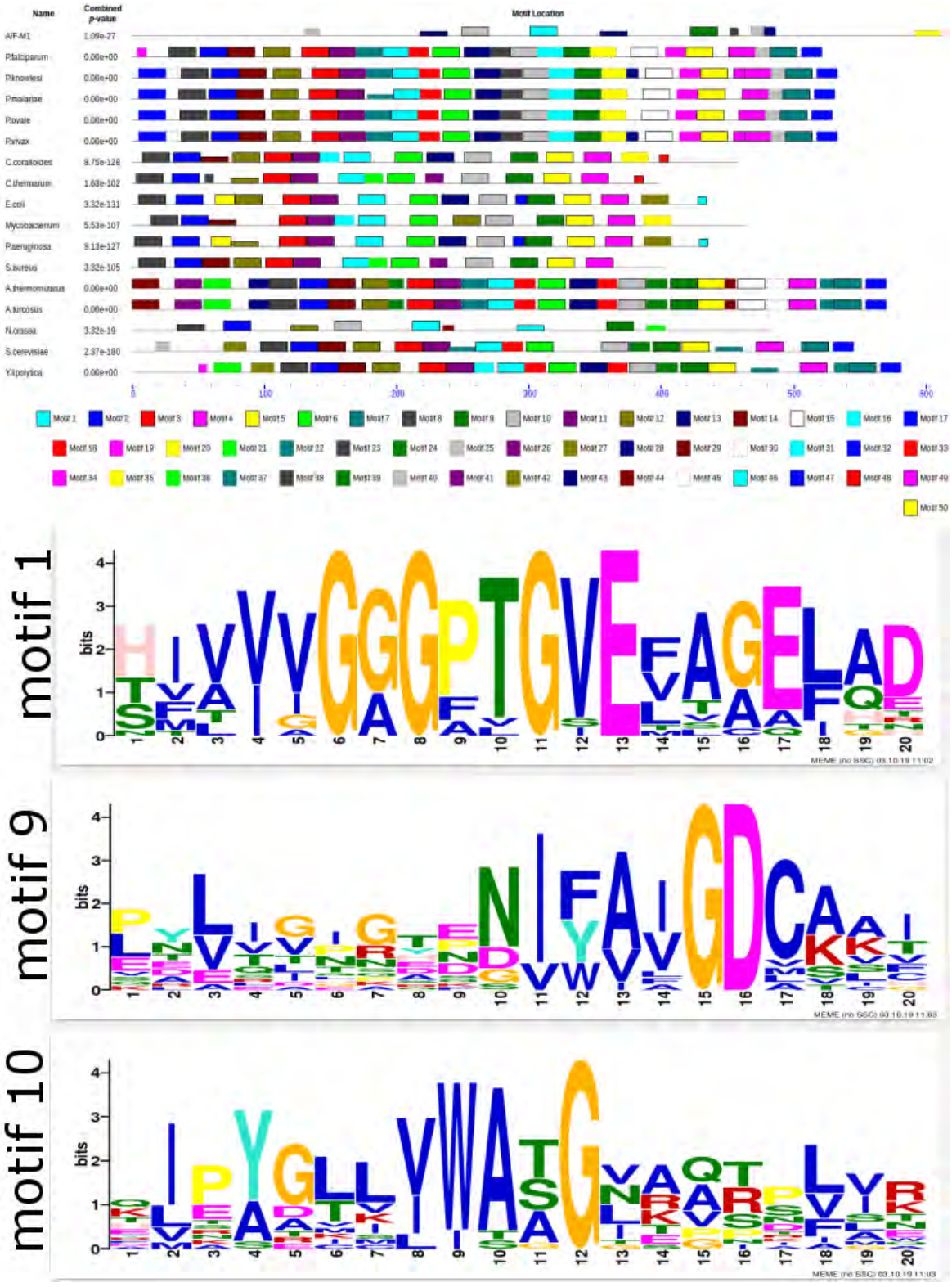

Figure 2.5 The map depiction of the 50 motifs that were calculated and the three Motif 1, 9 and 10 prominent motifs in all sequences

Figure 2.5 above shows the motif map, the name of the sequences as given in the FASTA file. The motif map showed the location of motif sites, with each block showing the position and strength of a motif site. Motif 1, 9 and 10 showed taller blocks proving the significance of the sites. The logo of the

motifs that appeared in all species was shown, the height of the letter indicates the relative frequency at the given position in the motif (Bailey et al., 2006).

The combined match *p*-value in the next column after the sequence name is defined as the probability that a random sequence (with the same length and conforming to the background) would have position *p*-values such that the product is smaller or equal to the value calculated for the sequence under test. The position *p*-value is defined as the probability that a random sequence (with the same length and conforming to the background) would have a match to the motif under test with a score greater or equal to the largest found in the sequence under test. The map shows the location of motif sites. Each block shows the position and strength of a motif site. The height of a block gives an indication of the significance of the site as taller blocks are more significant. The height is calculated to be proportional to the negative logarithm of the *p*-value of the site, truncated at the height for a *p*-value of 1e-10.

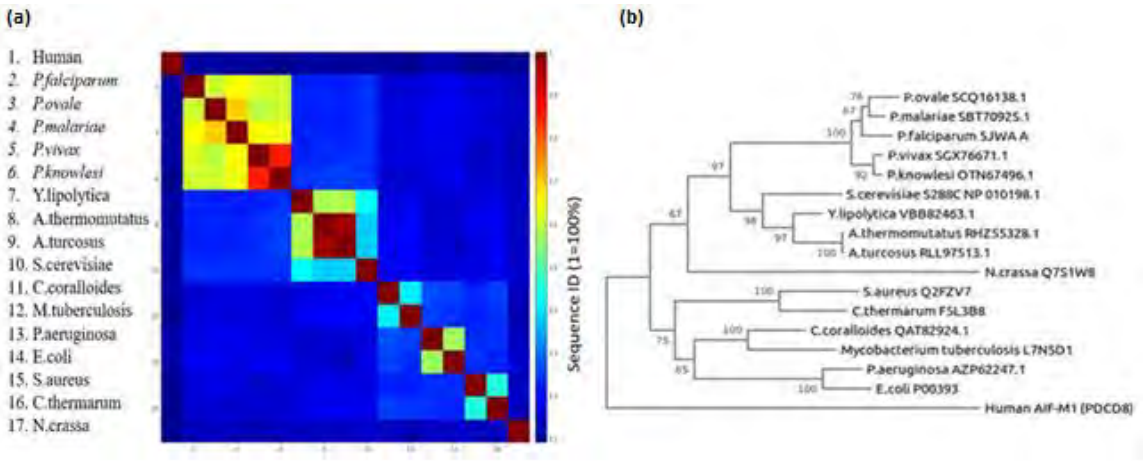

Figure 2.6 (a) Heatmap showing the sequence identity in all sequences and (b) showing the phylogenetic analysis of the sequences. Blue is low sequence identity and dark brown representing high sequence identity

In figure 2.6 above, there was a 65% identity amongst the *plasmodium* sequences, and a 90% identity between the *P. vivax* and *P. knowlesi. A. thermomutatus* had a sequence identity of 95% with *A. turcosus*. *P. aeruginosa* had a sequence identity of 55% with *E. coli*.

Phylogenetic trees contain a lot of information about the inferred evolutionary relationships between a set of sequences. The human species position has a longer branch in the horizonal dimension proving to be having large amount of change in comparison to the other species. The number close to the nodes indicated bootstrap value; it represented the phylogenetic confidence of the tree topology. Normally bootstrap value above 70 is considered a plausible score for branch validation. The number of bootstraps replicated was set at 100 and 65-90 number on the branch of the tree means that the tree is reliable.

Homology modelling is a technique that models the most accurate structure of a protein from its amino acid information based on the similar template (Moult et al., 2005). Pairwise sequence identity, secondary structural similarity and structural completeness characteristics were considered in selecting the most eligible template. 3-D structure modelling was performed using MODELLER tool, 100 models were created at very slow refinement. The model that had the lowest z-DOPE score was considered as the best model and was adopted for further studies and analysis. Homodimer models that had an FAD cofactor and four magnesium ions on each chain were created.

Table 1.1 The table below shows the models that were selected as the best from the 100 models that were created together with their z-DOPE score

Plasmodium species	3-D model selected	z-DOPE score	UniProtKB accession number
Plasmodium knowlesi	03	-0.78	OTN67496.1
Plasmodium malariae	01	-0.88	SBT70925.1
Plasmodium ovale	04	-0.83	SCQ16138.1
Plasmodium vivax	03	-0.88	SGX76671.1

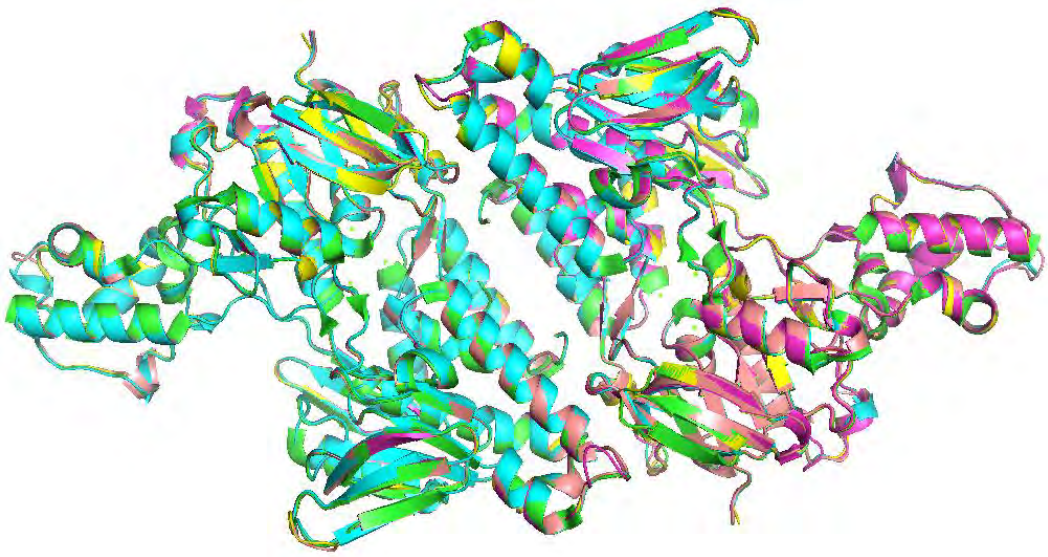

Figure 2.7 The four validated models, superimposed with the *Pf*NDH2 crystal structure. Green is the *Pf*NDH2, in yellow is the *Pk*NDH2, in cyan is the *Pm*NDH2, in pink is the *Po*NDH2 and in purple is the *Pv*NDH2 structure.

The quality of the models was assessed using Verfiy3D, PROCHECK and ProSA webservers as shown in table 2.2. In ProSA, the energy of the structure is evaluated after parsing the coordinates using a distance-based pair potential and a potential that captures the solvent exposure of protein residues (Sippl, 1993a). All calculations are carried out with C^{α} potentials, and the output on the web page displays a z-score and a plot of the residue energies. The *z*-score measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations and indicates overall model quality (Sippl, 1993b). The value for each of the models was displayed in a plot that contained the z-scores of all experimentally determined protein chains in the PDB, this plot was then used to survey if the z-score of the models were within the range of scores typically found for the proteins of similar size belong to one of the groups.

PROCHECK produces several output files in a default directory and provides assessment of the general quality of the model by comparing the model with well refined structures of the same resolution. The Ramachandran plots created classify the residues as favoured, allowed, or disallowed conformations, and in the table below majority of our residues were in the favoured or allowed regions validating the models. The quality of the crystal structure was also calculated using the same tools as those used for the 3-D models.

Table 2.1 Results obtained from the three webservers that were used to validate the quality of the models created, the quality of the crystal structure highlighted in green was also calculated

Protein structure	MODELLER Z-DOPE score	ProSA z-score	Verify3D (%)	PROCHECK (%)		
				Ramachandran residue locations		
				Favored	Allowed	Disallowed
***Pf*NDH2**	-1.56	-9.82	91.11	90.50	9.40	0.00
***Pk*NDH2**	-0.78	-10.03	88.15	90.70	8.60	0.20
***Pm*NDH2**	-0.88	-9.62	84.74	91.40	7.09	0.20
***Po*NDH2**	-0.83	-9.97	88.69	90.50	8.80	0.20
***Pv*NDH2**	-0.88	-9.98	88.76	91.80	8.50	0.00

2.4 Conclusion

This chapter focused on the sequence and structure analysis of *Pf*NDH2, *Pk*NDH2, *Pm*NDH2, *Po*NDH2, *Pv*NDH2 and AIF-M1 protein structures. NCBI and RCSB webservers were used to retrieve the crystal structures of the *plasmodium falciparum* (NDH2) and Apoptosis inducing factor (AIF-M1), and the former structure was used to create the 3-D models of the other four *plasmodium* species that causes malaria in human beings. To identify the conserved sequence patterns within the 17 sequences of different species, motif analysis was applied. Motifs 1, 9 and 10 were found in all 17 sequences. Motifs 1, 6, 10 and 11 were in the NADH binding site which form part of the NDH2 active site. ASP354 which formed conventional hydrogen bond with one of the oxygen in the FAD cofactor was present in motif 9, CYS355, Gly353, LEU316, THR211 and THR314 were some of the residues that form part of the enzyme‟s active site and were in motifs 1 and 9. Accurate 3-D models were constructed using MODELLER at a very slow refinement and three popular webservers were used to validate the quality of these structures. A low negative z-DOPE score value is considered a good native overlap.

Further experiments were later performed using these models and this was presented and discussed in the following chapters. The following chapter calculated and discussed molecular docking, molecular dynamics, PCA and DCC analysis of the structures without being bound to the membrane.

3

Structural analysis and compound search

Chapter overview

This portion of the study focuses on the mechanism of the NDH2, and how this protein interacts with the SANCDB compounds (Hatherley, Brown, Musyoka, Penkler, Faya, Lobb, & Tastan Bishop, 2015). Sequence analysis was performed in which we looked at the FASTA sequence of the NDH2, sequences of other species and the human homologue. In the previous section, multiple sequence alignment, motif analysis, phylogenetic analysis and sequence identity were the bioinformatics methods used to study the sequences. 3-D structure construction of the *plasmodium* species that causes malaria in humans but do not have crystal structures in the databases yet were also constructed. This section follows by describing the virtual screening of 623 geometrically refined and minimised SANCDB compounds against the *plasmodium* species and AIF-M1 protein structures. Molecular dynamics of the hit compound complexes were performed and analyzed to study the evolution of the systems at temperature; Principal component analysis (PCA) and Dynamic cross correlation (DCC) were the final methods applied in this study, looking at the residue behavior of the complexes, when ligand-bound or ligand-free.

3.1 INTRODUCTION

3.1.1 Virtual screening

In the early stages of drug discovery, virtual screening can provide an advantage to identify compounds with potential inhibitory activity towards the target of interest; optimisation of many parameters that combine the target properties (biological space), and ligand properties (chemical space) is essential to ensure the success of the drug discovery programme (Sandra et al., 2018). Autodock Vina requires receptor and ligand representations in a PDBQT file format, which is a modified protein data bank format containing atomic charges, atom type definitions and for ligands and topological information (rotatable bonds) (Berman et al., 2000). Molecular docking involves two conceptual stages: sampling and scoring; for sampling a large number of ligand poses are generated on the surface of a protein and these poses are then scored for their predicted binding affinities (Rentzsch & Renard, 2015). Natural compounds play a crucial role in drug discovery: they provide unique chemical structures, show more drug-like ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties and are well suited to phenotypic screening (Huggins et al., 2016). Some studies have fo-

cused on developing different rational drug design protocols including structural-based virtual screening for identifying novel candidates or have developed understanding of the most important chemical elements that guide ligand-protein interactions in relevant biological targets (Kellenberger et al., 2004). The accurate predictions of binding orientations and affinities are the principal issues of molecular docking, with docking algorithms exploring different ligand poses corresponding to different conformations and orientations within the receptor binding site and detecting the "best docking solution" among the orientations by using a scoring function which evaluates binding energies (Elokely & Doerksen, 2013). In order to optimize and rank results, scoring functions are used that are able to evaluate intermolecular binding affinity or binding free energy, choosing the best orientation after the docking procedure (Gutberlet & Rudolph, 1996). The scoring function defines the correct poses from incorrect poses or binders from inactive compounds during its computation (Meng et al., 2011). The formular below in Equation **3.1** shows the scoring function formula from AutoDock, for two atoms i, j; the pair-wise atomic energy is evaluated by the sum of van der Waals (vdw), hydrogen bond (hbond), coulomb energy (elec) and desolvation (sol). *W's* are weight factors used to calibrate the empirical free energy surface (Morris et al., 1998).

$$V = W_{vdw}\sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + W_{hbond}\sum_{i,j}E(t)\left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) + W_{elec}\sum_{i,j}\frac{q_i q_j}{e(r_{ij})r_{ij}} + W_{sol}\sum_{i,j}\left(S_i V_j + S_j V_i\right)e^{(-r_{ij}^2/2\sigma^2)}$$

Equation 3.1 Autodock formular, each of the pair-wise energetic terms includes evaluations for dispersion/repulsion, hydrogen bonding, electrostatics, and desolvation

3.1.2 METALizer

The NDH2 protein has four magnesium ions on each chain, hence it is important to know if these ions interact with the protein, in terms of if parameterization is required in preparation for molecular dynamics simulations (Y. Yang et al., 2017). METALizer predicts the coordination geometry of metals in metalloproteins, potential coordination geometries of metals are matched onto the found metal interactions in the examined structure (Schöning-Stierand et al., 2020). The predicted coordination geometries and the observed metal interaction distances can be compared interactively to statistics calculated on the PDB. METALizer is available on the ProteinsPlus server together with other tools such as SIENA (Bietz & Rarey, 2016). This tool allows the comparison of the predicted coordination geometries and metal interaction distances to statistics calculated only on related protein metal binding sites, and it also explores different binding models of ligands to the metal within the protein. Another available tool in this suite that is of use is the EDIA filter which detects atoms that are poorly supported by electron density (Meyder et al., 2017).

3.1.3 Molecular dynamics of protein structures without the membrane

Molecular dynamics simulations of the *apo* proteins, and simulations of the complexes with the hit compounds were performed using GROMACS (without embedding the complexes in a membrane) on the supercomputer at the Centre for High Performance Computing (CHPC) in Cape Town. For the protein structures, PDB text file of the *Pf*NDH2 crystal structure was downloaded from the RCSB website, while for the other structures the homology modelled structures were used. For the crystal structure, the crystal structure waters were stripped out using Discovery studio version 2020. All systems were visualised using the PyMOL program. GROMACS gmx pdb2gmx version 5.1.2 was used to prepare the protein structures for molecular dynamics simulations. As such, the five *plasmodium* species and the human homologue were prepared for molecular dynamics, and pdb2gmx was used to create the protein topology. The GlycoBioChem PRODRG2 server was used to prepare the ligand topology files (Schüttelkopf & Van Aalten, 2004). GROMACS was used to perform molecular dynamics with stages of optimization, equilibration, and production dynamics to produce simulation trajectories.

3.1.4 Principal component analysis (PCA) and dynamic cross correlation (DCC)

Principal component analysis (PCA) is the process of computing the principal components and using them to perform a change of basis on the data, sometimes using only the first few principal components and ignoring the rest. The principal components of a collection of points in a real coordinate space are a sequence of p unit vectors, where the i-th vector is the direction of a line that best fits the data while being orthogonal to the first i-1 vectors. This is applied to molecular dynamics trajectories in the following manner. Before applying the PCA algorithm, first frame of trajectory is superimposed onto a reference structure. A matrix of atomic coordinates can be diagonalised by either Eigenvalue Decomposition (EVD) or Singular Value Decomposition (SVD) to obtain the collective modes (eigen-vectors) and associated eigenvalues which characterize the motion of proteins during the MD simulation. The magnitude of an eigenvalue represents the variance of the data covered by its eigenvector (Ross et al., 2018). The most important motions are extracted by calculating the models from a covariance/correlation matrix constructed from atomic coordinates in case of EVD which are ranked based on their ability to explain the variance in the data. The projections on the selected modes called Principal components (PCs) represent the dynamics of a protein in terms of a reduced set of orthonormal modes.

To determine the degree to which any two atoms in a molecular dynamic simulation move together, dynamic cross correlation (DCC) is used (Borges et al., 2021). By extrapolating the motions along directions where major changes should occur, dynamic cross correlation can be used to understand the time-scale of a MD simulation (Karplus & Ichiye, 1995). The output is a matrix of all atom-wise cross correlations whose elements are displaced in a graphical representation.

3.2 METHODOLOGY

3.2.1 Ion and cofactor interaction with the protein

It was of importance to calculate and visualise the interaction between magnesium and the NDH2 protein, so to parameterize the ions prior to molecular dynamic simulations. The NDH2 crystal structure and the models were inputs in the webserver https://proteins.plus/, the option "METALizer Metal complex geometry prediction" was selected prior to initiation of calculation. Calculation was made for each magnesium ion on each chain.

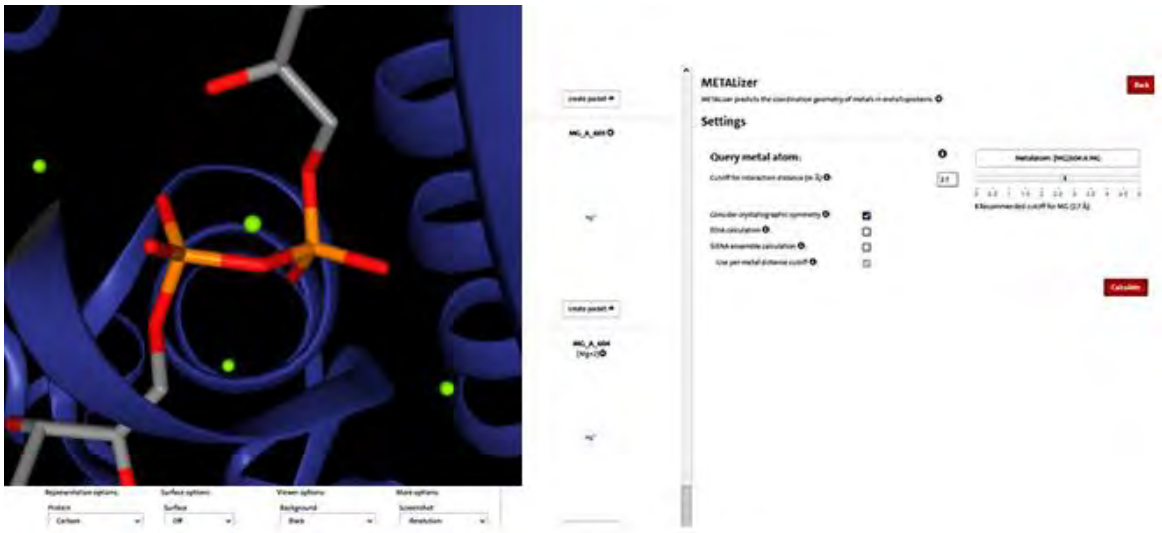

Figure 3.1 METALizer calculating the geometry of the magnesium ions represented as green spheres on the NDH2 protein structure and the stick is the FAD cofactor

3.2.2 Virtual screening

AutoDock Vina was used to perform molecular docking, and its implementation of the Lamarckian Genetic algorithm (LGA) is considered as one of the best docking methods available (Allouche, 2012). The PDB (ID:5JWA) structure was prepared first before performing molecular docking, in which to define the correct ionization and tautomeric states of the amino acid residues, hydrogen atoms were added to the enzyme using BIOVIA Discovery studio (Studio et al., 2020.) and to reduce complexities and steric hindrances, crystallographic water was removed using BIOVIA Discovery studio. A plugin tool in PyMol called AutoDock/Vina plugin was used to define the grid parameters and the grid for this study covered the interface region and part of the cofactor (FAD and NADH) binding site of the protein (Seeliger et al., 2010). The study aimed for a blind docking which had available as receptor the entire protein, so to allow for involvement of the interface region and the functioning or active sites.

The protein is a homodimer consisting of R39-K533 amino acids, FAD cofactor and four magnesium ions (Mg^{2+}) in each chain. FAD cofactor and the four magnesium ions that were present in the

*Pf*NDH2, ID: 5JWA crystal structure and present in the modelled structures were included in the docking calculations. Blind docking of the 623 South African Natural Compounds (SANCDB) on the *Plasmodium* species and the AIF-M1 protein, was performed using Autodock Vina. The prepare_ligand4.py script was used to prepare the ligands and the prepare_receptor4.py script was used to prepare the protein. Docking was performed identifying clusters with an energy range of 4, CPU of 4 and exhaustiveness of 96 and the box size parameters are shown in table 3.1 below. To eliminate any bias, the FAD cofactor was removed from the crystal structure and redocked to validate the docking results. Hit compounds were inspected visually for non-bonded interactions using BIOVIA Discovery Studio Visualizer. A rigid-protein flexible-ligand blind docking protocol was applied. Default torsional degrees of freedom were set in the ligands and Gasteiger atomic charges assigned.

Table 3.1 The parameters, dimensions and coordinates used for molecular docking

	Box dimensions (Å)			**Center coordinates (Å)**		
Plasmodium	**X**	**Y**	**Z**	**X**	**Y**	**Z**
P. falciparum	125	140	195	2.05	-35.95	-5.68
P. knowlesi	120	140	185	2.08	-39.87	-0.70
P. malariae	120	145	180	2.21	-40.87	0.41
P. ovale	120	145	185	2.20	-39.9	034
P. vivax	120	145	185	2.05	-38.87	-1.68
AIF-M1	70	70	80	6.0	54.0	24.0

LogP value, H-bond donors, H-bond acceptors, molecular weight, and rotational bonds which are the physicochemical properties of ligands were calculated using the free ADME/tox filtering tool and Lipinski rule of five. These properties are not absolute cut-offs but rather are viewed as guidelines; it is recommended that compounds should not have more than one violation of the properties for them to be considered safe for oral administration (Hossain et al., 2015). The five main parameters to test the drug likeness of compounds are absorption, distribution, metabolism, excretion, and toxicity which are calculated by the tools mentioned above. The ligands that bound to the desired regions, passed the physicochemical properties calculations, and those that had a lower binding energy were selected as hit compounds for further analysis.

3.2.3 Molecular dynamics simulation

Molecular dynamics simulation is well-established in the study of bio-molecular systems; this method provides key structural, energetic and thermo-dynamic information (Briones et al., 2019).

To elucidate the mechanism of action of the protein, and the structure exploration of the molecular nature of the protein, questions rose as to how the selected ligands could activate the protomers of the dimer. It is a challenging task to detect microscopic changes at the atomic level in the complex physiological environment through experimental techniques (Lei et al., 2020). To assist experimental investigations in protein systems molecular dynamic simulations provide insight into structures and dynamic behaviours at atomic levels (Fanelli & De Benedetti, 2011). GROMACS version 5.1.2 was used to perform the molecular dynamics and CHARMM36 (obtained from MacKerell lab website) was the force field used (Lindahl et al., 2021; Mackerell, 2004; Van Der Spoel et al., 2005). Pdb2gmx was used to prepare the protein topology and the GlycoBioChem PRODRG2 server was used to prepare the ligand topology files (Schüttelkopf & Van Aalten, 2004). There were no parameters available for the FAD cofactors in this force field, the cofactors were therefore extracted from the protein structures using BIOVIA Discovery studio and treated in the same way ligands during simulation. Hydrogen atoms were added to the ligands and the cofactors before processing them in PRODRG server. There were 4 complexes created and subjected to molecular dynamics for each *plasmodium* protein, the first being the *apo* and the remaining three for the hit compound complexes. Similarly, the AIF-M1 homology model was taken forward to dynamics with 4 systems, the first again being the *apo* and the remaining three were complexes with the three hit compounds (of *Pf*NDH2 proteins, but docked to AIF-M1).

Coordinates in the ligand.gro files were combined with those in the protein.gro files and the total number was adjusted according to the addition in the file. The parameters of the cofactors and the ligands were included into the topology file by adding "#include "ligand.itp" line into the topology file. Under the molecules section in the topology file the cofactor and ligand name were included to acknowledge new molecules in the protein.gro file. The unit cell was defined as cubic and an SPC/E solvent model was used to define the water model during solvation of the unit cell. The solvated.gro file was visualised in VMD before proceeding. Ions were added such that each system was neutralised with sufficient (Na+) or negative (Cl-) ions. The systems were simulated under periodic boundary conditions. The steepest descent method was used in the energy minimisation of the systems with a convergence criterion of 1000 $kJ.mol^{-1}.nm^{-1}$. Two consecutive equilibration steps were performed on the minimised systems at 300 K and a pressure of 1 bar and, including position restraints of heavy atoms of the entire system excluding those of ions and water molecules. The make_ndx command was used in the application of restraints to both the cofactors and ligands, to define the atoms to be restrained.

In the first step of equilibration, the number of particles, volume, and temperature were considered constant (NVT ensemble), while in the second step, the system was considered as isothermal-isobaric (NPT ensemble); this two-step equilibration allows the system to reach atom speeds compatible with the target temperature and pressure of the system (da Silva et al., 2020a). The position restraint of the protein-ligand atoms permits ions and water molecules to organize optimally along the surface of the

protein, forming more structured layers of solvation. Temperature control was achieved with the V-rescale thermostat (Bussi et al., 2007) and pressure coupling using Berendsen "pcoupl" (Berendsen et al., 1984). The linear constrained solver (LINCS) algorithm was used to constrain the bonds to hydrogen atoms in the complex (Hess et al., 1997). The long-distance electrostatic interactions were treated using the Particle-Mesh Ewald (PME) algorithm and the cut-off radius applied to the van der Waals and Coulomb interactions was 1.2 nm (Darden et al., 1993; Essmann et al., 1995). Production MD simulation was performed on the Centre for high performance computing (CHPC) supercomputer for 100ns of simulation using 2 fs integration time and considering the NPT ensemble without any position restraint.

3.2.4 Principal component Analysis (PCA) and dynamic cross correlation

Normal Mode Analysis (NMA) and Principal Component Analysis (PCA) are two popular methods used to quantify large-scale motions, and to identify the „essential motions", as conventional analysis of (MD) trajectories may not identify global motions of macromolecules (Nizami et al., 2016; Ross et al., 2018). To detect correlations in large data sets (PCA) was used, which allows for the summarisation of information content by means of a smaller set of summary indices that can be easily visualised and analysed. The MD trajectories were stripped off the Periodic Boundary Condition (PBC), water and saved in a (.xtc) file. Mode-task was downloaded and installed under the Anaconda environment on the YODA cluster; only the Cα atoms of the protein were used as input, the MD trajectories and a PDB file that had the same residues as in the trajectory file was used. MDM-Task "compare_essential_dynamics.py" tool implemented comparative essential dynamics run on the YODA cluster to assess the distribution of pocket conformational sampling across MD simulations of the protein structures (Sheik Amamuddy et al., 2021b). The output was the low energy basin and high energy PDB files and an essential dynamic plot picture showing the first (PC1) and second principal component (PC2) axes. Along the axes were the percentages of total variance explained by each PC. The principal components (PCs) represent the dynamics of a protein in terms of a reduced set of orthonormal modes (Ross et al., 2018).

Dynamic cross correlation (DCC) was calculated using the formula in equation 1.4 in chapter 1 (Di Marino et al., 2015).The MD-TASK program generates the heatmap showing the dynamic cross correlation (DCC) between the Cα atoms of the selected frames in an MD trajectory to identify relative residue movements (Brown et al., 2017b).

3.3 RESULTS

3.3.1 Molecular docking

The context of this study is the identification of compounds that could possibly alter the functioning of the Type II NADH: quinone oxidoreductase protein and possibly lead to the eradication of the *plasmodium* species that causes malaria in human beings. To this end, high throughput docking of

small natural compounds of the South African origin was performed on all the protein structures. Figure 3.2 below shows all 623 SANCDB (https://sancdb.rubi.ru.ac.za/) compounds docked on the homodimer proteins and the AIF-M1 protein. The proteins are oriented from chain A to chain H; compounds were binding mostly in the FAD and NADH binding domains, and in the interface region. Prior to screening, the FAD cofactor was redocked in the *Pf*NDH2 protein to validate the quality and accuracy of the docking results. We observed changes in conformation of the FAD cofactor when it was redocked to 5JWA protein; however similar and same interactions with residues were spotted. In the crystal structure we observed the Mg^{2+} ions interacting with the diphosphate region of the cofactor and, under redocking, these ions also interacted with both the diphosphate and adenine group (Figure 3.3).

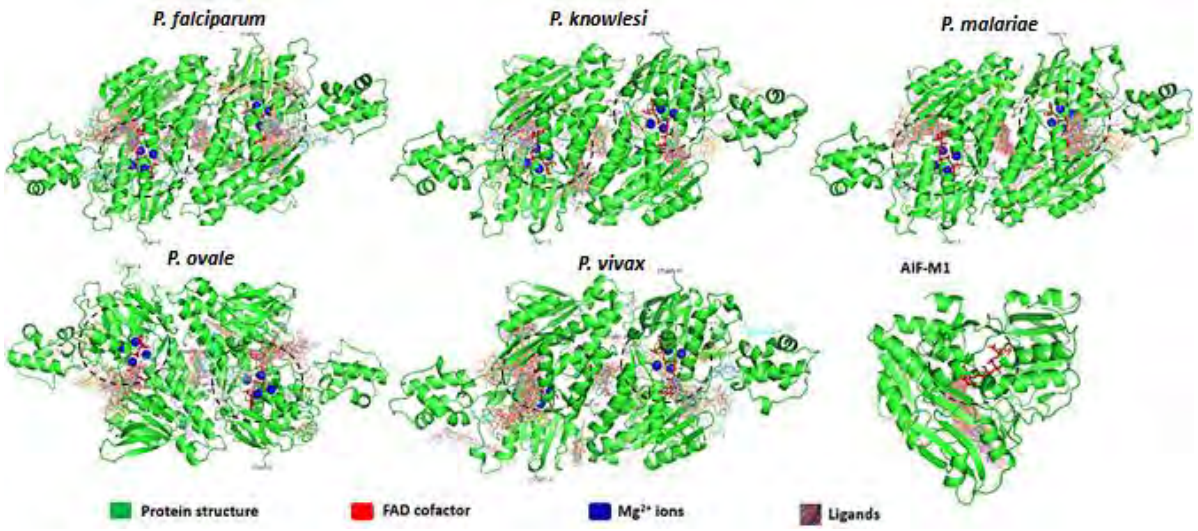

Figure 3.2 Docking results of the 623 SANCDB compounds viewed in PyMOL visualizer and the binding sites are circled in black broken lines

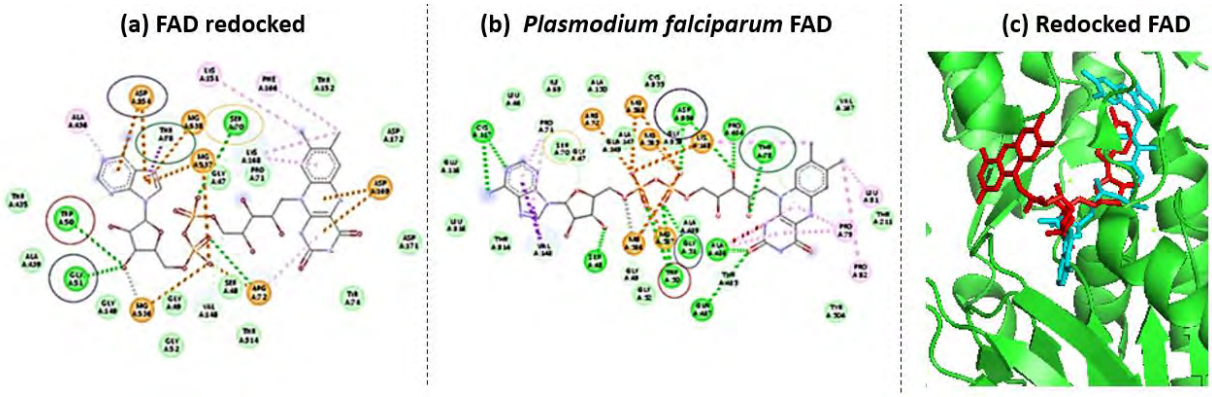

Figure 3.3 A picture of the redocked FAD cofactor to validate docking results, **(a)** is FAD cofactor redocked on 5JWA crystal structure, **(b)** the FAD cofactor interaction with the 5JWA crystal structure

and **(c)** The FAD cofactor binding in the FAD binding region, red (redocked FAD) and cyan (crystal structure FAD)

To get an overview of how well the SANCDB library was performing against each protein conformation, a binding energy heatmap of all protein structures versus the ligands was generated. R studio was used to map the binding energies of the protein and ligand interactions as shown in figure 3.3 (RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL Http://Www.Rstudio.Com/). The heatmap shows the AIF-M1, *P. falciparum, P. knowlesi, P. malariae, P. ovale* and *P. vivax*. The ligands in general exhibited lower binding energies when bound to AIF-M1 as compared to the other structures; while not concordant with the aim to find compounds binding favourably to the *plasmodium* species, it is individual compounds that are of interest rather than the dataset. Generally, the ligands had very low binding energies across all species.

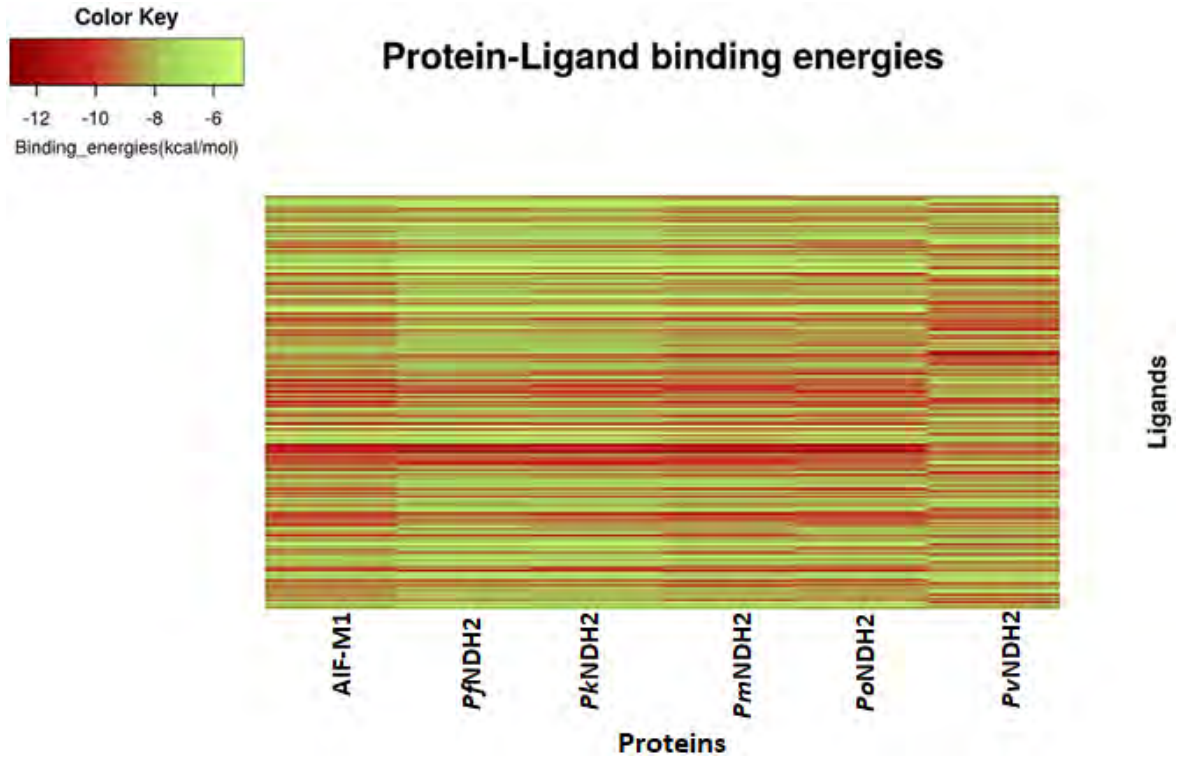

Figure 3.4 Binding energy heatmap of protein structures versus ligands, red depicts the lowest binding energy (kcal/mol) and green depicts the highest binding energy

The SANCDB is an accessible database that contains compounds isolated from the plant and marine life in and around South Africa, below are the 2-D representations of the most favorable hit compounds (Diallo et al., 2021a; Hatherley, Brown, Musyoka, Penkler, Faya, Lobb, & Bishop, 2015; SANCDB, URL https://sancdb.rubi.ru.ac.za/.). The ligands with the lowest binding energy, a small

sized ligand, binding in the active site of both chains and protein interface site and Lipinski rule of five score were considered when selecting the hit compounds shown in figure 3.5 below.

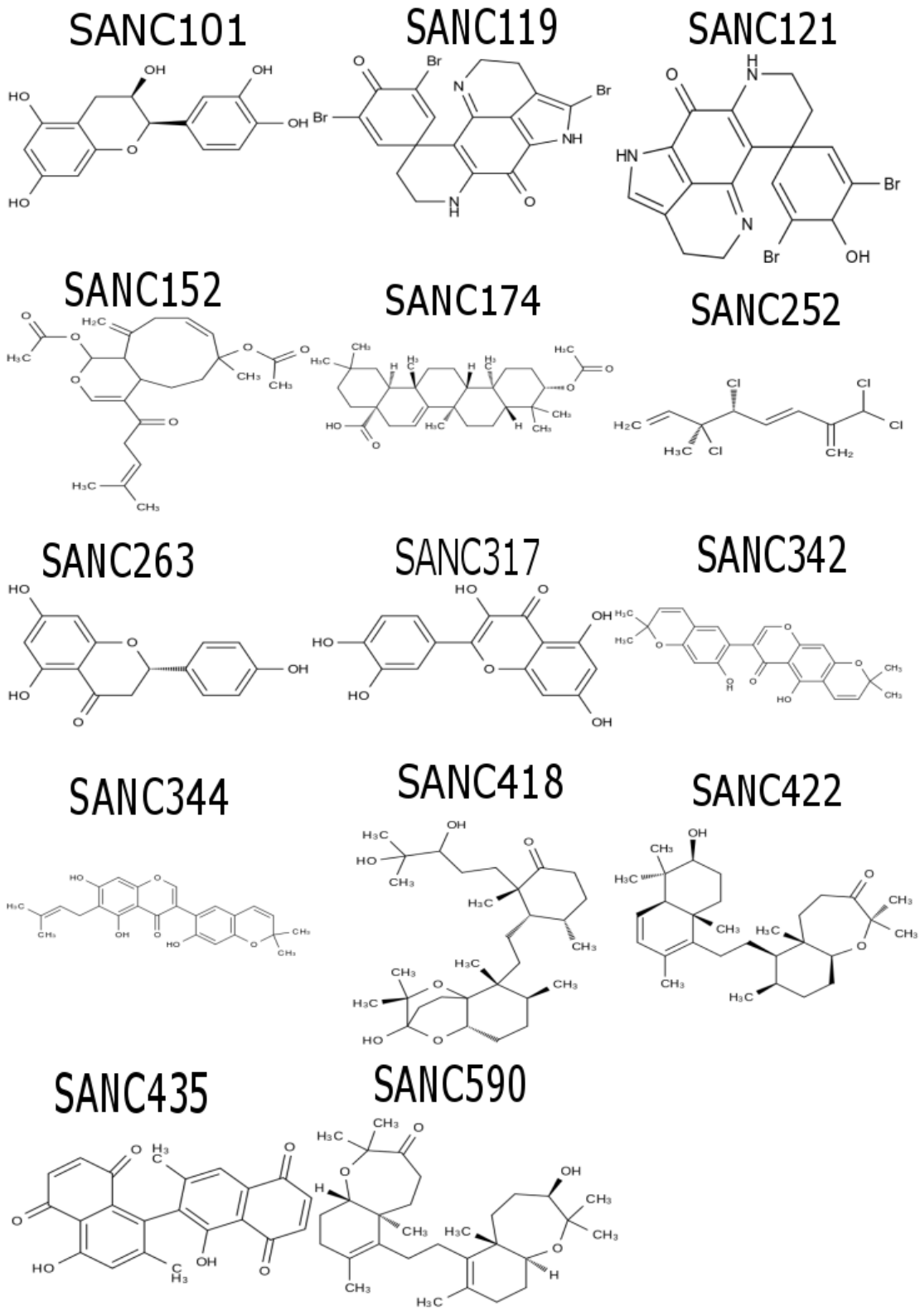

Figure 3.5 2D representation of the hit SANCDB compounds

It is of importance to calculate the oral availability and toxicity of the ligands before further analysis. As pioneered by Lipinski, analysis of the structures of orally administered drugs and of drug candidates has been the primary guide to correlating physical properties with successful drug development (Veber et al., 2002). The biological behaviour of the drug depends on several physiochemical properties such as water solubility, partition coefficient (LogP), melting point, boiling point, also including bio concentration factor (BCF); these properties affect various pharmacokinetic properties such as drug‟s bioavailability, transfer, permeability and others (Chandrasekaran et al., 2018a). The SCFBio Lipinski Rule of Five website (http://www.scfbio-iitd.res.in/software/drugdesign/lipinski.jsp) allows one to input data and calculate the druggability of the compounds (Lipinski, 2004). The lipinski rule of five helps to determine if the determined molecule has chemical and physical properties to be orally bioavailable. It is based on pharmacokinetic drug properties such as absorption, distribution, metabolism and excretion, but the specifics are: no more than 5 hydrogen bond donors, no more than 10 hydrogen bond acceptors, a molecular mass less than 500 Da and a partition coefficient not greater than 5. The violation of two or more of the above conditions suggests the molecule is not appropriate as an orally available drug. In the screening in this study, none of our hit compounds had more than 2 violations. Of course, molecular mass is the sum of the atomic masses of the atoms in a molecule. Lipophilicity refers to the ability of a chemical compound to dissolve in fats, oils, lipids, and non-polar solvents such as hexane or toluene, it also has a significant influence on various pharmacokinetic properties such as the absorption, distribution, permeability, as well as the routes of drugs clearance (Chandrasekaran et al., 2018b; van de Waterbeemd & Gifford, 2003). A hydrogen bond is an intermolecular force (IMF) that forms a special type of dipole-dipole attraction when a hydrogen atom bonded to a strongly electronegative atom exists in the vicinity of another electronegative atom with a lone pair of electrons. In biochemistry and medicinal chemistry the hydrogen bond is a key non-covalent interaction, a single hydrogen bond can decide the potency of drug-like molecules for a target when all other interactions stay constant (Bauer et al., 2019). In high throughput virtual screening the fitness is the total interaction energy of the ligand with the protein, and is evaluated using the energy function (Morris et al., 1998). Table 3.2 shows the physicochemical properties of the hit compounds from SANCDB.

Table 3.2 Lipinski rule of five results for the hit compounds together with the docking binding energy on the protein structures. *Pf*NDH2 shows the binding energy for *P. falciparum* and human homologue, respectively

Plasmodium	**SANCDB compound and their binding region**	**Molecular mass**	**High lipophilicity (logP)**	**Hydrogen bond donor**	**Hydrogen bond acceptor**	**Lipinski violations**	**Binding energy (Kcal/mol)**

***Pf*NDH2**	SANC101 (chain A)	290.27	0.51	5	6	0	-8.6 and -7.7
	SANC344 (chain H)	420.45	5.52	3	6	1	-10.0 and -9.1
	SANC418 (interface)	508.73	4.18	3	6	1	-8.8 and -8.2
***Pk*NDH2**	SANC263 (chain A)	272.25	2.52	3	5	0	-8.0
	SANC435 (chain H)	374.34	3.72	3	6	0	-10.5
	SANC342 (interface)	418.44	4.87	2	6	0	-10.2
***Pm*NDH2**	SANC101 (chain A)	290.27	0.51	5	6	0	-8.6
	SANC119 (chain H)	545.04	3.2	1	5	1	-9.4
	SANC422 (interface)	456.7	6.24	1	3	1	-10.4
***Po*NDH2**	SANC252 (chain A)	274.01	4.97	0	0	0	-6.1
	SANC174 (chain H)	498.74	8.4	1	4	1	-9.8
	SANC590 (interface)	472.7	4.2	1	4	0	-10.6
***Pv*NDH2**	SANC121 (chain A)	467.15	1.26	2	5	0	-8.8
	SANC317 (chain H)	302.24	1.54	5	7	0	-8.9
	SANC152 (interface)	416.51	3.65	0	6	0	-8.0

3.3.2 Molecular docking protein-ligand interaction

Molecular docking requires efficient sampling across the entire range of positional, orientational, and conformational possibilities and it is a difficult optimisation problem (Morris et al., 1998). AutoDock vina was the package used to perform high throughput virtual screening. Drug discovery programs

employ high throughput screening paradigms with the intent to identify high affinity binders to target macromolecules (Gleeson et al., 2011). In this study, the docked ligand poses in all six protein structures were visually inspected using PyMOL. The protein structures were oriented from chain A to chain H. The hit compounds as shown in table 3.2 above were mapped on the 3-D structures in their respective binding sites. Only compounds that were hit compounds in *Pf*NDH2 were visualised and analysed in the human homologue. Although the *plasmodium* species were similar structurally, we observed difference in ligand binding and interaction, hence compounds that were hit compounds in one structure were not necessarily hit compounds in the other.

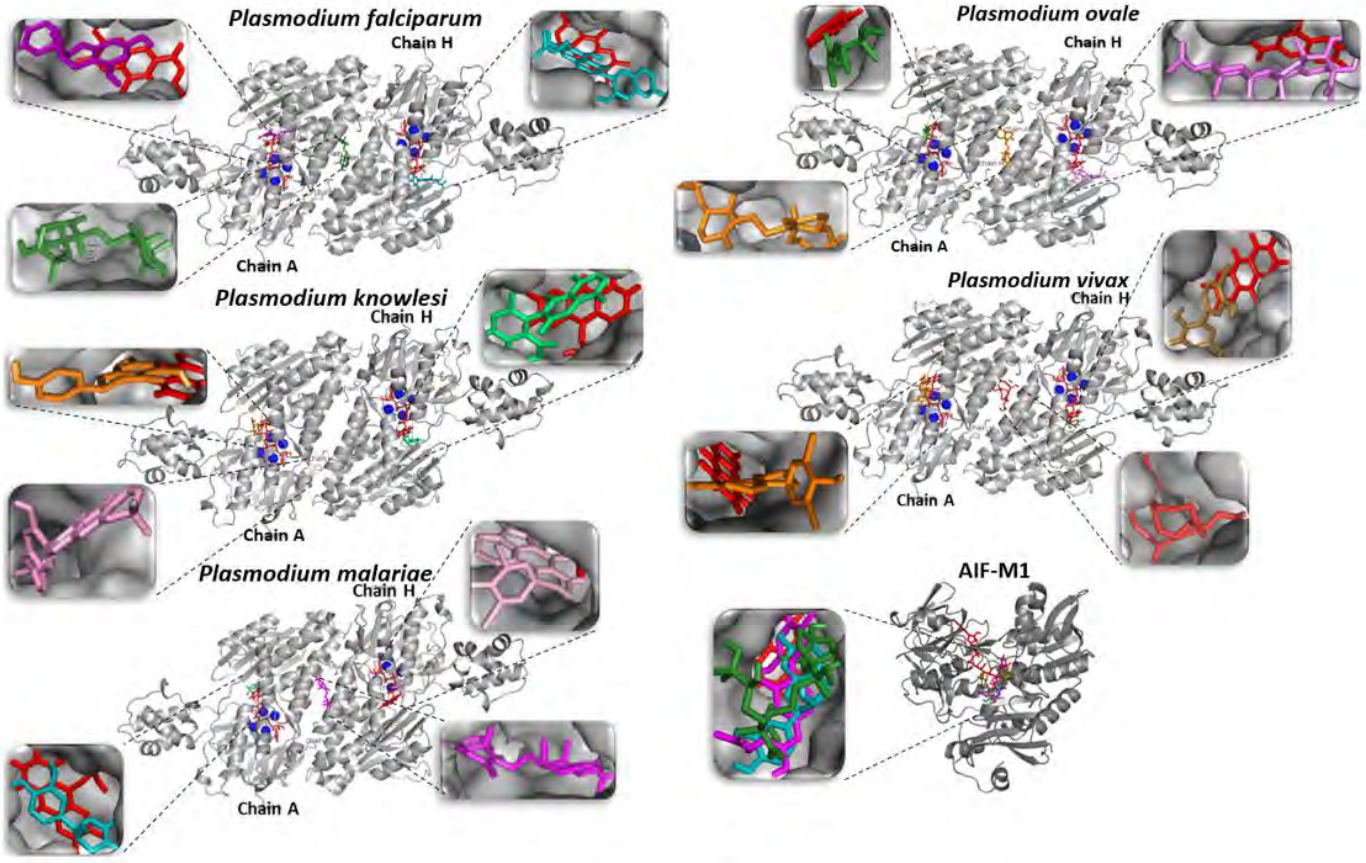

Figure 3.6 Visualization of the hit compounds docked on the protein structures. The magenta stick bound to chain A, the cyan stick bound to chain H and the green stick bound in the interface region of *P. falciparum*. The orange stick bound to chain A, the green stick bound to chain H and the pink stick bound in the interface region of *P. knowlesi.* The cyan stick bound to chain A, the pink stick bound to chain H and the dark pink stick bound in the interface region of *P. malariae*. The green stick bound to chain A, the pink stick bound to chain H and the orange stick bound in the interface region in *P. ovale*. The orange stick bound to chain A, the brown stick bound to chain H and the red stick bound in the interface region of *P. vivax*

BIOVIA discovery studio was used to analyse the protein and ligand interaction as shown in figures 3.7-3.9 below; we focused on hydrogen bonds (conventional hydrogen bond and carbon hydrogen bond) and hydrophobic bonds (Pi-Pi stacked, Pi-sigma, Pi-alkyl and alkyl). Van der Waals interactions are distance-dependent interactions between atoms, they are comparatively weak and can be easily broken (Nielsen, 2009). Conventional hydrogen bonds are a specific type of electrostatic interaction between a proton attached to an electronegative atom such as a nitrogen or an oxygen and a lone pair of electrons on an electronegative atom such as N, O, or F, they are very important in drug–receptor interactions as well as the structural integrity of many biological molecules, including proteins (Arunan et al., 2011). The carbon hydrogen bonds are non-covalent interactions that are weaker than heteroatom-hydrogen bonds, however when activated by electron withdrawing groups such as C-H groups that are covalently bound to a cationic nitrogen at (N+-C-H), they may become strong as heteroatom hydrogen bonds and could be important for molecular recognition (Itoh et al., 2019). Pi-Pi stacking interactions occur between aromatic rings containing π orbitals and can be used in broad applications such as immobilisation and specific recognition (Chen et al., 2018). Pi-alkyl interactions are non-covalent interactions involving Pi systems. Alkyl interactions have some relevance in biology and chemistry and are weak interactions (Giese & Albrecht, 2020). The hit compounds, possessing the lowest binding energies, were taken in complex with the target, to all atom molecular dynamics simulation studies.

*Pf*NDH2 interacted with the following ligands as follows SANC00101 formed conventional hydrogen bonds with Thr152, Pro434 and Thr211, and it had a much weaker carbon hydrogen bond with Ala308. SANC00344 formed a conventional hydrogen bond with Ser472 and carbon hydrogen bond with Thr211. SANC00418 formed conventional hydrogen bonds with Gly87, Thr88 and Leu89. For *Pk*NDH2 the ligand interactions were as follows: SANC00263 formed conventional hydrogen bonds with Tyr507; SANC00435 formed conventional hydrogen bonds with Lys167H and Gly310H; while SANC00342 formed conventional hydrogen bonds with Thr100A, Asn97A and Arg515. *Pm*NDH2 interacted with SANC00101 forming conventional hydrogen bonds with Thr437 and Gly312. However, for the SANC00119- *Pm*NDH2 there were conventional hydrogen bonds formed with Gly474 and Gln439, and SANC00422- *Pm*NDH2 showed conventional hydrogen bonds with Asn92A, Arg176H and Arg526H. The interaction between *Po*NDH2 and SANC00252 was without conventional hydrogen bonds; however there were alkyl bond interactions with Leu473, Val503, Tyr504, Ala436, Trp50, Pro79 and Leu507; with SANC00174 there were conventional hydrogen bonds with Ser433 and Ala308 and with SANC00590 conventional hydrogen bonds formed with Leu89. Finally, *Pv*NDH2 interacted with SANC00121 forming conventional hydrogen bonds with Gly474, and carbon hydrogen bonds with Leu510, Lys442 and Lys473; with SANC00317 there were conventional hydrogen bonds with Ala149 and Pro436, (although this ligand had an unfavourable metal-donor interaction with Gly310 and an unfavourable acceptor-acceptor interaction with

Mg^{2+}546, we still considered it as a hit compound worthy of further investigation); and SANC00152 formed conventional hydrogen bonds with Ser90H, Lys92A, Ala91A and Arg526.

We looked into the interaction of the *Pf*NDH2 hit compounds in AIF-M1 and, for SANC00101 observed an unfavourable donor-donor interaction with Leu311, but favourable Pi-Pi stacked (with Phe482) and Pi-Alky (with Ala494) interactions. For SANC00344-AIF-M1 we observed a conventional hydrogen bond with Glu453, and for SANC00418-AIF-M1 we observed a Pi-sigma interaction with Phe482. The ligands selected as hit compounds in *P. falciparum* had more favourable protein-ligand interactions and showed more conventional hydrogen bonds in comparison to when bound to the human homologue.

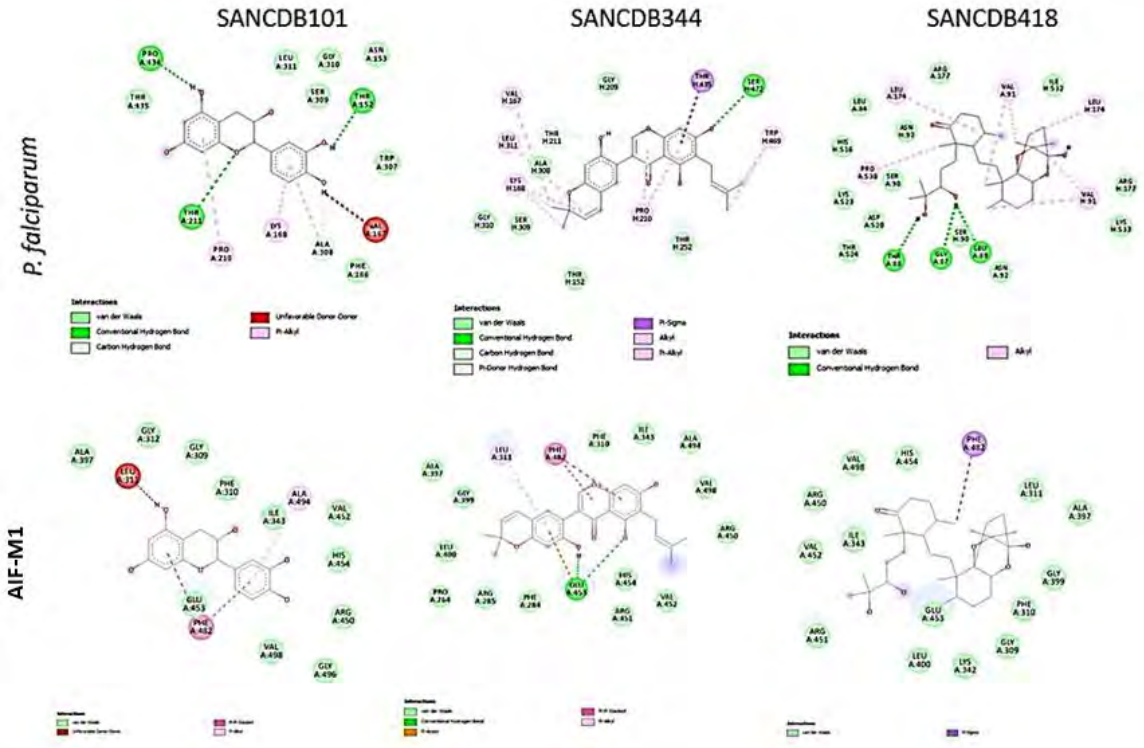

Figure 3.7 Protein-ligand interaction of the hit compounds in *P. falciparum* and AIF-M1 visualized in Discovery studio

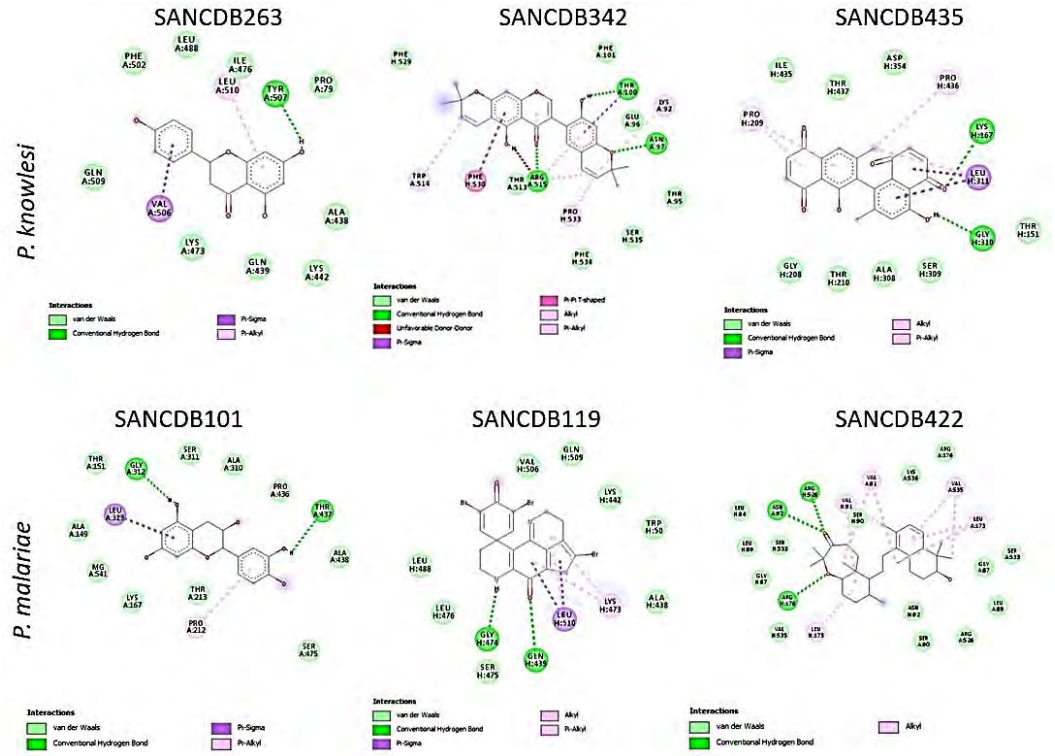

Figure 3.8 Protein-ligand interaction of the hit compounds in *P. knowlesi* and *P. malariae,* visualized in Discovery studio

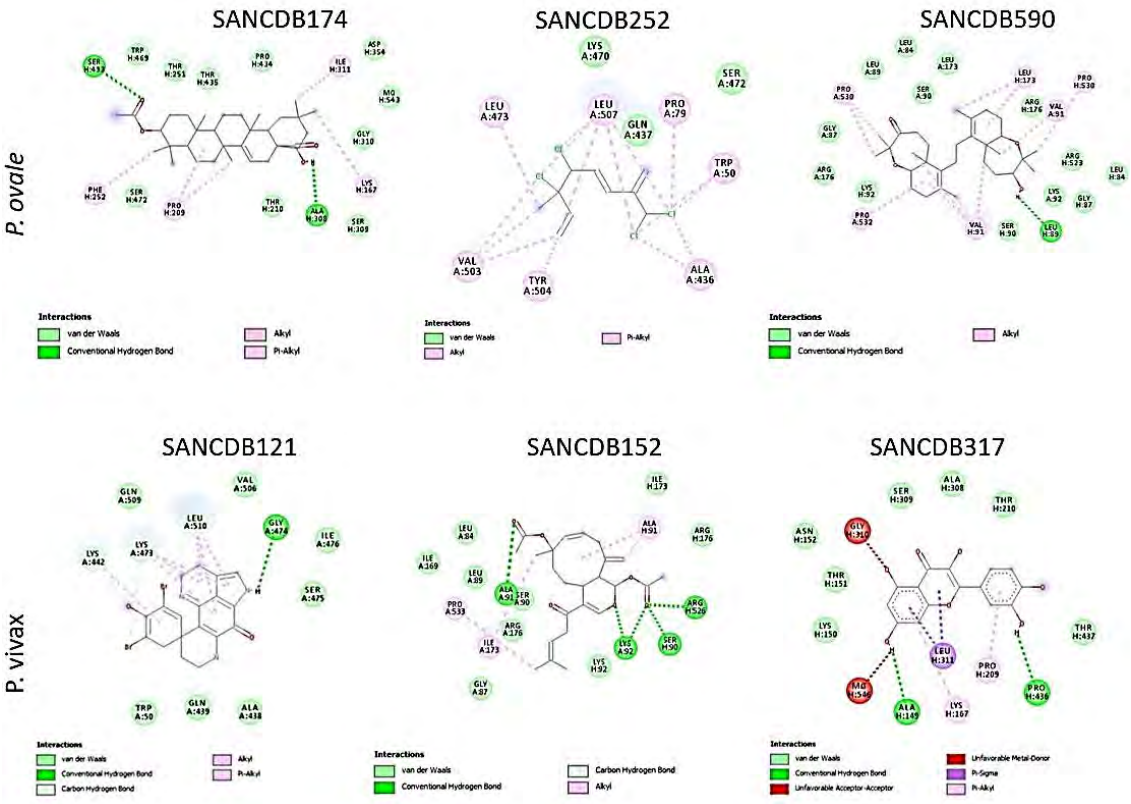

Figure 3.9 Protein-ligand interaction of the hit compounds in *P. ovale* and *P. vivax,* visualized in Discovery studio

3.3.2 METALizer

It is interesting to note that throughout the PDB an octahedral arrangement around Mg^{2+} is most common. However, from the output results when studying the Mg^{2+} in our *Plasmodium* NDH2 structures, the METALizer webserver suggests that in our case, the magnesium ions are free from the protein structure and only interact with the FAD cofactor. Due to this finding, it was of no need to parameterise the Mg^{2+} ions as coordinated to the protein before initiating molecular dynamic simulations.

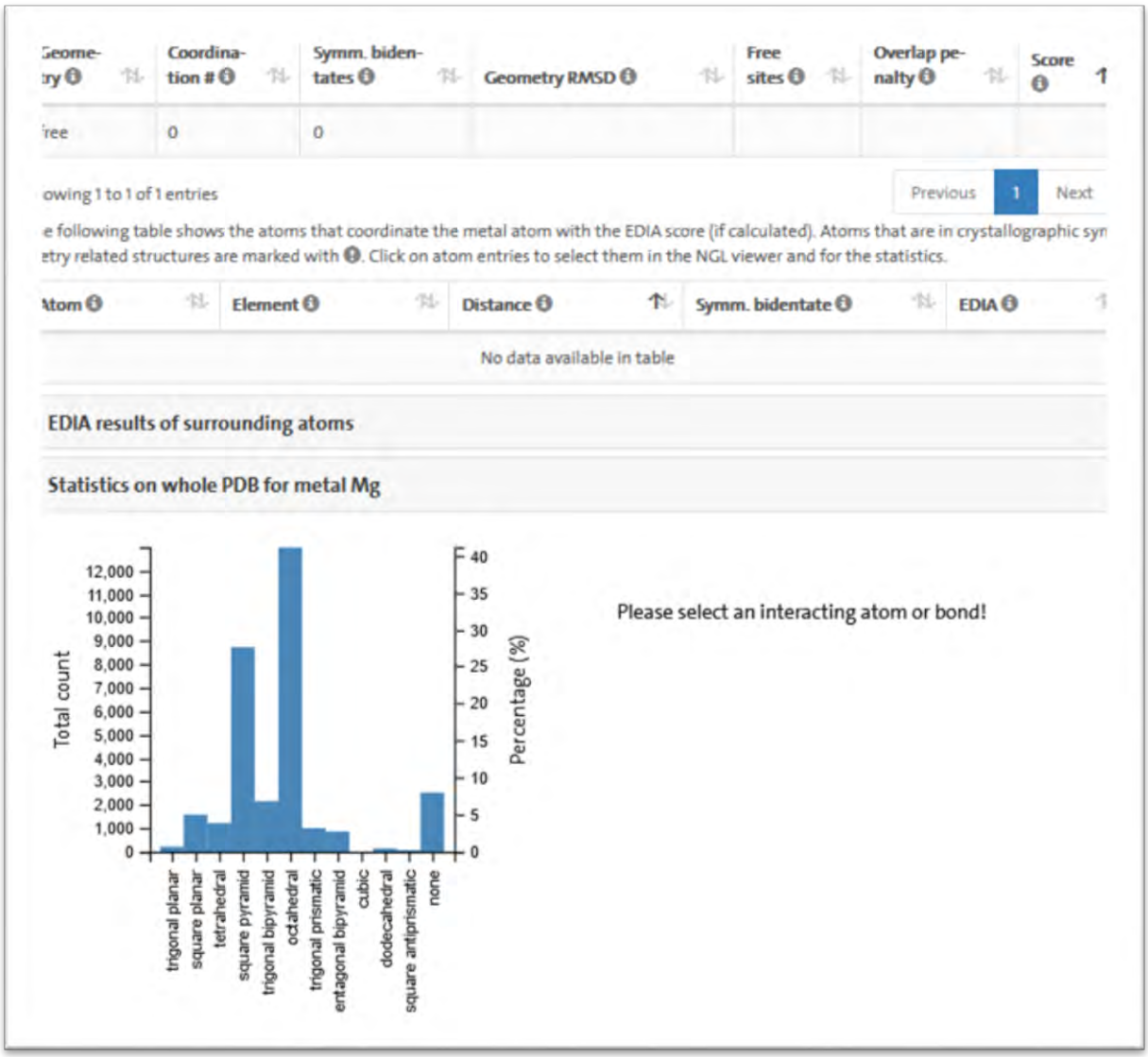

Figure 3.10 Results from the METALizer webserver showing the geometry value between the NDH2 protein structure and the magnesium ions

3.3.3 Molecular dynamic simulations without the membrane

Following production dynamics, the trajectories were analysed. The RMSD, RMSF and Rg was calculated from the MD simulation results. RMSD measures the structural distance between

coordinates, and for the protein this relates to the distance change during dynamics of the backbone atoms; RMSF on the other hand measures the average deviation of a particle over time from a reference position, allowing the identification of portions of structure that most fluctuate from their mean structure.

MD simulations for both the *apo* protein and the complex when bound to each of the hit compounds was calculated for a total simulation time of 100 ns. The MD trajectories were stripped of water and the protein was centred with respect to the periodic boundary conditions prior to analysis. The VMD tool was used to visualise the protein dynamics and changes occurring throughout simulation. The RMSD values for *P. falciparum* were below 0.5 nm (Figure 3.11), the deviation patterns for both the *apo* and complex protein were stable from 6 ns until the end of the simulation. Note that a stable RMSD for the complex is not necessarily favourable as we are interested also in compounds that destabilise the protein and potentially affect its functioning.

The RMSD values for the *P. knowlesi* (both *apo* and complexes) were below 2 nm; however, it took some time before these systems stabilised. The RMSD for this protein, *P*kNDH2 with ligands bound, showed a lower value in general compared to the *apo* case. In general, the RMSD values for the *P. malariae* simulations were below 1.0 nm; the systems stabilised after about 6 ns. For the *P. malariae* complex with ligand SANCDB119 (where the binding is to chain H), there was a greater deviation in this complex and stabilisation was observed after 80 ns. The RMSD values following simulation of *P. ovale* (*apo* and complexes) were below 2.6 nm and general stability in the complexes after 48 ns; however, the *apo* protein did not exhibit this stability. The RMSD values of *P. vivax* in all systems was below 5.0 nm; in general, a lower RMSD is better than a higher one - we observed a high RMSD in the SANCDB152 complex with *Pv*NDH2, while the other complexes and the *apo* protein showed stability. In this study only the *Pf*NDH2 hit compounds were docked on the AIF-M1 protein, since it is this *plasmodium* that is responsible for the more death in humans. The RMSD plots for AIF-M1 (*apo* and complexes) showed the complexes as stable with no significant difference in comparison to the *apo*. However, an interesting aspect of these systems is the effect of each ligand on the functioning of the homodimer.

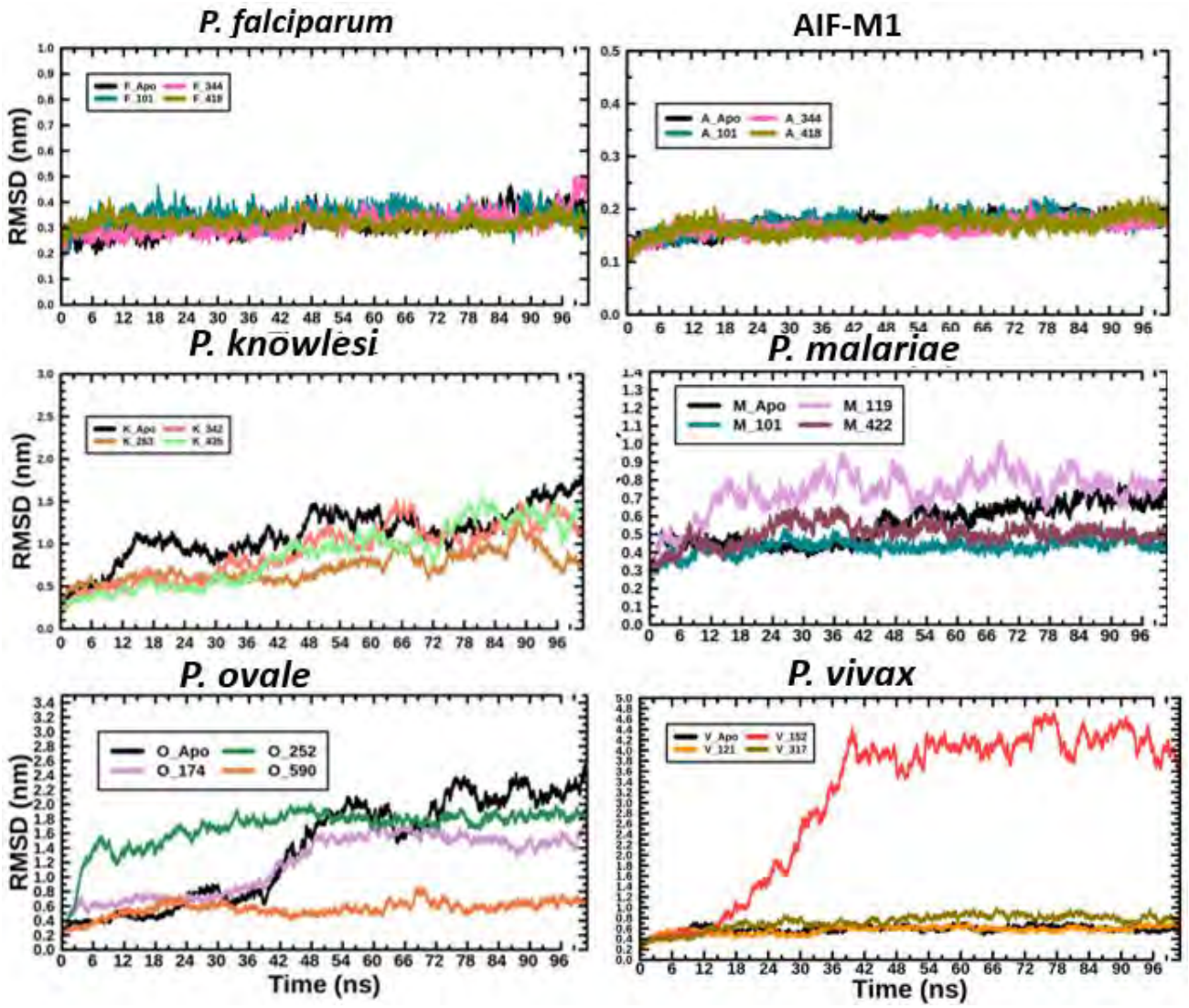

Figure 3.11 RMSD results from the complex and the *apo* proteins, F represents the *Pf*NDH2, A represents the AIF-M1, K represents *Pk*NDH2, M represents *Pm*NDH2, O represents *Po*NDH2 and V represents *Pv*NDH2

The protein at study is a homodimer and it was of interest to observe the residue behavior during simulation for both chains. From the RMSF calculations, differences in the residue fluctuations for each chain could be identified. Residues R39-I146 make up the FAD binding domain, residues D171-Q313 cover the NADH binding domain, residues T314-Q359 form the linker region, P360-N428 define domain C and residues A436-K533 cover the C-terminal domain. As the other protein structures were homology modelled from the *P. falciparum* crystal structure, this meant the description of the binding domains above were slightly different for these structures in terms of

change in residue number. The FAD and NADH binding domains are responsible for the association of cofactors, the biological function of domain C is not clearly defined and the C-terminal domain promotes the dimerization and immobilization of the enzyme on the inner mitochondrial membrane (Xie et al., 2019).

The RMSF of all residues for *P. falciparum* was below 0.5 nm in all simulations but observed were higher fluctuations between residues G343-N438 of both chains which are residues in the linker, domain C and C-terminal region; the SANC344 complex showed the higher fluctuations. The RMSF of *P. knowlesi* and *P. malariae* also showed high fluctuations, but in a narrower range between residues G343-H419 in both chains, these residues form part of the linker and Domain C region. There were higher fluctuations in the RMSF of the *apo P. ovale* system in comparison to when the ligands were bound to the protein, particularly between the residues N191-F286 found in the NADH binding domain in chain A and residues K40-Q114 found in the FAD binding domain of chain H.

There was high fluctuation in the RMSF of the *P. vivax* when SANCDB152 was bound to the interface region of the protein; meanwhile, the *apo* system and the other protein complexes had a low RMSF below 0.6 nm. We observed a higher RMSF value of 0.34 nm in the residue N583 when SANCDB418 was bound to AIF-M1 protein, the RMSF values in general in AIF-M1 were below 0.35 nm.

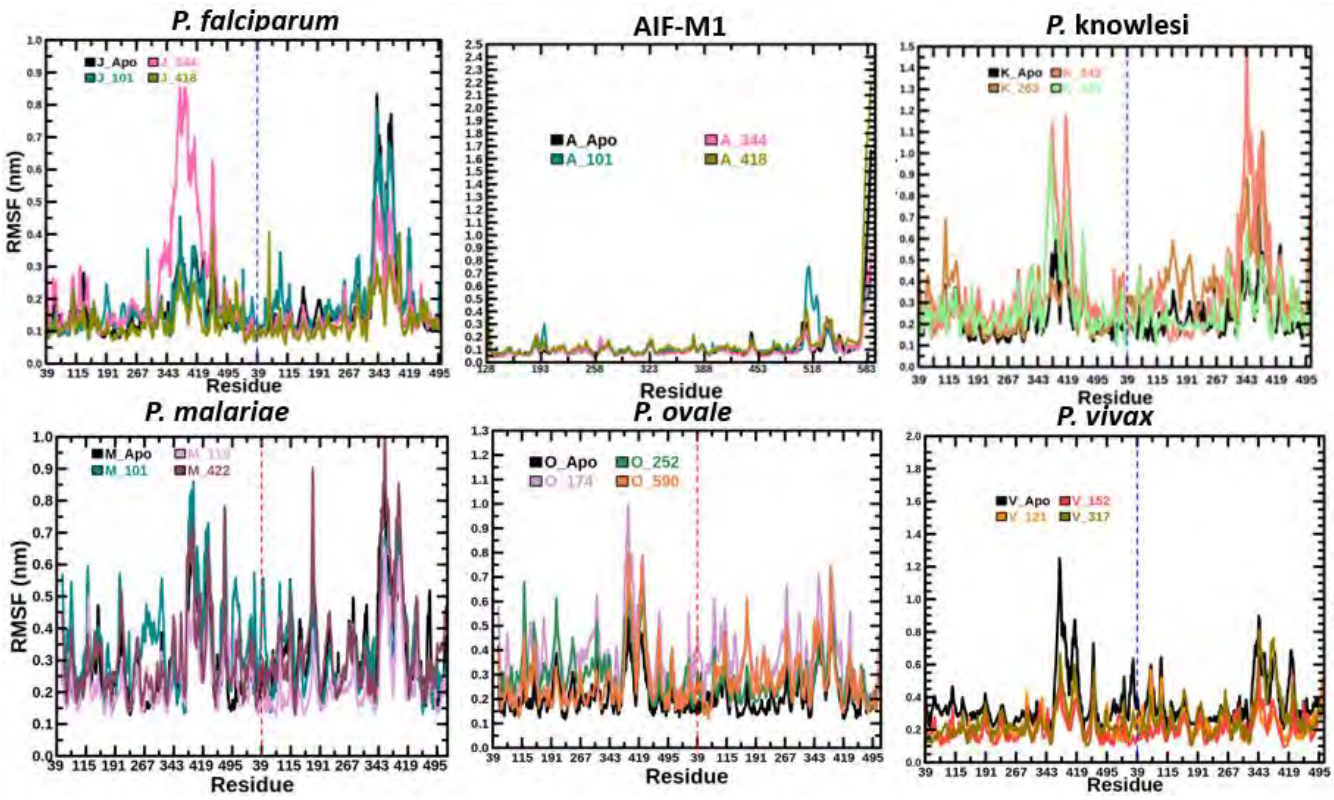

Figure 3.12 RMSF results of the complex and *apo* proteins, J represents the *Pf*NDH2, A represents the AIF-M1, K represents *Pk*NDH2, M represents *Pm*NDH2, O represents *Po*NDH2 and V represents *Pv*NDH2. The red and blue dotted lines separate the chains

3.3.4 Essential dynamic (ED) presentations and dynamic cross correlation (DCC) heatmaps

Essential dynamics (ED) is the process of applying PCA to a protein trajectory, where the "essential" motions are extracted from the set of sampled conformations (Tan et al., 2014). Molecular dynamics simulations produce large complex datasets showing the dynamic changes proteins go through and it is challenging to analyse all spatial variables and their relation to each other. PCA extracts the protein‟s most dominant modes of motion from the simulation trajectory, it is performed on the molecule‟s mass-weighed Cartesian coordinates. It accounts for the essential dynamics of the system, and this account is on a low-dimensional free energy landscape (Sittel et al., 2014). Comparative ED applies k-means to sample centroid conformations from the first 2 principal components in standard PCA and features auto-mated conformation extraction from lowest energy basins (Sheik Amamuddy et al., 2021b). Data shows that most of the conformational sampling is covered by PC1 and PC2 which represent the largest and second largest possible variances of the structures respectively. More stable protein conformations are expected to be associated with lower free energy.

The essential dynamics of the complexes were explored using standard PCA for the *apo* protein and for the protein-ligand complex MD trajectories, and for this only Cα atoms were selected. The dots on

the PCA plots produced (see Figure 3.13a PCA for example) correspond to individual protein conformations and are coloured by the time of sampling. In these plots, the kernel density contour plots are coloured from blue (lowest density) through yellow to red (highest density); this is a visual guide for the energy surface and is independently scaled (Sheik Amamuddy et al., 2021b). Further, these plots include both red (high energy basins) and blue labels (low energy basins) with data obtained from the k-means algorithm, and these are obtained from the probability density maxima. The k-means algorithm is non-deterministic, even though sampling is repeated a moderately large number of times to mitigate this effect.

The MD simulation was 100 ns long for all protein complexes and the PCA plots are shown in figures below (Figures 3.13, and 3.15-3.19). PCA was done to analyse, through the 3-D Cα, the conformations; while the RMSD showed the conformational changes in a 1-D representation, the PCA shows information from a multidimensional analysis. (Table 3.3) below shows the low energy basin for *Pf*NDH2 protein together with their PC1 and PC2 percentages. The protein-ligand complexes showed change in structure conformation at low energy basins at the middle of the simulation in protein-ligand structures, and in the *apo* protein low energy basin was towards the end of simulation. The PCA results agreed with the RMSD results as the lowest energy basins showed a low RMSD and a stable deviation.

Table 3.3 The PCA low energy basin results together with their PC1 and PC2 percentage ranges for the human homologue and *plasmodium* species

***Plasmodium* complex**	**Low energy basin (ps)**	**PC1 (%)**	**PC2 (%)**
*Pf*NDH2_*apo*	71550	28.29%	12.30%.
*Pf*NDH2_101	56350	23.54%	19.35%
*Pf*NDH2_344	5380 and 25000	34.76%	11.82%
*Pf*NDH2_418	40150	26.36%	11.32%
AIF-M1-*apo*	73520	23.81%	7.12%
AIF-M1-101	102470	20.34%	7.83%
AIF-M1-344	75810	24.33%	7.69%
AIF-M1-418	65840	18.20%	10.24%
*Pk*NDH2_*apo*	86 050	73.55%	9.20%
*Pk*NDH2_263	31040	66.39%	9.36%
*Pk*NDH2_342	21240 and 87030	76.15%	6.12%
*Pk*NDH2_435	14720 and 44060	76.60%	14.51%
*Pm*NDH2_*apo*	64500	57.38%	8.45%

*Pm*NDH2_101	13350	42.47%	16.14%
*Pm*NDH2_119	79820	56.17%	15.02%
*Pm*NDH2_422	42920	29.55%	18.81%
*Po*NDH2_*apo*	19270	89.53%	4.73%
*Po*NDH2_174	92380	80.91%	7.71%
*Po*NDH2_252	54960 and 91490	69.96%	9.40%
*Po*NDH2_590	39550	38.17%	24.63%
*Pv*NDH2_a*po*	78800	32.32%	17.10%
*Pv*NDH2_121	50400 and 89140	27.66%	21.58%
*Pv*NDH2_152	43810 and 13480	73.90%	14.92%
*Pv*NDH2_317	47350	60.34%	9.68%

To calculate the extent in which protein residues interact together, dynamic cross correlation was used. The dynamic cross correlation plots (in Figure 3.13b) show the *apo*, *Pf*NH2-101, *Pf*NDH2-344 and *Pf*NDH2-418 results respectively. The DCC is represented from a range of -1(blue) to 1(red) in which 1(red) is high correlation, the X and Y axes show the residue index from 0-500 residues; the relationship between this number and the residues Arg39-Lys533 is as follows: 0 = R39, 100 = L139, 200 = S239, 300 = L339, 400 = A439 and 500 = 539. A high correlation was noted between residues Lys299-His479 (260-440) and Lys299-His479 (260-440) that form part of the NADH domain, the linker region and part of the C domain, in both axes in the *Pf*NDH2-*apo* and protein complexes. There were slight differences in all DCC results, the FAD (R39-I146) and NADH (D171 and Q313 binding domain had DCC values ranging from 0.25-1 and the DCC for the C-terminal had a value from 0.2-1. The residues G343-H419 that showed high RMSF fluctuations in *Pf*NDH2 also showed positive correlation at positions 382 and 458 on the DCC heatmap. It is apparent that the ligands did not change the residue interaction in the protein as they have similar DCC results as in the *apo* protein, a result that is reflected in the protein RMSD.

The low energy basin of the essential dynamic calculations were superimposed against the PCA for all including the *Pf*NDH2-*apo* low energy basin. The green cartoon structures in figure 3.13c below show the *apo* protein, and this is overlayed (after superimposition) with the protein from the final frame of dynamics of each of the complexes. This shows the change in conformation during dynamics in certain regions of the protein. In all complexes there was good alignment with the *apo*, yet flexibility at the arms of the protein (or the outer regions), which is in good agreement with the b-factor representations (shown in the figure 3.14 below), where the mid-section of the protein shows enhanced rigidity.

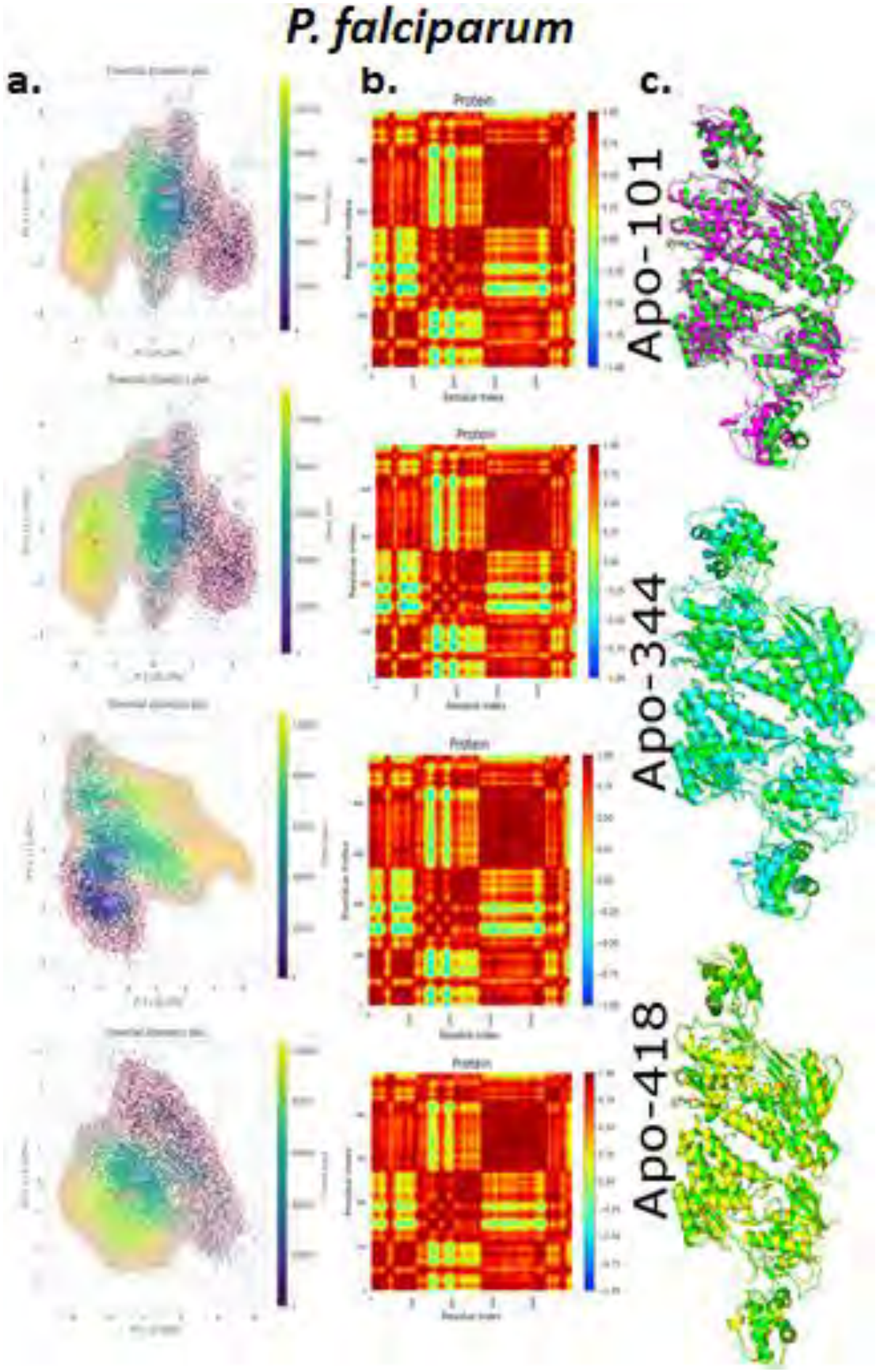

Figure 3.13 (a) Essential dynamic (PCA) analysis in the first column with *Pf*NDH2-*apo* in the first row, *Pf*NDH2-101 in the second row, *Pf*NDH2-344 in the third row and *Pf*NDH2-418 in the last row,

with the simulation ranging from 0 to 100000 (ps). **(b)** Dynamic cross correlation (DCC) analysis in the middle column with *Pf*NDH2-*apo* in the first row, *Pf*NDH2-101 in the second row, *Pf*NDH2-344 in the third row and P*f*NDH2-418 in the last row, with the correlation ranging from -1 to 1.**(c)** P*f*NDH2 protein-ligand superimposed structures with the green cartoon structure being the final frame from the dynamic simulations of the *apo* protein.

The b-factor of the 5JWA crystal structure (in Figure 3.14) was calculated using the PyMOL visualising program. The b-factor shows the displacement of atoms from their mean position in a crystal structure, which diminishes the scattered X-ray intensity; this displacement may be the result of temperature-dependent atomic vibrations or as a result of static disorder in a crystal lattice (Carugo, 2018). These thermal motion paths may indicate dynamic and transient channels that allow molecules to enter or exit from protein internal cavities (Carugo & Argos, 1998). As shown in the figure 3.14 below, blue represents rigid residues and yellow to red represents the residues that are likely flexible.

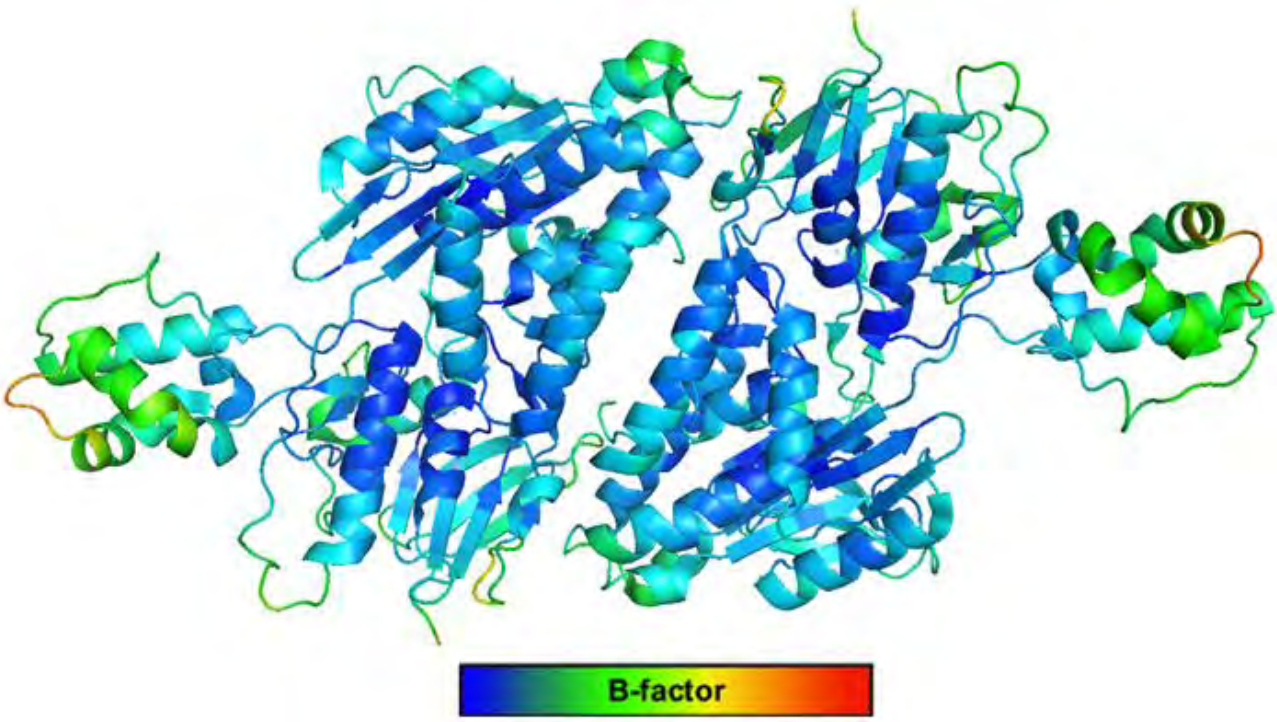

Figure 3.14 *P. falciparum* (5JWA) b-factor representation visualized in PyMOL visualizer, blue being rigid and red being flexible

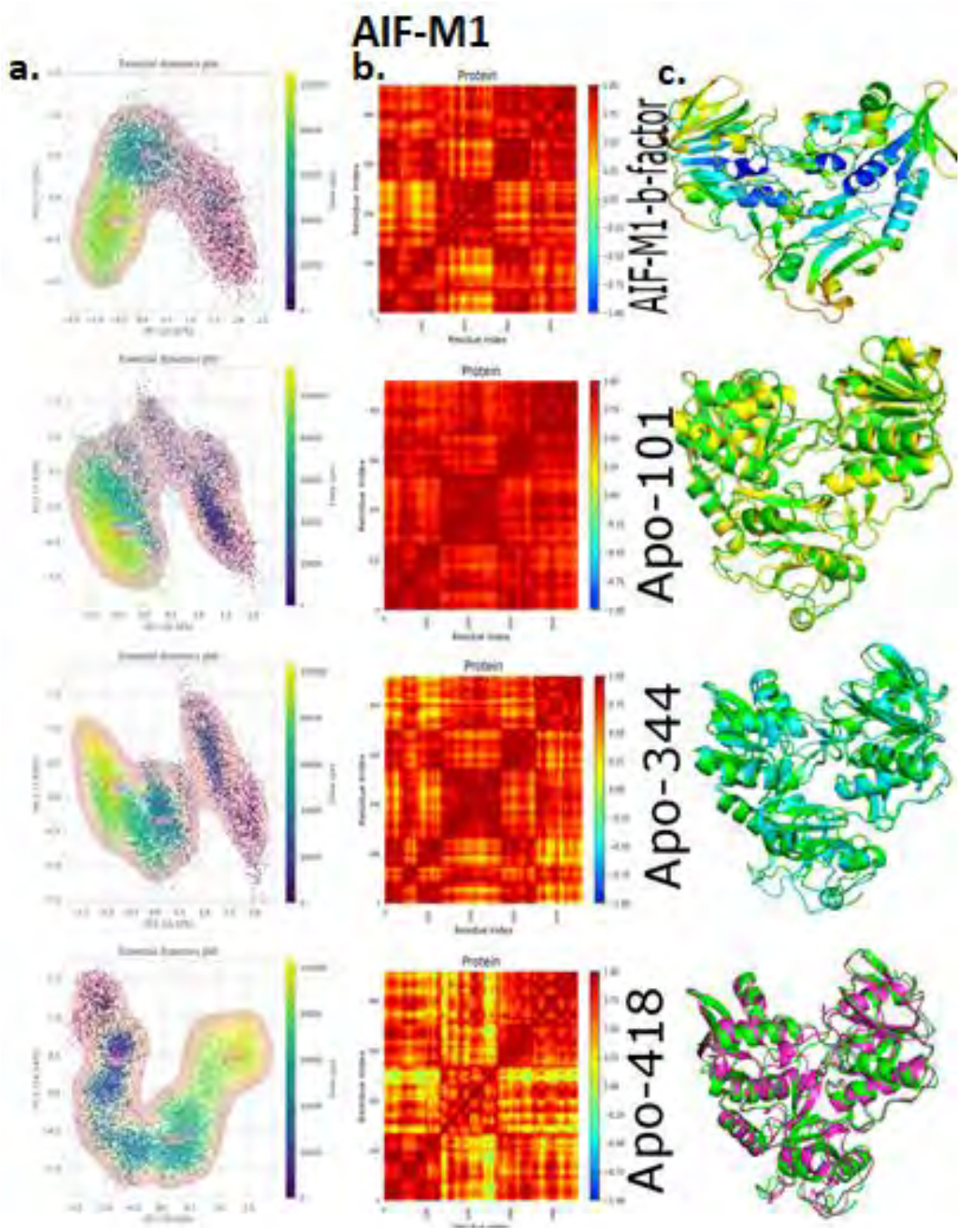

Figure 3.15 (a) Essential dynamic (PCA) analysis in the first column with AIF-M1-*apo* in the first row, AIF-M1-101 in the second row, AIF-M1-344 in the third row and AIF-M1-418 in the last row, with the simulation ranging from 0 to 100000 (ps). **(b)** Dynamic cross correlation (DCC) analysis in the middle column with AIF-M1-*apo* in the first row, AIF-M1-101 in the second row, AIF-M1-344 in the third row and AIF-M1-418 in the last row, with the correlation ranging from -1 to 1 **(c)**AIF-M1 protein-ligand superimposed structures with the green cartoon structure being the final frame from the dynamic simulations of the *apo* protein and the AIF-M1-b-factor representation

(Table 3.3 and Figure 3.15a) above shows the low energy basin for AIF-M1 protein together with their PC1 and PC2 percentages. Change in structure conformation at low energy basin occurred towards the end of the simulation in all structures.

The dynamic cross correlation plots (Figure 3.15b) show the *apo*, AIF-M1-101, AIF-M1-344 and AIF-M1-418 results respectively. The DCC is represented from a range of -1(blue) to 1(red) in which 1 is high correlation, the X and Y axes show the residue index from 0-500 residues and this will be translated from Ala128-Phe608 to correctly number the residues, meaning 0 = A128, 100 = L228, 200 = T328, 300 = Q428, 400 = I528 and 500 = 628. There was a high correlation in the AIF-M1-101 complex in all residues, and correlation ranged between 0.2-1 in the apo and the AIF-M1-344 and AIF-M1-418 complexes. The residues that showed high fluctuations in the protein‟s RMSF had a positive correlation of 1. There was more residue positive correlation when the ligands were bound to the human homologue in comparison to when bound to P*f*NDH2.

The b-factor value of the crystallised (AIF-M1) structure visualized in PyMOL showed the flexible regions of the protein as red (Figure 3.15c). The low energy basin PDB structures from the Essential Dynamics calculations were superimposed against the *apo* protein, and the difference in conformation is clearly visible, particularly around loop and some beta sheet regions.

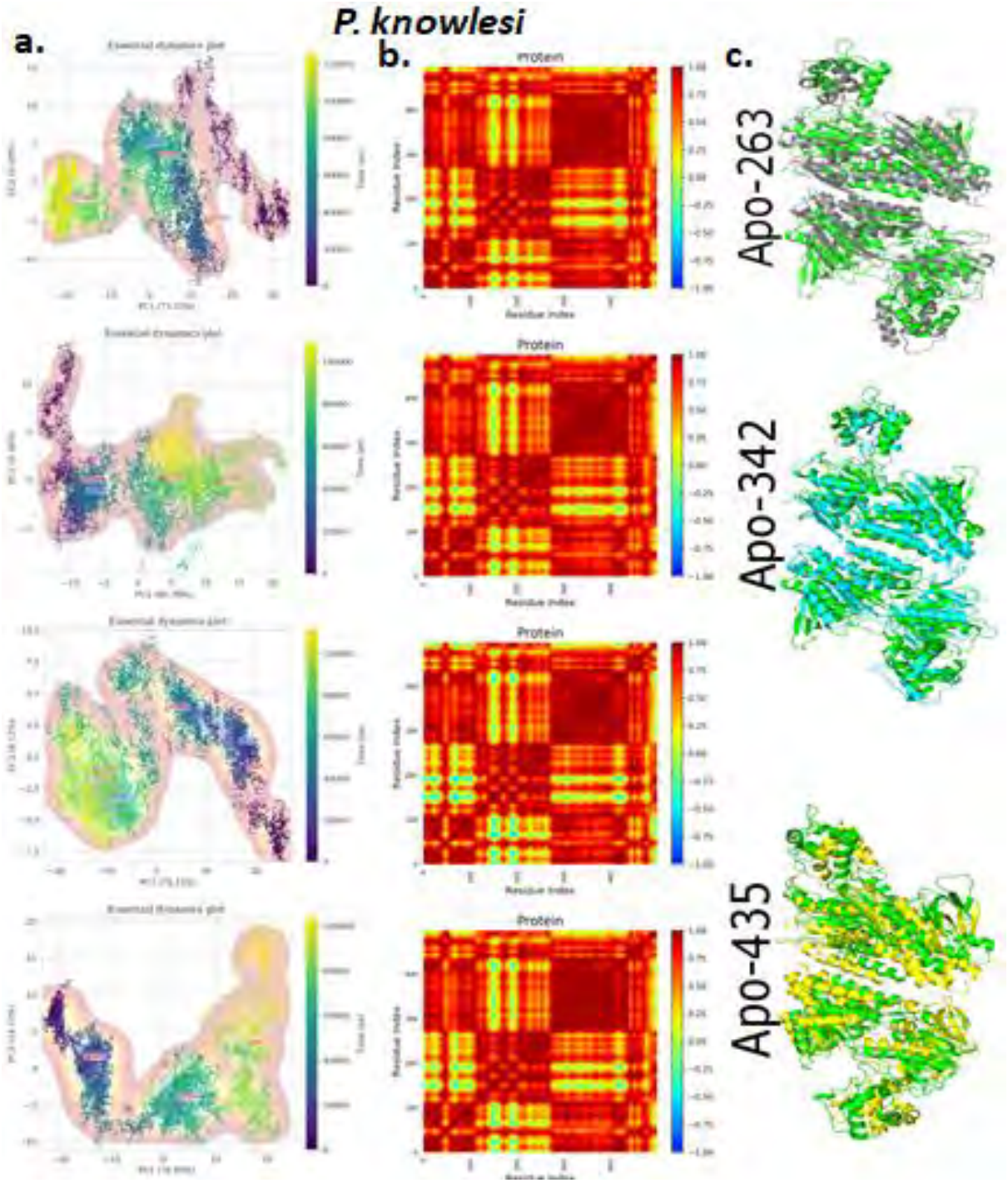

Figure 3.16 (a) Essential dynamic (PCA) analysis in the first column with *Pk*NDH2-*apo* in the first row, *Pk*NDH2-263 in the second row, *Pk*NDH2-342 in the third row and *Pk*NDH2-435 in the last row, with the simulation ranging from 0 to 100000 (ps). **(b)** Dynamic cross correlation (DCC) analysis in the middle column with *Pk*NDH2-*apo* in the first row, *Pk*NDH2-263 in the second row, *Pk*NDH2-342 in the third row and *Pk*NDH2-435 in the last row, with the correlation ranging from -1 to 1. **(c)** *Pk*NDH2 protein-ligand superimposed structures with the green cartoon structure being the final frame from the dynamic simulations of the *apo* protein

(Table 3.3 and figure 3.16a) above shows the low energy basin for *Pk*NDH2 protein together with their PC1 and PC2 percentages. The change in conformation at low energy basin occurred early in simulation of protein-complex structures in comparison to the *apo* protein. The low energy basins in the PCA calculations, showed a low RMSD with unstable deviations.

The dynamic cross correlation plots (figure 3.16b) show the *apo*, *Pk*NH2-263, *Pk*NDH2-342 and *Pk*NDH2-435 results respectively. The DCC is represented from a range of -1(blue) to 1(red) in which 1 is high correlation, the X and Y axes show the residue index from 0-500 residues and this will be translated from R39-K533 to correctly number the residues, meaning 0 = R39, 100 = A139, 200 = V239, 300 = L339, 400 = Q439 and 500 = 539. High correlation was observed in the residues Q299-I479, the plot showed correlation from a range of -0.5-1. The region Q299-I479 of high correlation showed high fluctuations in the RMSF results.

The low energy basin PDB outputs from the PCA calculations were superimposed against the *Pk*NDH2 *apo* protein and we observed difference in conformations even in rigid regions, the green cartoon structure was the *apo* protein.

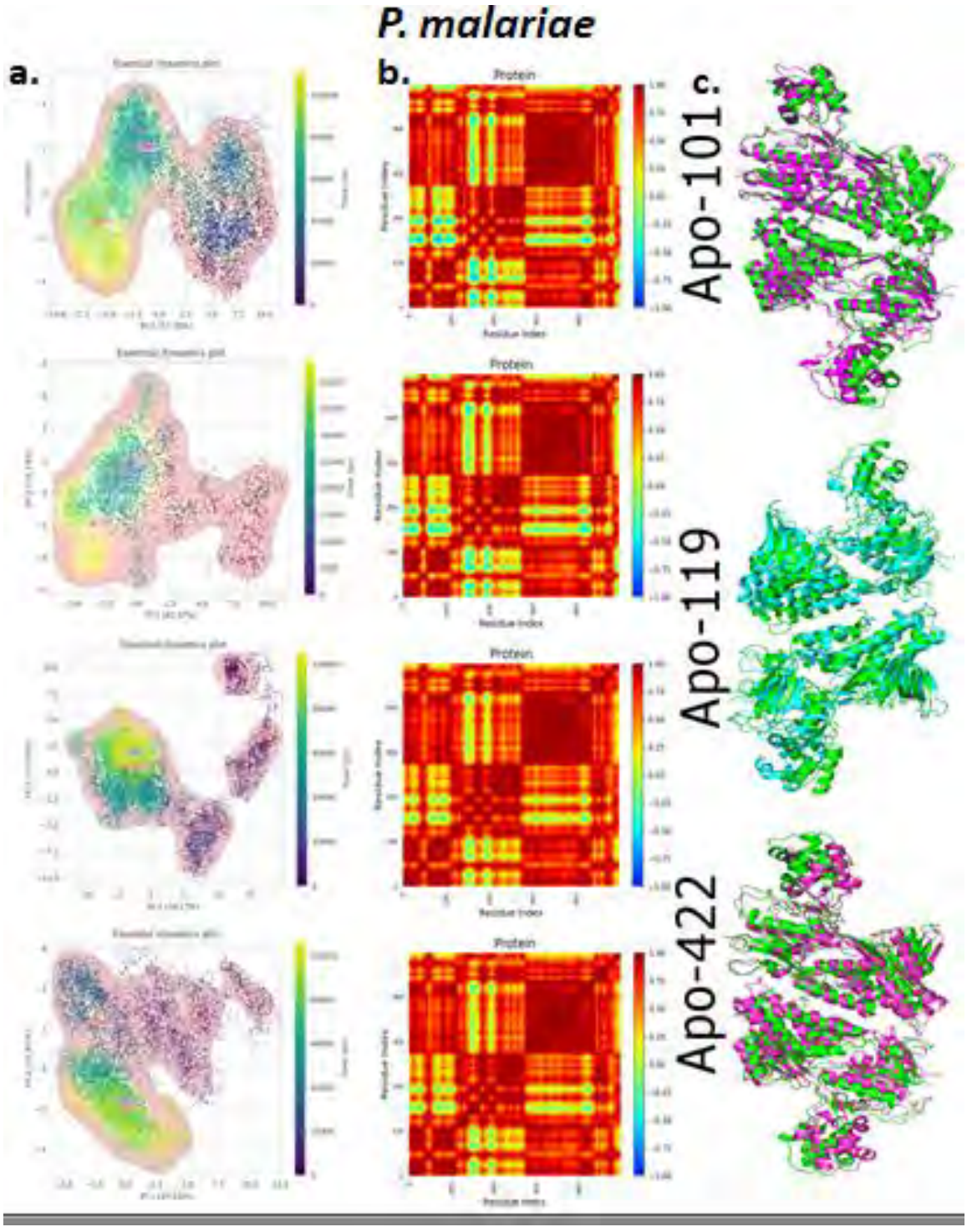

Figure 3.17 (a) Essential dynamic (PCA) analysis in the first column with *Pm*NDH2-*apo* in the first row, *Pm*NDH2-101 in the second row, *Pm*NDH2-119 in the third row and *Pm*NDH2-422 in the last row, with the simulation ranging from 0 to 100000 (ps). **(b)** Dynamic cross correlation (DCC) analysis in the middle column with *Pm*NDH2-*apo* in the first row, *Pm*NDH2-101 in the second row, *Pm*NDH2-119 in the third row and *Pm*NDH2-422 in the last row, with the correlation ranging from -1 to 1. **(c)** *Pm*NDH2 protein-ligand superimposed structures with the green cartoon structure being the final frame from the dynamic simulations of the *apo* protein.

Table 3.3 and figure 3.17 show the low energy basin for *Pm*NDH2 protein together with their PC1 and PC2 percentages. *Pm*NDH2-119 and *apo* protein showed change in conformation at low energy basin towards the end of simulation, and the *Pm*NDH2-101 and *Pm*NDH2-422 showed change in conformation at low energy basin at the beginning of the simulation. The low energy basins in *P. malariae* had a low RMSD in the specific time intervals.

The dynamic cross correlation plots (figure 3.17) show the *apo*, *Pm*NH2-101, *Pm*NDH2-119 and *Pm*NDH2-422 results respectively. The DCC is represented from a range of -1(blue) to 1(red) in which 1 is high correlation, the X and Y axes show the residue index from 0-500 residues and this will be translated from R39-K533 to correctly number the residues, meaning 0 = R39, 100 = M139, 200 = L239, 300 = S339, 400 = Q439 and 500 = 539. Correlation ranged from -0.5-1, with residues K299-I479 showing high correlation. The residues between K299-I479 that had a positive correlation covered the region with high fluctuations in the *P. malariae* RMSF. In general, the DCC values showed a positive correlation between residues.

The low energy basins of the PCA results were superimposed with the *apo* protein. The green cartoon structure is the *Pm*NDH2 *apo* protein and it was superimposed with the ligand complexes to observe and understand any changes in conformation. There was difference in conformation at the arms or outer residues of the protein.

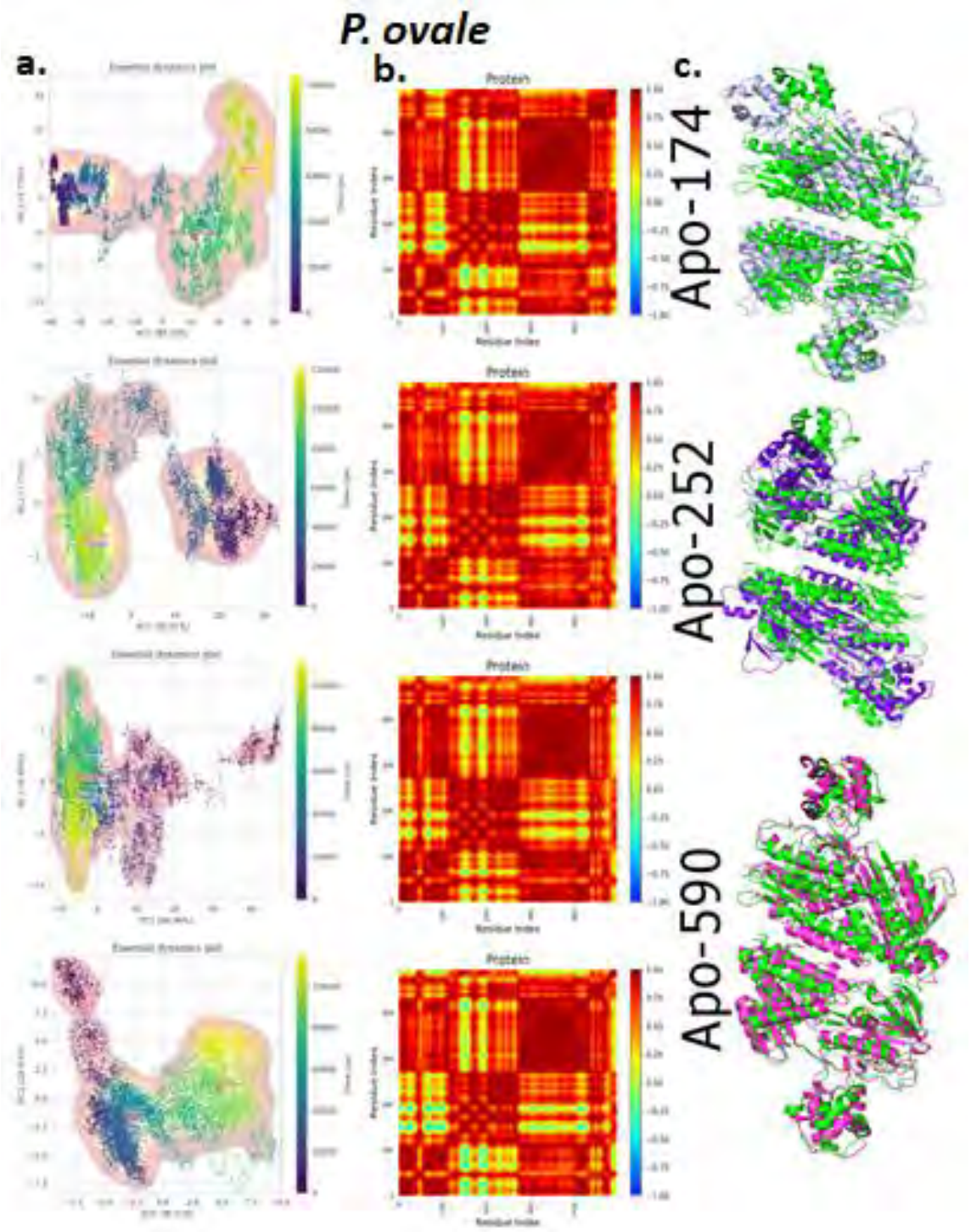

Figure 3.18 (a) Essential dynamic (PCA) analysis in the first column with *Po*NDH2-*apo* in the first row, *Po*NDH2-174 in the second row, *Po*NDH2-252 in the third row and *Po*NDH2-590 in the last row, with the simulation ranging from 0 to 100000 (ps). **(b)** Dynamic cross correlation (DCC) analysis in the middle column with *Po*NDH2-*apo* in the first row, *Po*NDH2-174 in the second row, *Po*NDH2-252 in the third row and *Po*NDH2-590 in the last row, with the correlation ranging from -1 to 1. **(c)** *Po*NDH2 protein-ligand superimposed structures with the green cartoon structure being the final frame from the dynamic simulations of the *apo* protein.

Table 3.3 and figure 3.18 shows the low energy basin for *Po*NDH2 protein together with their PC1 and PC2 percentages. *Po*NDH2-174 and *Po*NDH2-252 had change in conformation at low energy basin towards the end of the simulation and the *apo* protein and *Po*NDH2-590 showed change in conformation at low energy basin at the beginning of the simulation. The RMSD of the protein showed low values in low energy time intervals presents in the *P. ovale* PCA plots.

The dynamic cross correlation plots show the *apo*, *Po*NDH2-174, P*o*NDH2-252 and *Po*NDH2-590 results respectively. The DCC is represented from a range of -1(blue) to 1(red) in which 1 is high correlation, the X and Y axes show the residue index from 0-500 residues and this will be translated from K39-K533 to correctly number the residues, meaning 0 = K39, 100 = F139, 200 = I239, 300 = L339, 400 = A439 and 500 = 539. High correlation was observed between the residues K299-H479 which was also similar in the other *plasmodium* DCC plots. The region K299-H479 had peaks in the RMSF, however not necessarily the residues of highest fluctuations in the protein.

The low energy basins of the PCA results were superimposed with the *apo* protein. The green cartoon structure is the *Po*NDH2 *apo* protein and it was superimposed with the ligand complexes to observe and understand any changes in conformation. There was difference in conformation even at regions considered rigid by the b-factor calculation of the 5JWA protein. Correlation was between -0.5-1 in all plots.

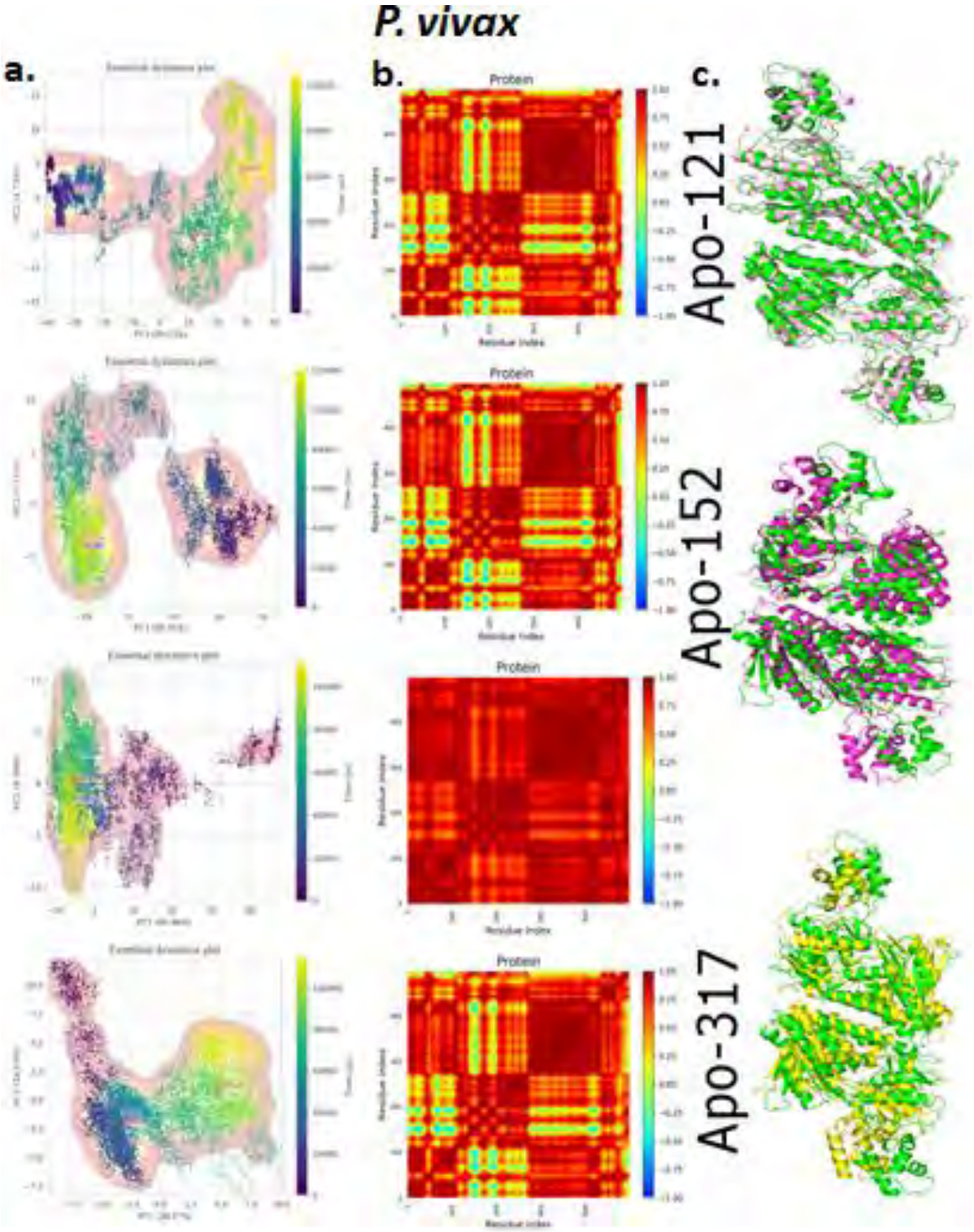

Figure 3.19 (a) Essential dynamic (PCA) analysis in the first column with *Pv*NDH2-*apo* in the first row, *Pv*NDH2-121 in the second row, *Pv*NDH2-152 in the third row and *Pv*NDH2-317 in the last row, with the simulation ranging from 0 to 100000 (ps). **(b)** Dynamic cross correlation (DCC) analysis in the middle column with *Pv*NDH2-*apo* in the first row, *Pv*NDH2-121 in the second row, *Pv*NDH2-152 in the third row and *Pv*NDH2-317 in the last row, with the correlation ranging from -1 to 1. **(c)** *Pv*NDH2 protein-ligand superimposed structures with the green cartoon structure being the final frame from the dynamic simulations of the *apo* protein.

Table 3.3 and figure 3.19 shows the low energy basin for *Pv*NDH2 protein together with their PC1 and PC2 percentages. *Pv*NDH2-*apo* and *Pv*NDH2-121 had change in conformation at low energy basin towards the end of simulation, and *Pv*NDH2-152 and P*v*NDH2-317 had change in conformation at low energy basin towards the beginning of simulation. The low energy basins in the PCA plots showed low RMSD values.

The dynamic cross correlation plots show the *apo*, *Pv*NH2-121, *Pv*NDH2-152 and *Po*NDH2-590 results respectively. The DCC is represented from a range of -1(blue) to 1(red) in which 1 is high correlation, the X and Y axes show the residue index from 0-500 residues and this will be translated from G39-K533 to correctly number the residues, meaning 0 = G39, 100 = S139, 200 = V239, 300 = L339, 400 = Q439 and 500 = 539. There was high correlation in residues H299-I479 and *Pv*NDH2-152 showed positive correlation of 1 in all residues and in other complexes DCC ranged between -0.5-1. *Pv*NDH2-152 showed high correlation in all residues in comparison to other complexes which agreed with its RMSD and RMSF values as they were very distinct.

The low energy basins were superimposed with the a*po P. vivax*. The green cartoon structure is the *apo* protein superimposed with the ligand complexes. We observed flexibility in the outer regions the arms of *Po*NDH2-121 and the mid-section was rigid. There was flexibility when SANC152 and SANC317 were bound to the protein even in residues that are considered rigid.

3.3.6 Chapter conclusion

This chapter focused on high throughput virtual screening of 623 SANCDB natural compounds on the 5 NDH2 protein structures and the human homologue, hit compound selection, molecular dynamic simulation without the membrane, the PCA calculations and DCC analysis of the MD trajectories. We aimed to study the effect of the ligands on the protein function. It was observed that the ligands bind on different sites in all four *plasmodium* species, hence the hit compounds were also different for each structure.

The ligands docked mainly in the FAD and NADH binding domain, and at the interface of the homodimer. Hit compounds were selected for further analysis based on their binding energy, size, binding site, ligand interaction with the protein and Lipinski rule of five score. The results obtained from the ProteinsPlus (METALizer) webserver showed that the Mg^{2+} ions were not interacting with the protein, hence there was no need to parameterize the ions prior to MD calculations. MD calculations were run at 100 ns using GROMACS package, using the CHARMM36 force field; we focused on the change in conformation caused by ligands bound to the proteins. The RMSD values of *Pf*NDH2 were generally stable in the *apo* protein and ligand complexes throughout the simulation. In general, it took time for the *Pk*NDH2 *apo* and complexes to stabilise and we observed in the RMSD results, the protein deviation, both when ligands were present or absent. The *Pm*NDH2 protein had a

more stable RMSD when SANC101 and SANC422 were bound. *Po*NDH2 RMSD stabilised after 48 ns when the ligands were bound, and we observed the *apo* protein deviating until towards the end of the simulation. SANC152 bound in the interface region caused a high RMSD of 4.8 nm, tending to destabilize the protein relative to the *apo* and other ligand bound complexes with the *Pv*NDH2 protein. The *Pf*NDH2 hit compounds were analysed in the AIF-M1 homologue and they caused a stable RMSD for the complex throughout the simulation, the ligands seem to change slightly the conformation of the protein from the structure superimpositions, and the *apo* and complex proteins had a stable RMSD deviation throughout simulation, which was favourable as the ligands are not supposed to affect or change the AIF-M1 protein.

In general, the RMSF of the structure‟s trajectories showed high fluctuations between residues V343-Q419 in both chain A and chain H, except for *Po*NDH2 which showed high peaks between residues S153-K229, and where the highest RMSF peaks were in the *apo* protein simulation. We observed a high RMSF at residue N583 when SANC418 was bound to the human protein AIF-M1. There was no significant difference between the RMSD values on AIF-M1 and *Pf*NDH2, the ligands had the same deviation as the human homolog *apo* meaning they did not change the conformation of the human protein which was favourable. The ligands bound close to the FAD binding region of AIF-M1 and the protein-ligand complexes when superimposed showed less alignment in loop regions and outer parts of the protein.

The PCA calculations showed the time frames in which there was change to conformations within low energy basins. The DCC calculations showed residues that correlated, and we observed high correlation in most of the protein residues (39-145 and 329-459) regions (0-110 and 280-420) of the structures. In this chapter we have identified strong binders to the range of NDH2 targets, while also exploring their interaction with the human protein. In the next chapter we will explore the ligand performance in the presence of the protein membrane. This will more accurately show the effect of the ligand on the protein within the expected environment for this trans-membrane protein.

4

Molecular dynamics and compound interaction study

Chapter overview

4.1 INTRODUCTION

4.1.1 Molecular dynamics

Proteins perform various essential cellular functions, and they are able to recognise small molecules with specificity and bind them (Guterres et al., 2021). Computational structure-based techniques have amazingly transformed *in silico* drug discovery studies, helping lead compounds identification, optimisation, protein-ligand interaction analysis at a lower cost (Allouche, 2012). In this study ligands bound to the local regions of the protein surface when molecular docking was performed, and complexes with these hit compounds were taken for further investigations in molecular dynamic simulations.

MD simulation have also become an important technique in studying lipid membrane systems, computational resources and appropriate empirical force fields have also advanced to allow for meaningful simulations of these systems to better understand their physical properties and functions (Lee et al., 2016). To investigate the structure of membrane proteins in ways that are complementary to experimental procedures the MD simulation approach is used (Goossens & De Winter, 2018). Membrane proteins within the membrane are within a complex environment. The heterogeneity in the spatial distribution of lipids and proteins in the cell membrane and between bilayer leaflets lead to the complexity of biological membranes, and some of these complex chemical systems further contain sugars such as glycolipids or lipoglycans (J. Lee et al., 2019).

Empirical force fields (FF) have been refined and developed, such that they are an accurate mathematical description of the relationship of conformation to energy, both in and between molecules, and these are used extensively in MD simulation. MD is a simulative technique that has been traditionally perceived to be difficult to implement and automatize, because it relies on a complex theoretical background. Similarly, analysis of long MD trajectories can be a difficult task that in some cases fails to return clear and specific insights into how to design and synthesize new and more effective com-

pounds (Cavalli, 2018). The CHARMM-GUI Membrane builder a web-based tool (http://www.charmm-gui.org/input/membrane) released in 2007, that has been developed and updated to support heterogeneous bilayers using more than 400 lipid types (such as phospholipids, phosphoinositides, cardiolipin, sphingolipids, sterols and fatty acids), allows users to build biologically realistic and experimentally comparable membrane systems (Additive et al., 2016). The membrane builder tool provides well-validated topology files or simulation input files for various molecular dynamics simulation programs such as CHARMM (B. R. Brooks et al., 2009), GROMACS (Abraham et al., 2015), AMBER (Biller et al., 2013), GENESIS (Jung et al., 2015), NAMD (Phillips et al., 2016) and OpenMM (Eastman et al., 2017). An initial system configuration must be given prior to simulation, and this includes initial positions, velocities and topologies and, further, conversion of elements to suit the native format of the simulation package is essential before simulation commences (Vermaas et al., 2016). The CHARMM-GUI graphical user interface is an important tool as it provides various functional modules with broad capability, used to prepare complex biomolecular systems and input files for molecular simulations (Park et al., 2019). The CHARMM-GUI platform has modules such as Solution builder, Membrane builder, Nano disc builder, HMMM builder, Monolayer builder, and Micelle builder (Qi et al., 2020). To allow general usage of CHARMM36 FF, the CHARMM-GUI Membrane builder, for example, incorporates the option to generate minimisation, equilibration, and production inputs with simulation parameters for each simulation program.

There is an increased usage of molecular simulations, with software able to make use of a wide range of hardware to achieve this purpose,, and the accessibility to multicore CPU and GPU technologies has allowed users to extend simulation times of many different experiments. MD simulations provide the dynamics and energetics of membrane-associated proteins at the atomic level, which is generally difficult to obtain from experiments (Jo et al., 2009). Simulations continue to be a powerful method to comprehend the detailed lipid-protein interactions, responsible for protein activity in signal transduction and disease (Ash et al., 2004). The membrane is a thin double layer that keeps the cells intact and controls the transport of matter to and from the cells; it is also involved in other important processes such as protein anchoring and cell fusion and division. The membrane in this study was constructed in the study using CHARMM-GUI as the NDH2 is a peripheral membrane protein (Wiederstein & Sippl, 2007).

4.1.2 CHARMM36 force field and POPC lipid

As has already been mentioned, computer simulations offer molecular insight into the structural organisation of membrane systems including those of lipid-protein complexes. However, it is the quality of the force field used that determines the quality and accuracy of simulation (Allouche, 2012). GROMOS (Oostenbrink et al., 2004), OPLS (Jorgensen et al., 1996), AMBER (Hornak et al., 2006) and CHARMM (A. D. MacKerell et al., 1998) are the force fields mostly used in simulation studies of biomolecules and they differ in systems and protocols used in parameterisation stage. Force fields are developed by fitting parameters describing the molecular interactions to the results of highly detailed quantum-mechanics (QM) calculations of small peptides molecules, to obtain the parameters necessary to match experimental results (Smith et al., 2015).

It is very important to consider certain factors when choosing the force field that will satisfy the needs of one‟s system. Biomolecular force fields have different functional forms, different combination rules for non-bonded interaction parameters, different ways to treat non-bonded interaction exclusions, diverse ways of handling long-range electrostatic forces and diverse ways of defining torsional-angle interactions (Gebhardt et al., 2018).

The CHARMM force field parameters in recent years have provided for good properties of lipid bilayers, although derivation and validation of the forcefield for certain molecules may be monotonous and time consuming (in particular the derivation of the torsional potential), yet these parameters are a key factor of molecular structure in simulations (Kulig et al., 2015). The empirical force field is an accurate mathematical description of the relationship between conformation and energy in molecules and is important in MD simulations. In particular, the CHARMM36 FF was designed to accurately represent experimental bilayer properties using the constant particle number, pressure and temperature (NPT) ensemble (J. Lee et al., 2016). The CHARMM community developed CHARMM36 FF applies the long-range non-bonded interactions as follows: the Lennard-jones (LJ) 6-12 (van der Waals) interactions smoothed over 10-12 Å using the force-based switching function; the particle mesh Ewald (PME) method used for long-range electrostatic interactions; and 1-4 non-bonded interactions (coulombic and LJ interactions) (B. R. Brooks et al., 2009). The GROMACS package has the functionality needed to work with the CHARMM36 FF and its single-core code is faster with mixed point precision (previously called single-point precision), than in CHARMM and NAMD.

In this study the CHARMM36 FF was used, which covers biological systems and many important lipid types and is well validated for properties such as bilayer areas, compressibility, spontaneous curvature and bending constants (Venable et al., 2015). Found in the CHARMM36 force field are 73 distinct sugar residues, 5 nucleic acids, 20 amino acids and 35 lipids (Guvench et al., 2008). This force field was chosen for this study as it is well known to produce excellent membrane properties, together with a high quality of simulation of globular proteins (Domański et al., 2018). The potential

energy function of the CHARMM36 force field is composed by bonded and nonbonded terms; thus, it is necessary to add to it the force constant K, the equilibrium values b_0 (bond), θ_0 (angle) and φ_0 (dihedral) for the bonded terms, the dihedral phase δ, the partial atomic charges q, and the Lennard-Jones terms ε and ζ (da Silva et al., 2020b; Alexander D. MacKerell et al., 1995) (Equation 4.1).

Bonded equation

$$V_{bonded} = \sum_{bonds} K_b(b-b_0)^2 + \sum_{angles} K_\theta(\theta-\theta_0)^2 + \sum_{n=1}^{n=6} \sum_{dihedrals} K_{\phi n}(1+\cos(n\phi-\delta)) + \sum_{impropers} K_\varphi(\varphi-\varphi_0)^2$$

Non-bonded equation

$$V_{nonbonded} = \sum_{\substack{pair\ ij,\\ charges}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_{\substack{pairs\ ij,\\ Lennard-Jones}} \epsilon_{ij} ij \left[\left(\frac{r_{min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{min,ij}}{r_{ij}}\right)^{6}\right]$$

Equation 4.1 The potential energy function of CHARMM36 force field

1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), a phosphatidylcholine that is a diacylglycerol and phospholipid, is naturally present in eukaryotic cell membranes and is important for biophysical experiments, used in systems mimicking the cell membrane. POPC is a zwitterionic phospholipid, with a single C=C double bond in the longest (oleoyl acyl chain) of its two hydrocarbon tails (Bingham & Ballone, 2012). We chose POPC to be our model bilayer system because most of eukaryotic cells of PC/cholesterol mixtures often have a high concentration of unsaturated lipids, several studies of PC-bilayer systems have led to the conclusion that different chain conformations create a fluid phase micro-immiscibility leading to the cholesterol-rich and cholesterol-poor microdomains, rich in unsaturated chains and may be important for the stabilisation of specific protein functions in biological membranes (Rappolt et al., 2003). POPC is a synthetic monounsaturated phospholipid with a phosphocholine head group and two fatty acid chains, one of which is saturated (16:0), while the other is unsaturated (18:1) combined with two ester carbonyls as shown in the figure 4.1 below (Wanderlingh et al., 2017).

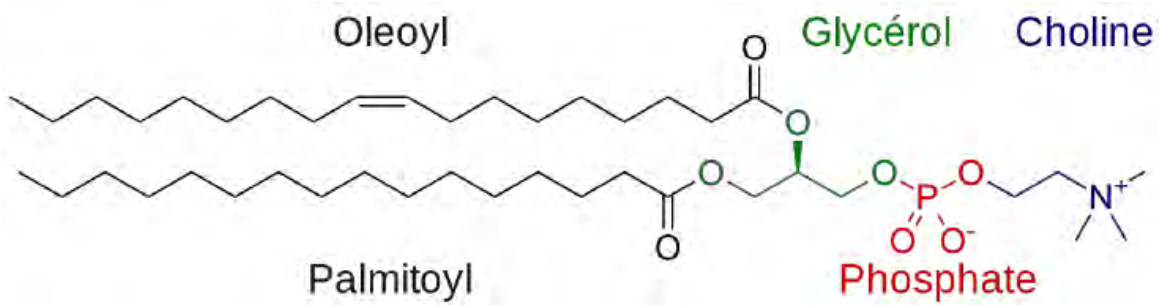

Figure 4.1 Chemical structure of POPC lipid, with an empirical formular of $C_{42}H_{82}NO_8P$ and a molecular weight of 760.08 Da

4.2 METHODOLOGY

4.2.1 Molecular dynamics

CHARMM-GUI (Chemistry at HARvard Macromolecular Mechanics – Graphic user interface) Membrane Builder, http://www.charmm-gui.org/input/membrane, is a web-based user interface designed to interactively build all-atom protein/membrane or membrane only systems for molecular dynamic simulations through an automated optimized process (Wu et al., 2014).

The membrane builder (Bilayer builder) option which incorporates the Glycolipid Modeler and LPS Modeler into the lipid selection was used to construct the membrane, embed the protein onto the membrane and to create files that were later used in the simulation. Residues were protonated as per their expected states at pH 7. The membrane builder performed the job in 6 steps: **Step 1:** reading the protein structure and manipulating it. The OPM (Orientations of Proteins in Membranes) server, which orients the protein with respect to membrane normal (the Z axis by definition) and which determines the rotational and translational position of the transmembrane region (M. A. Lomize et al., 2012) was used. For this entire step the PDB format for files was used, and accordingly this step was managed by the PDB reader & Manipulator. Through reading the PDB files in CHARMM-GUI, there were options to select the chains, cofactors, ions, or ligands to be included in the job run.

Step 2: orienting the protein. The membrane builder provides the option to change orientation and position of the protein along the membrane. Accordingly, we used the PPM server orientation option that submits the input structure to the PPM server (https://opm.phar.umich.edu/ppm_server). The PPM server is the sister server to the OPM database and is used to calculate the rotational and translational positions of transmembrane (peripheral) proteins in membranes using their 3-D structure (PDB coordinate file) as input (M. A. Lomize et al., 2012). Figure 4.2 shows the resulting orientation of the crystal structure PfNDH2 (5JWA), as a result of processing through this server.

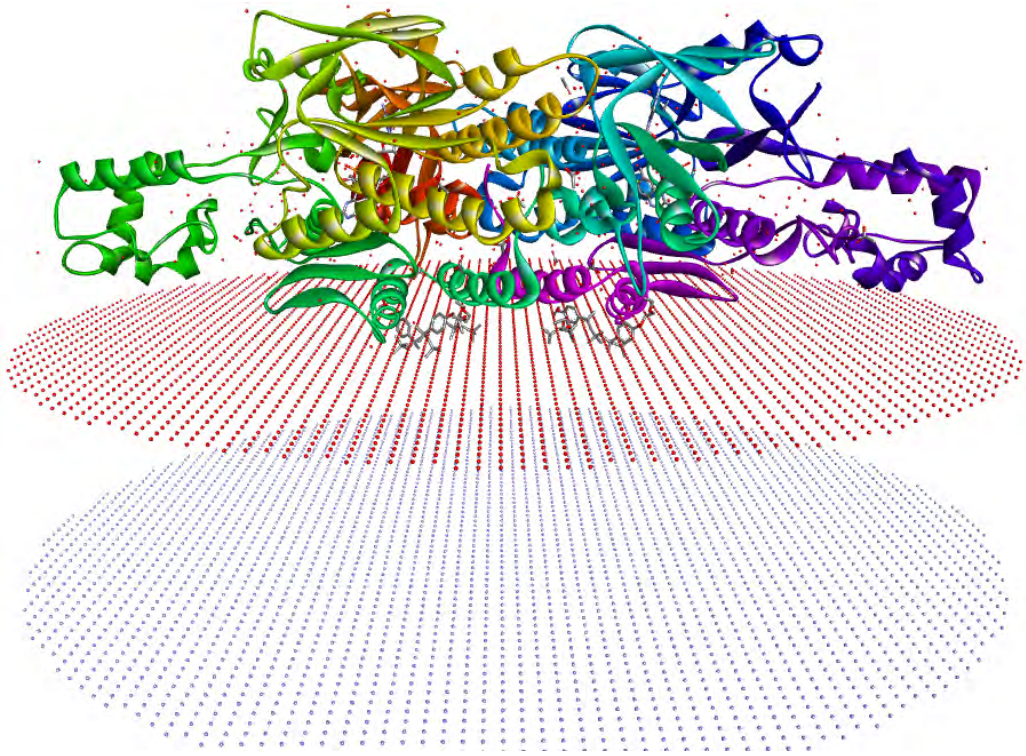

Figure 4.2 The OPM server output showing the orientation of 5jwa protein on the membrane, with a hydrophobic thickness or depth of 7.6 Å, tilt angle of 90 and $\Delta G_{transfer}$ of -11.3 kcal/mol. The residues in contact with the membrane are PHE495-VAL502 in chain A and GLY493-PHE518 in chain H

Step 3: determining the system size. This is based on specified parameters, these being the number of lipids in the bilayer membrane, system shape and the water thickness on top and bottom of the entire system. The cross-sectional area was calculated and the plot of Area (x100) versus the Z- axis were shown in the figure 4.3 below,

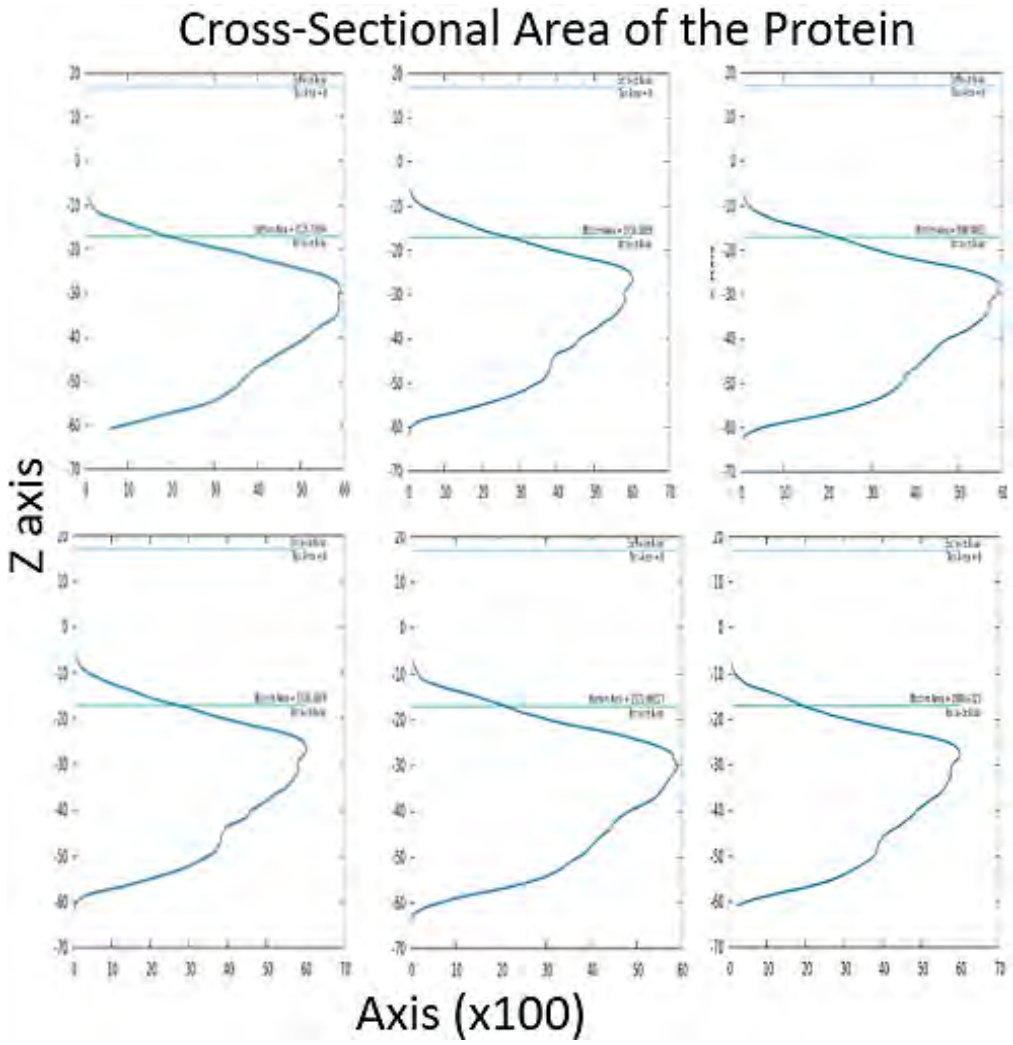

Figure 4.3 The cross-sectional area of the protein structures, Aif-M1, *P. falciparum*, *P. knowlesi, P. malariae*, *P. ovale* and *P. vivax*

A heterogenous lipid bilayer was used, with a water thickness of 22.5 Å, this being the minimum water height on the top and the bottom of the system for all the proteins in this study. The system size along the X and the Y axis was set to be identical, and these were set to be 145 Å for the *Plasmodium* species, while they were set to be length of 140 Å for the human homologue. Having an accurate value of the surface area is important because the surface area of a lipid type is directly related to the system size calculation. 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) a phosphatidylcholine which is a class of phospholipids that incorporate choline as a headgroup was used in the study. For all protein structures a 1:1 upper leaflet and lower leaflet was used respectively, with a surface area of 68.3 $Å^2$. The calculated area of lipid contact and XY system size for the lipid layer is shown in the table 4.1 below.

Table 4.1 The calculated lipid contact area (Å^2) and XY (Å) system size for all the protein structures

Protein	Top area	Bottom area	X extent	Y extent	Z extent
AIF-M1	0	453.07	-32 to 36	-29 to 36	-68 to -17
P. falciparum	0	3015.74	-69 to 69	-38 to 38	-62 to -8
P. knowlesi	0	3736.39	-72 to 70	-38 to 36	-62 to -6
P. malariae	0	3000.88	-69 to 69	-39 to 35	-63 to -8
P. ovale	0	2922.80	-71 to 69	-38 to 38	-65 to -8
P. vivax	0	2888.43	-70 to 72	-35 to 334	-61 to -8

Step 4: the building of components. Components of the system such as the lipid bilayer, bulk water, and ions were generated. A replacement method was used for system building and an ion concentration of 0.15 M of KCI was used; neutralizing ions were added to make the system neutral for long-range electrostatic interactions using the particle-mesh-Ewald method (Essmann et al., 1995). Through a Monte Carlo simulation, the initial configuration of ions was then determined using a primitive model including, for example, van der Waals and scaled Coulombic interactions. As mentioned above a heterogeneous lipid bilayer was used which was generated by the replacement method. This method uses a lipid-like pseudo atoms to generate lipid packing around a protein, and these pseudo atoms are replaced with lipid molecules from the lipid conformer library (T. B. Woolf & Roux, 1994; Thomas B. Woolf & Roux, 1996).

Lipid penetration was checked; in this the protein surface penetration check finds the lipids that go beyond the protein surface and the lipid ring penetration check detects the lipid tails that passed through the cyclic groups in the simulation systems. No protein surface penetration or lipid ring penetration was found. The membrane components were generated and then assembled. The GROMACS input generator was selected, the grid information for PME FFT was automatically generated, and, following equilibration, an NPT (constant particle number, pressure, and temperature) ensemble and a temperature of 303.15 K was applied for molecular dynamics. Table 4.2 below shows the setup restrains for protein and lipids; we equilibrated the system to reduce the possible problems with the numerical integration with the uncorrelated systems.

A significantly improved new building algorithm was used, which did not replace pseudo atoms randomly but rather replaced those that were closer to the protein first, using a systematic translation (XY plane) and rotation (around the Z-axis) rigid-body search for each lipid molecule. This search continued until the optimal orientation was found. Following this the algorithm tried and rejected or accepted different lipid conformers (Wu et al., 2014). These algorithms are known to significantly optimize the lipid packing around the embedded protein.

Table 4.2 The calculated XY system size for each protein structure, the lower and upper leaflet number and the lipid type that was used in the study

Aif-M1

Lipid Type	Upperleaflet Number	Lowerleaflet Number
POPC	287	281

Calculated XY System Size:

	Upperleaflet	Lowerleaflet
Protein Area	0	453.074917
Lipid Area	19602.1	19192.3
# of Lipids	287	281
Total Area	19602.1	19645.374917
Protein X Extent	35.85	
Protein Y Extent	35.93	
Average Area	19623.74	
A	140.08	
B	140.08	

P. falciparum

Lipid Type	Upperleaflet Number	Lowerleaflet Number
POPC	287	243

Calculated XY System Size:

	Upperleaflet	Lowerleaflet
Protein Area	0	3015.73964
Lipid Area	19602.1	16596.9
# of Lipids	287	243
Total Area	19602.1	19612.63964
Protein X Extent	68.94	
Protein Y Extent	37.49	
Average Area	19607.37	
A	140.03	
B	140.03	

P. knowlesi

Lipid Type	Upperleaflet Number	Lowerleaflet Number
POPC	308	254

Calculated XY System Size:

	Upperleaflet	Lowerleaflet
Protein Area	0	3736.3869
Lipid Area	21036.4	17348.2
# of Lipids	308	254
Total Area	21036.4	21084.5869
Protein X Extent	71.13	
Protein Y Extent	37.64	
Average Area	21060.49	
A	145.12	
B	145.12	

P. malariae

Lipid Type	Upperleaflet Number	Lowerleaflet Number
POPC	287	244

Calculated XY System Size:

	Upperleaflet	Lowerleaflet
Protein Area	0	3000.6833
Lipid Area	19602.1	16665.2
# of Lipids	287	244
Total Area	19602.1	19666.0833
Protein X Extent	68.53	
Protein Y Extent	38.45	
Average Area	19634.09	
A	140.12	
B	140.12	

P. ovale

Lipid Type	Upperleaflet Number	Lowerleaflet Number
POPC	308	266

Calculated XY System Size:

	Upperleaflet	Lowerleaflet
Protein Area	0	2922.80227
Lipid Area	21036.4	18167.8
# of Lipids	308	266
Total Area	21036.4	21090.60227
Protein X Extent	70.54	
Protein Y Extent	37.56	
Average Area	21063.50	
A	145.13	
B	145.13	

P. vivax

Lipid Type	Upperleaflet Number	Lowerleaflet Number
POPC	308	266

Calculated XY System Size:

	Upperleaflet	Lowerleaflet
Protein Area	0	2888.4323
Lipid Area	21036.4	18167.8
# of Lipids	308	266
Total Area	21036.4	21056.2323
Protein X Extent	71.35	
Protein Y Extent	34.06	
Average Area	21046.32	
A	145.07	
B	145.07	

Step 5 and 6: this stage assembled the components and equilibrated the system. Membrane builder does not provide the equilibrated structure due to the significant computing resources required. Six Consecutive CHARMM input files for system equilibration and production are produced. In this case CHARMM-GUI produced GROMACS formatted CHARMM36 force fields, a pre-optimized PDB structure and GROMACS inputs. The GROMACS input files consisted of .pdb/.gro files that contained the atomic coordinates, the "top" system topology file which had force field parameters, ".mdp" files that had the simulation parameters and the ".tpr" file which was the binary run input file (Vermaas et al., 2016). The Verlet cut-off scheme was applied for all minimization, equilibration, and production steps since this is faster and more accurate than the group scheme. Energy minimisation of the system was performed locally using 5 000 steps to remove any steric clashes using a single precision version of GROMACS, followed by an equilibration protocol which included periodic boundary conditions. The equilibration protocol was provided by the CHARMM-GUI web server and consisted of 6 steps with harmonic positional restraints applied, while the production run were performed in

tend steps accounting for 50 000 ps simulation. Equilibration and production runs were performed at the Centre for High Performance (CHPC) cluster in Cape Town South Africa. An in-house script was used to extend the simulations from 500 000 ps to 75 000 000 ps (150 ns). Acpype was used to parameterise the ligands in their docked regions on the protein and were submitted to CHARMM-GUI as input files and the procedure as mentioned above was followed.

Each system was equilibrated, using the equilibration input files generated by Membrane Builder. The MD trajectories of each system were generated, using the all-atom parameter set of optimized lipid parameters and the modified TIP3P water model (Jorgensen et al., 1983), Nose´-Hoover temperature control (Kassir et al., 1985) maintained a constant temperature of 303.15 K, and the Langevin-piston algorithm (Feller et al., 1995) was used to maintain constant pressure at 1 atm along the Z direction, while the XY area varied, with a constant surface tension g of ¼ 0 dyn/cm for NPgT (constant pressure, surface tension, and temperature) dynamics. The production runs produced (trr) files which contain information on the particle positions, velocities, and forces, as well as on the size of the simulation box with high precision. The (trr) files were used to perform MD analysis and trajectory calculations. Figure 4.4 summarizes the full setup of the system using CHARMM-GUI.

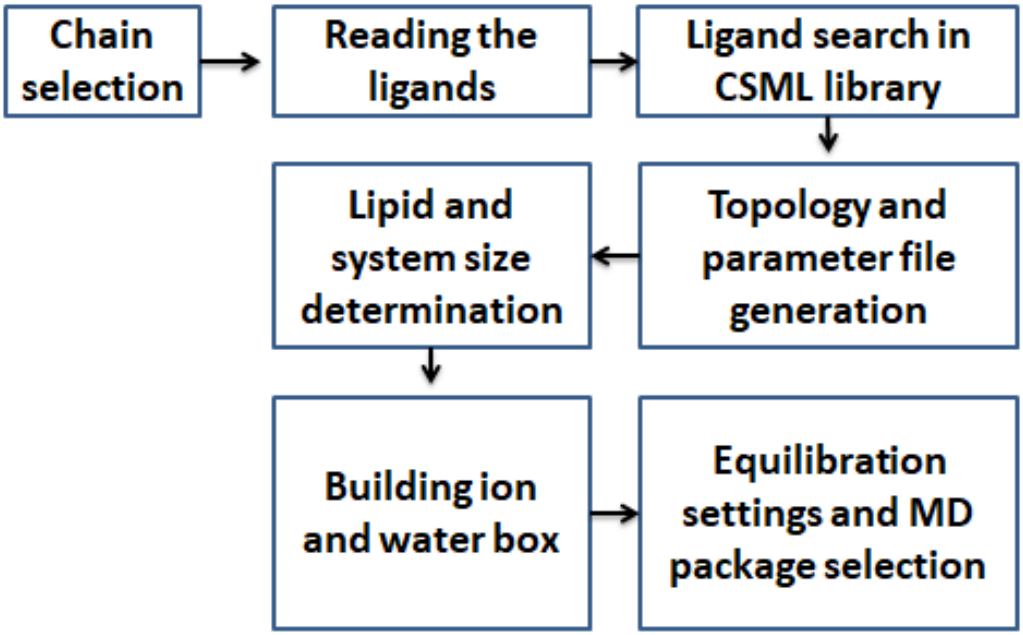

Figure 4.4 Summary of the CHARMM-GUI workflow

4.2.2 **Essential dynamic calculations and dynamic cross correlation analysis**

To detect correlations in large data sets a standard mathematical tool called Principal Component Analysis (PCA) is used; it also allows one to summarise the information content by means of a smaller set of summary indices that can be easily visualised and analysed. PCA is a popular method performed on the molecular dynamics simulation trajectories, accounting for the essential dynamics of the system in a low-dimensional free energy landscape (Sittel et al., 2014). This technique represents a

linear transformation that diagonalizes the covariance matrix and thus removes the instantaneous linear correlations among the coordinates (Jollife & Cadima, 2016).

PCA was performed on *apo* and protein-ligand complexes to comprehensively assess the effects of the ligands on NDH2 conformation. The MD trajectories were stripped of the Periodic Boundary Condition (PBC) (proteins were centred without discontinuities), water was removed and the resultant system saved in a (.xtc) file. Mode-task was downloaded and installed under the Anaconda environment on the YODA cluster (the in-house Linux cluster associated with the Rhodes University Bioinformatics Research Unit, RUBi). Only the Cα atoms of the protein were used for analysis; the MD trajectories and the PDB file that had the same residues as in the trajectory file were used for input. The MDM-Task "compare_essential_dynamics.py" tool (implementing comparative essential dynamics) was run on the YODA cluster to assess the distribution of pocket conformational sampling across MD simulations of the protein structures (Sheik Amamuddy et al., 2021b). The output was the low and high energy basins (together with associated PDB files) and an essential dynamic plot heatmap showing the first (PC1) and second principal component (PC2) axes. Along the axes is the percentage of total variance explained by each PC. The principal components (PCs) represent the dynamics of a protein in terms of a reduced set of orthonormal modes (Ross et al., 2018). Dynamic cross correlation (DCC) is calculated using the formular in equation 1.4 in chapter 1 (Di Marino et al., 2015). The MD-Task program generates the heatmap showing the dynamic cross correlation between the Cα atoms of the selected frames in an MD trajectory to identify relative residue movements (Brown et al., 2017b).

4.3 RESULTS

4.3.1 Molecular dynamics trajectories analysis

MD simulations of membranes and membrane proteins have produced insights into these biomolecular structures and their processes; simulations are considered a standard tool to guide, interpret and understand the atomistic details of molecular structure and kinetics (Y. Yu et al., 2021). In this study, we seek to understand the dynamics of the bound hit compounds to the protein structures, and the influence this binding has on the motions of the protein within the context of the membrane.

The MD simulations were run for 150 ns with the protein structures embedded in the membrane, using the CHARMM-GUI (membrane builder) generated GROMACS topology files. The GROMACS folder contained the GROMACS formatted CHARMM force fields, pre-optimised PDB structure and input files optimised for GROMACS 2019.2 or above. The Verlet cut-off scheme was used for all minimisation, equilibration, and production steps, as it ran parallelised simulation. A single precision in GROMACS was used in minimisation. MD simulation for systems embedded in the membrane was run for 50 ns longer than previous simulations in chapter 3 due to the size of the system.

Subsequent to dynamics simulations, the proteins were centred in the simulation box, water and the periodic boundary box was removed, and the trajectories were visualised using the VMD tool. The PDB files at the beginning and end of the simulation were generated for comparison, root mean square deviation (RMSD), root mean square fluctuation (RMSF) and radius of gyration (Rg) of the proteins were calculated using the gmx_mpi rms command.

The RMSD results in the study when the membrane was included in the simulation are shown in the figure 4.5 below.

4.3.1.1 RMSD analysis

RMSD is commonly used for comparing different conformations of given molecular systems, this measure is essential for analysing and validating simulation results (Velázquez-Libera et al., 2020).

*Pf*NDH2 within the membrane, with and without coordinating ligands, had an RMSD below 0.6 nm; the structures started stabilising after 20 ns. However, deviation occurred again after 50 ns in the *apo* protein. The protein stabilised rapidly when SANC418 was bound; however, it took 110 ns for the SANC101 complex to stabilise. Further, it appeared that the SANC101 and SANC344 ligands destabilised the protein, which is positive in terms of the aims of this project. We believe these induced conformational changes will affect the functioning of the NDH2 protein. There were significant difference between the RMSD of the protein when the membrane was present and absent.

Unexpectedly, there was stability in the complexes in the absence of the membrane and deviations when the membrane was present, although this could be influenced by the size of the system.

Only the *P. falciparum* hit compounds were analysed in the human homologue and we observed an increase in terms of RMSD of the AIF-M1 *apo* protein after 100 ns; the complex AIF-M1-SANC101 had an increase in RMSD after 20 ns, and AIF-M1-SANC344 and AIF-M1-SANC418 had a RMSD decrease after 90 ns and 140 ns respectively and RMSDs observed was below 0.5 nm. As the human homologue is not a membrane protein, we extended the simulations on the non-membrane bound systems to 150 ns to match the time interval of that of the plasmodium complex with a membrane. As a matter of interest, we ran the AIF-M1 protein through the CHARMM-GUI membrane builder; however, this was rejected since the membrane builder only assigns a membrane, to membrane proteins.

The RMSD of all simulations with *Pk*NDH2 was below 1.0 nm; however, and the *apo* protein was more stable than the complexes, stabilising after 55 ns. Conversely, the protein-ligand structures showed high RMSD deviations when the complexes were bound to the membrane. This ligand-based destabilisation of the protein was evident as deviations in RMSD right through to the end of simulation. Although the *apo* structure was stabilized, the results for the complexes were similar to those in the absence of the membrane.

In *Pm*NDH2 the RMSD for all simulations was below 1.0 nm; the membrane-bound complex with SANC119 showed a greater relative stability in comparison to the rest of the complexes. With observed stabilisation after 85 ns. However, the complexes for *P. malariae* were more stable in simulation when free from the membrane. The *apo* protein stabilised after 80 ns, SANC101 and SANC422 proved to be destabilising the protein as we observed deviations in RMSD right to the end of simulation.

The RMSD of all simulations of *Po*NDH2 structures in the presence of the membrane were below 1.2 nm. The membrane-bound complex with SANC174 had a higher RMSD, stabilizing late in the simulation (after 120 ns). In any case the generally the protein structures appeared to show greater stability when the ligands were bound. The simulation of *P. ovale* structures in complexes resulted in a low observed RMSD when these complexes were bound to the membrane.

Simulations of *Pv*NDH2 in the presence of a membrane resulted in a protein RMSD below 0.9nm. However, in the *apo* case there was deviation in observed RMSD until the end of simulation. This is in contrast to the ligand complexes. In the absence of a membrane the complexes with SANC121 and SANC317 were stable, however we observed change in the process of stabilization in the presence of a membrane. The complex with SANC152 was similar in terms of the progression of RMSD (in the presence of a membrane) to that of the membrane bound complex with SANC121.

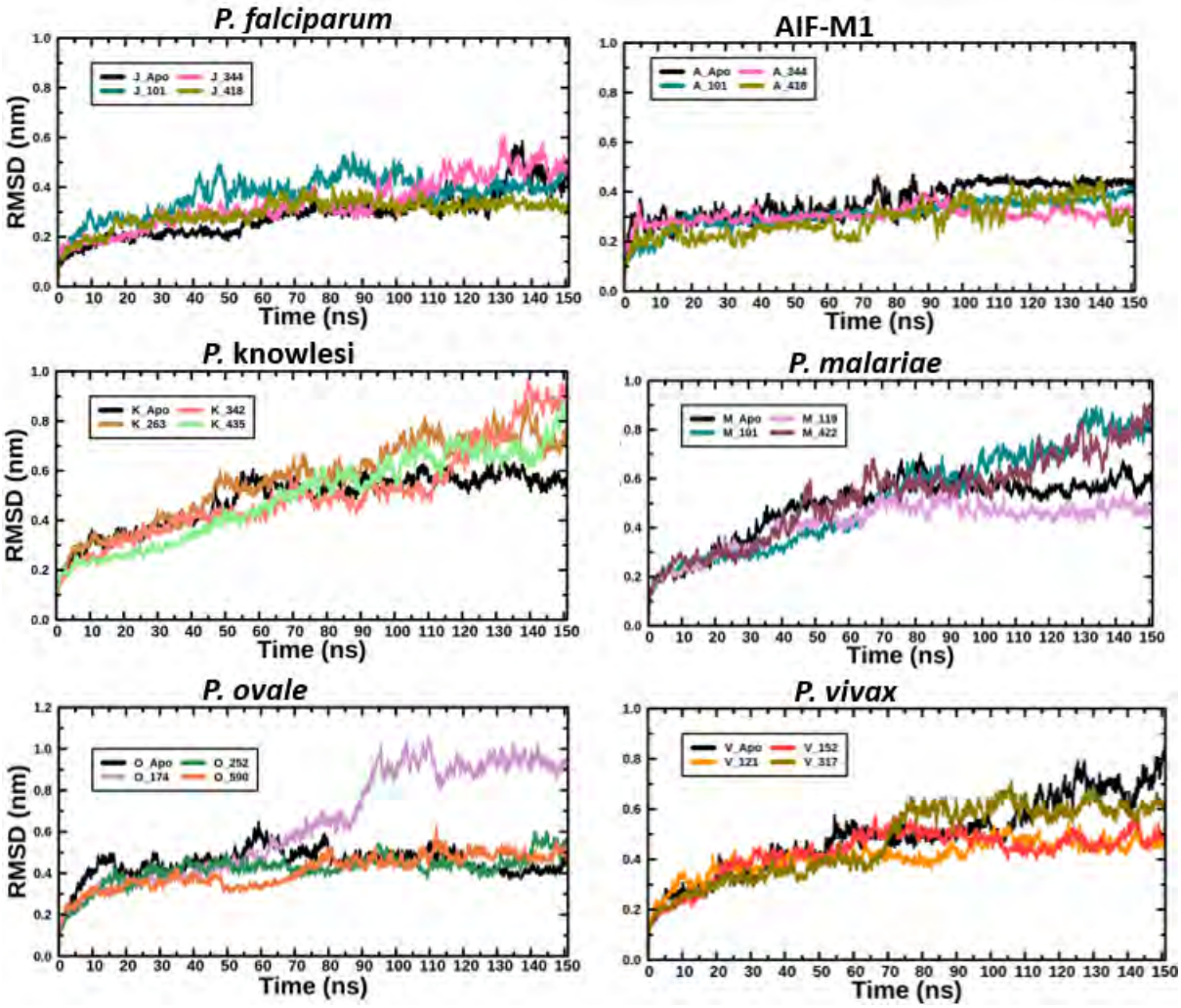

Figure 4.5 RMSD plots of the protein structures; J represents the *Pf*NDH2, A represents the AIF-M1, K represents *Pk*NDH2, M represents *Pm*NDH2, O represents *Po*NDH2 and V represents *Pv*NDH2

R studio was used to plot the graphs (RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL Http://Www.Rstudio.Com/. - Search, 2020)

The helix, beta-sheets and loops of the secondary protein elements can have an impact on the dynamic behaviour of structural and conformational aspects of the structure (Mirza et al., 2018). The *apo* RMSD results were different from the protein-ligand RMSD results, proving the ligands can have a conformational impact on the proteins and therefore, can possibly affect their functioning.

Regions of the protein that go through conformational changes to facilitate ligand binding are known as ligand-binding-sites (Guterres et al., 2019). However, to evaluate the binding stability of the ligands the ligand RMSD was calculated, in order to understand further the behaviour of the hit compounds in the binding cavity of the target proteins. This is important as many studies have used small molecules binding in the active sites of proteins, where mechanism is either by inhibition or activation, to cure a range of acute diseases.

In figure 4.6a, the RMSD value of the ligands throughout simulation was below 1 nm, *Pf*NDH2-SANC101, *Pf*NDH2-SANC3444, *Pk*NDH2-SANC263 and *Pk*NDH2-SANC435, showed a gradual increase in deviation after 105 ns time interval of the simulation. *Pf*NDH2-SANC418, AIF-SANC101, AIF-SANC344, AIF-SANC418, had stable ligand RMSDs; generally, the ligands showed a stable RMSD, evidence that they did not change in conformation during the MD simulation, proving the quality of the docked pose and justifying further investigation. *Pk*NDH2-SANC342 showed increase in ligand RMSD deviation after 120 ns, signifying a change in conformation. In figure 4.6b, the ligand RMSDs of *Pm*NDH2, *Po*NDH2 and *Pv*NDH2 were stable showing no significant conformational changes during simulation.

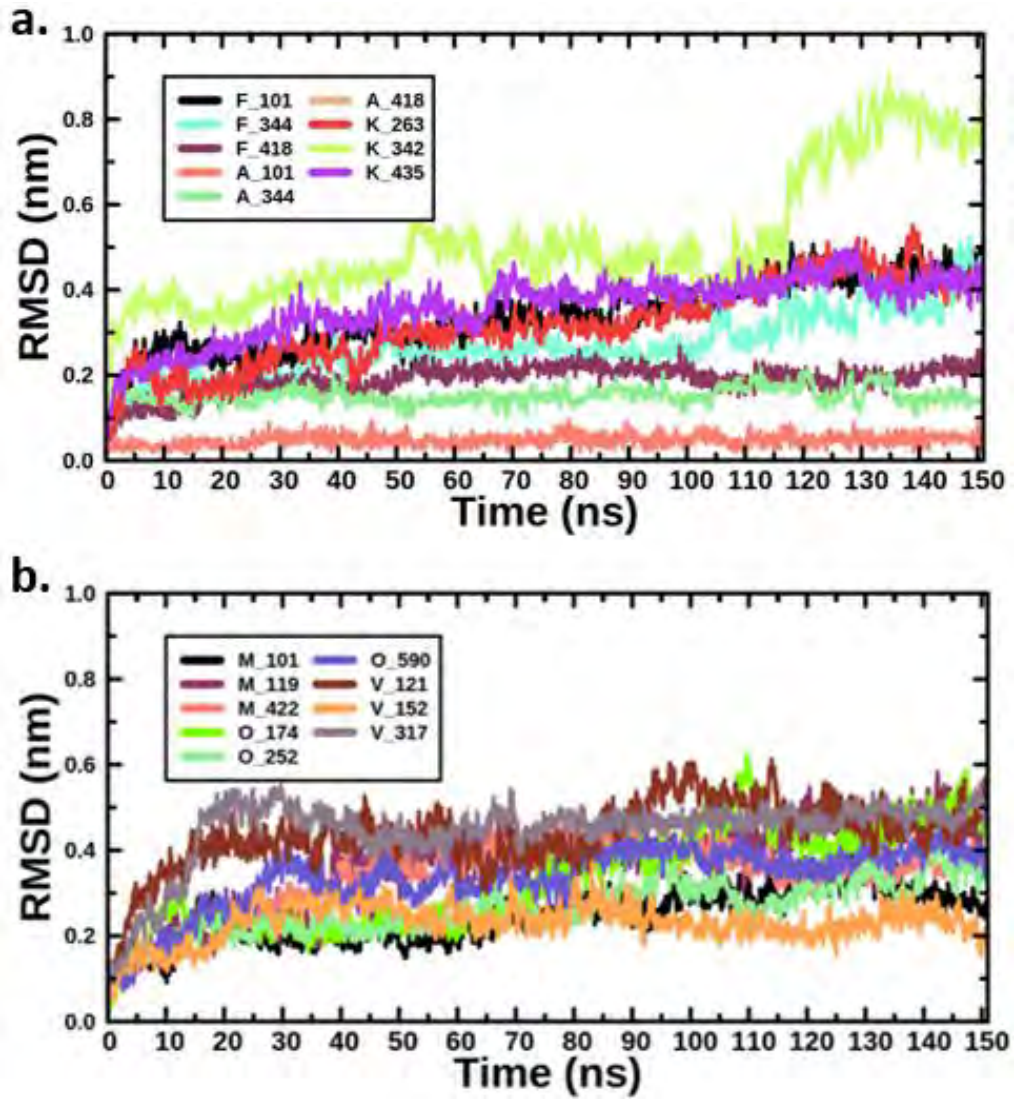

Figure 4.6 RMSD results of the ligands, F represents *Pf*NDH2, A represents AIF-M1, K represents *Pk*NDH2, M represents *Pm*NDH2, O represents *Po*NDH2, V represents *Pv*NDH2 and the numbers are representing the specific ligand

4.3.1.2 RMSF analysis

The NDH2 protein is a homodimer and both chains A and H were included in the MD simulations. The RMSF results in the study, of the protein in the context of a membrane present during simulation, are shown in the figure 4.7 below. The RMSF of the residues in NDH2 complexes for *P. falciparum* were below 0.9 nm; there were high fluctuations between residues G343-N438 of both chains which were in the linker, domain C and C-terminal region, in particular with both the *apo* protein and the SANC344 complex showing high fluctuations in both chains A and in H. In general, the residues had high fluctuations when the protein was bound to the membrane, with particularly notable fluctuations in the same G343-N438 region.

The RMSF of all residues for AIF-M1 complexes was below 2.3 nm; the complex with SANC418 had a high RMSF fluctuation of 2.2 nm for N583. In *Pk*NDH2 complexes the RMSF for protein residues was below 1.5 nm and high fluctuations were seen between residues G343-Q419 in both chains which form part of the linker and Domain C region. In *Pk*NDH2 complex with SANC342 within the membrane had a high RMSF in both NDH2 chains. Similar behavior was observed in complexes of *P. knowlesi* with low fluctuations in the absence of the membrane.

The RMSF of residues in the membrane-bound *Pm*NDH2 were all below 1.0 nm; however high fluctuations were observed between residues V343-Q419 in both chains which form part of the linker and Domain C region. The SANC422 complex had high fluctuations; although in general the membrane bound complexes had higher RMSF values compared to the absence of the membrane. The greatest fluctuations were observed in chain A for the SANC101-bound complex and in chain H for the SANC422-bound complex. For the SANC119 complex high fluctuations were observed in both chains.

The RMSF of residues of *Po*NDH2 across all simulations, *apo* and complexed, was below 1.1 nm, but the highest fluctuations were between residues G343-E419 in both chains which form part of the linker and Domain C region. The SANC174 complex showed high fluctuations in chain A, while for the complex with SANC590 high fluctuations were observed in chain H.

The RMSF for residues across all *Pv*NDH2 simulations as membrane-bound systems was below 1.4 nm, with the highest fluctuations being observed between residues G343-Q419 in both chains (which form part of the linker and Domain C region). The membrane bound *apo* protein demonstrated relatively high fluctuations in both chains, which was a change from the membrane free simulation.

Generally, the residues 343-419 showed high RMSF fluctuation in plasmodium species both in the presence and absence of the membrane.

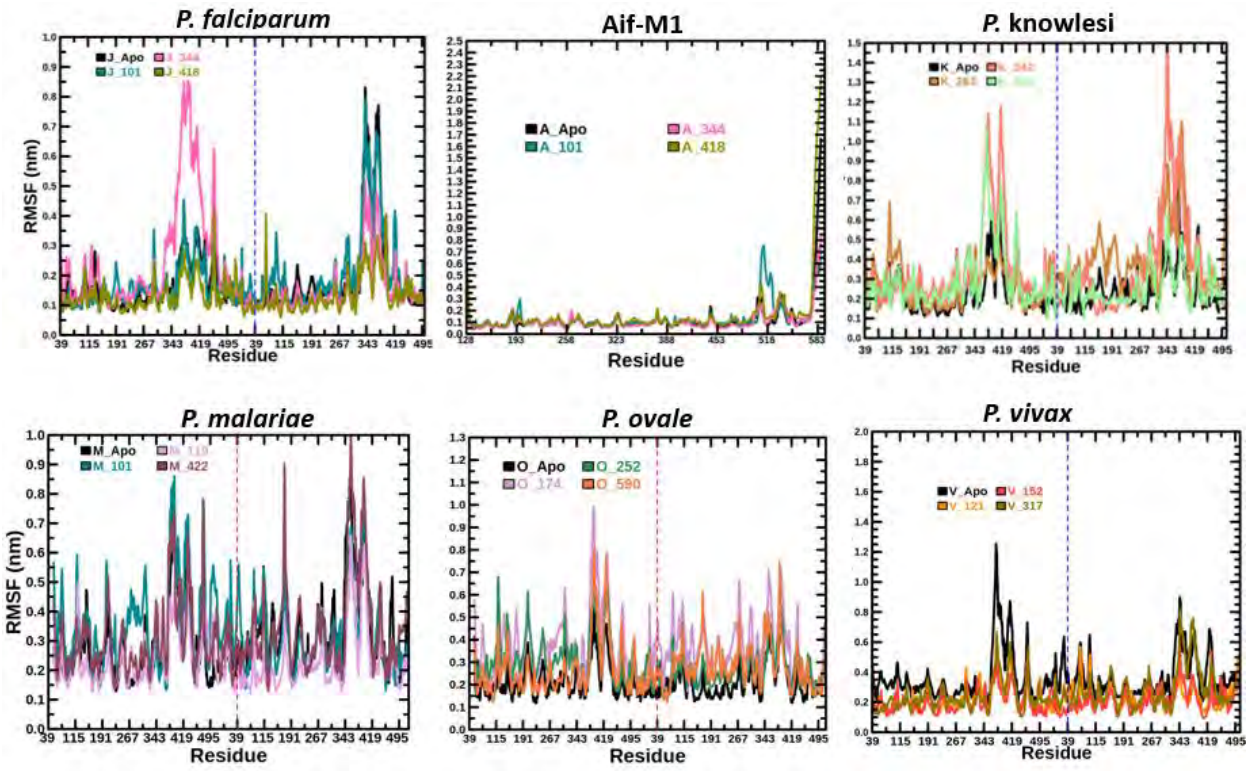

Figure 4.7 RMSF plots of the protein structures, J represents the *Pf*NDH2, A represents the AIF-M1, K represents *Pk*NDH2, M represents *Pm*NDH2, O represents *Po*NDH2 and V represents *Pv*NDH2

R studio was used to plot the graphs

4.3.1.3 Rg analysis

The radius of gyration calculates the compactness of the protein, increase in protein compactness would result in key residues moving closer together, and conversely decrease in compactness would result in protein residues moving away from each other. Calculating the dimensions of the protein in terms of radius of gyration (Rg) allows for initial assessment of the level of compactness that contributes to conformational selection. The distance between the axis of rotation and a point where total mass of the system is supposed to be concentrated, so that the moment of inertia around the axis may stay the same is known as the radius of gyration (Pirhadi & Amani, 2020). In essence, gyration shows the distribution of components of an object. The effect of the ligands on the radius of gyration of the NDH2 protein was studied.

The total simulation time was 150 ns, and through this time the Rg plots show the change in compactness of the complexes (figure 4.8). The Rg for all *Pf*NDH2 complexes (in the context of the membrane) was below 3.75 nm; however, there was gradual increase noticed in the compactness of the *apo* protein, The SANC101 complex was stable between 50-130 ns, SANC344 was stable until 90 ns, after which an increase in Rg was observed. The SANC418 complex proved to be stable in terms of a relatively constant Rg throughout simulation. For simulations involving AIF-M1 the Rg remained

below 2.38 nm, interestingly with the *apo* protein decreasing the Rg value after 90 ns. However, for the SANC101 complex an increase in gyration was observed, while the Rg in the SANC344 complex continued to decrease until the end of the simulation. Conversely the SANC418 complex showed relative stability although there was a decrease in Rg after 100 ns. The hit compound complexes in *P. falciparum* NDH2 proved to increase the compactness of the system; this contrasts with the behavior of the AIF-M1 complexes.

The Rg for the *Pk*NDh2 complexes remained below 4.0 nm, and while the *apo* protein and the SANC263 and SANC342 complexes showed an increase in Rg value, the SANC435 complex showed relative stability with a decrease in Rg after 50 ns. On the other hand, the Rg of the *Pm*NDH2 was below 3.9 nm, with the *apo* protein showing an increase and then a decrease of compactness (the latter towards the end of the simulation), but the three SANC101, SANC119 and SANC422 complexes all showed an increase in compactness during simulation. The radius of gyration of *Po*NDH2 was below 4.1 nm, with both the *apo* protein and the SANC590 and SANC252 complexes showing a stable plot (although the Rg for the SANC252 complex increases after 125 ns), but with the SANC174 complex increasing in Rg through simulation. Finally, the radius of gyration in *Pv*NDH2 complexes was consistently below 4.0 nm, the *apo* protein, SANC121 and SANC317 complex showed increase in radius of gyration, while the SANC152 complex decreased in compactness after 50 ns.

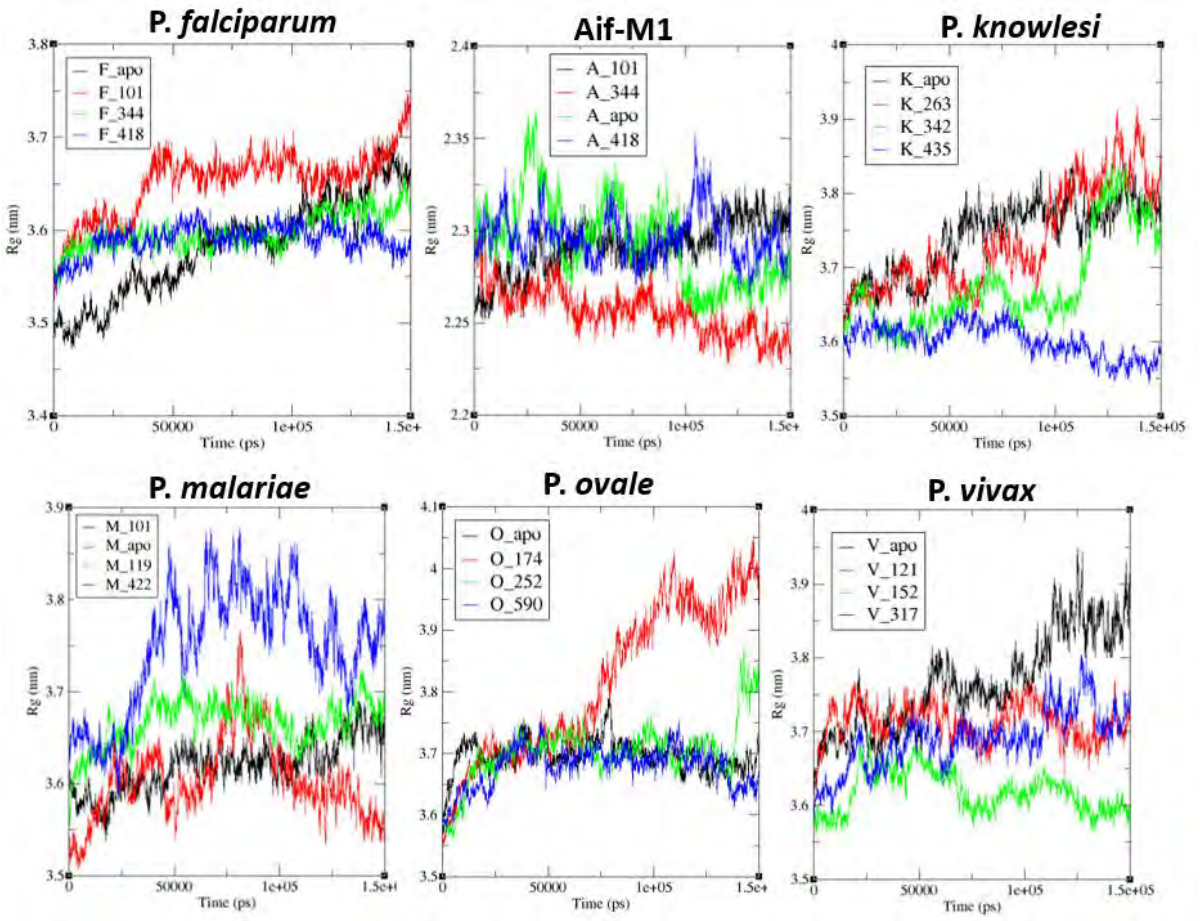

Figure 4.8 The radius of gyration for all structures, F represents the *Pf*NDH2, A represents the AIF-M1, K represents *Pk*NDH2, M represents *Pm*NDH2, O represents *Po*NDH2 and V represents *Pv*NDH2

4.3.2 Essential Dynamic calculations (PCA) and Dynamic cross correlation (DCC)

RMSD projections represent conformational changes in a single dimension; on the other hand, PCA is used to show and analyse 3-D conformational sampling through multidimensional analysis, as such through the calculation of PCA internal dynamics, and association of Cα protein conformations can be explored. The PCA calculation is a multivariate statistical approach to excerpt the important motions from the MD trajectories, by removing the rotational and translational movements (Nizami et al., 2016). The principal component calculations were calculated on the MD trajectories and conformation populations were analysed based on the first two principal components (PC1 and PC2), since the combination of both components explained majority of total variance across the models.

By diagonalising the covariance matrix of atomic fluctuation, eigenvalues and eigenvectors are obtained. The eigenvector with the largest possible variance in the dataset is the first principal component (PC1). The eigenvalues represent the extent of motion, while the eigenvector defines the direction of motion (Amadei et al., 1993). The plots show the conformational sampling covered by PC1 and PC2 (that represent the largest and second largest possible variances of the structures). In the PCA plots created for each 150 ns molecular dynamics simulation (figures 4.9-4.14), the energy plot with the low energy basin, superimposed, shows the more stable protein conformations and their positioning during molecular dynamics. In the essential dynamic plots, the dots correspond to individual protein conformations and are coloured by the time of sampling. The kernel density contour plots are from blue (lowest density) to yellow (highest density) and serve as a visual guide for the energy surface and are independently scaled based on the respective samples. The red and blue labels are results of two separate methods for extracting conformations found at the estimated energy minima.

Dynamic cross correlation was performed to analyse the extent to which protein residues move together, and the protein-ligand complexes were also visualized to monitor the positioning of the ligands with respect to the target during simulation.

In the membrane-bound *apo Pf*NDH2_we observed low energy conformations (in the lowest energy basin) at three points during the simulation (20100 ps, 29900 ps and 73300 ps). In the PCA plot, there was observed a percentage variance of 17.32% along PC1, and along PC2 a total variance of 9.99%. In the simulation of complex *Pf*NDH2-101, the conformation in the low energy basin was observed once at 9000 ps, while the percentage variances in the PCA plot of the first two components were 14.87% along PC1 and 10.04% along PC2. Similarly, for the *Pf*NDH2-344 complex we observed a conformation from the lowest energy at 114700 ps, while percentage variance of 13.59% along PC1

and 10.48% along PC2 was evident. In the final simulation for this structure, that of complex PfNDH2- 418 we observed the lowest energy conformation early in the simulation (at 1200 ps), and in its PCA plot 14.29% variance accounted by PC1 and 10.19% for PC2 (figure 4.9a). It was interesting to relate the PCA results with the RMSD results, as the lowest energy basins correlated with a low RMSD and stability during dynamics.

To calculate the extent to which protein residues interact together, dynamic cross correlation was used, and this produced heatmaps that are shown in figure 4.9b below. The DCC is represented from a range of -1(blue) to 1(red) in which 1 is high correlation, the X and Y axes show the residue index from 0- 500 residues and this will be translated from Arg39-Lys533 to correctly number the residues, meaning 0 = R39, 100 = L139, 200 = S239, 300 = L339, 400 = A439 and 500 = 539. There were slight differences in all DCC heatmaps, the FAD, and part of the NADH binding showed high correlation in the heatmap. The DCC for *P. falciparum* complexes ranged from -0.25 to 1.0; the SANC101 and SANC344 complexes reduced residue correlation in comparison to the membrane-bound *apo* protein; conversely the SANC418 complex exhibited a high residue correlation. Of note is that residues that showed high fluctuations in the RMSF (the range G343-N438) showed a low cross correlation. In comparison to the simulations in the absence of the membrane, it was apparent that residue correlation was reduced in the presence of the membrane.

In figure 4.9c below, the complete system of *Pf*NDH2 within the membrane is shown, showing the orientation chosen by the OPM server within the workflow in CHARMM-GUI. In figure 4.9c below, the first image shows the peripheral membrane protein embedded on the membrane, the second image shows the complex structure solvated and the third image shows the top view of the complex.

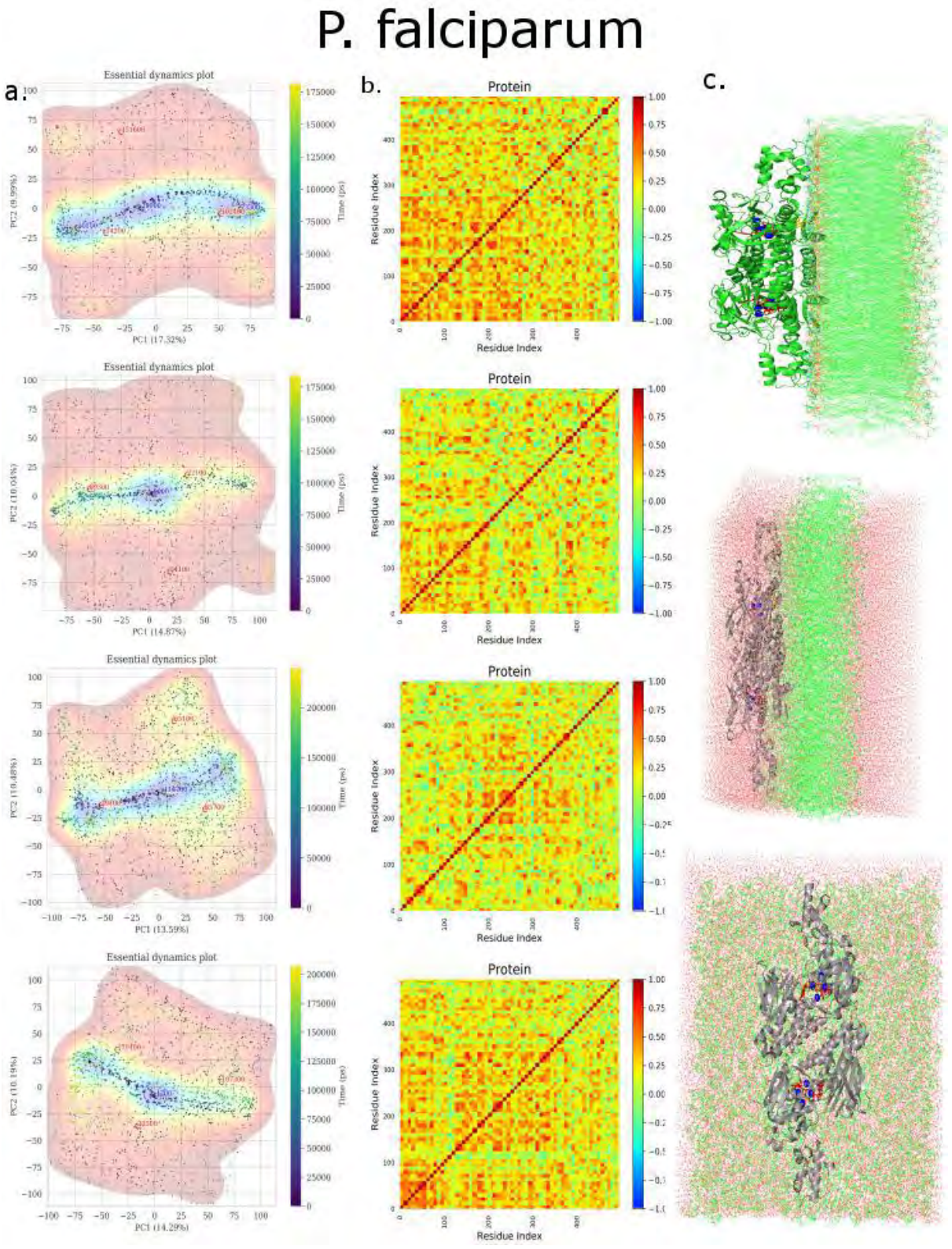

Figure 4.9 (a) Essential dynamic (PCA) analysis in the first column with *Pf*NDH2-*apo* in the first row, *Pf*NDH2-101 in the second row, *Pf*NDH2-344 in the third row and *Pf*NDH2-418 in the last row, with the simulation ranging from 0 (purple) to 150000 (yellow) ps. **(b)** Dynamic cross correlation (DCC) analysis in the middle column with *Pf*NDH2-apo in the first row, *Pf*NDH2-101 in the second

row, *Pf*NDH2-344 in the third row and *Pf*NDH2-418 in the last row, with the correlation ranging from -1(blue) to 1(red). **(c)** *Pf*NDH2 oriented on the membrane

Comparison of the complexes of *Pf*NDH2 with the human homologue are of particular interest. In the membrane-bound AIF-M1_*apo* we observed the lowest energy conformation about halfway through the simulation at 73520 ps. For the *apo* AIF-M1, in the PCA analysis of the simulation, PC1 accounted for 23.81% of the total variance, while for PC2 this was 7.12% (figure 4.10a). For the complex AIF-M1- 101 the lowest-energy conformation was observed slightly later in simulation at 99479 ps, and the PCA analysis provided a percentage variance of 20.34% along PC1, and 7.83% along PC2. For the second complex, In AIF-M1-344 again we observed the lowest-energy conformation about half-way through simulation at 75810 ps; for PCA the percentage of total variances was 24.33% along PC1, and 7.69% along PC2. For the final AIF-M1 complex, AIF-M1-418, the lowest energy conformation was observed slightly earlier than in other simulations with the membrane-bound AIF-M1, at 65840 ps, while the PCA plots showed percentage of variance of 18.20% and 10.24% along PC1 and PC2 respectively. In figure 4.10a the low energy basins are superimposed on the PCA; these correlate with the RMSD results in terms of stability. As expected, we observed differences when the ligands were bound to *P. falciparum* and apoptosis inducing factor (AIF-M1), in terms of the PCA behaviour, and in terms of the appearance of low-energy conformations during dynamics.

Again, the DCC is represented (figure 4.10b) with a range of -1(blue) to 1(red) in which 1 is high correlation; in this figure the X and Y axes show the residue index from 0-500 residues and corresponds to Ala128-Phe608, meaning 0 = A128, 100 = L228, 200 = T328, 300 = Q428, 400 = I528 and 500 = 628. In the DCC, there was a high correlation in the AIF-M1-101 complex in all residues, and the correlation ranged between 0.2 and 1 in the three simulations: *apo* AIF-M1, AIF-M1-344 and AIF-M1- 418 complexes. Of note was that the residues in the protein that showed high fluctuations in the RMSF analysis had a positive correlation close to 1 in the DCC. There was a greater residue positive correlation in the DCC when the ligands were bound to the human homologue in comparison to when bound to *Pf*NDH2. We observed a correlation value of 1 on residue N583 that had a high fluctuation in RMSF.

In figure 4.10c the membrane/solvent/AIF-M1 construct is shown, as produced by CHARMM-GUI (Membrane builder) webserver. Note that the human homologue was used as a control, as it is not a transmembrane protein. Figure 4.10c illustrates that the AIF-M1 protein is not bound to the membrane and is neither interacting nor touching the membrane.

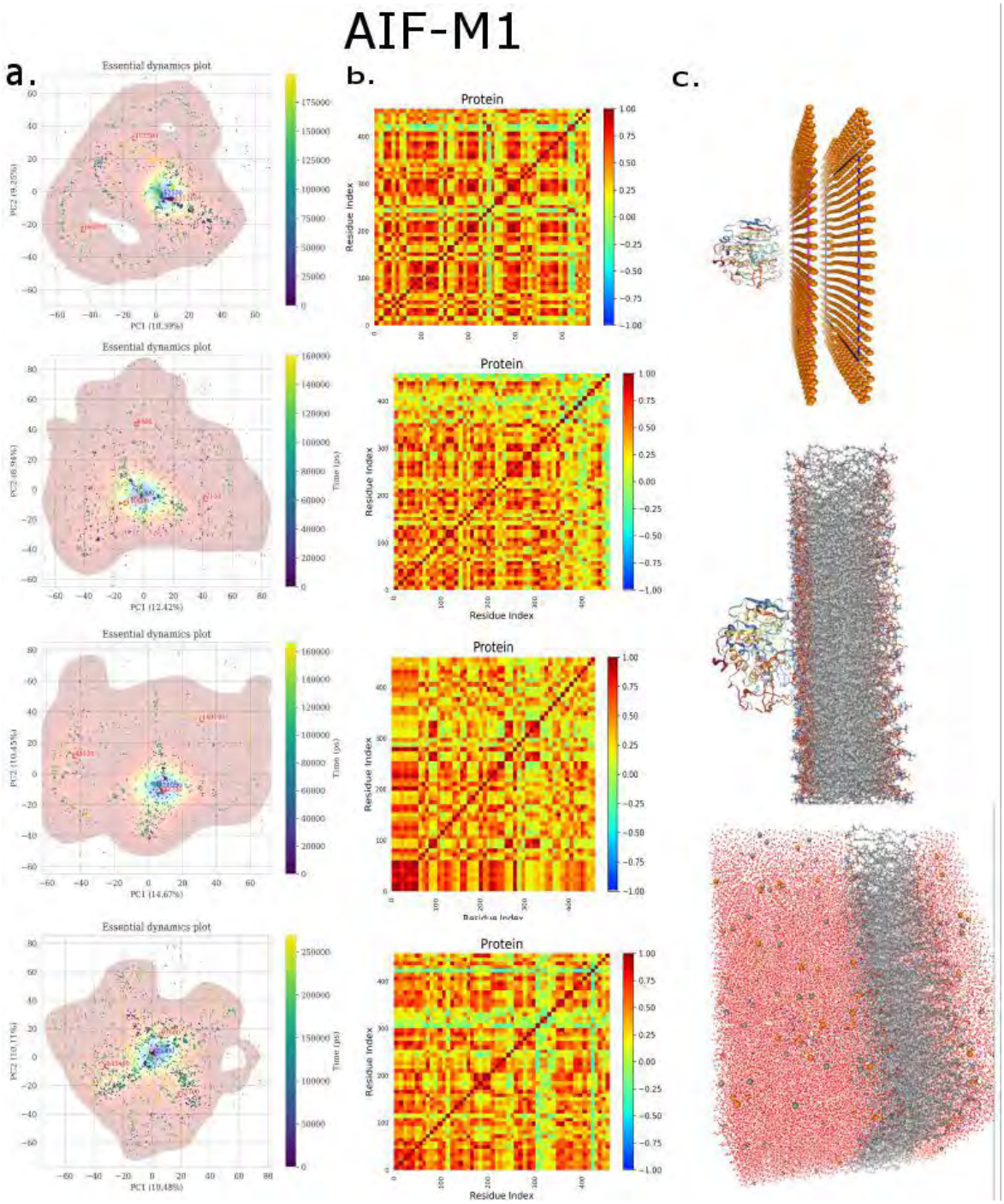

Figure 4.10 (a) Essential dynamic (PCA) analysis in the first column with AIF-M1-*apo* in the first row, AIF-M1-101 in the second row, AIF-M1-344 in the third row and AIF-M1-418 in the last row, with the simulation ranging from 0 (purple) to 150000 (yellow) ps. **(b)** Dynamic cross correlation (DCC) analysis in the middle column with AIF-M1-*apo* in the first row, AIF-M1-101 in the second row, AIF-M1-344 in the third row and AIF-M1-418 in the last row, with the correlation ranging from -1 (blue) to 1 (red). **(c)** AIF-M1 oriented on the membrane

For two of the *Pk*NDH2 simulations the lowest energy conformation occurred at the start of simulation (figure 4.11a). For *Pk*NDH2-*apo* we observed a conformation from the lowest energy

basin at 2500 ps, while for *Pk*NDH2-263 the lowest conformation was observed at 6900 ps. For *Pk*NDH2-apo and *Pk*NDH2-263, the observed percentage variances along PC1 and PC2 were 15.97% and 9.21%, and 17.37% and 8.01% respectively. For the membrane bound *Pk*NDH2-342 complex, the lowest energy conformations were observed at 64400 ps and 105400 ps; and PCA analysis of this simulation resulted in PC1 and PC2 describing 15.98% and 11.31% of the total variance. Finally, for this system, for the complex *Pk*NDH2-435 the lowest energy conformation was observed at 15450 ps, and the respective variance percentages in the PCA plot were 17.36% and 8.56%. In this last case, it is of note that an increase in RMSD was observed well after the appearance of the low energy conformation.

In figure 4.11b the DCC is represented from a range of -1(blue) to 1(red) in which 1 is high correlation; for *Pk*NDH2 the X and Y axes show the residue index from 0-500 and this corresponds to R39-K533, meaning 0 = R39, 100 = A139, 200 = V239, 300 = L339, 400 = Q439 and 500 = 539. The dynamic cross correlation of the residues ranged from -0.25 to 1 and the correlation pattern was similar across simulations for the *Pk*NDH2 complexes. In terms of relation to the RMSF, residues in membrane bound *Pk*NDH2 complexes with high RMSF fluctuations tended to have low DCC correlations when the ligands were bound.

Figure 4.11c shows the membrane-bound *Pk*NDH2 as prepared in CHARMM-GUI webserver; this visualization showing the protein orientation on the membrane was created using the PyMOL tool.

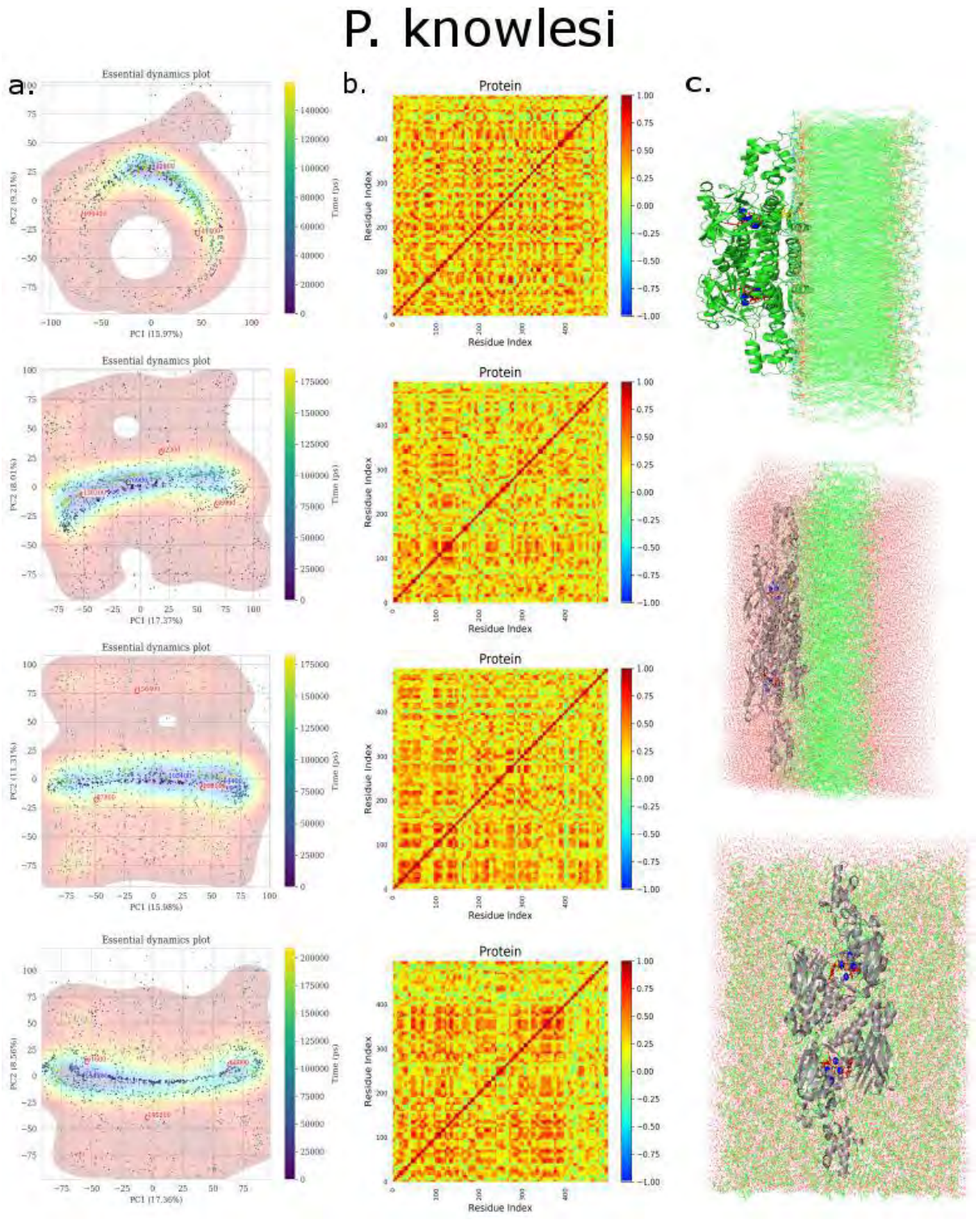

Figure 4.11 **(a)** Essential dynamic (PCA) analysis in the first column with *Pk*NDH2-*apo* in the first row, *Pk*NDH2-263 in the second row, *Pk*NDH2-342 in the third row and *Pf*NDH2-435 in the last row, with the simulation ranging from 0 (purple) to 150000 (yellow) ps. **(b)** Dynamic cross correlation (DCC) analysis in the middle column with *Pk*NDH2-*apo* in the first row, *Pk*NDH2-263 in the second row, *Pk*NDH2-342 in the third row and *Pk*NDH2-435 in the last row, with the correlation ranging from -1 (blue) to 1 (red). **(c)** *Pk*NDH2 oriented on the membrane

Figure 4.12a shows the PCA analysis from simulations of membrane bound *Pm*NDH2. For the *apo* system *Pm*NDH2-*apo* there were two instances of conformations in the lowest energy basin at 23400 ps and 127300 ps. These low energy occurrences were at 75700 ps for *Pm*NDH2_101, at 16010 ps for *Pm*NDH2-119 and quite late in simulation at 100400 ps for *Pm*NDH2-422. For the observed variances in PC1 and PC2, these were 15.10% and 11.32% for *Pm*NDH2, 14.73% and 9.79% for *Pm*NDH2-101, 17.57% and 9.73% for *Pm*NDH2-119 and 19.43% and 9.60% for *Pm*NDH2-422. As was the case with previous sets of simulations, the low energy basins correlated with a stable protein RMSD.

For the membrane bound *Pm*NDH2 complexes the DCC is also presented in a range of -1(blue) to 1(red) in which 1 is high correlation. In figure 4.12b the X and Y axes show the residue index from 0-500 residues and now corresponds to residues R39-K533 in *Pm*NDH2, where 0 = R39, 100 = M139, 200 = L239, 300 = S339, 400 = Q439 and 500 = 539. The DCC values in all these *Pm*NDH2 complexes ranged from -0.26 to 1 and we observed a greater residue correlation in SANC101 and SANC119 complexes. Further, the membrane-bound systems again appeared to have lower DCC compared to the membrane-free systems. In residues with high RMSF fluctuation, we observed high residue correlation particularly when SANC101 and SANC119 were bound. Figure 4.12c below shows the full systems for membrane-bound *Pm*NDH2 as produced through the use of the CHARMM-GUI webserver.

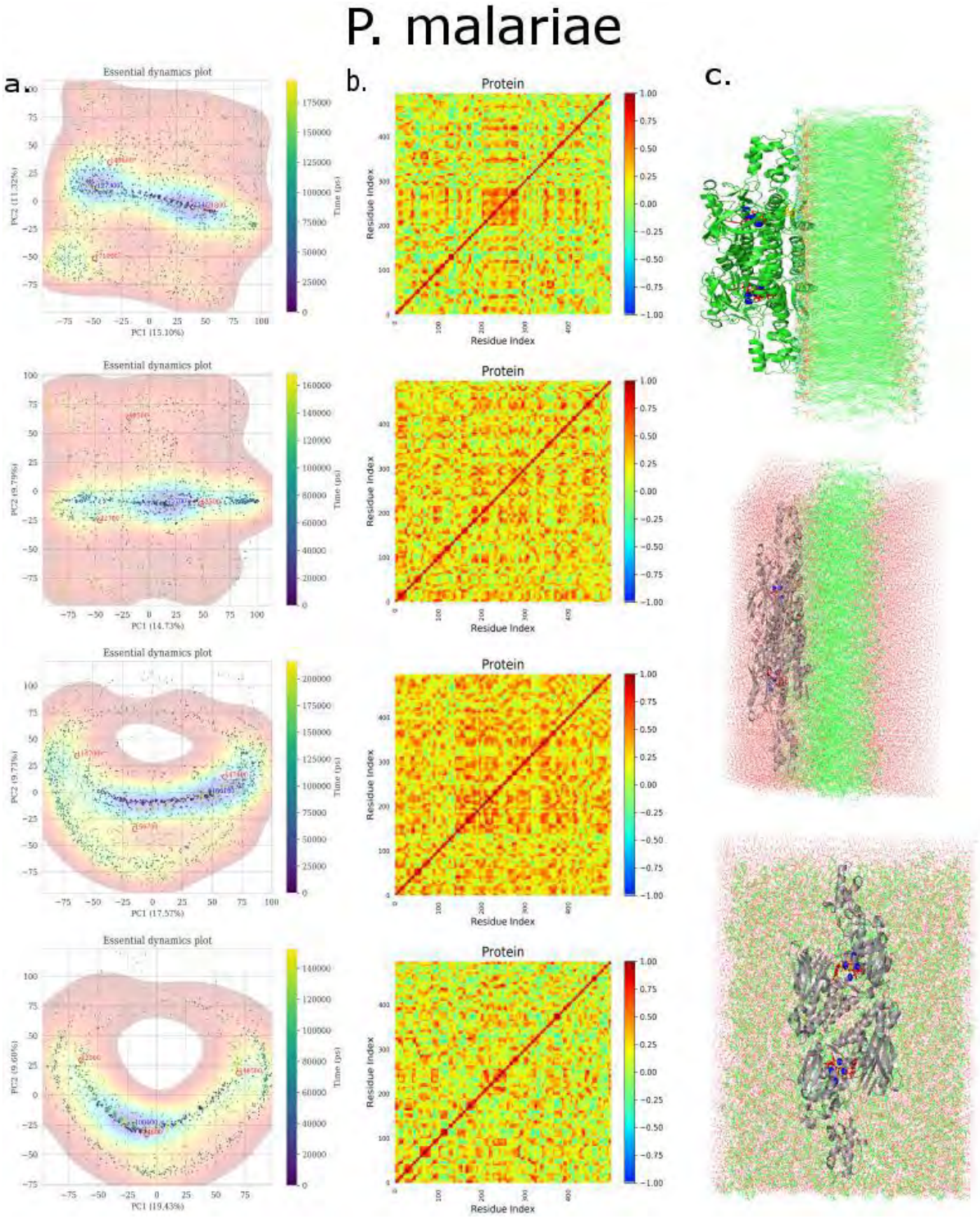

Figure 4.12 (a) Essential dynamic (PCA) analysis in the first column with *Pm*NDH2-*apo* in the first row, *Pm*NDH2-101 in the second row, *Pm*NDH2-119 in the third row and *Pm*NDH2-422 in the last row, with the simulation ranging from 0 (purple) to 150000 (yellow) ps. **(b)** Dynamic cross correlation (DCC) analysis in the middle column with *Pm*NDH2-*apo* in the first row, *Pm*NDH2-101 in the second row, *Pm*NDH2-119 in the third row and *Pm*NDH2-422 in the last row, with the correlation ranging from -1 (blue) to 1 (red). **(c)** *Pm*NDH2 oriented on the membrane

Figure 4.13a shows the PCA analysis for the membrane bound *Po*NDH2 complexes together with the energy basins, in which the lowest energy structures are at 131500 ps for *Po*NDH2-*apo*, 25700 ps and 67200 ps for *Po*NDH2-174, 11200 ps for *Po*NDH2-252 and 7800 ps for *Po*NDH2-590. Also evident is the percentage variance accounted for by the respective PC1 and PC2, viz, 14.14% and 9.33% for *Po*NDH2-*apo*, 15.91% and 9.72% for *Po*NDH2-174, 15.77% and 10.24% for *Po*NDH2-252 and 14.98% and 8.89% for *Po*NDH2-590. Change in conformation occurred towards the end of simulation in *Po*NDH2-590 and *Po*NDH2-*apo*.

Figure 4.13b again shows the DCC with the same scales as before where 1(red) indicates high correlation. For *Po*NDH2 the X and Y axes show the residue index from 0-500 residues and this corresponds to residues K39-K533 (i.e., 0 = K39, 100 = F139, 200 = I239, 300 = L339, 400 = A439 and 500 = 539). The correlation values in the complexes ranged from -0.25 to 1, where of particular note the residues in the FAD binding domain showed high correlation. In terms of RMSF, residues which had high RMSF also had low correlation here, although the correlation was higher when the membrane was absent. The CHARMM-GUI outputs are shown in figure 4.13c with the *P. ovale* protein oriented on the membrane.

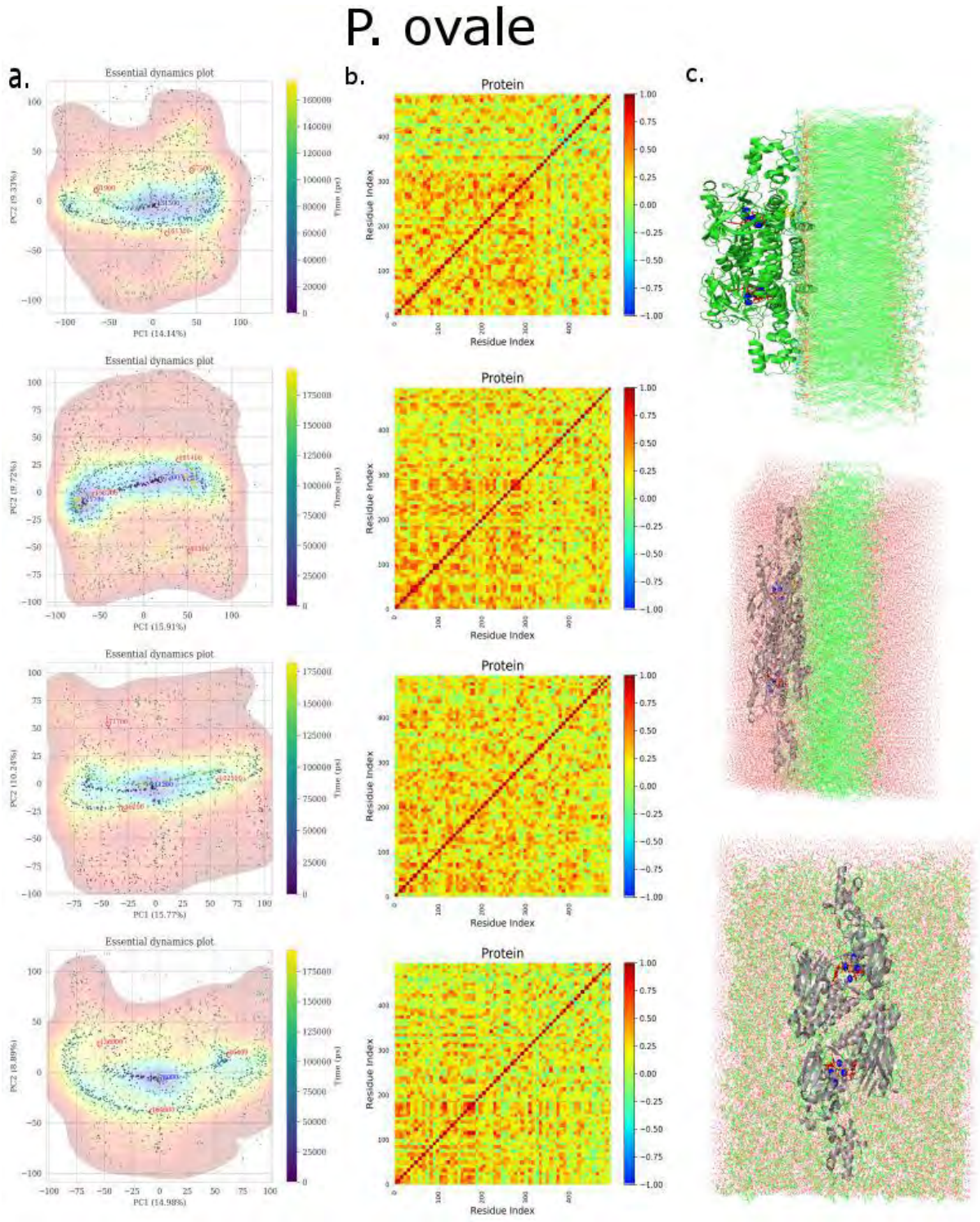

Figure 4.13 (a) Essential dynamic (PCA) analysis in the first column with *Po*NDH2-*apo* in the first row, *Po*NDH2-174 in the second row, *Po*NDH2-252 in the third row and *Po*NDH2-590 in the last row, with the simulation ranging from 0 (purple) to 150000 (yellow) ps. (b) Dynamic cross correlation (DCC) analysis in the middle column with *Po*NDH2-*apo* in the first row, *Po*NDH2-174 in the second row, *Po*NDH2-252 in the third row and *Po*NDH2-590 in the last row, with the correlation ranging from -1 (blue) to 1 (red). (c) *Po*NDH2 oriented on the membrane

For the membrane bound *Pv*NDH2-*apo* system we observed conformations from the lowest energy basin at 39700 ps and 64800 ps, along the axis we observed a percentage variance of 17.06% along PC1, and along PC2 a total variance of 11.48% (figure 4.14a). In terms of the complexes, for *Pv*NDH2- 121, *Pv*NDH2-152 and *Pv*NDH2-317 the lowest energy conformation occurred early during simulation at 6200 ps, 400 ps and 5300 ps respectively. For *Pv*NDH2-121 the PCA provided percentage of variance of 17.24% and 9.80% for PC1 and PC2 respectively. For *Pv*NDH2-152 and *Pv*NDH2-317 these percentages were 14.31% and 10.13%, and 15.06% and 11.81% respectively.

Figure 4.14b shows the DCC for these systems, where 1(red) is high correlation. For *Pv*NDH2 the X and Y axes show the residue index from 0-500 residues and corresponds to G39-K533where 0 = G39, 100 = S139, 200 = V239, 300 = L339, 400 = Q439 and 500 = 539. The DCC ranged from -0.25 to 1, and we observed more residue interaction when the ligands were bound to the protein. Residues which had high fluctuations in RMSF showed low correlation and there was higher residue correlation when the membrane was absent. The CHARMM-GUI output for *P. vivax* was shown in figure 4.14c with the protein oriented on the membrane.

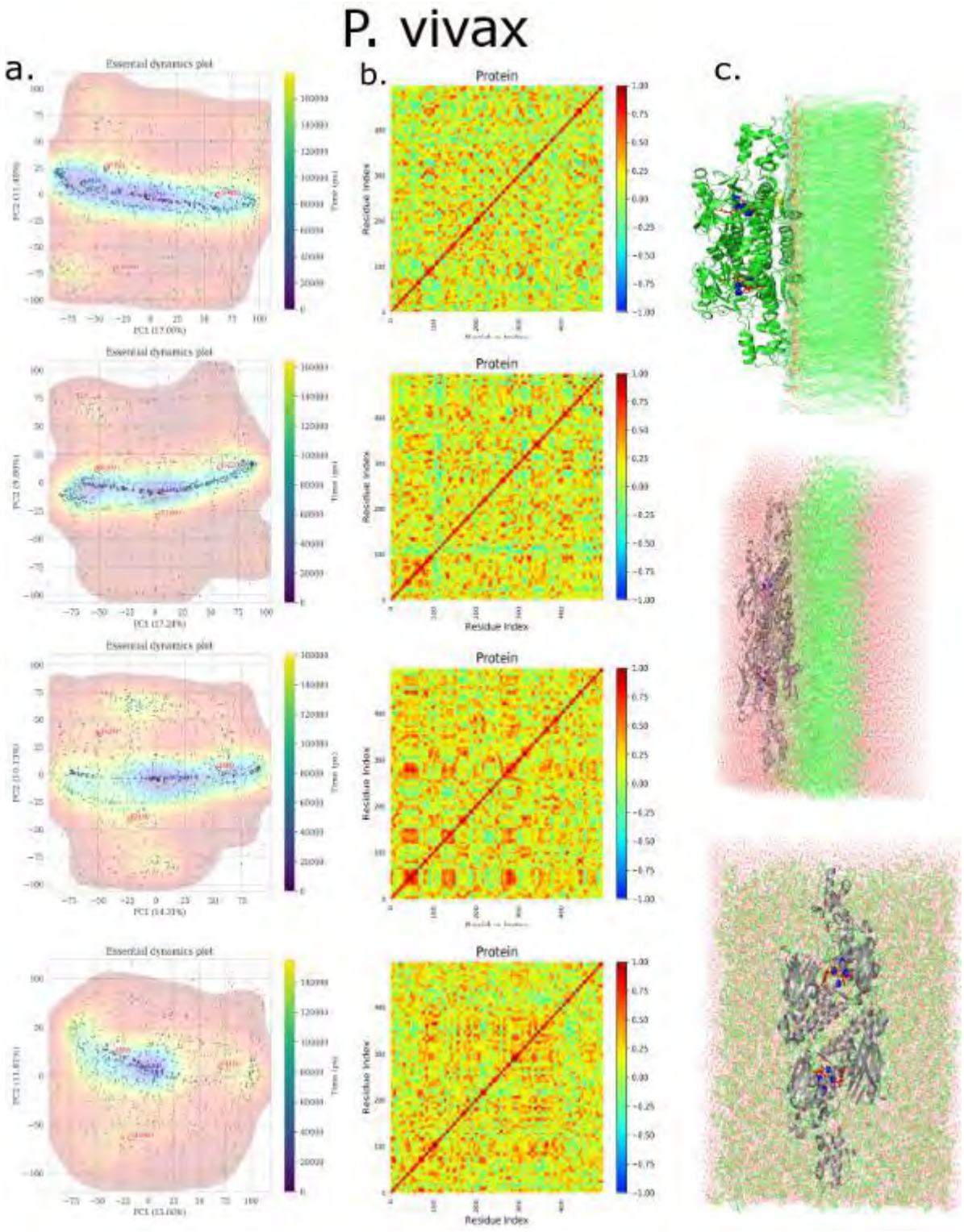

Figure 4.14 (a) Essential dynamic (PCA) analysis in the first column with *Pv*NDH2-*apo* in the first row, *Pv*NDH2-121 in the second row, *Pv*NDH2-152 in the third row and *Pv*NDH2-317 in the last row, with the simulation ranging from 0 (purple) to 150000 (yellow) ps. (b) Dynamic cross correlation (DCC) analysis in the middle column with *Pv*NDH2-*apo* in the first row, *Pv*NDH2-121 in the second row, *Pv*NDH2-152 in the third row and *Pv*NDH2-317 in the last row, with the correlation ranging from -1 (blue) to 1 (red). (c) *Pv*NDH2 oriented on the membrane

4.3.3 Chapter conclusion

This chapter focused on MD simulations of the type II NADH: quinone oxidoreductase (NDH2) protein structures and the Apoptosis inducing factor (AIF-M1), both *apo* and ligand complexed systems, in the context of a membrane. The analysis of these sets of MD trajectories was through the use of RMSD, Rg, and RMSF information, and through the use of Principal component analysis (PCA) and Dynamic cross correlation (DCC).

The CHARMM_GUI (Membrane builder) was used to construct the membrane bound systems and to produce the geometry and topology files for GROMACS that were used to perform these molecular dynamic simulations. In *in silico* drug discovery studies, it is important to mimic the exact protein environment and OPM server was used to correctly position the NDH2 protein on the membrane. The FAD cofactors and four magnesium ions in both chains were included in the study. The system took time to equilibrate due to its size, but for production dynamics the total simulation time was 150 ns. Residues between 343-419 of the protein showed high RMSF fluctuations; this was expected as this region covers the linker region. In many cases the ligand influenced and changed the dynamic behaviour of NDH2 in the presence of the membrane, in which some compounds clearly destabilised the protein. In this project it was our aim to identify compounds that could disturb the electron transfer process or affect the functioning of NDH2 leading to the death of the *plasmodium* without affecting the human. It was interesting that the 5 *plasmodium* species had different hit compounds, and that these behaved differently, influencing the dynamics of NDH2 in a variety of ways. It is further interesting that compounds interacting well with *P. falciparum* NDH2 are not necessarily going to interact with NDH2 for other *Plasmodium* species, proving the need to also study these species individually.

In essential dynamics (PCA) calculations, the highest probability density conformations (shown on our plots in blue) were extracted from the centroids of the highest contour level using the k-nearest neighbour algorithm. K = 1 for the points within that level, while the k-means algorithm was independently used to estimate other n possible levels of interest as specified from the number of clusters. These conformations with the highest probability densities, for all the protein complexes were represented by the conformations observed at the time (ps) intervals (in blue) on the plots. The PCA results showed different low energy basins in for the *apo* membrane dynamics, and for when the ligands were bound in all protein structures, proving that the ligands had an impact on conformational changes of the proteins. The dynamic cross correlation results showed a greater residue correlation in the absence of the membrane. Altogether, the findings confirmed the relevance of mimicking the correct protein environment, for which the inclusion of a membrane was necessary.

5

Conclusion

CHAPTER 1

In the first chapter we focused on general functioning and mechanism of spreading of the disease, understanding the protein under investigation and description, and learning of the computational techniques applied to reach the aim of the study. The lifecycle and spreading mechanism of malaria was explained in this chapter. Understanding the malarial disease and how it is spread to humans is important, as it leads to development of alternative ways people can adapt to protect and treat this disease. The mechanism and protein structure studies of the species leads to discovery of effective anti-malarial drugs.

Malaria continues to be a burden worldwide claiming the lives of many, especially children under the ages of 5, and there is need to do further drug search studies with aim to eradicate this disease. The *plasmodium* species have become resilient to current anti-malarial drugs leading to the need of new and effective drugs that are affordable, can be easily accessed and have less to no side effects on the human. Drug-resistant malarial strains have been continuously emerging through mutations, causing a great global challenge on health. There are several avenues of research including focus on the mutations and how they function; but aims are to find new anti-malarial drug targets such as the NDH2 protein which is a viable target for anti-malarial drug development, and to find active agents against these targets. The *Plasmodium falciparum, Plasmodium knowlesi, Plasmodium malariae, Plasmodium ovale* and *Plasmodium vivax* NDH2 proteins were included in the project as the target. It was of importance to study all five species as they all cause malaria in human beings, so to understand similarities and difference in structure and functional mechanism. It was also vital to understand if there was need for drug design for a specific plasmodium or if a single drug could affect all *plasmodium* species. As the NDH2 is not present in the human, apoptosis inducing factor (AIF-M1) protein was used as the human homologue.

The general, aim of this project was to study and use bioinformatics techniques on the NDH2 protein, identify hit compounds and to analyse protein-ligand interactions of the five *plasmodium* species that causes malaria in human beings. The study looked at the importance and difference of including or excluding a membrane for this transmembrane protein. Details are given of the webservers, computational methods, tools, and packages which were used to achieve the goal of this study.

CHAPTER 2

Data retrieval, sequence and structure analysis of the proteins were investigated. The beginning of many *in silico* studies begins with data retrieval, in which the sequence of the protein paves a way to multiple sequence analysis, phylogenetic analysis, motif discovery and 3-D structure creation of proteins whose crystal structure has not yet been solved. RCSB, NCBI and UNiProtKB webservers were used to retrieve the 5JWA (NDH2) *P. falciparum* crystal structure, together with the 1M6I apoptosis inducing factor (AIF-M1) crystal structure of the human homologue and the sequences of the other four *plasmodium* species.

To assess sequence conservation and conservation of protein domains multiple sequence alignment was used; this allowed us to analyse the similarities and differences between the *plasmodium* species and the human homologue. Other species sequences were added to reduce bias and add volume or data to the alignment. In multiple sequence analysis 3 webservers (MAFFT, MUSCLE and PROMALS3D) were used for comparison sake and we observed conserved regions in all 17 sequences, the MSA outputs of the 3 tools were similar with MUSCLE and PROMALS3D output having more gaps. The MEGA tool was used to calculate the phylogenetic trees and we observed the relationship of the species, with indications that the human homologue had the same ancestry as the *plasmodium* species proving a distance relation. The MAFFT MSA output was used as the input file to calculate the phylogenetic trees. The sequence identity heatmap that was calculated using an in-house MATLAB script, and this correlated with the MSA results and the phylogenetic tree as it showed portions in the sequence that were highly conserved.

A sequence motif is an amino acid sequence pattern that is widespread and has been proven or assumed to have a biological significance. The motifs were identified, and this was performed to understand if there were similar sequence patterns in the five *plasmodium* species and the human homologue, that might have been conserved during evolution to perform certain functions. The MEME tool was used to perform this calculation, and this was performed in the YODA cluster. 50 motifs were calculated. In the output a consensus sequence (the string of most frequent letters) was produced as a graphical representation called the sequence logo. In this logo the residues were presented in different heights and the height of different letters at the same position was proportional to their frequency in the motifs, the higher the letters the better the base conservation in that position. Motif 1, 9 and 10 were common in all sequences and these sequence patterns were manually scouted for on the sequences and mapped on the protein structures.

In the absence of the *P. knowlesi, P. malariae, P. ovale* and *P. vivax* NDH2 crystal structures in the databases, the crystal structure (PDB ID: 5JWA) of the *Pf*NDH2 protein was used as a template in the 3-D construction of these four NDH2 models. The MODELLER tool was used to create 100 models under very slow refinement and the best model with the lowest z-score was picked for further

investigations. The FAD cofactor and the four magnesium ions in each chain were modelled. To validate the quality of the models PROSA, Verify3D and PROCHECK webservers were used and they accepted the four models as of good quality. The models were superimposed with the 5JWA crystal structure to identify overlapping regions and assess the alignment. The sequence analysis of the protein structures was successfully performed and validated.

CHAPTER 3

An essential tool in drug discovery is molecular docking which is a computational method to predict the structure of a protein-ligand complex using conformational search and scoring techniques. Molecular docking of the 623 SANCDB compounds on the NDH2 structures (together with their human homologue) was performed, and the hit compounds were selected based on ligand size, binding energy, binding site and its interaction with the protein. The SANCDB is a free database that contains compounds isolated from the plant and marine life in and around South Africa, and this provided the ligands for screening. Natural products from this database have been shown to display antimicrobial, anticancer, against various neurological disorders and anti-diabetic activity. This was the first study to use this library against the type II NADH: quinone oxidoreductase. In future studies other ligand library databases can be used, to see how they affect NDH2 protein.

AutoDock Vina was used to perform molecular docking in a high throughput virtual screening approach. To this end, an Autodock plugin tool in PyMOL was used to calculate the box dimensions and size. As NDH2 is a homodimer the compounds were docked on all chains. The study was interested in finding hit compounds that bound to sites both on the two chains, and compounds that bound to the interface region. Each hit was analysed individually. To validate the docking method used, the FAD cofactor was redocked to the 5JWA protein structure, which resulted in very similar protein interactions. The compounds screened bound mainly in the FAD and NADH binding domains, and the linker region. Each of the five NDH2 protein structures had different hit compounds, as the ligands docked differently and with different binding energies. Hit compounds in the *Pf*NDH2 were analysed by molecular docking to the AIF-M1 protein, as this parasite is the most prevalent malarial parasite in causing disease and death. BIOVIA discovery studio was used to calculate the 2D protein-ligand interaction; for this we had as focus, an interest in conventional hydrogen bonds evident as a result of docking. The ligands screened both against *plasmodium* and human proteins showed better protein-ligand interactions in the parasite than the human homologue. METALizer, a ProteinPlus tool was used to calculate the interaction of the Mg2+ ions with the protein, so to determine if there was need for ion parameterisation before MD simulations.

Molecular dynamics simulations of the protein complexes without a membrane were performed as a series of controls. The GROMACS package and the CHARMM36 force field were used. The molecular dynamics was performed for a simulation time of 100 ns using the CHPC cluster in Cape

town. The ligands and protein structures were prepared for production run using standard GROMACS methods. Upon completion of simulation, the trajectories were analysed using Root mean square deviation (RMSD), Root mean square fluctuation (RMSF), Radius of gyration (Rg), Principal component analysis (PCA) and Dynamic cross correlation (DCC).

The MD simulation included the two chains, the FAD cofactor, four Mg2+ ions and the ligand of selection. The RMSD showed stability in general for *Pf*NDH2. In general, the RMSD results of the *plasmodium* proteins showed deviations when the ligands were bound. For the compounds bound to AIF-M1 stability was evident, indicating that the hit compounds lent some stability to this system; this matched the *apo* AIF-M1 simulation which had a very similar RMSD to that of its complexes through simulation. Across all simulations in this study, the RMSD for the protein was low. There were similar RMSF fluctuations in all parasite proteins between residues 343-419; this region covers the linker region and part of domain C region, and this RMSF fluctuation was observed in both chains. Similarly, in general, the RMSF in the protein structures across all simulations was low. In the human homologue the residue 583 showed a high RMSF fluctuation.

The essential dynamics (PCA) study showed the development of protein conformation during simulation, and the position during the simulation in which the system was in the low energy basin – this information correlated with the RMSD results. The PC1 and PC2 percentage values are detailed for all dynamics simulations and plotted accordingly. The structures from low energy basins were captured and superimposed against the protein *apo* protein, this superimposition is in harmony with the b-factor calculations of the crystal structures. The proteins showed greater flexibility around the peripheral regions of the protein, while the centre of the proteins remained aligned and rigid. The Dynamic cross correlation (DCC) of the plasmodium species resulted in values with a range from -0.50 to 1, where most of the residues had a positive correlation even when ligands were bound. AIF-M1-SANC101 had a positive correlation in all residues. Residue 583 which had a high RMSF fluctuation in AIF-M1 had a positive correlation of 1. In the parasite structures, the residues 343-419 that showed high RMSF fluctuations had a positive correlation of 1.

CHAPTER 4

Molecular dynamics simulations can capture the movement and flexibility of both protein and the ligand. In this chapter the MD simulations for the proteins and their protein-ligand complexes were performed in the presence of a membrane, CHARMM-GUI was used to create the membrane with the help of the OPM server to orient the NDH2 protein on the membrane. GROMACS topology files were created by CHARMM-GUI and were used as input files for the MD simulation. The simulation was performed over a timeframe of 150 ns due to the size of the system. The trajectories were also analysed visually in VMD to see the relative positions of protein, ligand, and membrane during

simulation. From the visualisation the protein did not detach from the membrane and the protein-ligand complexes remained intact throughout simulation. Due to the size of the trajectory (trr and xtc) files, the periodic boundary box together with water was removed prior to further analyses.

RMSD, RMSF, Rg, PCA and DCC calculations were performed as analysis of the MD trajectories. These results showed interesting differences from the simulations which were in the absence of membrane. The hit compounds in *Pf*NDH2 in general showed a destabilising influence on the protein, evident in the observed RMSD throughout simulation; results that were different from the AIF-M1 protein. The ligands in the *plasmodium* species caused a deviation from the behaviour observed by the membrane-bound *apo* proteins, generally causing destabilisation of the protein. In the *plasmodium* protein RMSF high fluctuations were also observed between the residues 343-419, a similar observation as to when the membrane was absent. A high RMSF fluctuation of residue 583 in AIF-M1 was also observed and the human homologue was not embedded in a membrane. In general, across other residues the RMSF of the protein structures were low when the membrane was present. To indicate the stability of the ligands in terms of conformation in the binding site, ligand RMSD was calculated. This showed general ligand stability across simulations in this dataset, with exception of SANC342 bound to membrane embedded *Pk*NH2. The protein RMSD provided information on the ability of ligands to cause conformational change in the protein structures. The radius of gyration calculations showed generally an increase in Rg and decrease of protein compactness towards the end of simulations. However, for the AIF-M1 protein, compactness remained even when the ligands were bound.

Also explored were the simulation time per system to reach low energy basins; this information was included as an overlay on the PCA plots; these results correlated with the observed protein RMSD during this set of simulations. In terms of dynamics cross-correlation, the residue correlation in the proteins was in the range of -0.25 to 1 and there was negative residue correlation observed between some residues in some simulations in the presence of the membrane.

Future studies will focus on docking other ligand library datasets and will concentrate on identifying allosteric sites (looking for aspects such as selectivity and druggability) and further work will also study the NDH2 mutations. We seek also to perform *in vitro* tests of the hit compounds, since this will complement the computational work, potentially giving further insights into binding and relevant biological effects.

References

1. Access, O. (n.d.). *We are IntechOpen , the world ' s leading publisher of Open Access books Built by scientists , for scientists TOP 1 %*.

2. Additive, C., Field, F., Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., Jo, S., Pande, V. S., Case, D. A., Iii, C. L. B., Jr, A. D. M., Klauda, B., & Im, W. (2016). *CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field*. https://doi.org/10.1021/acs.jctc.5b00935

3. Altschup, S. F., Gish, W., Pennsylvania, T., & Park, U. (1990). *Basic Local Alignment Search Tool 2Department of Computer Science*. 403–410.

4. Alzohairy, A. M. (2014). *همدقتملا هيويجلا هيتامولعملا) 8 (هددعتملا تاعباتتلا نزاوت ءانب) 3 (ويڤ لاچلا مادختسأب نيولتلاو ليدعتلا Building a Multiple Sequence Alignment (8) Jalview Basics in More details بريهزلا روصنم دمحأ / د. November*. https://doi.org/10.13140/2.1.3365.8561

5. Aneja, B., Kumar, B., Jairajpuri, A., & Abid, M. (2016). *RSC Advances A structure guided drug-discovery approach towards identi fi cation of Plasmodium inhibitors*. *3*, 18364–18406. https://doi.org/10.1039/c5ra19673f

6. Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., & Noble, W. S. (2009). *MEME S UITE: tools for motif discovery and searching*. *37*(May), 202–208. https://doi.org/10.1093/nar/gkp335

7. Bailey, T. L., Gribskov, M., Diego, S., Box, P. O., & Diego, S. (1998). *Combining evidence using p-values: application to sequence homology searches*. *14*(1), 48–54.

8. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). *The Protein Data Bank*. *28*(1), 235–242.

9. Bradley, P., Misura, K. M. S., & Baker, D. (2005). Biochemistry: Toward high-resolution de novo structure prediction for small proteins. *Science*, *309*(5742), 1868–1871. https://doi.org/10.1126/science.1113801

10. Cavalli, A. (2018). *BiKi Life Sciences: A New Suite for Molecular Dynamics and Related Methods in Drug Discovery*. https://doi.org/10.1021/acs.jcim.7b00680

11. Consortium, T. U. (2019). *UniProt : a worldwide hub of protein knowledge*. *47*(November 2018), 506–515. https://doi.org/10.1093/nar/gky1049

12. Edgar, R. C., Drive, R. M., & Valley, M. (2004). *MUSCLE: multiple sequence alignment with high accuracy and high throughput*. *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

13. Eisenberg, D. (1997). *VERIFY3D: Assessment of Protein Models with Three-Dimensional Profiles*. *6879*(February). https://doi.org/10.1016/S0076-6879(97)77022-8

14. Elguindy, M. M., & Nakamaru-ogiso, E. (2015). *Apoptosis-inducing Factor (AIF) and Its Family Member Protein, AMID, Are Rotenone-sensitive NADH: Ubiquinone Oxidoreductases (NDH-2) *. *290*(34), 20815–20826. https://doi.org/10.1074/jbc.M115.641498

15. Elokely, K. M., & Doerksen, R. J. (2013). Docking challenge: Protein sampling and molecular docking performance. *Journal of Chemical Information and Modeling*, *53*(8), 1934–1945. https://doi.org/10.1021/ci400040d

16. Eswar, N., Webb, B., Marti-renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M., Pieper, U., & Sali, A. (2014). *NIH Public Access*.

https://doi.org/10.1002/0471250953.bi0506s15.Comparative

17. Fiser, A. (2004). Protein structure modeling in the proteomics era. *Expert Review of Proteomics*, *1*(1), 97–110. https://doi.org/10.1586/14789450.1.1.97

18. Fitch, W. M., & Margoliash, E. (1967). Construction of phylogenetic trees. *Science*, *155*(760), 279–284. https://doi.org/10.1126/science.155.3760.279

19. Gotoh, O. (1995). A weighting system and aigorithm for aligning many phylogenetically related sequences. *Bioinformatics*, *11*(5), 543–551. https://doi.org/10.1093/bioinformatics/11.5.543

20. Hashim, F. A., Mabrouk, M. S., & Al-atabany, W. (2019). *Review of Different Sequence Motif Finding Algorithms*. *11*(2).

21. Henikoff, S., & Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(22), 10915–10919. https://doi.org/10.1073/pnas.89.22.10915

22. Hirosawa, M., Totoki, Y., Hoshida, M., & Ishikawa, M. (1995). Comprehensive study on iterative algorithms of multiple sequence alignment. *Bioinformatics*, *11*(1), 13–18. https://doi.org/10.1093/bioinformatics/11.1.13

23. Hossain, M., Roy, P. K., Mosnaz, A. T. M. J., & Shakil, S. K. (2015). *Structural analysis and molecular docking of potential ligands with chorismate synthase of Listeria monocytogenes: A novel antibacterial drug target*. *52*(February), 45–59.

24. Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297–307. https://doi.org/10.1093/biomet/76.2.297

25. Katoh, K., Kuma, K., Toh, H., & Miyata, T. (2005). *MAFFT version 5: improvement in accuracy of multiple sequence alignment*. *33*(2), 511–518. https://doi.org/10.1093/nar/gki198

26. Katoh, K., Misawa, K., Kuma, K. I., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

27. Kellenberger, E., Rodrigo, J., Muller, P., & Rognan, D. (2004). Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins: Structure, Function and Genetics*, *57*(2), 225–242. https://doi.org/10.1002/prot.20149

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

29. Manuscript, A. (2016). *Europe PMC Funders Group Rational Methods for the Selection of Diverse Screening Compounds*. *6*(3), 208–217. https://doi.org/10.1021/cb100420r.Rational

30. Misura, K. M. S., Chivian, D., Rohl, C. A., Kim, D. E., & Baker, D. (2006). Physically realistic homology models built with ROSETTA can be more accurate than their templates. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(14), 5361–5366. https://doi.org/10.1073/pnas.0509355103

31. Pei, J., Tang, M., & Grishin, N. V. (2008). *PROMALS3D web server for accurate multiple protein sequence and structure alignments*. *36*(May), 30–34. https://doi.org/10.1093/nar/gkn322

32. Petrey, D., & Honig, B. (2005). Protein structure prediction: Inroads to biology. *Molecular*

Cell, *20*(6), 811–819. https://doi.org/10.1016/j.molcel.2005.12.005

33. Posada, D., & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic Biology*, *53*(5), 793–808. https://doi.org/10.1080/10635150490522304

34. Rentzsch, R., & Renard, B. Y. (2015). *Docking small peptides remains a great challenge: an assessment using AutoDock Vina*. *16*(November 2014), 1045–1056. https://doi.org/10.1093/bib/bbv008

35. Rose, P. W., Beran, B., Bi, C., Bluhm, W. F., Dimitropoulos, D., Goodsell, D. S., Prlić, A., Quesada, M., Quinn, G. B., Westbrook, J. D., Young, J., Yukich, B., Zardecki, C., Berman, H. M., & Bourne, P. E. (2011). The RCSB Protein Data Bank: Redesigned web site and web services. *Nucleic Acids Research*, *39*(SUPPL. 1), 392–401. https://doi.org/10.1093/nar/gkq1021

36. Sali, A. (2013). MODELLER: A Program for Protein Structure Modeling Release 9.12, r9480. *Rockefeller University*. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:MODELLER+A+Program+for+Protein+Structure+Modeling#6

37. Sandra, Á., Pavlopoulou, N., Adams, J., Gilsenan, J., & Tabernero, L. (2018). *VSpipe , an Integrated Resource for Virtual Screening and Hit Selection : Applications to Protein Tyrosine Phospahatase Inhibition*. 1–14. https://doi.org/10.3390/molecules23020353

38. Seeliger, D., Groot, B. L. De, & Pymol, V. (2010). *Ligand docking and binding site analysis with PyMOL and Autodock / Vina*. 417–422. https://doi.org/10.1007/s10822-010-9352-6

39. Shoemaker, J. S., & Fitch, W. M. (1989). Evidence from nuclear sequences that invariable sites should be considered when sequence divergence is calculated. *Molecular Biology and Evolution*, *6*(3), 270–289. https://doi.org/10.1093/oxfordjournals.molbev.a040550

40. Studio, D., Discovery, T., Visualizer, S., & Studio, D. (n.d.). *Introduction to the Discovery Studio Visualizer*.

41. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). *MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods Research resource*. *28*(10), 2731–2739. https://doi.org/10.1093/molbev/msr121

42. Vej, G. W., & Telephone, D. (2007). *Bioinformatics Explained Bioinformatics Explained*.

43. 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), 407–410. https://doi.org/10.1093/nar/gkm290

44. Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *Journal of Molecular Evolution*, *39*(3), 306–314. https://doi.org/10.1007/BF00160154

45. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I., & Lomize, A. L. (2012). OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Research*, *40*(D1), 370–376. https://doi.org/10.1093/nar/gkr703

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

47. Jo, S., Lim, J. B., Klauda, J. B., & Im, W. (2009). CHARMM-GUI membrane builder for mixed bilayers and its application to yeast membranes. *Biophysical Journal*, *97*(1), 50–58. https://doi.org/10.1016/j.bpj.2009.04.013

48. Woolf, T. B., & Roux, B. (1994). Molecular dynamics simulation of the gramicidin channel in a phospholipid bilayer. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(24), 11631–11635. https://doi.org/10.1073/pnas.91.24.11631

49. Woolf, Thomas B., & Roux, B. (1996). Structure, energetics, and dynamics of lipid-protein interactions: A molecular dynamics study of the gramicidin A channel in a DMPC bilayer. *Proteins: Structure, Function and Genetics*, *24*(1), 92–114. https://doi.org/10.1002/(SICI)1097-0134(199601)24:1<92: AID-PROT7>3.0.CO;2-Q

50. Wu, E. L., Cheng, X., Jo, S., Rui, H., Song, K. C., Dávila-Contreras, E. M., Qi, Y., Lee, J., Monje-Galvan, V., Venable, R. M., Klauda, J. B., & Im, W. (2014). CHARMM-GUI membrane builder toward realistic biological membrane simulations. *Journal of Computational Chemistry*, *35*(27), 1997–2004. https://doi.org/10.1002/jcc.23702

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

52. Feller, S. E., Zhang, Y., Pastor, R. W., & Brooks, B. R. (1995). Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics*, *103*(11), 4613–4621. https://doi.org/10.1063/1.470648

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

54. Kassir, Y., Kupiec, M., Shalom, A., & Simchen, G. (1985). Cloning and mapping of CDC40, a Saccharomyces cerevisiae gene with a role in DNA repair. *Current Genetics*, *9*(4), 253–257. https://doi.org/10.1007/BF00419952

55. Vermaas, J. V., Hardy, D. J., Stone, J. E., Tajkhorshid, E., & Kohlmeyer, A. (2016). TopoGROMACS: Automated Topology Conversion from CHARMM to GROMACS within VMD. *Journal of Chemical Information and Modeling*, *56*(6), 1112–1116. https://doi.org/10.1021/acs.jcim.6b00103

56. Ahmed, M. A., & Cox-Singh, J. (2015). Plasmodium knowlesi - an emerging pathogen. *ISBT Science Series*, *10*(S1), 134–140. https://doi.org/10.1111/voxs.12115

57. Alkadi, H. O. (2007). Antimalarial drug toxicity: A review. *Chemotherapy*, *53*(6), 385–391. https://doi.org/10.1159/000109767

58. Butler, A. R., Khan, S., & Ferguson, E. (2010). A brief history of malaria chemotherapy. *Journal of the Royal College of Physicians of Edinburgh*, *40*(2), 172–177. https://doi.org/10.4997/JRCPE.2010.216

59. Cerqueira, G. C., Cheeseman, I. H., Schaffner, S. F., Nair, S., Mcdew-white, M., Phyo, A. P., Ashley, E. A., Melnikov, A., Rogov, P., Birren, B. W., Nosten, F., Anderson, T. J. C., & Neafsey, D. E. (2017). *Longitudinal genomic surveillance of Plasmodium falciparum malaria parasites reveals complex genomic architecture of emerging artemisinin resistance*. 1–13. https://doi.org/10.1186/s13059-017-1204-4

60. Consortium, T. U. (2019). *UniProt : a worldwide hub of protein knowledge*. *47*(November 2018), 506–515. https://doi.org/10.1093/nar/gky1049

61. Cowman, A. F., Healer, J., Marapana, D., & Marsh, K. (2016). Review Malaria: Biology and Disease. *Cell*, *167*(3), 610–624. https://doi.org/10.1016/j.cell.2016.07.055

62. Dungworth, D. L., Kawakami, T. G., Munn, R. J., Ward, M., Harrold, J. B., Essex, M., Klein, G., Synder, S. P., Harrold, J. B., Essex, M., Synder, S. P., Jarrett, W., Mackey, L., Laird, H., Essex, M., Essex, M., Snyder, P., Klein, G., Dutcher, R. M., … Zuckerman, E. (1975). *ci ~~~~~~ C*. *190*, 792–794.

63. Eastman, R., & Fidock, D. (2009). ACTs: a vital tool in efforts to eliminate malaria. *Nature Reviews. Microbiology*, *7*(12), 864–874. https://doi.org/10.1038/nrmicro2239.Artemisinin-based

64. Edwards, G., McGrath, C., Ward, S., Supanaranond, W., Pukrittayakamee, S., Davis, T., & White, N. (1993). Interactions among primaquine, malaria infection and other antimalarials in Thai subjects. *British Journal of Clinical Pharmacology*, *35*(2), 193–198. https://doi.org/10.1111/j.1365-2125.1993.tb05685.x

65. Ezzet, F., & Karbwang, J. (1998). Population pharmacokinetics and therapeutic response of CGP 56697 (artemether + benflumetol) in malaria patients. *British Journal of Clinical Pharmacology*, *46*(6), 553–561. https://doi.org/10.1046/j.1365-2125.1998.00830.x

66. Foth, B. J., Ralph, S. A., Tonkin, C. J., Struck, N. S., Fraunholz, M., Roos, D. S., Cowman, A. F., & Mcfadden, G. I. (2003). *Dissecting Apicoplast Targeting in the Malaria Parasite Plasmodium falciparum*. *299*(January), 705–708.

67. Harbut, A. M. B., Liu, R., Yang, B., Yano, T., Vilcheze, C., Lockner, J., Guo, H., Scott, G., Petrassi, M., Jacobs, W. R., Rubin, H., Chatterjee, A. K., Wang, F., Harbut, M. B., Yang, B., Liu, R., Yano, T., Vilchèze, C., Cheng, B., … Wang, F. (n.d.). *Chemie*. https://doi.org/10.1002/anie.201800260

68. Kelm, S., Shi, J., & Deane, C. M. (2010). *MEDELLER: homology-based coordinate generation for membrane proteins*. *26*(22), 2833–2840. https://doi.org/10.1093/bioinformatics/btq554

69. Matz, J. M., Goosmann, C., Matuschewski, K., Matz, J. M., Goosmann, C., Matuschewski, K., Wilhelmus, T., & Kooij, A. (2018). An Unusual Prohibitin Regulates Malaria Parasite Mitochondrial Membrane Potential Article an Unusual Prohibitin Regulates Malaria Parasite Mitochondrial Membrane Potential. *CellReports*, *23*(3), 756–767. https://doi.org/10.1016/j.celrep.2018.03.088

70. Mita, T., Tanabe, K., & Kita, K. (2009). Spread and evolution of Plasmodium falciparum drug resistance. *Parasitology International*, *58*(3), 201–209. https://doi.org/10.1016/j.parint.2009.04.004

71. Naing, C., Whittaker, M. A., Nyunt Wai, V., & Mak, J. W. (2014). Is Plasmodium vivax Malaria a Severe Malaria? A Systematic Review and Meta-Analysis. *PLoS Neglected Tropical Diseases*, *8*(8). https://doi.org/10.1371/journal.pntd.0003071

72. Petersen, I., Eastman, R., & Lanzer, M. (2011). Drug-resistant malaria: Molecular mechanisms and implications for public health. *FEBS Letters*, *585*(11), 1551–1562. https://doi.org/10.1016/j.febslet.2011.04.042

73. Raynes, K. (1999). Bisquinoline antimalarials: Their role in malaria chemotherapy. *International Journal for Parasitology*, *29*(3), 367–379. https://doi.org/10.1016/S0020-7519(98)00217-3

74. Sá, J. M., Twu, O., Hayton, K., Reyes, S., Fay, M. P., Ringwald, P., & Wellems, T. E. (2009). Geographic patterns of Plasmodium falciparum drug resistance distinguished by differential

responses to amodiaquine and chloroquine. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(45), 18883–18889. https://doi.org/10.1073/pnas.0911317106

75. Saggu, G. S., Pala, Z. R., Garg, S., & Saxena, V. (2016). *New Insight into Isoprenoids Biosynthesis Process and Future Prospects for Drug Designing in Plasmodium*. *7*(September), 1–14. https://doi.org/10.3389/fmicb.2016.01421

76. Siregar, J. E., Kurisu, G., Kobayashi, T., Matsuzaki, M., & Sakamoto, K. (2015). Parasitology International Direct evidence for the atovaquone action on the Plasmodium cytochrome bc 1 complex. *Parasitology International*, *64*(3), 295–300. https://doi.org/10.1016/j.parint.2014.09.011

77. Smith, R. A., Cokkinides, V., & Eyre, H. J. (2004). *American Cancer Society Guidelines for the Early Detection of Cancer, 2004*. *54*(1), 41–52.

78. Srivastava, I. K., Rottenberg, H., & Vaidya, A. B. (1997). Atovaquone, a broad spectrum antiparasitic drug, collapses mitochondrial membrane potential in a malarial parasite. *Journal of Biological Chemistry*, *272*(7), 3961–3966. https://doi.org/10.1074/jbc.272.7.3961

79. Srivastava, I. K., & Vaidya, A. B. (1999). A mechanism for the synergistic antimalarial action of atovaquone and proguanil. *Antimicrobial Agents and Chemotherapy*, *43*(6), 1334–1339. https://doi.org/10.1128/aac.43.6.1334

80. Stepniewska, K., & White, N. J. (2008). Pharmacokinetic determinants of the window of selection for antimalarial drug resistance. *Antimicrobial Agents and Chemotherapy*, *52*(5), 1589–1596. https://doi.org/10.1128/AAC.00903-07

81. Tilley, L., Straimer, J., Gnädig, N. F., Ralph, S. A., & Fidock, D. A. (2016). Artemisinin Action and Resistance in Plasmodium falciparum. *Trends in Parasitology*, *32*(9), 682–696. https://doi.org/10.1016/j.pt.2016.05.010

82. Wells, T. N. C., Burrows, J. N., & Baird, J. K. (2010). Targeting the hypnozoite reservoir of Plasmodium vivax: the hidden obstacle to malaria elimination. *Trends in Parasitology*, *26*(3), 145–151. https://doi.org/10.1016/j.pt.2009.12.005

83. Cogswell, F. B. (1992). The hypnozoite and relapse in primate malaria. *Clinical Microbiology Reviews*, *5*(1), 26–35. https://doi.org/10.1128/cmr.5.1.26-35.1992

84. Bridges, H. R., Petri, J., Shimaki, Y., Jiao, W., Bridges, H. R., & Russell, E. R. (2018). Structure of the NDH-2 – HQNO inhibited complex provides molecular insight into quinone-binding site inhibitors BBA - Bioenergetics Structure of the NDH-2 – HQNO inhibited complex provides molecular insight into quinone-binding site inhibitors. *BBA - Bioenergetics*, *1859*(7), 482–490. https://doi.org/10.1016/j.bbabio.2018.03.014

85. Murugesan, D., Ray, P. C., Bayliss, T., Prosser, G. A., Harrison, J. R., Green, K., Soares De Melo, C., Feng, T. S., Street, L. J., Chibale, K., Warner, D. F., Mizrahi, V., Epemolu, O., Scullion, P., Ellis, L., Riley, J., Shishikura, Y., Ferguson, L., Osuna-Cabello, M., … Boshoff, H. I. M. (2018). 2-Mercapto-Quinazolinones as Inhibitors of Type II NADH Dehydrogenase and Mycobacterium tuberculosis: Structure-Activity Relationships, Mechanism of Action and Absorption, Distribution, Metabolism, and Excretion Characterization. *ACS Infectious Diseases*, *4*(6), 954–969. https://doi.org/10.1021/acsinfecdis.7b00275

86. Heikal, A., Nakatani, Y., Dunn, E., Weimar, M. R., Day, C. L., Baker, E. N., Lott, J. S., Sazanov, L. A., & Cook, G. M. (2014). *Structure of the bacterial type II NADH dehydrogenase : a monotopic membrane protein with an essential role in energy generation*. *91*(January), 950–

964. https://doi.org/10.1111/mmi.12507

87. Lencina, A. M., Koepke, J., Preu, J., Muenke, C., Gennis, R. B., Michel, H., & Schurig-Briccio, L. A. (2019). Characterization and X-ray structure of the NADH-dependent coenzyme A disulfide reductase from Thermus thermophilus. *Biochimica et Biophysica Acta - Bioenergetics*, *1860*(11), 148080. https://doi.org/10.1016/j.bbabio.2019.148080

88. Mori, T., Miyashita, N., Im, W., Feig, M., & Sugita, Y. (2016). Biochimica et Biophysica Acta Molecular dynamics simulations of biological membranes and membrane proteins using enhanced conformational sampling algorithms ☆. *BBA - Biomembranes*, *1858*(7), 1635–1651. https://doi.org/10.1016/j.bbamem.2015.12.032

89. Fagerberg, L., Jonasson, K., Von Heijne, G., Uhlén, M., & Berglund, L. (2010). Prediction of the human membrane proteome. *Proteomics*, *10*(6), 1141–1149. https://doi.org/10.1002/pmic.200900258

90. Goossens, K., & De Winter, H. (2018). Molecular Dynamics Simulations of Membrane Proteins: An Overview. *Journal of Chemical Information and Modeling*, *58*(11), 2193–2202. https://doi.org/10.1021/acs.jcim.8b00639

91. Kozma, D., Simon, I., & Tusnády, G. E. (2013). PDBTM: Protein data bank of transmembrane proteins after 8 years. *Nucleic Acids Research*, *41*(D1), 524–529. https://doi.org/10.1093/nar/gks1169

92. White, S. H. (2004). The progress of membrane protein structure determination. *Protein Science*, *13*(7), 1948–1949. https://doi.org/10.1110/ps.04712004

93. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindah, E. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001

94. Access, O. (2010). Therapeutic efficacy of Artemether / Lumefantrine (Coartem ®) against Plasmodium falciparum in. *Parasites & Vectors*, *3*(1), 1–9.

95. Additive, C., Field, F., Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., Jo, S., Pande, V. S., Case, D. A., Iii, C. L. B., Jr, A. D. M., Klauda, B., & Im, W. (2016). *CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field*. https://doi.org/10.1021/acs.jctc.5b00935

96. Adjei, G. O., Kurtzhals, J. A. L., Rodrigues, O. P., Alifrangis, M., Hoegberg, L. C. G., Kitcher, E. D., Badoe, E. V., Lamptey, R., & Goka, B. Q. (2008). Amodiaquine-artesunate vs artemether-lumefantrine for uncomplicated malaria in Ghanaian children: A randomized efficacy and safety trial with one year follow-up. *Malaria Journal*, *7*, 1–11. https://doi.org/10.1186/1475-2875-7-127

97. Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bollin, C., Bolton, E., Bourexis, D., Brister, J. R., Bryant, S. H., Canese, K., Cavanaugh, M., Charowhas, C., Clark, K., Dondoshansky, I., Feolo, M., Fitzpatrick, L., Funk, K., Geer, L. Y., Gorelenkov, V., … Zbicz, K. (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *46*(D1), D8–D13. https://doi.org/10.1093/nar/gkx1095

98. Aggarwal, R., & Koes, D. R. (2020). Learning RMSD to improve Protein-Ligand Scoring and Pose Selection. *ChemRxiv*, 1–8. https://doi.org/10.26434/chemrxiv.11910870

99. Aliyu, A. W., Mohd, K., Mustaffa, F., Chong, L., & Low, Y. (2021). *Chloroquine bioconjugates*

and hybrid compounds : past and recent developments in combatting chloroquine resistant malaria. *20*(December), 2663–2674.

100. Allouche, A. (2012). Software News and Updates Gabedit — A Graphical User Interface for Computational Chemistry Softwares. *Journal of Computational Chemistry*, *32*, 174–182. https://doi.org/10.1002/jcc

101. Altschup, S. F., Gish, W., Pennsylvania, T., & Park, U. (1990). *Basic Local Alignment Search Tool 2Department of Computer Science*. 403–410.

102. Alzohairy, A. M. (2014). [illegible] *) 8 (* [illegible] *) 3 (* [illegible] *Building a Multiple Sequence Alignment (8) Jalview Basics in More details* [illegible] /. *November*. https://doi.org/10.13140/2.1.3365.8561

103. Amir, A., Cheong, F. W., Ryan De Silva, J., Wee, J. K., & Lau, Y. L. (2018a). Infection and Drug Resistance Dovepress Plasmodium knowlesi malaria: current research perspectives. *Infection Drug Resistance*, 1145–1155. http://dx.doi.org/10.2147/IDR.S148664

104. Amir, A., Cheong, F. W., Ryan De Silva, J., Wee, J. K., & Lau, Y. L. (2018b). Infection and Drug Resistance Dovepress Plasmodium knowlesi malaria: current research perspectives. *Infection Drug Resistance*, 1145–1155.

105. Arnittali, M., Rissanou, A. N., & Harmandaris, V. (2019). Structure of Biomolecules Through Molecular Dynamics Simulations. *Procedia Computer Science*, *156*, 69–78. https://doi.org/10.1016/j.procs.2019.08.181

106. Arunan, E., Desiraju, G. R., Klein, R. A., Sadlej, J., Scheiner, S., Alkorta, I., Clary, D. C., Crabtree, R. H., Dannenberg, J. J., Hobza, P., Kjaergaard, H. G., Legon, A. C., Mennucci, B., & Nesbitt, D. J. (2011). Defining the hydrogen bond: An account (IUPAC Technical Report). *Pure and Applied Chemistry*, *83*(8), 1619–1636. https://doi.org/10.1351/PAC-REP-10-01-01

107. Ashley, E. A., Recht, J., & White, N. J. (2014). Primaquine: The risks and the benefits. *Malaria Journal*, *13*(1), 1–7. https://doi.org/10.1186/1475-2875-13-418

108. Autzen, H. E., Julius, D., & Cheng, Y. (2019). Membrane mimetic systems in CryoEM: keeping membrane proteins in their native environment. *Current Opinion in Structural Biology*, *58*, 259–268. https://doi.org/10.1016/j.sbi.2019.05.022

109. Bailey, T. L., Williams, N., Misleh, C., & Li, W. W. (2006). MEME: Discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research*, *34*(WEB. SERV. ISS.), 369–373. https://doi.org/10.1093/nar/gkl198

110. Baird, J. K., & Hoffman, S. L. (2004). Primaquine therapy for malaria. *Clinical Infectious Diseases*, *39*(9), 1336–1345. https://doi.org/10.1086/424663

111. Bakshi, R. P., Tatham, L. M., Savage, A. C., Tripathi, A. K., Mlambo, G., Ippolito, M. M., Nenortas, E., Rannard, S. P., Owen, A., & Shapiro, T. A. (2018). Long-acting injectable atovaquone nanomedicines for malaria prophylaxis. *Nature Communications*, *9*(1), 1–8. https://doi.org/10.1038/s41467-017-02603-z

112. Bano, D., & Prehn, J. H. M. (2018). Apoptosis-Inducing Factor (AIF) in Physiology and Disease: The Tale of a Repented Natural Born Killer. *EBioMedicine*, *30*, 29–37. https://doi.org/10.1016/j.ebiom.2018.03.016

113. Barber, B. E., Rajahram, G. S., Grigg, M. J., William, T., & Anstey, N. M. (2017). World Malaria Report: time to acknowledge Plasmodium knowlesi malaria. *Malaria Journal*, *16*(1),

13–15. https://doi.org/10.1186/s12936-017-1787-y

114. Barber, B. E., William, T., Grigg, M. J., Menon, J., Auburn, S., Marfurt, J., Anstey, N. M., & Yeo, T. W. (2013). A prospective comparative study of Knowlesi, falciparum, and vivax malaria in sabah, Malaysia: High proportion with severe disease from plasmodium knowlesi and plasmodium vivax but No mortality with early Referral and artesunate therapy. *Clinical Infectious Diseases*, *56*(3), 383–397. https://doi.org/10.1093/cid/cis902

115. Battle, K. E., Lucas, T. C. D., Nguyen, M., Howes, R. E., Nandi, A. K., Twohig, K. A., Pfeffer, D. A., Cameron, E., Rao, P. C., Casey, D., Gibson, H. S., Rozier, J. A., Dalrymple, U., Keddie, S. H., Collins, E. L., Harris, J. R., Guerra, C. A., Thorn, M. P., Bisanzio, D., … Gething, P. W. (2019). Mapping the global endemicity and clinical burden of Plasmodium vivax, 2000–17: a spatial and temporal modelling study. *The Lancet*, *394*(10195), 332–343. https://doi.org/10.1016/S0140-6736(19)31096-7

116. Bauer, C. A., Schneider, G., & Göller, A. H. (2019). Machine learning models for hydrogen bond donor and acceptor strengths using large and diverse training data generated by first-principles interaction free energies. *Journal of Cheminformatics*, *11*(1), 1–16. https://doi.org/10.1186/s13321-019-0381-4

117. Benson, N. C., & Daggett, V. (2012). A comparison of multiscale methods for the analysis of molecular dynamics simulations. *Journal of Physical Chemistry B*, *116*(29), 8722–8731. https://doi.org/10.1021/jp302103t

118. Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., Dinola, A., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, *81*(8), 3684–3690. https://doi.org/10.1063/1.448118

119. Berendsen, H. J. C., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, *91*(1–3), 43–56. https://doi.org/10.1016/0010-4655(95)00042-E

120. Bergman, M. E., Davis, B., & Phillips, M. A. (2019). *Occurrence , and Mechanism of Action*. 3961.

121. Bhatt, S., Weiss, D. J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K. E., Moyes, C. L., Henry, A., Eckhoff, P. A., Wenger, E. A., Briët, O., Penny, M. A., Smith, T. A., Bennett, A., Yukich, J., Eisele, T. P., Griffin, J. T., Fergus, C. A., … Gething, P. W. (2015). The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature*, *526*(7572), 207–211. https://doi.org/10.1038/nature15535

122. Biagini, G. A., Viriyavejakul, P., Neill, P. M. O., Bray, P. G., & Ward, S. A. (2006). *Functional Characterization and Target Validation of Alternative Complex I of Plasmodium falciparum Mitochondria*. *50*(5), 1841–1851. https://doi.org/10.1128/AAC.50.5.1841

123. Bietz, S., & Rarey, M. (2016). SIENA: Efficient Compilation of Selective Protein Binding Site Ensembles. *Journal of Chemical Information and Modeling*, *56*(1), 248–259. https://doi.org/10.1021/acs.jcim.5b00588

124. Biller, J. R., Elajaili, H., Meyer, V., Rosen, G. M., Eaton, S. S., & Eaton, G. R. (2013). The Amber biomolecular simulation programs. *Journal of MAgnetic Resonance*, *236*(16), 47–56.

125. Bingham, R. J., & Ballone, P. (2012). Computational study of room-temperature ionic liquids interacting with a POPC phospholipid bilayer. *Journal of Physical Chemistry B*, *116*(36), 11205–11216. https://doi.org/10.1021/jp306126q

126. Blaza, J. N., Bridges, H. R., Aragão, D., Dunn, E. A., Heikal, A., Cook, G. M., Nakatani, Y., & Hirst, J. (2017). The mechanism of catalysis by type-II NADH : quinone oxidoreductases. *Nature Publishing Group*, 1–11. https://doi.org/10.1038/srep40165

127. Boes, D. M., Godoy-Hernandez, A., & McMillan, D. G. G. (2021). Peripheral membrane proteins: Promising therapeutic targets across domains of life. *Membranes*, *11*(5). https://doi.org/10.3390/membranes11050346

128. Borges, B., Gallo, G., Coelho, C., Negri, N., Maiello, F., Hardy, L., & Würtele, M. (2021). Dynamic cross correlation analysis of Thermus thermophilus alkaline phosphatase and determinants of thermostability. *Biochimica et Biophysica Acta - General Subjects*, *1865*(7), 129895. https://doi.org/10.1016/j.bbagen.2021.129895

129. Bouth Denis, M., Davis, T. M. E., Hewitt, S., Incardona, S., Nimol, K., Fandeur, T., Poravuth, Y., Lim, C., & Socheat, D. (2002). Efficacy and safety of dihydroartemisinin-piperaquine (Artekin) in Cambodian children and adults with uncomplicated falciparum malaria. *Clinical Infectious Diseases*, *35*(12), 1469–1476. https://doi.org/10.1086/344647

130. Briones, R., Blau, C., Kutzner, C., de Groot, B. L., & Aponte-Santamaría, C. (2019). GROmaps: A GROMACS-Based Toolset to Analyze Density Maps Derived from Molecular Dynamics Simulations. *Biophysical Journal*, *116*(1), 4–11. https://doi.org/10.1016/j.bpj.2018.11.3126

131. 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

132. Brooks, Bernard R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, *4*(2), 187–217. https://doi.org/10.1002/jcc.540040211

133. Brosey, C. A., Ho, C., Long, W. Z., Singh, S., Burnett, K., Hura, G. L., Nix, J. C., Bowman, G. R., Ellenberger, T., & Tainer, J. A. (2016). Defining NADH-Driven Allostery Regulating Apoptosis-Inducing Factor. *Structure*, *24*(12), 2067–2079. https://doi.org/10.1016/j.str.2016.09.012

134. Brown, D. K., Penkler, D. L., Amamuddy, O. S., Ross, C., Atilgan, A. R., & Atilgan, C. (2017a). *Structural bioinformatics MD-TASK : a software suite for analyzing molecular dynamics trajectories*. *May*. https://doi.org/10.1093/bioinformatics/btx349

135. Brown, D. K., Penkler, D. L., Amamuddy, O. S., Ross, C., Atilgan, A. R., & Atilgan, C. (2017b). *Structural bioinformatics MD-TASK : a software suite for analyzing molecular dynamics trajectories*. *33*(May), 2768–2771. https://doi.org/10.1093/bioinformatics/btx349

136. Brown, D. K., & Tastan Bishop, Ö. (2017). Role of Structural Bioinformatics in Drug Discovery by Computational SNP Analysis: Analyzing Variation at the Protein Level. *Global Heart*, *12*(2), 151–161. https://doi.org/10.1016/j.gheart.2017.01.009

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

138. C, S., S, D. K., Ragunathan, V., Tiwari, P., A, S., & Brindha Devi, B. D. (2020). Molecular docking, validation, dynamics simulations, and pharmacokinetic prediction of

natural compounds against the SARS-CoV-2 main-protease. *Journal of Biomolecular Structure and Dynamics*, *0*(0), 1–27. https://doi.org/10.1080/07391102.2020.1815584

139. Cameron Ritchie, E., Block, J., & Lee Nevin, R. (2013). Psychiatric side effects of mefloquine: Applications to forensic psychiatry. *Journal of the American Academy of Psychiatry and the Law*, *41*(2), 224–235.

140. Campos, S. R. R., MacHuqueiro, M., & Baptista, A. M. (2010). Constant-pH molecular dynamics simulations reveal a β-rich form of the human prion protein. *Journal of Physical Chemistry B*, *114*(39), 12692–12700. https://doi.org/10.1021/jp104753t

141. Carrière, F., Longhi, S., & Record, M. (2020). The endosomal lipid bis(monoacylglycero) phosphate as a potential key player in the mechanism of action of chloroquine against SARS-COV-2 and other enveloped viruses hijacking the endocytic pathway. *Biochimie*, *179*, 237–246. https://doi.org/10.1016/j.biochi.2020.05.013

142. Carugo, O. (2018). How large B-factors can be in protein crystal structures. *BMC Bioinformatics*, *19*(1), 1–9. https://doi.org/10.1186/s12859-018-2083-8

143. Carugo, O., & Argos, P. (1998). Accessibility to internal cavities and ligand binding sites monitored by protein crystallographic thermal factors. *Proteins: Structure, Function and Genetics*, *31*(2), 201–213. https://doi.org/10.1002/(SICI)1097-0134(19980501)31:2<201::AID-PROT9>3.0.CO;2-O

144. Cavasotto, C. N., & Phatak, S. S. (2009). Homology modeling in drug discovery: current trends and applications. *Drug Discovery Today*, *14*(13–14), 676–683. https://doi.org/10.1016/j.drudis.2009.04.006

145. Chandrasekaran, B., Abed, S. N., Al-Attraqchi, O., Kuche, K., & Tekade, R. K. (2018a). Computer-Aided Prediction of Pharmacokinetic (ADMET) Properties. In *Dosage Form Design Parameters* (Vol. 2). Elsevier Inc. https://doi.org/10.1016/B978-0-12-814421-3.00021-X

146. Chandrasekaran, B., Abed, S. N., Al-Attraqchi, O., Kuche, K., & Tekade, R. K. (2018b). Computer-Aided Prediction of Pharmacokinetic (ADMET) Properties. In *Dosage Form Design Parameters* (Vol. 2). Elsevier Inc. https://doi.org/10.1016/B978-0-12-814421-3.00021-X

147. Chen, T., Li, M., & Liu, J. (2018). π-π Stacking Interaction: A Nondestructive and Facile Means in Material Engineering for Bioapplications. *Crystal Growth and Design*, *18*(5), 2765–2783. https://doi.org/10.1021/acs.cgd.7b01503

148. Cheng, H., Schaeffer, R. D., Liao, Y., Kinch, L. N., Pei, J., Shi, S., Kim, B. H., & Grishin, N. V. (2014). ECOD: An Evolutionary Classification of Protein Domains. *PLoS Computational Biology*, *10*(12). https://doi.org/10.1371/journal.pcbi.1003926

149. Coggill, P., Finn, R. D., & Bateman, A. (2008). Identifying protein domains with the Pfam database. *Current Protocols in Bioinformatics*, *SUPPL. 23*, 1–17. https://doi.org/10.1002/0471250953.bi0205s23

150. Collins, W. E., & Jeffery, G. M. (2005). Plasmodium ovale: Parasite and disease. *Clinical Microbiology Reviews*, *18*(3), 570–581. https://doi.org/10.1128/CMR.18.3.570-581.2005

151. Collins, W. E., & Jeffery, G. M. (2007). Plasmodium malariae: Parasite and disease. *Clinical Microbiology Reviews*, *20*(4), 579–592. https://doi.org/10.1128/CMR.00027-07

152. Corey, R. A., Stansfeld, P. J., & Sansom, M. S. P. (2020). The energetics of protein-lipid interactions as viewed by molecular simulations. *Biochemical Society Transactions*,

48(1), 25–37. https://doi.org/10.1042/BST20190149

153. Corradi, V., Mendez-Villuendas, E., Ingólfsson, H. I., Gu, R. X., Siuda, I., Melo, M. N., Moussatova, A., Degagné, L. J., Sejdiu, B. I., Singh, G., Wassenaar, T. A., Delgado Magnero, K., Marrink, S. J., & Tieleman, D. P. (2018). Lipid-Protein Interactions Are Unique Fingerprints for Membrane Proteins. *ACS Central Science*, *4*(6), 709–717. https://doi.org/10.1021/acscentsci.8b00143

154. Cournia, Z., Allen, B. K., Beuming, T., Pearlman, D. A., Radak, B. K., & Sherman, W. (2020). Rigorous free energy simulations in virtual screening. *Journal of Chemical Information and Modeling*, *60*(9), 4153–4169. https://doi.org/10.1021/acs.jcim.0c00116

155. Cox-Singh, J., Davis, T. M. E., Lee, K. S., Shamsul, S. S. G., Matusop, A., Ratnam, S., Rahman, H. A., Conway, D. J., & Singh, B. (2008). Plasmodium knowlesi malaria in humans is widely distributed and potentially life threatening. *Clinical Infectious Diseases*, *46*(2), 165–171. https://doi.org/10.1086/524888

156. Cytochrome, E. (2019). Tafenoquine versus Primaquine to Prevent Relapse of Plasmodium vivax Malaria . *New England Journal of Medicine*, *380*(19), 1875–1876. https://doi.org/10.1056/nejmc1902327

157. da Silva, T. U., Pougy, K. de C., Albuquerque, M. G., da Silva Lima, C. H., & Machado, S. de P. (2020a). Development of parameters compatible with the CHARMM36 force field for [Fe4S4]2+ clusters and molecular dynamics simulations of adenosine-5'-phosphosulfate reductase in GROMACS 2019. *Journal of Biomolecular Structure and Dynamics*, *0*(0), 1–11. https://doi.org/10.1080/07391102.2020.1847687

158. da Silva, T. U., Pougy, K. de C., Albuquerque, M. G., da Silva Lima, C. H., & Machado, S. de P. (2020b). Development of parameters compatible with the CHARMM36 force field for [Fe4S4]2+ clusters and molecular dynamics simulations of adenosine-5'-phosphosulfate reductase in GROMACS 2019. *Journal of Biomolecular Structure and Dynamics*, *0*(0), 1–11. https://doi.org/10.1080/07391102.2020.1847687

159. Dalal, A., & Atri, A. (2014). An Introduction to Sequence and Series. *International Journal of Research*, *1*(10), 1286–1292. https://doi.org/10.1002/0471250953.bi0301s42.An

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

161. Davis, R. L. (2020). Mechanism of Action and Target Identification: A Matter of Timing in Drug Discovery. *IScience*, *23*(9), 101487. https://doi.org/10.1016/j.isci.2020.101487

162. Delettre, C., Yuste, V. J., Moubarak, R. S., Bras, M., Lesbordes-Brion, J. C., Petres, S., Bellalou, J., & Susin, S. A. (2006). AIFsh, a novel apoptosis-inducing factor (AIF) pro-apoptotic isoform with potential pathological relevance in human cancer. *Journal of Biological Chemistry*, *281*(10), 6413–6427. https://doi.org/10.1074/jbc.M509884200

163. Di Marino, D., D'Annessa, I., Coletta, A., Via, A., & Tramontano, A. (2015). Characterization of the differences in the cyclopiazonic acid binding mode to mammalian and P. Falciparum Ca2+ pumps: A computational study. *Proteins: Structure, Function and Bioinformatics*, *83*(3), 564–574. https://doi.org/10.1002/prot.24734

164. Diallo, B. N., Glenister, M., Musyoka, T. M., Lobb, K., & Tastan Bishop, Ö. (2021). SANCDB: an update on South African natural compounds and their readily available analogs.

Journal of Cheminformatics, *13*(1), 1–14. https://doi.org/10.1186/s13321-021-00514-2

165. Ding, X. C., Ade, M. P., Baird, J. K., Cheng, Q., Cunningham, J., Dhorda, M., Drakeley, C., Felger, I., Gamboa, D., Harbers, M., Herrera, S., Lucchi, N., Mayor, A., Mueller, I., Sattabongkot, J., Ratsimbason, A., Richards, J., Tanner, M., & González, I. J. (2017). Defining the next generation of Plasmodium vivax diagnostic tests for control and elimination: Target product profiles. *PLoS Neglected Tropical Diseases*, *11*(4), 1–15. https://doi.org/10.1371/journal.pntd.0005516

166. Domański, J., Sansom, M. S. P., Stansfeld, P. J., & Best, R. B. (2018). Balancing Force Field Protein-Lipid Interactions to Capture Transmembrane Helix-Helix Association. *Journal of Chemical Theory and Computation*, *14*(3), 1706–1715. https://doi.org/10.1021/acs.jctc.7b00983

167. Douglas, N. M., Simpson, J. A., Phyo, A. P., Siswantoro, H., Hasugian, A. R., Kenangalem, E., Poespoprodjo, J. R., Singhasivanon, P., Anstey, N. M., White, N. J., Tjitra, E., Nosten, F., & Price, R. N. (2013). Gametocyte dynamics and the role of drugs in reducing the transmission potential of plasmodium vivax. *Journal of Infectious Diseases*, *208*(5), 801–812. https://doi.org/10.1093/infdis/jit261

168. Dreher, M., Piuzzi, M., Turki, A., Chavent, M., Baaden, M., Férey, N., Limet, S., Raffin, B., & Robert, S. (2013). Interactive molecular dynamics: Scaling up to large systems. *Procedia Computer Science*, *18*, 20–29. https://doi.org/10.1016/j.procs.2013.05.165

169. Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., Wang, L. P., Simmonett, A. C., Harrigan, M. P., Stern, C. D., Wiewiora, R. P., Brooks, B. R., & Pande, V. S. (2017). OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Computational Biology*, *13*(7), 1–17. https://doi.org/10.1371/journal.pcbi.1005659

170. Egwu, C. O., Tsamesidis, I., Pério, P., Augereau, J. M., Benoit-Vical, F., & Reybier, K. (2021). Superoxide: A major role in the mechanism of action of essential antimalarial drugs. *Free Radical Biology and Medicine*, *167*(March), 271–275. https://doi.org/10.1016/j.freeradbiomed.2021.03.001

171. Eisenberg, D. (1997). *VERIFY3D : Assessment of Protein Models with Three-Dimensional Profiles*. *6879*(February). https://doi.org/10.1016/S0076-6879(97)77022-8

172. El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., & Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, *47*(D1), D427–D432. https://doi.org/10.1093/nar/gky995

173. Engel, A., & Gaub, H. E. (2008). Structure and mechanics of membrane proteins. *Annual Review of Biochemistry*, *77*, 127–148. https://doi.org/10.1146/annurev.biochem.77.062706.154450

174. Er, M., Abounakhla, A. M., Tahtaci, H., Bawah, A. H., Çinaroğlu, S. S., Onaran, A., & Ece, A. (2018). An integrated approach towards the development of novel antifungal agents containing thiadiazole: Synthesis and a combined similarity search, homology modelling, molecular dynamics and molecular docking study. *Chemistry Central Journal*, *12*(1), 1–21. https://doi.org/10.1186/s13065-018-0485-3

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

176. Eswar, N., John, B., Mirkovic, N., Fiser, A., Ilyin, V. A., Pieper, U., Stuart, A. C., Marti-Renom, M. A., Madhusudhan, M. S., Yerkovich, B., & Sali, A. (2003). Tools for comparative protein structure modeling and analysis. *Nucleic Acids Research*, *31*(13), 3375–3380. https://doi.org/10.1093/nar/gkg543

177. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M., Eramian, D., Shen, M., Pieper, U., & Sali, A. (2006). Comparative Protein Structure Modeling Using Modeller Modeling Structure from Sequence 5.6.1. *Current Protocols in Bioinformatics*, 1–30. http://www.currentprotocols.com.

178. Evers, A., Gohlke, H., & Klebe, G. (2003). Ligand-supported homology modelling of protein binding-sites using knowledge-based potentials. *Journal of Molecular Biology*, *334*(2), 327–345. https://doi.org/10.1016/j.jmb.2003.09.032

179. Fanelli, F., & De Benedetti, P. G. (2011). Update 1 of: Computational modeling approaches to structure - Function analysis of G protein-coupled receptors. *Chemical Reviews*, *111*(12). https://doi.org/10.1021/cr100437t

180. Fang, J., & Beattie, D. S. (2002). Rotenone-insensitive NADH dehydrogenase is a potential source of superoxide in procyclic Trypanosoma brucei mitochondria. *Molecular and Biochemical Parasitology*, *123*(2), 135–142. https://doi.org/10.1016/S0166-6851(02)00139-1

181. Feachem, R. G. A., Phillips, A. A., Hwang, J., Cotter, C., Wielgosz, B., Greenwood, B. M., Sabot, O., Rodriguez, M. H., Abeyasinghe, R. R., Ghebreyesus, T. A., & Snow, R. W. (2010). Shrinking the malaria map: Progress and prospects. *The Lancet*, *376*(9752), 1566–1578. https://doi.org/10.1016/S0140-6736(10)61270-6

182. Feng, Y., Li, W., Li, J., Wang, J., Ge, J., Xu, D., Liu, Y., Wu, K., Zeng, Q., Wu, J. W., Tian, C., Zhou, B., & Yang, M. (2012). Structural insight into the type-II mitochondrial NADH dehydrogenases. *Nature*, *491*(7424), 478–482. https://doi.org/10.1038/nature11541

183. Fischer, E. (1894). 5.70, N 4.20 Gef. *Ber. Dtsch. Chem. Ges.*, *27*, 2985–2993.

184. Fiser, A. (n.d.). *Chapter 6*. *673*(3). https://doi.org/10.1007/978-1-60761-842-3

185. Fisette, O., Päslack, C., Barnes, R., Isas, J. M., Langen, R., Heyden, M., Han, S., & Schäfer, L. V. (2016). Hydration Dynamics of a Peripheral Membrane Protein. *Journal of the American Chemical Society*, *138*(36), 11526–11535. https://doi.org/10.1021/jacs.6b07005

186. Fox, N. K., Brenner, S. E., & Chandonia, J. M. (2014). SCOPe: Structural Classification of Proteins - Extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Research*, *42*(D1), 304–309. https://doi.org/10.1093/nar/gkt1240

187. Frauenfelder, H., & McMahon, B. (1998). Dynamics and function of proteins: The search for general concepts. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(9), 4795–4797. https://doi.org/10.1073/pnas.95.9.4795

188. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., & Shenkin, P. S. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry*, *47*(7), 1739–1749. https://doi.org/10.1021/jm0306430

189. Gabler, F., Nam, S. Z., Till, S., Mirdita, M., Steinegger, M., Söding, J., Lupas, A. N., & Alva, V. (2020). Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Current Protocols in Bioinformatics*, *72*(1), 1–30. https://doi.org/10.1002/cpbi.108

190. Gebhardt, J., Kleist, C., Jakobtorweihen, S., & Hansen, N. (2018). Validation and Comparison of Force Fields for Native Cyclodextrins in Aqueous Solution. *Journal of Physical Chemistry B*, *122*(5), 1608–1626. https://doi.org/10.1021/acs.jpcb.7b11808

191. Gething, P. W., Patil, A. P., Smith, D. L., Guerra, C. A., Elyazar, I. R. F., Johnston, G. L., Tatem, A. J., & Hay, S. I. (2011). A new world malaria map: Plasmodium falciparum endemicity in 2010. *Malaria Journal*, *10*, 1–16. https://doi.org/10.1186/1475-2875-10-378

192. Ghosh, D. K., Kumar, A., & Ranjan, A. (2021). Cellular targets of mefloquine. *Toxicology*, *464*(October), 152995. https://doi.org/10.1016/j.tox.2021.152995

193. Gibney, G., & Baxevanis, A. D. (2011). Searching NCBI databases using Entrez. *Current Protocols in Bioinformatics*, *SUPPL. 34*, 1–25. https://doi.org/10.1002/0471250953.bi0103s34

194. Giese, M., & Albrecht, M. (2020). Alkyl-Alkyl Interactions in the Periphery of Supramolecular Entities: From the Evaluation of Weak Forces to Applications. *ChemPlusChem*, *85*(4), 715–724. https://doi.org/10.1002/cplu.202000077

195. Gimeno, A., Ojeda-Montes, M. J., Tomás-Hernández, S., Cereto-Massagué, A., Beltrán-Debón, R., Mulero, M., Pujadas, G., & Garcia-Vallvé, S. (2019). The light and dark sides of virtual screening: What is there to know? *International Journal of Molecular Sciences*, *20*(6). https://doi.org/10.3390/ijms20061375

196. Gisselmann, G., Alisch, D., Welbers-Joop, B., & Hatt, H. (2018). Effects of quinine, quinidine and chloroquine on human muscle nicotinic acetylcholine receptors. *Frontiers in Pharmacology*, *9*(NOV), 1–7. https://doi.org/10.3389/fphar.2018.01339

197. Gleeson, M. P., Hersey, A., Montanari, D., & Overington, J. (2011). Probing the links between in vitro potency, ADMET and physicochemical parameters. *Nature Reviews Drug Discovery*, *10*(3), 197–208. https://doi.org/10.1038/nrd3367

198. Gordiz, K., Singh, D. J., & Henry, A. (2015). Ensemble averaging vs. time averaging in molecular dynamics simulations of thermal conductivity. *Journal of Applied Physics*, *117*(4). https://doi.org/10.1063/1.4906957

199. Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: Scanning for occurrences of a given motif. *Bioinformatics*, *27*(7), 1017–1018. https://doi.org/10.1093/bioinformatics/btr064

200. Gregori-Puigjané, E., Setola, V., Hert, J., Crews, B. A., Irwin, J. J., Lounkine, E., Marnett, L., Roth, B. L., & Shoichet, B. K. (2012). Identifying mechanism-of-action targets for drugs and probes. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(28), 11178–11183. https://doi.org/10.1073/pnas.1204524109

201. Grigg, M. J., William, T., Barber, B. E., Rajahram, G. S., Menon, J., Schimann, E., Piera, K., Wilkes, C. S., Patel, K., Chandna, A., Drakeley, C. J., Yeo, T. W., & Anstey, N. M. (2018). Age-related clinical spectrum of plasmodium knowlesi malaria and predictors of severity. *Clinical Infectious Diseases*, *67*(3), 350–359. https://doi.org/10.1093/cid/ciy065

202. Gutberlet, K., & Rudolph, R. (1996). Angiosis carcinomatosa bei Mammatumoren der Hündin - Häufigkeit und Verbindung mit prognostisch wichtigen Faktoren. *Kleintierpraxis*, *41*(7), 473–482.

203. Guterres, H., Lee, H. S., & Im, W. (2019). Ligand-Binding-Site Structure Refinement Using Molecular Dynamics with Restraints Derived from Predicted Binding Site Templates. *Journal of Chemical Theory and Computation*, *15*(11), 6524–6535.

https://doi.org/10.1021/acs.jctc.9b00751

204. Guterres, H., Park, S. J., Zhang, H., & Im, W. (2021). CHARMM-GUI LBS Finder & Refiner for Ligand Binding Site Prediction and Refinement. *Journal of Chemical Information and Modeling*, *61*(8), 3744–3751. https://doi.org/10.1021/acs.jcim.1c00561

205. Guvench, O., Greenr, S. N., Kamath, G., Brady, J. W., Venable, R. M., Pastor, R. W., & Mackerell, A. D. (2008). Additive empirical force field for hexopyranose monosaccharides. *Journal of Computational Chemistry*, *29*(15), 2543–2564. https://doi.org/10.1002/jcc.21004

206. Hammerstad, M., & Hersleth, H. P. (2021). Overview of structurally homologous flavoprotein oxidoreductases containing the low Mr thioredoxin reductase-like fold – A functionally diverse group. *Archives of Biochemistry and Biophysics*, *702*(February), 108826. https://doi.org/10.1016/j.abb.2021.108826

207. 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

208. 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

209. Hatherley, R., Brown, D. K., Musyoka, T. M., Penkler, D. L., Faya, N., Lobb, K. A., & Bishop, Ö. T. (2015). SANCDB : a South African natural compound database. *Journal of Cheminformatics*, 1–9. https://doi.org/10.1186/s13321-015-0080-8

210. Hatherley, R., Brown, D. K., Musyoka, T. M., Penkler, D. L., Faya, N., Lobb, K. A., & Tastan Bishop, Ö. (2015). SANCDB: A South African natural compound database. *Journal of Cheminformatics*, *7*(1). https://doi.org/10.1186/s13321-015-0080-8

211. Hess, B., Bekker, H., Berendsen, H. J. C., & Fraaije, J. G. E. M. (1997). LINCS: A Linear Constraint Solver for molecular simulations. *Journal of Computational Chemistry*, *18*(12), 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

212. Hill, D. R., Baird, J. K., Parise, M. E., Lewis, L. S., Ryan, E. T., & Magill, A. J. (2006). Primaquine: Report from CDC expert meeting on malaria chemoprophylaxis I. *American Journal of Tropical Medicine and Hygiene*, *75*(3), 402–415. https://doi.org/10.4269/ajtmh.2006.75.402

213. Hirasawa, T., Kotani, S., Suzuki, T., Sato, K., Sakakibara, M., & Tokimasa, T. (2000). Effects of lanthanides on voltage-dependent potassium currents in bullfrog sympathetic neurons. *Neuroscience Letters*, *290*(2), 97–100. https://doi.org/10.1016/S0304-3940(00)01335-5

214. Hirosawa, M., Totoki, Y., Hoshida, M., & Ishikawa, M. (1995). Comprehensive study on iterative algorithms of multiple sequence alignment. *Bioinformatics*, *11*(1), 13–18. https://doi.org/10.1093/bioinformatics/11.1.13

215. Hoi, K. K., Bada Juarez, J. F., Judge, P. J., Yen, H. Y., Wu, D., Vinals, J., Taylor, G. F., Watts, A., & Robinson, C. V. (2021). Detergent-free Lipodisq Nanoparticles Facilitate High-Resolution Mass Spectrometry of Folded Integral Membrane Proteins. *Nano Letters*, *21*(7), 2824–2831. https://doi.org/10.1021/acs.nanolett.0c04911

216. 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

217. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., & Simmerling, C. (2006). Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function and Genetics*, *65*(3), 712–725. https://doi.org/10.1002/prot.21123

218. Howes, R. E., Piel, F. B., Patil, A. P., Nyangiri, O. A., Gething, P. W., Dewi, M., Hogg, M. M., Battle, K. E., Padilla, C. D., Baird, J. K., & Hay, S. I. (2012). G6PD Deficiency Prevalence and Estimates of Affected Populations in Malaria Endemic Countries: A Geostatistical Model-Based Map. *PLoS Medicine*, *9*(11). https://doi.org/10.1371/journal.pmed.1001339

219. Huang, S. Y., & Zou, X. (2010). Advances and challenges in Protein-ligand docking. *International Journal of Molecular Sciences*, *11*(8), 3016–3034. https://doi.org/10.3390/ijms11083016

220. Humphrey, W., Dalke, A., & Schulten, K. (1996). Sartorius products. *Journal of Molecular Graphics*, *14*(October 1995), 33–38. https://www.tapbiosystems.com/tap/products/index.htm

221. Hyde, J. E. (2002). Parasite resistance. *Journal of Equine Veterinary Science*, *22*(8), 351.

222. Hyochol Ahn, PhD, Michael Weaver, PhD, Debra Lyon, PhD, Eunyoung Choi, RN, and Roger B. Fillingim, P. (2017). 乳鼠心肌提取 HHS Public Access. *Physiology & Behavior*, *176*(10), 139–148. https://doi.org/10.1016/j.physbeh.2017.03.040

223. Issa, N. T., Badiavas, E. V., & Schürer, S. (2019). Research Techniques Made Simple: Molecular Docking in Dermatology - A Foray into In Silico Drug Discovery. *Journal of Investigative Dermatology*, *139*(12), 2400-2408.e1. https://doi.org/10.1016/j.jid.2019.06.129

224. Itoh, Y., Nakashima, Y., Tsukamoto, S., Kurohara, T., Suzuki, M., Sakae, Y., Oda, M., Okamoto, Y., & Suzuki, T. (2019). N+-C-H···O Hydrogen bonds in protein-ligand complexes. *Scientific Reports*, *9*(1), 1–5. https://doi.org/10.1038/s41598-018-36987-9

225. Iwata, M., Lee, Y., Yamashita, T., Yagi, T., & Iwata, S. (2012). *The structure of the yeast NADH dehydrogenase (Ndi1) reveals overlapping binding sites for water- and lipid-soluble substrates*. *109*(38). https://doi.org/10.1073/pnas.1210059109

226. Jain, A. N. (2003). Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *Journal of Medicinal Chemistry*, *46*(4), 499–511. https://doi.org/10.1021/jm020406h

227. Jiang, X. Z., Feng, M., Luo, K. H., & Ventikos, Y. (2018). Large-scale molecular dynamics simulation of flow under complex structure of endothelial glycocalyx. *Computers and Fluids*, *173*, 140–146. https://doi.org/10.1016/j.compfluid.2018.03.014

228. Jo, S., Lim, J. B., Klauda, J. B., & Im, W. (2009). CHARMM-GUI membrane builder for mixed bilayers and its application to yeast membranes. *Biophysical Journal*, *97*(1), 50–58. https://doi.org/10.1016/j.bpj.2009.04.013

229. Jollife, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065). https://doi.org/10.1098/rsta.2015.0202

230. Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, *267*(3), 727–748. https://doi.org/10.1006/jmbi.1996.0897

231. Jongwutiwes, S., Putaporntip, C., Iwasaki, T., Sata, T., & Kanbara, H. (2004). Naturally acquired Plasmodium knowlesi malaria in human, Thailand. *Emerging Infectious Diseases*, *10*(12), 2211–2213. https://doi.org/10.3201/eid1012.040293

232. Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, *118*(45), 11225–11236. https://doi.org/10.1021/ja9621760

233. Jung, J., Mori, T., Kobayashi, C., Matsunaga, Y., Yoda, T., Feig, M., & Sugita, Y. (2015). GENESIS: A hybrid-parallel and multi-scale molecular dynamics simulator with enhanced sampling algorithms for biomolecular and cellular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *5*(4), 310–323. https://doi.org/10.1002/wcms.1220

234. Kadam, A., Mehta, D., Jubin, T., Mansuri, M. S., & Begum, R. (2020). Apoptosis inducing factor: Cellular protective function in Dictyostelium discoideum. *Biochimica et Biophysica Acta - Bioenergetics*, *1861*(5–6), 148158. https://doi.org/10.1016/j.bbabio.2020.148158

235. Kalman, M., & Ben-Tal, N. (2010). Quality assessment of protein model-structures using evolutionary conservation. *Bioinformatics*, *26*(10), 1299–1307. https://doi.org/10.1093/bioinformatics/btq114

236. Karle, J. M., Karle, I. L., Gerena, L., & Milhous, W. K. (1992). Stereochemical evaluation of the relative activities of the cinchona alkaloids against Plasmodium falciparum. *Antimicrobial Agents and Chemotherapy*, *36*(7), 1538–1544. https://doi.org/10.1128/AAC.36.7.1538

237. Karplus, M., & Ichiye, T. (1995). Comment on a “Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations.” *Journal of Molecular Biology*, *252*(4), 492–503. https://doi.org/10.1006/jmbi.1995.0514

238. Kasahara, K., Fukuda, I., & Nakamura, H. (2014). A novel approach of dynamic cross correlation analysis on molecular dynamics simulations and its application to Ets1 dimer-DNA complex. *PLoS ONE*, *9*(11). https://doi.org/10.1371/journal.pone.0112419

239. Kerscher, S. J. (2000). Diversity and origin of alternative NADH:ubiquinone oxidoreductases. *Biochimica et Biophysica Acta - Bioenergetics*, *1459*(2–3), 274–283. https://doi.org/10.1016/S0005-2728(00)00162-6

240. Kevin Baird, J. (2013). Evidence and implications of mortality associated with acute plasmodium vivax malaria. *Clinical Microbiology Reviews*, *26*(1), 36–57. https://doi.org/10.1128/CMR.00074-12

241. Kim, W., Chen, B., Kim, J., Pan, Y., & Park, H. (2011). Sparse nonnegative matrix factorization for protein sequence motif discovery. *Expert Systems with Applications*, *38*(10), 13198–13207. https://doi.org/10.1016/j.eswa.2011.04.133

242. Klim, J., Gladki, A., Kucharczyk, R., Zielenkiewicz, U., & Kaczanowski, S. (2018). Ancestral state reconstruction of the apoptosis machinery in the common ancestor of eukaryotes. *G3: Genes, Genomes, Genetics*, *8*(6), 2121–2134. https://doi.org/10.1534/g3.118.200295

243. Kotepui, M., Kotepui, K. U., Milanez, G. D., & Masangkay, F. R. (2020). Severity and mortality of severe plasmodium ovale infection: A systematic review and meta-analysis. *PLoS*

ONE, *15*(6). https://doi.org/10.1371/journal.pone.0235014

244. Kotepui, M., Masangkay, F. R., Kotepui, K. U., & De Jesus Milanez, G. (2020). Misidentification of Plasmodium ovale as Plasmodium vivax malaria by a microscopic method: a meta-analysis of confirmed P. ovale cases. *Scientific Reports*, *10*(1), 1–13. https://doi.org/10.1038/s41598-020-78691-7

245. Krishna, S., & White, N. J. (1996). Pharmacokinetics of Quinine, Chloroquine and Amodiaquine. *Clinical Pharmacokinetics*, *30*(4), 263–299.

246. Kufareva, I., & Abagyan, R. (2012). Methods of protein structure comparison. *Methods in Molecular Biology*, *857*, 231–257. https://doi.org/10.1007/978-1-61779-588-6_10

247. Kuhlman, B., & Bradley, P. (2019). Advances in protein structure prediction and design. *Nature Reviews Molecular Cell Biology*, *20*(11), 681–697. https://doi.org/10.1038/s41580-019-0163-x

248. Kulig, W., Pasenkiewicz-Gierula, M., & Róg, T. (2015). Topologies, structures and parameter files for lipid simulations in GROMACS with the OPLS-aa force field: DPPC, POPC, DOPC, PEPC, and cholesterol. *Data in Brief*, *5*, 333–336. https://doi.org/10.1016/j.dib.2015.09.013

249. Kumar, A. H. (2020). Pharmacology of Chloroquine: Potential Mechanism of Action against Coronavirus. *Biology, Engineering, Medicine and Science Reports*, *6*(1), 9–10. https://doi.org/10.5530/bems.6.1.3

250. Kumar, S., Stecher, G., & Tamura, K. (2016). *MEGA7 : Molecular Evolutionary Genetics Analysis Version 7 . 0 for Bigger Datasets Brief communication*. *33*(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

251. Kutateladze, T., & Overduin, M. (2001). Structural mechanism of endosome docking by the FYVE domain. *Science*, *291*(5509), 1793–1796. https://doi.org/10.1126/science.291.5509.1793

252. Ladunga, I. (2017). Finding homologs in amino acid sequences using network blast searches. *Current Protocols in Bioinformatics*, *2017*, 3.4.1-3.4.24. https://doi.org/10.1002/cpbi.34

253. Lalève, A., Vallières, C., Golinelli-Cohen, M. P., Bouton, C., Song, Z., Pawlik, G., Tindall, S. M., Avery, S. V., Clain, J., & Meunier, B. (2016). The antimalarial drug primaquine targets Fe-S cluster proteins and yeast respiratory growth. *Redox Biology*, *7*, 21–29. https://doi.org/10.1016/j.redox.2015.10.008

254. Lamoureux, G., & Roux, B. (2003). Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *Journal of Chemical Physics*, *119*(6), 3025–3039. https://doi.org/10.1063/1.1589749

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

256. Laury, M. L., Wang, L. P., Pande, V. S., Head-Gordon, T., & Ponder, J. W. (2015). Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *Journal of Physical Chemistry B*, *119*(29), 9423–9437. https://doi.org/10.1021/jp510896n

257. Lee, A. G. (2018). A Database of Predicted Binding Sites for Cholesterol on

Membrane Proteins, Deep in the Membrane. *Biophysical Journal*, *115*(3), 522–532. https://doi.org/10.1016/j.bpj.2018.06.022

258. Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., Jo, S., Pande, V. S., Case, D. A., Brooks, C. L., MacKerell, A. D., Klauda, J. B., & Im, W. (2016). CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation*, *12*(1), 405–413. https://doi.org/10.1021/acs.jctc.5b00935

259. Lee, J., Patel, D. S., Ståhle, J., Park, S. J., Kern, N. R., Kim, S., Lee, J., Cheng, X., Valvano, M. A., Holst, O., Knirel, Y. A., Qi, Y., Jo, S., Klauda, J. B., Widmalm, G., & Im, W. (2019). CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *Journal of Chemical Theory and Computation*, *15*(1), 775–786. https://doi.org/10.1021/acs.jctc.8b01066

260. Leffler, E. M., Band, G., Busby, G. B. J., Kivinen, K., Le, Q. S., Clarke, G. M., Bojang, K. A., Conway, D. J., Jallow, M., Sisay-joof, F., Bougouma, E. C., Mangano, V. D., Modiano, D., Sirima, S. B., Achidi, E., Apinjoh, T. O., Marsh, K., Ndila, C. M., Peshu, N., … Spencer, C. C. A. (2017). *Resistance to malaria through structural variation of red blood cell invasion receptors*. *1139*(June). https://doi.org/10.1126/science.aam6393

261. Lei, T., Hu, Z., Ding, R., Chen, J., Li, S., Zhang, F., Pu, X., & Zhao, N. (2020). Exploring the Activation Mechanism of a Metabotropic Glutamate Receptor Homodimer via Molecular Dynamics Simulation. *ACS Chemical Neuroscience*, *11*(2), 133–145. https://doi.org/10.1021/acschemneuro.9b00425

262. Lemkul, J. A. (2019). *From Proteins to Perturbed Hamiltonians : A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1 . 0]*. *1*(July 2018), 1–53.

263. Letkovska, K., Babal, P., Cierna, Z., Schmidtova, S., Liskova, V., Kalavska, K., Miskovska, V., Horak, S., Rejlekova, K., Chovanec, M., Mardiak, J., Janega, P., & Mego, M. (2021). Prognostic value of apoptosis-inducing factor (AIF) in germ cell tumors. *Cancers*, *13*(4), 1–17. https://doi.org/10.3390/cancers13040776

264. 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

265. Lichtenthaler, S. F., Lemberg, M. K., & Fluhrer, R. (2018). Proteolytic ectodomain shedding of membrane proteins in mammals—hardware, concepts, and recent developments. *The EMBO Journal*, *37*(15), 1–24. https://doi.org/10.15252/embj.201899456

266. Lienhart, W. D., Gudipati, V., & MacHeroux, P. (2013). The human flavoproteome. *Archives of Biochemistry and Biophysics*, *535*(2), 150–162. https://doi.org/10.1016/j.abb.2013.02.015

267. Liguori, N., Croce, R., Marrink, S. J., & Thallmair, S. (2020). Molecular dynamics simulations in photosynthesis. *Photosynthesis Research*, *144*(2), 273–295. https://doi.org/10.1007/s11120-020-00741-y

268. Lindahl, Abraham, Hess, & Spoel, van der. (2021). *GROMACS Documentation*. http://manual.gromacs.org/documentation/5.1.4/%0Ahttp://files/1243/5.1.html%0Ahttp://manual.gromacs.org/2019-current/manual-2019.6.pdf

269. 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

270. Liu, X., Shi, D., Zhou, S., Liu, H., Liu, H., & Yao, X. (2018). Molecular dynamics simulations and novel drug discovery. *Expert Opinion on Drug Discovery*, *13*(1), 23–37. https://doi.org/10.1080/17460441.2018.1403419

271. Lomize, A. L., Pogozheva, I. D., & Mosberg, H. I. (2011). Anisotropic solvent model of the lipid bilayer. 2. Energetics of insertion of small molecules, peptides, and proteins in membranes. *Journal of Chemical Information and Modeling*, *51*(4), 930–946. https://doi.org/10.1021/ci200020k

272. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I., & Lomize, A. L. (2012). OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Research*, *40*(D1), 370–376. https://doi.org/10.1093/nar/gkr703

273. Ma, N., Zhang, Z., Liao, F., Jiang, T., & Tu, Y. (2020). The birth of artemisinin. *Pharmacology and Therapeutics*, *216*, 107658. https://doi.org/10.1016/j.pharmthera.2020.107658

274. Ma, R., Guo, D. X., Li, H. F., Liu, H. X., Zhang, Y. R., Ji, J. B., Xing, J., & Wang, S. Q. (2019). Spectroscopic methodologies and molecular docking studies on the interaction of antimalarial drug piperaquine and its metabolites with human serum albumin. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, *222*, 117158. https://doi.org/10.1016/j.saa.2019.117158

275. Mackerell, A. D. (2004). Empirical force fields for biological macromolecules: Overview and issues. *Journal of Computational Chemistry*, *25*(13), 1584–1604. https://doi.org/10.1002/jcc.20082

276. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., … Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B*, *102*(18), 3586–3616. https://doi.org/10.1021/jp973084f

277. MacKerell, Alexander D., Wiórkiewicz-Kuczera, J., Karplus, M., & MacKerell, A. D. (1995). An All-Atom Empirical Energy Function for the Simulation of Nucleic Acids. *Journal of the American Chemical Society*, *117*(48), 11946–11975. https://doi.org/10.1021/ja00153a017

278. Madhusudhan, M. S., Marti-Renom, M. A., Sanchez, R., & Sali, A. (2006). Variable gap penalty for protein sequence-structure alignment. *Protein Engineering, Design and Selection*, *19*(3), 129–133. https://doi.org/10.1093/protein/gzj005

279. Mahittikorn, A., Masangkay, F. R., Kotepui, K. U., Milanez, G. D. J., & Kotepui, M. (2021). Comparison of Plasmodium ovale curtisi and Plasmodium ovale wallikeri infections by a meta-analysis approach. *Scientific Reports*, *11*(1), 1–15. https://doi.org/10.1038/s41598-021-85398-w

280. Maia, E. H. B., Assis, L. C., de Oliveira, T. A., da Silva, A. M., & Taranto, A. G. (2020). Structure-Based Virtual Screening: From Classical to Artificial Intelligence. *Frontiers in Chemistry*, *8*(April). https://doi.org/10.3389/fchem.2020.00343

281. Marti-Renom, M. A. (2004). Alignment of protein sequences by their profiles. *Protein Science*, *13*(4), 1071–1087. https://doi.org/10.1110/ps.03379804

282. Martin Vabulas, R. (2021). Ferroptosis-related flavoproteins: Their function and stability. *International Journal of Molecular Sciences*, *22*(1), 1–14. https://doi.org/10.3390/ijms22010430

283. Martínez, L. (2015). Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. *PLoS ONE*, *10*(3), 1–10. https://doi.org/10.1371/journal.pone.0119264

284. Martins, A. C., Paoliello, M. M. B., Docea, A. O., Santamaria, A., Tinkov, A. A., Skalny, A. V., & Aschner, M. (2021). Review of the mechanism underlying mefloquine-induced neurotoxicity. *Critical Reviews in Toxicology*, *51*(3), 209–216. https://doi.org/10.1080/10408444.2021.1901258

285. Mattea, S., Baptista, M., Reichert, P., Spinale, A., Wu, J., Allaire, M., Fiske, B., & Knapp, S. (2018). *Crystallizing the {Parkinson}'s {Disease} {Protein} {LRRK}2 {Under} {Microgravity} {Conditions}. epub*, 11 PP. https://doi.org/10.1101/259655

286. Meng, X.-Y., Zhang, H.-X., Mezei, M., & Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Current Computer-Aided Drug Design*, *7*(2), 146–157. https://doi.org/10.1109

287. Meyder, A., Nittinger, E., Lange, G., Klein, R., & Rarey, M. (2017). Estimating Electron Density Support for Individual Atoms and Molecular Fragments in X-ray Structures. *Journal of Chemical Information and Modeling*, *57*(10), 2437–2447. https://doi.org/10.1021/acs.jcim.7b00391

288. Mills, D. B., Francis, W. R., Vargas, S., Larsen, M., Elemans, C. P. H., Canfield, D. E., & Wörheide, G. (2018). The last common ancestor of animals lacked the HIF pathway and respired in low-oxygen environments. *ELife*, *7*, 1–17. https://doi.org/10.7554/eLife.31176

289. Minale, A. S., & Alemu, K. (2018). Mapping malaria risk using geographic information systems and remote sensing: The case of Bahir dar city, Ethiopia. *Geospatial Health*, *13*(1), 157–163. https://doi.org/10.4081/gh.2018.660

290. Miramar, M. D., Costantini, P., Ravagnan, L., Saraiva, L. M., Haouzi, D., Brothers, G., Penninger, J. M., Peleato, M. L., Kroemer, G., & Susin, S. A. (2001). NADH Oxidase Activity of Mitochondrial Apoptosis-inducing Factor. *Journal of Biological Chemistry*, *276*(19), 16391–16398. https://doi.org/10.1074/jbc.M010498200

291. Mirza, S. B., Lee, R. C. H., Chu, J. J. H., Salmas, R. E., Mavromoustakos, T., & Durdagi, S. (2018). Discovery of selective dengue virus inhibitors using combination of molecular fingerprint-based virtual screening protocols, structure-based pharmacophore model development, molecular dynamics simulations and in vitro studies. *Journal of Molecular Graphics and Modelling*, *79*, 88–102. https://doi.org/10.1016/j.jmgm.2017.10.010

292. Mitchell, C. L., Brazeau, N. F., Keeler, C., Mwandagalirwa, M. K., Tshefu, A. K., Juliano, J. J., & Meshnick, S. R. (2021). Under the Radar: Epidemiology of Plasmodium ovale in the Democratic Republic of the Congo. *Journal of Infectious Diseases*, *223*(6), 1005–1014. https://doi.org/10.1093/infdis/jiaa478

293. Montoya, M. C., Beattie, S., Alden, K. M., & Krysan, D. J. (2019). Derivatives of the antimalarial drug mefloquine are broad spectrum antifungal molecules with activity against drug-resistant clinical isolates. *BioRxiv*, *December 2019*, 1–18. https://doi.org/10.1101/851816

294. Moraes, F., & Góes, A. (2016). A decade of human genome project conclusion:

Scientific diffusion about our genome knowledge. *Biochemistry and Molecular Biology Education*, *44*(3), 215–223. https://doi.org/10.1002/bmb.20952

295. Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S., & Stewart, P. D. S. (2014). Membrane protein structure determination - The next generation. *Biochimica et Biophysica Acta - Biomembranes*, *1838*(1 PARTA), 78–87. https://doi.org/10.1016/j.bbamem.2013.07.010

296. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, *19*(14), 1639–1662. https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B

297. Moult, J., Fidelis, K., Rost, B., Hubbard, T., & Tramontano, A. (2005). Critical assessment of methods of protein structure prediction (CASP) - Round 6. *Proteins: Structure, Function and Genetics*, *61*(SUPPL. 7), 3–7. https://doi.org/10.1002/prot.20716

298. Muhammed, M. T., & Aki-Yalcin, E. (2019). Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chemical Biology and Drug Design*, *93*(1), 12–20. https://doi.org/10.1111/cbdd.13388

299. Mulenga-cilundika, P., Ekofo, J., Bagalwa, D., Kabuya, M., & Chenge, F. (2020). *Adverse Effects of Anti-malarial Drugs Used in the Treatment of Malaria Cases Caused by Species Other than Plasmodium falciparum : A scoping Review*. *December*.

300. Muller, M. P., Jiang, T., Sun, C., Lihan, M., Pant, S., Mahinthichaichan, P., Trifan, A., & Tajkhorshid, E. (2019). Characterization of Lipid-Protein Interactions and Lipid-Mediated Modulation of Membrane Protein Function through Molecular Simulation. *Chemical Reviews*, *119*(9), 6086–6161. https://doi.org/10.1021/acs.chemrev.8b00608

301. Nabarro, L. E. B., Nolder, D., Broderick, C., Nadjm, B., Smith, V., Blaze, M., Checkley, A. M., Chiodini, P. L., Sutherland, C. J., & Whitty, C. J. M. (2018). Geographical and temporal trends and seasonal relapse in Plasmodium ovale spp. and Plasmodium malariae infections imported to the UK between. *BMC Medicine*, *16*(1), 1–8. https://doi.org/10.1186/s12916-018-1204-6

302. Nakatani, Y., Jiao, W., Aragão, D., Shimaki, Y., Petri, J., Parker, E. J., & Cook, G. M. (2017). Crystal structure of type II NADH:quinone oxidoreductase from Caldalkalibacillus thermarum with an improved resolution of 2.15 Å. *Acta Crystallographica Section F:Structural Biology Communications*, *73*(10), 541–549. https://doi.org/10.1107/S2053230X17013073

303. Nakatani, Y., Shimaki, Y., Dutta, D., Muench, S. P., Ireton, K., Cook, G. M., & Jeuken, L. J. C. (2020). Unprecedented Properties of Phenothiazines Unraveled by a NDH-2 Bioelectrochemical Assay Platform. *Journal of the American Chemical Society*, *142*(3), 1311–1320. https://doi.org/10.1021/jacs.9b10254

304. Neves, B. J., Braga, R. C., Melo-Filho, C. C., Moreira-Filho, J. T., Muratov, E. N., & Andrade, C. H. (2018). QSAR-based virtual screening: Advances and applications in drug discovery. *Frontiers in Pharmacology*, *9*(NOV), 1–7. https://doi.org/10.3389/fphar.2018.01275

305. Nielsen, P. (2009). Coastal and estuarine processes. In *Coastal And Estuarine Processes* (pp. 1–360). https://doi.org/10.1142/7114

306. Nixon, G. L., Moss, D. M., Shone, A. E., Lalloo, D. G., Fisher, N., O'neill, P. M., Ward, S. A., & Biagini, G. A. (2013). Antimalarial pharmacology and therapeutics of atovaquone. *Journal of Antimicrobial Chemotherapy*, *68*(5), 977–985. https://doi.org/10.1093/jac/dks504

307. Nizami, B., Sydow, D., Wolber, G., & Honarparvar, B. (2016). Molecular insight on the binding of NNRTI to K103N mutated HIV-1 RT: Molecular dynamics simulations and dynamic pharmacophore analysis. *Molecular BioSystems*, *12*(11), 3385–3395. https://doi.org/10.1039/c6mb00428h

308. Nwanochie, E., & Uversky, V. N. (2019). Structure determination by single-particle cryo-electron microscopy: Only the sky (and intrinsic disorder) is the limit. *International Journal of Molecular Sciences*, *20*(17). https://doi.org/10.3390/ijms20174186

309. O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., … Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*(D1), D733–D745. https://doi.org/10.1093/nar/gkv1189

310. O'Neill, P. M., Barton, V. E., & Ward, S. A. (2010). The molecular mechanism of action of artemisinin - The debate continues. *Molecules*, *15*(3), 1705–1721. https://doi.org/10.3390/molecules15031705

311. Oostenbrink, C., Villa, A., Mark, A. E., & Van Gunsteren, W. F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry*, *25*(13), 1656–1676. https://doi.org/10.1002/jcc.20090

312. Österberg, F., Morris, G. M., Sanner, M. F., Olson, A. J., & Goodsell, D. S. (2002). Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in autodock. *Proteins: Structure, Function and Genetics*, *46*(1), 34–40. https://doi.org/10.1002/prot.10028

313. Pagadala, N. S., Syed, K., & Tuszynski, J. (2017). Software for molecular docking: a review. *Biophysical Reviews*, *9*(2), 91–102. https://doi.org/10.1007/s12551-016-0247-1

314. Park, S. J., Lee, J., Qi, Y., Kern, N. R., Lee, H. S., Jo, S., Joung, I., Joo, K., Lee, J., & Im, W. (2019). CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology*, *29*(4), 320–331. https://doi.org/10.1093/glycob/cwz003

315. Parsanathan, R., & Jain, S. K. (2020). Glucose-6-phosphate dehydrogenase (G6PD) deficiency is linked with cardiovascular disease. *Hypertension Research*, *43*(6), 582–584. https://doi.org/10.1038/s41440-020-0402-8

316. Pearson, W. R., & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(8), 2444–2448. https://doi.org/10.1073/pnas.85.8.2444

317. Pei, J., & Grishin, N. V. (2014). PROMALS3D: Multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods in Molecular Biology*, *1079*, 263–271. https://doi.org/10.1007/978-1-62703-646-7_17

318. Petri, J., Shimaki, Y., Jiao, W., Bridges, H. R., Russell, E. R., Parker, E. J., Aragão, D., Cook, G. M., & Nakatani, Y. (2018). Structure of the NDH-2 – HQNO inhibited complex provides molecular insight into quinone-binding site inhibitors. *Biochimica et Biophysica*

Acta - Bioenergetics, *1859*(7), 482–490. https://doi.org/10.1016/j.bbabio.2018.03.014

319. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

320. Phillips, J. C., Hardy, D. J., Maia, J. D. C., Stone, J. E., Ribeiro, J. V., Bernardi, R. C., Buch, R., Fiorin, G., Hénin, J., Jiang, W., McGreevy, R., Melo, M. C. R., Radak, B. K., Skeel, R. D., Singharoy, A., Wang, Y., Roux, B., Aksimentiev, A., Luthey-Schulten, Z., … Tajkhorshid, E. (2020). Scalable molecular dynamics on CPU and GPU architectures with NAMD. *Journal of Chemical Physics*, *153*(4), 1–33. https://doi.org/10.1063/5.0014475

321. Phillips, J. C., Schulten, K., Bhatele, A., Mei, C., Sun, Y., Bohm, E. J., & Kale, L. V. (2016). Scalable molecular dynamics with NAMD. *Parallel Science and Engineering Applications: The Charm++ Approach*, *26*(16), 60–76. https://doi.org/10.1201/b16251-15

322. Pillat, M. M., Krüger, A., Guimarães, L. M. F., Lameu, C., de Souza, E. E., Wrenger, C., & Ulrich, H. (2020). Insights in Chloroquine Action: Perspectives and Implications in Malaria and COVID-19. *Cytometry Part A*, *97*(9), 872–881. https://doi.org/10.1002/cyto.a.24190

323. Ponting, C. P., & Russell, R. B. (2000). Identification of distant homologues of fibroblast growth factors suggests a common ancestor for all β-trefoil proteins. *Journal of Molecular Biology*, *302*(5), 1041–1047. https://doi.org/10.1006/jmbi.2000.4087

324. Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., & Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Research*, *46*(W1), W200–W204. https://doi.org/10.1093/nar/gky448

325. Pousibet-Puerto, J., Salas-Coronas, J., Sánchez-Crespo, A., Molina-Arrebola, M. A., Soriano-Pérez, M. J., Giménez-López, M. J., Vázquez-Villegas, J., & Cabezas-Fernández, M. T. (2016). Impact of using artemisinin-based combination therapy (ACT) in the treatment of uncomplicated malaria from Plasmodium falciparum in a non-endemic zone. *Malaria Journal*, *15*(1), 1–7. https://doi.org/10.1186/s12936-016-1408-1

326. Prakash, A., Jeffryes, M., Bateman, A., & Finn, R. D. (2017). The HMMER Web Server for Protein Sequence Similarity Search. *Current Protocols in Bioinformatics*, *60*(1), 3.15.1-3.15.23. https://doi.org/10.1002/cpbi.40

327. Price, R. N., Commons, R. J., Battle, K. E., Thriemer, K., & Mendis, K. (2020). Plasmodium vivax in the Era of the Shrinking P. falciparum Map. *Trends in Parasitology*, *36*(6), 560–570. https://doi.org/10.1016/j.pt.2020.03.009

328. *prime*. (n.d.).

329. Pundir, S., Martin, M. J., & O'Donovan, C. (2016). UniProt Tools. *Current Protocols in Bioinformatics*, *53*(1), 1.29.1-1.29.15. https://doi.org/10.1002/0471250953.bi0129s53

330. Punihaole, D., Workman, R. J., Upadhyay, S., Van Bruggen, C., Schmitz, A. J., Reineke, T. M., & Frontiera, R. R. (2018). New Insights into Quinine-DNA Binding Using Raman Spectroscopy and Molecular Dynamics Simulations. *Journal of Physical Chemistry B*, *122*(43), 9840–9851. https://doi.org/10.1021/acs.jpcb.8b05795

331. Qi, Y., Lee, J., Cheng, X., Shen, R., Islam, S. M., Roux, B., & Im, W. (2020). CHARMM-GUI DEER facilitator for spin-pair distance distribution calculations and preparation of restrained-ensemble molecular dynamics simulations. *Journal of Computational Chemistry*, *41*(5), 415–420. https://doi.org/10.1002/jcc.26032

332. Rajahram, G. S., Barber, B. E., William, T., Grigg, M. J., Menon, J., Yeo, T. W., & Anstey, N. M. (2016). Falling Plasmodium knowlesi Malaria Death Rate among Adults despite Rising Incidence, Sabah, Malaysia, 2010-2014. *Emerging Infectious Diseases*, *22*(1), 41–48. https://doi.org/10.3201/eid2201.151305

333. Rappolt, M., Fernández Vidal, M., Kriechbaum, M., Steinhart, M., Amenitsch, H., Bernstortf, S., & Laggner, P. (2003). Structural, dynamic and mechanical properties of POPC at low cholesterol concentration studied in pressure/temperature space. *European Biophysics Journal*, *31*(8), 575–585. https://doi.org/10.1007/s00249-002-0253-z

334. Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology*, *261*(3), 470–489. https://doi.org/10.1006/jmbi.1996.0477

335. Reigosa, M. J., Gonzalez, L., Sanches-Moreiras, A., Duran, B., Puime, D., Fernadez, D. A., & Bolano, J. C. (2001). Comparison of physiological effects of allelochemicals and commercial herbicides. *Allelopathy Journal*, *8*(2), 211–220.

336. Remmert, M., Biegert, A., Hauser, A., & Söding, J. (2012). HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods*, *9*(2), 173–175. https://doi.org/10.1038/nmeth.1818

337. Resende, L. A., da Silva, P. H. R., & Fernandes, C. (2019). Quantitative determination of the antimalarials artemether and lumefantrine in biological samples: A review. *Journal of Pharmaceutical and Biomedical Analysis*, *165*, 304–314. https://doi.org/10.1016/j.jpba.2018.12.021

338. Richard, J., John, T., Manuel, N., Sean, L., & Ian, M. (2019). *LA-UR-19-29136 Title : Intended for : Issued : 11*(Scipy 2016).

339. Risner, N., Martinez, D., Rosenberg, J., Stuhlmueller, C., Goldsmith, B., Develop, H. O. W. T. O., Objectives, L., Instructions, T., Departement, E., Omartian, S., Dna, H. B. V, Beauboeuf-Lafontant, T., Form, O., 아더케이, 김경섭최복순, Audience, A. A., Smith, R. E. R. C., Masters, C. F. O. R., Students, D., … Bourne, P. E. (2018). No Title. *Clinical Neurology and Neurosurgery*, *169*(1), 1–11. https://doi.org/10.1017/CBO9781107415324.004

340. Rodriguez, J., Xie, C., Li, T., Sun, Y., Wang, Y., Xu, Y., Li, K., Zhang, S., Zhou, K., Wang, Y., Mallard, C., Hagberg, H., Doti, N., Wang, X., & Zhu, C. (2020). Inhibiting the interaction between apoptosis-inducing factor and cyclophilin A prevents brain injury in neonatal mice after hypoxia-ischemia. *Neuropharmacology*, *171*(March), 108088. https://doi.org/10.1016/j.neuropharm.2020.108088

341. Romero-Tamayo, S., Laplaza, R., Velazquez-Campoy, A., Villanueva, R., Medina, M., & Ferreira, P. (2021). W196 and the β -Hairpin Motif Modulate the Redox Switch of Conformation and the Biomolecular Interaction Network of the Apoptosis-Inducing Factor. *Oxidative Medicine and Cellular Longevity*, *2021*(Im). https://doi.org/10.1155/2021/6673661

342. Ross, C., Nizami, B., Glenister, M., Amamuddy, O. S., Atilgan, A. R., Atilgan, C., & Bishop, Ö. T. (2018). MODE-TASK: Large-scale protein motion tools. *Bioinformatics*, *34*(21), 3759–3763. https://doi.org/10.1093/bioinformatics/bty427

343. Rozewicki, J., Li, S., Amada, K. M., Standley, D. M., & Katoh, K. (2019). MAFFT-DASH: Integrated protein sequence and structural alignment. *Nucleic Acids Research*, *47*(W1), W5–W10. https://doi.org/10.1093/nar/gkz342

344. *RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston,*

MA URL http://www.rstudio.com/. - Search. (n.d.). Retrieved August 2, 2022, from https://www.bing.com/search?q=RStudio+Team+(2020).+RStudio%3A+Integrated+Development+for+R.+RStudio%2C+PBC%2C+Boston%2C+MA+URL+http%3A%2F%2Fwww.rstudio.com%2F.&cvid=32525a953f994367920a9d6671a437bc&aqs=edge..69i57.986j0j1&pglt=171&FORM=ANNTA1&DAF0=1&PC=ED

345. Ruiz-Carmona, S., Alvarez-Garcia, D., Foloppe, N., Garmendia-Doval, A. B., Juhos, S., Schmidtke, P., Barril, X., Hubbard, R. E., & Morley, S. D. (2014). rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Computational Biology*, *10*(4), 1–7. https://doi.org/10.1371/journal.pcbi.1003571

346. Rutledge, G. G., Böhme, U., Sanders, M., Reid, A. J., Cotton, J. A., Maiga-Ascofare, O., Djimdé, A. A., Apinjoh, T. O., Amenga-Etego, L., Manske, M., Barnwell, J. W., Renaud, F., Ollomo, B., Prugnolle, F., Anstey, N. M., Auburn, S., Price, R. N., McCarthy, J. S., Kwiatkowski, D. P., … Otto, T. D. (2017). Plasmodium malariae and P. ovale genomes provide insights into malaria parasite evolution. *Nature*, *542*(7639), 101–104. https://doi.org/10.1038/nature21038

347. Rutledge, G. G., Marr, I., Huang, G. K. L., Auburn, S., Marfurt, J., Sanders, M., White, N. J., Berriman, M., Newbold, C. I., Anstey, N. M., Otto, T. D., & Price, R. N. (2017). Genomic characterization of recrudescent plasmodium malariae after treatment with artemether/lumefantrine. *Emerging Infectious Diseases*, *23*(8), 1300–1307. https://doi.org/10.3201/eid2308.161582

348. Sacchi, P., Loconte, L., Macetti, G., Rizzato, S., & Lo Presti, L. (2019). Correlations of Crystal Structure and Solubility in Organic Salts: The Case of the Antiplasmodial Drug Piperaquine. *Crystal Growth and Design*, *19*(2), 1399–1410. https://doi.org/10.1021/acs.cgd.8b01794

349. Sahu, P., Mohan, K. V., Aggarwal, S., Arindkar, S., Mahesh Kumar, J., Kumar Upadhyay, P., Ramakrishna, G., & Nagarajan, P. (2021). Apoptosis-inducing factor deficient mice fail to develop hepatic steatosis under high fat high fructose diet or bile duct ligation. *Cell Biochemistry and Function*, *39*(2), 296–307. https://doi.org/10.1002/cbf.3579

350. Šali, A., & Blundell, T. L. (1993). 1-s2.0-S0022283683716268-main.pdf. In *Journal of Molecular Biology* (Vol. 234, pp. 779–815).

351. Salmaso, V., & Moro, S. (2018). Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview. *Frontiers in Pharmacology*, *9*(AUG), 1–16. https://doi.org/10.3389/fphar.2018.00923

352. Salomon-Ferrer, R., Case, D. A., & Walker, R. C. (2013). An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *3*(2), 198–210. https://doi.org/10.1002/wcms.1121

353. *SANCDB*. (n.d.). Retrieved August 2, 2022, from https://sancdb.rubi.ru.ac.za/

354. Santos, B. M., Dias, B. K. M., Nakabashi, M., & Garcia, C. R. S. (2021). The Knockout for G Protein-Coupled Receptor-Like PfSR25 Increases the Susceptibility of Malaria Parasites to the Antimalarials Lumefantrine and Piperaquine but Not to Medicine for Malaria Venture Compounds. *Frontiers in Microbiology*, *12*(March). https://doi.org/10.3389/fmicb.2021.638869

355. Schaeffer, R. D., Liao, Y., & Grishin, N. V. (2018). Searching ECOD for Homologous Domains by Sequence and Structure. *Current Protocols in Bioinformatics*, *61*(1), 1–11. https://doi.org/10.1002/cpbi.45

356. Schöning-Stierand, K., Diedrich, K., Fährrolfes, R., Flachsenberg, F., Meyder, A., Nittinger, E., Steinegger, R., & Rarey, M. (2020). ProteinsPlus: Interactive analysis of protein–ligand binding interfaces. *Nucleic Acids Research*, *48*(W1), W48–W53. https://doi.org/10.1093/NAR/GKAA235

357. Schrezenmeier, E., & Dörner, T. (2020a). Mechanisms of action of hydroxychloroquine and chloroquine: implications for rheumatology. *Nature Reviews Rheumatology*, *16*(3), 155–166. https://doi.org/10.1038/s41584-020-0372-x

358. Schrezenmeier, E., & Dörner, T. (2020b). Mechanisms of action of hydroxychloroquine and chloroquine: implications for rheumatology. *Nature Reviews Rheumatology*, *16*(3), 155–166. https://doi.org/10.1038/s41584-020-0372-x

359. Schuetz, D. A., Bernetti, M., Bertazzo, M., Musil, D., Eggenweiler, H. M., Recanatini, M., Masetti, M., Ecker, G. F., & Cavalli, A. (2019). Predicting Residence Time and Drug Unbinding Pathway through Scaled Molecular Dynamics. *Journal of Chemical Information and Modeling*, *59*(1), 535–549. https://doi.org/10.1021/acs.jcim.8b00614

360. Schüttelkopf, A. W., & Van Aalten, D. M. F. (2004). PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica Section D: Biological Crystallography*, *60*(8), 1355–1363. https://doi.org/10.1107/S0907444904011679

361. Sevrioukova, I. F. (2011). Apoptosis-inducing factor: Structure, function, and redox regulation. *Antioxidants and Redox Signaling*, *14*(12), 2545–2579. https://doi.org/10.1089/ars.2010.3445

362. Sheik Amamuddy, O., Glenister, M., Tshabalala, T., & Tastan Bishop, Ö. (2021a). MDM-TASK-web: MD-TASK and MODE-TASK web server for analyzing protein dynamics. *Computational and Structural Biotechnology Journal*, *19*, 5059–5071. https://doi.org/10.1016/j.csbj.2021.08.043

363. Sheik Amamuddy, O., Glenister, M., Tshabalala, T., & Tastan Bishop, Ö. (2021b). MDM-TASK-web: MD-TASK and MODE-TASK web server for analyzing protein dynamics. *Computational and Structural Biotechnology Journal*, *19*, 5059–5071. https://doi.org/10.1016/j.csbj.2021.08.043

364. Sippl, M. J. (1993a). Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *Journal of Computer-Aided Molecular Design*, *7*(4), 473–501. https://doi.org/10.1007/BF02337562

365. Sippl, M. J. (1993b). Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function, and Bioinformatics*, *17*(4), 355–362. https://doi.org/10.1002/prot.340170404

366. Sittel, F., Jain, A., & Stock, G. (2014). Principal component analysis of molecular dynamics: On the use of Cartesian vs. internal coordinates. *Journal of Chemical Physics*, *141*(1). https://doi.org/10.1063/1.4885338

367. Smith, M. D., Rao, J. S., Segelken, E., & Cruz, L. (2015). Force-Field Induced Bias in the Structure of Aβ21-30: A Comparison of OPLS, AMBER, CHARMM, and GROMOS Force Fields. *Journal of Chemical Information and Modeling*, *55*(12), 2587–2595. https://doi.org/10.1021/acs.jcim.5b00308

368. Snow, B. R. W., Craig, M. H., Deichmann, U., & Le Sueur, D. (1999). A preliminary continental risk map for malaria mortality among African children. *Parasitology Today*, *15*(3), 99–104. https://doi.org/10.1016/S0169-4758(99)01395-2

369. Söding, J. (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics*, *21*(7), 951–960. https://doi.org/10.1093/bioinformatics/bti125

370. Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, *33*(SUPPL. 2), 244–248. https://doi.org/10.1093/nar/gki408

371. Sofi, F., Capalbo, A., Pucci, N., Giuliattini, J., Condino, F., Alessandri, F., Abbate, R., Gensini, G. F., & Califano, S. (2008). Cardiovascular evaluation, including resting and exercise electrocardiography, before participation in competitive sports: Cross sectional study. *Bmj*, *337*(7661), 88–92. https://doi.org/10.1136/bmj.a346

372. Srivastava, I. K., Rottenberg, H., & Vaidya, A. B. (1997). Atovaquone, a broad spectrum antiparasitic drug, collapses mitochondrial membrane potential in a malarial parasite. *Journal of Biological Chemistry*, *272*(7), 3961–3966. https://doi.org/10.1074/jbc.272.7.3961

373. Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S. J., & Söding, J. (2019). HH-suite3 for fast remote homology detection and deep protein annotation. *BioRxiv*, 1–15. https://doi.org/10.1101/560029

374. Studer, G., Tauriello, G., Bienert, S., Biasini, M., Johner, N., & Schwede, T. (2021). ProMod3 - A versatile homology modelling toolbox. *PLoS Computational Biology*, *17*(1), 1–18. https://doi.org/10.1371/JOURNAL.PCBI.1008667

375. Sun, P. D., Foster, C. E., & Boyington, J. C. (2004). Overview of protein structural and functional folds. *Current Protocols in Protein Science / Editorial Board, John E. Coligan ... [et Al.]*, *Chapter 17*, 1–189. https://doi.org/10.1002/0471140864.ps1701s35

376. Swinney, D. C. (2011). Molecular Mechanism of Action (MMoA) in Drug Discovery. In *Annual Reports in Medicinal Chemistry* (1st ed., Vol. 46). Elsevier Inc. https://doi.org/10.1016/B978-0-12-386009-5.00009-6

377. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). *MEGA5 : Molecular Evolutionary Genetics Analysis Using Maximum Likelihood , Evolutionary Distance , and Maximum Parsimony Methods Research resource*. *28*(10), 2731–2739. https://doi.org/10.1093/molbev/msr121

378. Tan, Y., Hanson, J. A., Chu, J., & Yang, H. (2014). VIP MSM UBER WIP PCA ESSENTIAL Protein Dynamics. In *Proteins* (Vol. 1084). https://doi.org/10.1007/978-1-62703-658-0

379. Thompson, A. P., Aktulga, H. M., Berger, R., Bolintineanu, D. S., Brown, W. M., Crozier, P. S., in 't Veld, P. J., Kohlmeyer, A., Moore, S. G., Nguyen, T. D., Shan, R., Stevens, M. J., Tranchida, J., Trott, C., & Plimpton, S. J. (2022). LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications*, *271*, 108171. https://doi.org/10.1016/j.cpc.2021.108171

380. Thomsen, R., & Christensen, M. H. (2006). MolDock: A new technique for high-accuracy molecular docking. *Journal of Medicinal Chemistry*, *49*(11), 3315–3321. https://doi.org/10.1021/jm051197e

381. Tönnesmann, E., Stroehmann, I., Kandolf, R., Wolburg, H., Strach, K., Musshoff, F., Tiemann, K., & Lewalter, T. (2012). Cardiomyopathy caused by longterm treatment with chloroquine: A rare disease or a rare diagnosis? (Journal of Rheumatology (2012) 39, (1099-1104)). *Journal of Rheumatology*, *39*(7), 1496. https://doi.org/10.3899/jrheum.110959.C1

382. Trisolini, L., Gambacorta, N., Gorgoglione, R., Montaruli, M., Laera, L., Colella, F.,

Volpicella, M., De Grassi, A., & Pierri, C. L. (2019). FAD/NADH Dependent Oxidoreductases: From Different Amino Acid Sequences to Similar Protein Shapes for Playing an Ancient Function. *Journal of Clinical Medicine*, *8*(12), 2117. https://doi.org/10.3390/jcm8122117

383. Trott, O., & Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, *31*(2), NA-NA. https://doi.org/10.1002/jcc.21334

384. Tse, E. G., Korsik, M., & Todd, M. H. (2019). The past, present and future of anti-malarial medicines. *Malaria Journal*, *18*(1), 1–21. https://doi.org/10.1186/s12936-019-2724-z

385. Twohig, K. A., Pfeffer, D. A., Baird, J. K., Price, R. N., Zimmerman, A., Id, S. I. H., Gething, P. W., Battle, K. E., & Id, E. H. (2019). Twohig KA, Pfeffer DA, Baird JK, Price RN, Zimmerman PA, Hay SI, et al. Growing evidence of Plasmodium vivax across malaria-endemic Africa. *PLoS Negl Trop Dis*, *13*(1), e0007140.

386. Urner, L. H., Liko, I., Yen, H. Y., Hoi, K. K., Bolla, J. R., Gault, J., Almeida, F. G., Schweder, M. P., Shutin, D., Ehrmann, S., Haag, R., Robinson, C. V., & Pagel, K. (2020). Modular detergents tailor the purification and structural analysis of membrane proteins including G-protein coupled receptors. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-14424-8

387. Vale, N., Moreira, R., & Gomes, P. (2009). Primaquine revisited six decades after its discovery. *European Journal of Medicinal Chemistry*, *44*(3), 937–953. https://doi.org/10.1016/j.ejmech.2008.08.011

388. van de Waterbeemd, H., & Gifford, E. (2003). ADMET in silico modelling: Towards prediction paradise? *Nature Reviews Drug Discovery*, *2*(3), 192–204. https://doi.org/10.1038/nrd1032

389. 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

390. Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry*, *45*(12), 2615–2623. https://doi.org/10.1021/jm020017n

391. Velázquez-Libera, J. L., Durán-Verdugo, F., Valdés-Jiménez, A., Valdés-Jiménez, A., Núñez-Vivanco, G., & Caballero, J. (2020). LigRMSD: A web server for automatic structure matching and RMSD calculations among identical and similar compounds in protein-ligand docking. *Bioinformatics*, *36*(9), 2912–2914. https://doi.org/10.1093/bioinformatics/btaa018

392. Vénien-Bryan, C., Li, Z., Vuillard, L., & Boutin, J. A. (2017). Cryo-electron microscopy and X-ray crystallography: Complementary approaches to structural biology and drug discovery. *Acta Crystallographica Section:F Structural Biology Communications*, *73*(4), 174–183. https://doi.org/10.1107/S2053230X17003740

393. Venkatachalam, C. M., Jiang, X., Oldfield, T., & Waldman, M. (2003). LigandFit: A novel method for the shape-directed rapid docking of ligands to protein active sites. *Journal of Molecular Graphics and Modelling*, *21*(4), 289–307. https://doi.org/10.1016/S1093-3263(02)00164-X

394. Verdaguer, I. B., Crispim, M., Zafra, C. A., Sussmann, R. A. C., Buriticá, N. L., Melo, H. R., Azevedo, M. F., Almeida, F. G., Kimura, E. A., & Katzin, A. M. (2021). Exploring ubiquinone

biosynthesis inhibition as a strategy for improving atovaquone efficacy in Malaria. *Antimicrobial Agents and Chemotherapy*, *65*(4). https://doi.org/10.1128/AAC.01516-20

395. Vermaas, J. V., Hardy, D. J., Stone, J. E., Tajkhorshid, E., & Kohlmeyer, A. (2016). TopoGromacs: Automated Topology Conversion from CHARMM to GROMACS within VMD. *Journal of Chemical Information and Modeling*, *56*(6), 1112–1116. https://doi.org/10.1021/acs.jcim.6b00103

396. Von Grotthuss, M., Pas, J., Wyrwicz, L., Ginalski, K., & Rychlewski, L. (2003). Application of 3D-Jury, GRDB, and Verify3D in Fold Recognition. *Proteins: Structure, Function and Genetics*, *53*(SUPPL. 6), 418–423. https://doi.org/10.1002/prot.10547

397. Vyas, V. K., Ukawala, R. D., Ghate, M., & Chintha, C. (2012). Homology modeling a fast tool for drug discovery: Current perspectives. *Indian Journal of Pharmaceutical Sciences*, *74*(1), 1–17. https://doi.org/10.4103/0250-474X.102537

398. Wallace, I. M., Blackshields, G., & Higgins, D. G. (2005). Multiple sequence alignments. *Current Opinion in Structural Biology*, *15*(3 SPEC. ISS.), 261–266. https://doi.org/10.1016/j.sbi.2005.04.002

399. Wanderlingh, U., Branca, C., Crupi, C., Nibali, V. C., La Rosa, G., Rifici, S., Ollivier, J., & D'Angelo, G. (2017). Molecular dynamics of POPC phospholipid bilayers through the gel to fluid phase transition: An incoherent quasi-elastic neutron scattering study. *Journal of Chemistry*, *2017*. https://doi.org/10.1155/2017/3654237

400. Wang, H., Chen, T., Wan, L., Lu, J., Wei, H., Deng, K. Y., Wei, J., & Xin, H. B. (2020). Attenuated Salmonella engineered with an apoptosis-inducing factor (AIF) eukaryotic expressing system enhances its anti-tumor effect in melanoma in vitro and in vivo. *Applied Microbiology and Biotechnology*, *104*(8), 3517–3528. https://doi.org/10.1007/s00253-020-10485-3

401. Wang, J., Xu, C., Wong, Y. K., Li, Y., Liao, F., Jiang, T., & Tu, Y. (2019). Artemisinin, the Magic Drug Discovered from Traditional Chinese Medicine. *Engineering*, *5*(1), 32–39. https://doi.org/10.1016/j.eng.2018.11.011

402. Wångdahl, A., Wyss, K., Saduddin, D., Bottai, M., Ydring, E., Vikerfors, T., & Färnert, A. (2019). Severity of Plasmodium falciparum and Non-falciparum Malaria in Travelers and Migrants: A Nationwide Observational Study over 2 Decades in Sweden. *Journal of Infectious Diseases*, *220*(8), 1335–1345. https://doi.org/10.1093/infdis/jiz292

403. Watanabe, Y. S., Fukunishi, Y., & Nakamura, H. (2004). 1P047 Modeling of Loops in Protein Structures. *Seibutsu Butsuri*, *44*(supplement), S41. https://doi.org/10.2142/biophys.44.s41_3

404. 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

405. WHO. (2020). *World malaria report 2020- WHO*.

406. 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), 407–410. https://doi.org/10.1093/nar/gkm290

407. Windle, S. T., Lane, K. D., Gadalla, N. B., Liu, A., Mu, J., Caleon, R. L., Rahman, R. S., Sá, J. M., & Wellems, T. E. (2020a). Evidence for linkage of pfmdr1, pfcrt, and pfk13

polymorphisms to lumefantrine and mefloquine susceptibilities in a Plasmodium falciparum cross. *International Journal for Parasitology: Drugs and Drug Resistance*, *14*(October), 208–217. https://doi.org/10.1016/j.ijpddr.2020.10.009

408. Windle, S. T., Lane, K. D., Gadalla, N. B., Liu, A., Mu, J., Caleon, R. L., Rahman, R. S., Sá, J. M., & Wellems, T. E. (2020b). Evidence for linkage of pfmdr1, pfcrt, and pfk13 polymorphisms to lumefantrine and mefloquine susceptibilities in a Plasmodium falciparum cross. *International Journal for Parasitology: Drugs and Drug Resistance*, *14*(October), 208–217. https://doi.org/10.1016/j.ijpddr.2020.10.009

409. Woodford, J., Collins, K. A., Odedra, A., Wang, C., Jang, I. K., Domingo, G. J., Watts, R., Marquart, L., Berriman, M., Otto, T. D., & McCarthy, J. S. (2020). An Experimental Human Blood-Stage Model for Studying Plasmodium malariae Infection. *Journal of Infectious Diseases*, *221*(6), 948–955. https://doi.org/10.1093/infdis/jiz102

410. Xia, J., Wu, D., Sun, L., Zhu, H., Li, K., Zhang, J., Lin, W., Wan, L., Zhang, H., & Liu, S. (2020). Characteristics of imported Plasmodium ovale spp. And Plasmodium malariae in Hubei Province, China, 2014-2018. *Malaria Journal*, *19*(1), 1–8. https://doi.org/10.1186/s12936-020-03337-y

411. Xie, T., Wu, Z., Gu, J., Guo, R., Yan, X., Duan, H., Liu, X., Liu, W., Liang, L., Wan, H., Luo, Y., Tang, D., Shi, H., & Hu, J. (2019). The global motion affecting electron transfer in: Plasmodium falciparum type II NADH dehydrogenases: A novel non-competitive mechanism for quinoline ketone derivative inhibitors. *Physical Chemistry Chemical Physics*, *21*(33), 1805–18118. https://doi.org/10.1039/c9cp02645b

412. Xu, X., Wang, J., Han, K., Li, S., Xu, F., & Yang, Y. (2018). Antimalarial drug mefloquine inhibits nuclear factor kappa B signaling and induces apoptosis in colorectal cancer cells. *Cancer Science*, *109*(4), 1220–1229. https://doi.org/10.1111/cas.13540

413. Yamashita, T., Inaoka, D. K., Shiba, T., Oohashi, T., & Iwata, S. (2018). Ubiquinone binding site of yeast NADH dehydrogenase revealed by structures binding novel competitive- and mixed-type inhibitors. *Scientific Reports*, *November 2017*, 1–2. https://doi.org/10.1038/s41598-018-20775-6

414. Yang, G. Z., Zhu, J. K., Yin, X. D., Yan, Y. F., Wang, Y. L., Shang, X. F., Liu, Y. Q., Zhao, Z. M., Peng, J. W., & Liu, H. (2019). Design, Synthesis, and Antifungal Evaluation of Novel Quinoline Derivatives Inspired from Natural Quinine Alkaloids. *Journal of Agricultural and Food Chemistry*, *67*(41), 11340–11353. https://doi.org/10.1021/acs.jafc.9b04224

415. Yang, M., Derbyshire, M. K., Yamashita, R. A., & Marchler-Bauer, A. (2020). NCBI's Conserved Domain Database and Tools for Protein Domain Analysis. *Current Protocols in Bioinformatics*, *69*(1), 1–25. https://doi.org/10.1002/cpbi.90

416. Yang, Y., Yu, Y., Li, X., Li, J., Wu, Y., Yu, J., Ge, J., Huang, Z., Jiang, L., Rao, Y., & Yang, M. (2017). Target Elucidation by Cocrystal Structures of NADH-Ubiquinone Oxidoreductase of Plasmodium falciparum (PfNDH2) with Small Molecule To Eliminate Drug-Resistant Malaria. *Journal of Medicinal Chemistry*, *60*(5), 1994–2005. https://doi.org/10.1021/acs.jmedchem.6b01733

417. Yman, V., Wandell, G., Mutemi, D. D., Miglar, A., Asghar, M., Hammar, U., Karlsson, M., Lind, I., Nordfjell, C., Rooth, I., Ngasala, B., Homann, M. V., & Färnert, A. (2019). Persistent transmission of plasmodium malariae and plasmodium ovale species in an area of declining plasmodium falciparum transmission in Eastern Tanzania. *PLoS Neglected Tropical Diseases*, *13*(5), 1–16. https://doi.org/10.1371/journal.pntd.0007414

418. Yu, H., & Dalby, P. A. (2020). A beginner's guide to molecular dynamics simulations and the identification of cross-correlation networks for enzyme engineering. In *Methods in Enzymology* (1st ed., Vol. 643). Elsevier Inc. https://doi.org/10.1016/bs.mie.2020.04.020

419. Yu, S.-W., Wang, Y., Frydenlund, D. S., Ottersen, O. P., Dawson, V. L., & Dawson, T. M. (2009). Outer Mitochondrial Membrane Localization of Apoptosis-Inducing Factor: Mechanistic Implications for Release. *ASN Neuro*, *1*(5), AN20090046. https://doi.org/10.1042/an20090046

420. Yu, Y., Krämer, A., Venable, R. M., Simmonett, A. C., Mackerell, A. D., Klauda, J. B., Pastor, R. W., & Brooks, B. R. (2021). Semi-automated Optimization of the CHARMM36 Lipid Force Field to Include Explicit Treatment of Long-Range Dispersion. *Journal of Chemical Theory and Computation*, *17*(3), 1562–1580. https://doi.org/10.1021/acs.jctc.0c01326

421. Yuan, S., Chan, H. C. S., Filipek, S., & Vogel, H. (2016). PyMOL and Inkscape Bridge the Data and the Data Visualization. *Structure*, *24*(12), 2041–2042. https://doi.org/10.1016/j.str.2016.11.012

422. Zhang, X., Xu, J., & Xiao, W.-X. (n.d.). *A New Method for the Discovery of Essential Proteins*. https://doi.org/10.1371/journal.pone.0058763

423. Zhuang, Y., Zhou, X., & Wang, S. (2012). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Plant Systematics and Evolution = Entwicklungsgeschichte Und Systematik Der Pflanzen*, *298*(7), 3389–3402. https://doi.org/10.2503/jjshs.58.977

424. Zou, L., Xue, Y., Jones, M., Heinbockel, T., Ying, M., & Zhan, X. (2018). The Effects of Quinine on Neurophysiological Properties of Dopaminergic Neurons. *Neurotoxicity Research*, *34*(1), 62–73. https://doi.org/10.1007/s12640-017-9855-1

Webservers

UniProtKB. https://www.uniprot.org/help/uniprotkb. Accessed 15 March 2019.

RCSB PDB. https://www.rcsb.org/. Accessed 17 February 2019.

MODELLER. https://salilab.org/modeller/download_installation.html. Accessed 26 February 2019.

National Center for Biotechnology Information. https://blast.ncbi.nlm.nih.gov/Blast.cgi. Accessed 09 February 2019.

PROMALS3D multiple sequence and structure alignment server. http://prodata.swmed.edu/promals3d/promals3d.php. Accessed 05 March 2019.

PROCHECK. https://saves.mbi.ucla.edu/. Accessed 07 March 2019.

3-D Structure Viewer. https://servicesn.mbi.ucla.edu/viewer/?job=9678&p=verify3d. Accessed 08 March 2019.

ProSA-web Protein Structure Analysis. https://prosa.services.came.sbg.ac.at/prosa.php. Accessed 10 March 2019.

Supplementary data

```
>AIF-M1 Human (PDCD8) [AIF-M1]
MFRCGGLAAGALKQKLVPLVRTVCVRSPRQRNRLPGNLFQRWHVPLELQMTRQMASSGASGGKIDNSVLV
LIVGLSTVGAGAYAYKTMKEDEKRYNERISGLGLTPEQKQKKAALSASEGEEVPQDKAPSHVPFLLIGGG
TAAFAAARSIRARDPGARVLIVSEDPELPYMRPPLSKELWFSDDPNVTKTLRFKQWNGKERSIYFQPPSF
YVSAQDLPHIENGGVAVLTGKKVVQLDVRDNMVKLNDGSQITYEKCLIATGGTPRSLSAIDRAGAEVKSR
TTLFRKIGDFRSLEKISREVKSITIIGGGFLGSELACALGRKARALGTEVIQLFPEKGNMGKILPEYLSN
WTMEKVRREGVKVMPNAIVQSVGVSSGKLLIKLKDGRKVETDHIVAAVGLEPNVELAKTGGLEIDSDFGG
FRVNAELQARSNIWVAGDAACFYDIKLGRRRVEHHDHAVVSGRLAGENMTGAAKPYWHQSMFWSDLGPDV
GYEAIGLVDSSLPTVGVFAKATAQDNPKSATEQSGTGIRSESETESEASEITIPPSTPAVPQAPVQGEDY
GKGVIFYLRDKVVVGIVLWNIFNRMPIARKIIKDGEQHEDLNEVAKLFNIHED

>P.falciparum 5JWA_A
MRGSHHHHHHGSNVAKNNLKNNKDIERKEKIIILGSGWGGFNFLLNIDFKKYDVTLISPRNYFTFTPLLP
CLCSGTLSVNVCTESIRNFLRKKNGYCGNYLQLECTDVFYEDKYINCIDIENNKVKLFYDYLIIAVGAKT
NTFNINGVDKYAYFVKDIDDALKIRKKFLDILEKCTLPNISNEEKKKMLHVAVVGGGPTGVEVTAEFADF
INKEVKINYKDIFNFISISIIEGGNNLLPTFTQNISDFTKENFHNLNINVLTNYYVIDVDKHSFHIQSSL
NKNEKKKLSYGLLIWASGLAQTTLIQKFLKTIPVQANNAILKVDEKLRVIGIPSNNIYAIGDCKKIQPKL
LHEHTNEIIKILTGNKLTSEALKLKQSELTKTFPQLSISKWDYEKNKKGEMTPQQFHDYLFEIDKNYKSP
TPTAQNAKQEAYYLSNVFNNFIHTNQKFNIPSFIEKWKGSLAYIGNHQVVADLPYYELKGGRFSSTFWKV
VYIQLLLSWKSRFHFFIDFIKTKWYGRPFIK

>P.knowlesi OTN67496.1
MSMKVRRNGVHSVLKNVKYVYESSRDISTSRIYKDRKEKVVILGSGWGGIHFFINIDFKKYDVTLISPRS
YFTFTPLLPCLCSGTLSAKVCTENVSTFLKKKGSSGKYLQMECTDISPEERQVICRDNKNNEVKIAYDHL
VISVGAKTNSFNIKGVDKHAFFVKDIEGVINIRKRFLDVLDICCTDKISNEEKKKLLHVVVVGGGPTGVE
VAGEFADFINKDVKKKYKNIFPLISVSIIEGGKNLLPTFTQNISDFTKRTFHTANINVLTNYYVKEVDED
TICVQSSLDQNEKKKQIPYGLLIWASGLAQTPLITNFLKKIPEQVNNRILNVNGHLAVIGIKEQNIYAIG
DCKKIQPLQLHQNFHEVLDYFSSSSTTFSSDLLKSKANELSKKFPQVSQSKWDYKKNKKTQMDKHQFCEY
LKEIDENYKSPIPTAQNAKQEAYFLSNLFNTLMDKKADGHQFPSFVEKWKGSIAYIGSHQVVAHLPFFEI
TGGLFSFTFWKMVYIQLLLTWRSRFAFIMDFLRIKFFGRPFSK

>P.malariae SBT70925.1
MLMNVRKNKINNLLRKIKYTYEKNWMNDGKENKKKEKVLVLGSGWGGINFLLNIDFNKYDVTLISPRNYF
TFTPLLPCLCCGTLSVNACRESVSRFLKKNGCSGNYLQLECTDIMYKEKYVKCVDSNNEEVNIMYDYLII
AVGAKTNSFNIKGVEKYALFVKDVDDALRIRRKFLENLEMCTFEKNNVISDEKKKYMLHVVIVGGGPTGV
EVAGEFADFVNKEIRKKYEDIFPLISISIIEGGNNLLPTFTQSISDFTKNNFKNLNINVLTNYYVTEVDE
QYFTIQSSVNKNEKKKIPYGLIIWASGLAQTPLVNNFLKKIPEQVNNKILNVNSQLNVIGINTNNIFAIG
DCKKIQPLQLHQHTNEIIQFLNCSKLTSDTLKIKAEKLSMTFPQVSESKWDYIKNKTGDMNEQQFHDYLY
QIDQHYKSPSPTAQNAKQEARYLSHVFNHILPGKNDQLLIPSFQEKWKGSLAYIGNHQAVANLPFYEIKG
GLFSFTFWRMVYIQLLLTWKTRYFFFVDSLRTRICGRSFVK

>P.ovale SCQ16138.1
MSIRMRKGQFSDIFRKIKYACEKKWKSDKNNGEKKKENVVVLGSGWGGIHFMLNIDFEKYNVTLVSPRNY
FTFTPLLPCLCSGTLSVKSCKENVKNFLKKNNCTSNYLQYECTDIMYNDKYIICKDKNDKQVKIFYDYLV
IAVGAKSNSFNIKGVDKYAFFVKDIEDALKIRNKFLQNLEKCKIKNYSEEKKKNLLHIVVVGGGPTGVEV
AGEFADFVNKEIKNKYKNIYPYISISIIESGNKLLPTFTENISNFTKNNFNILNINVYTNYCVTEIDEYY
FFIQSSLNQNEKKKKIPYGLIIWASGIKQTILINNFLKKIPEQVHNNILNVNSHLNVIGLNTFNIFAIGD
CKIIKPLQLHKYTYEIINFLNTNNLNSNCLKLKSKELCKIFPQVSENKWDYFNNKKDNMNEIEFENYLHK
IDLNYKTPSPTAQNAKQEAYYLSYIFNNYLQNKNNVLIPSFLDTWKGSLAYIGNHQVVAHLPFCEIKGGL
FSFTFWRIVYIQLLLTWKSRLHFLLDTFRTRFYGRPVPK

>P.vivax SGX76671.1
MSLKVRQNGVQSAVKNVKCLCENKREISTSRIYKGGKEKVVILGSGWGGIHFFISIDFKKYDVTLISPRS
YFTFTPLLPCLCSGTLSAKVCTENISTFLRKKGSSGSYLQMECTDILPEERQVICRDNQNNQVKISYDYL
IISVGAKTNSFNIKGVEKYAFFVKDIQGVINIRRRFLDILSICSTDRISNEEKKKLLHIVVVGGGPTGVE
VAGEFADFINKDVKRKYKSIFPFISVSIIEGGNNLLPTFTQNISDFTRKTFRRSNINVLTNYYVTEVDEH
NICVQSSVDTNEERKHIPYGILIWASGLAQTPLITNFLKKIPEQVNNKILNVNGHLAVIGIRQKNIYAIG
```

```
DCKKIQPLQLHEHLNEVLHHFSSSSTTFSSDLLKSKASELSKKFPQVSQSKWDYRKNKRAQMDKQQFWEY
LKQIDQNYKSPTPTAQNAKQEAYFLSNLFNTLVEKKADENHFPSFVEKWKGSIAYIGNHQVVAHLPFFEI
RGGLFSFTFWKMVYIQLLLTWRSRFAFILDFLRTKICGRPFAK

>C.thermarum F5L3B8
MSKPSIVILGAGYGGIVAALGLQKRLNYNEADITLVNKNDYHYITTELHQPAAGTMHHDQARVGIKELID
EKKIKFVKDTVVAIDREQQKVTLQNGELHYDYLVVGLGSEPETFGIEGLREHAFSINSINSVRIIRQHIE
YQFAKFAAEPERTDYLTIVVGGAGFTGIEFVGELADRMPELCAEYDVDPKLVRIINVEAAPTVLPGFDPA
LVNYAMDVLGGKGVEFKIGTPIKRCTPEGVVIEVDGEEEEIKAATVVWTGGVRGNSIVEKSGFETMRGRI
KVDPYLRAPGHENIFIVGDCALIINEENNRPYPPTAQIAIQHGENVAANLAALIRGGSMTPFKPHIRGTV
ASLGRNDAIGIVGGRKVYGHAASWLKKLIDMRYLYLIGGLSLVLKKGRF

>E.coli P00393
MTTPLKKIVIVGGGAGGLEMATQLGHKLGRKKKAKITLVDRNHSHLWKPLLHEVATGSLDEGVDALSYLA
HARNHGFQFQLGSVIDIDREAKTITIAELRDEKGELLVPERKIAYDTLVMALGSTSNDFNTPGVKENCIF
LDNPHQARRFHQEMLNLFLKYSANLGANGKVNIAIVGGGATGVELSAELHNAVKQLHSYGYKGLTNEALN
VTLVEAGERILPALPPRISAAAHNELTKLGVRVLTQTMVTSADEGGLHTKDGEYIEADLMVWAAGIKAPD
FLKDIGGLETNRINQLVVEPTLQTTRDPDIYAIGDCASCPRPEGGFVPPRAQAAHQMATCAMNNILAQMN
GKPLKNYQYKDHGSLVSLSNFSTVGSLMGNLTRGSMMIEGRIARFVYISLYRMHQIALHGYFKTGLMMLV
GSINRVIRPRLKLH

>Mycobacterium tuberculosis L7N5D1
MSPQQEPTAQPPRRHRVVIIGSGFGGLNAAKKLKRADVDIKLIARTTHHLFQPLLYQVATGIISEGEIAP
PTRVVLRKQRNVQVLLGNVTHIDLAGQCVVSELLGHTYQTPYDSLIVAAGAGQSYFGNDHFAEFAPGMKS
IDDALELRGRILSAFEQAERSSDPERRAKLLTFTVVGAGPTGVEMAGQIAELAEHTLKGAFRHIDSTKAR
VILLDAAPAVLPPMGAKLGQRAAARLQKLGVEIQLGAMVTDVDRNGITVKDSDGTVRRIESACKVWSAGV
SASRLGRDLAEQSRVELDRAGRVQVLPDLSIPGYPNVFVVGDMAAVEGVPGVAQGAIQGAKYVASTIKAE
LAGANPAEREPFQYFDKGSMATVSRFSAVAKIGPVEFSGFIAWLIWLVLHLAYLIGFKTKITTLLSWTVT
FLSTRRGQLTITDQQAFARTRLEQLAELAAEAQGSAASAKVAS

>P.aeruginosa AZP62247.1
MSHRIVIVGGGAGGLELATRLGRTLGKRGKAQVTLVDANLTHIWKPLLHEVAAGSLNSSEDELNYVAQAK
WNHFEFQLGRMQGLDRAGKRVRLAALVDENGVELVPERDIAYDTLVIAVGSTTNDFGTEGAAEHCIFLDT
RAQAERFHRQLLGQYLRAHTGHDANNNAVSIAIIGAGATGVELAAELHHAAQMLAAYGLDRIRPEDLHIS
LVEAGPRVLPALPERISQPVHQTLSKLGVRVMTGSAVSKVDADGLWTGDGEFVPATLKVWAAGIRAPAFL
KELDGLESNRINQLVVRPTLQTTRDDDIFAFGDCAACPQPAQPGEQPRNVPPRAQAAHQQASLLVKSLRA
RLEGKPLGEYHYRDYGSLISLSRFSAVGNLMGNLMGSVMLEGWLARMFYVSLYRMHQMALYGPLRTAMLM
LSGRLRSTEPRLKLH

>S.aureus Q2FZV7
MAQDRKKVLVLGAGYAGLQTVTKLQKAISTEEAEITLINKNEYHYEATWLHEASAGTLNYEDVLYPVESV
LKKDKVNFVQAEVTKIDRDAKKVETNQGIYDFDILVVALGFVSETFGIEGMKDHAFQIENVITARELSRH
IEDKFANYAASKEKDDNDLSILVGGAGFTGVEFLGELTDRIPELCSKYGVDQNKVKITCVEAAPKMLPMF
SEELVNHAVSYLEDRGVEFKIATPIVACNEKGFVVEVDGEKQQLNAGTSVWAAGVRGSKLMEESFEGVKR
GRIVTKQDLTINGYDNIFVIGDCSAFIPAGEERPLPTTAQIAMQQGESVAKNIKRILNGESTEEFEYVDR
GTVCSLGSHDGVGMVFGKPIAGKKAAFMKKVIDTRAVFKIGGIGLAFKKGKF
```

```
>A.thermomutatus RHZ55328.1
MASPSVMQLSSLSRRSLSTRSPALALRSSRPLRSGNFSQQSVQRTFRRSYADAAPAPKPKKRFRFLRWTW
RLTWLSGVGLAGTLIYSIYEQRHPIEQIEPDPTKKTLVILGTGWGSVSLLKKLDTENYNVVVISPRNYFL
FTPLLPSCTTGQVEHRSIMEPIRNILRQKKAHVKFYEAEATKIDYEKRVVYISDDSEIKGDISHTEVPFD
MLVVGVGAENATFGIKGVKEHSCFLKEVGDAQKIRKRIMDCVETAMFKDQPEEEVNRLLHMVVVGGGPTG
VEFAGELQDFFNEDLRKWIPEIKDHFHVTLVEALPNVLPMFSKQLIDYTESTFKEEAITIRTKTMVKNVT
DKYIEAEVTKPDGTKELETIPYGLLVWATGNAVRNVVRDLMNQIPAQKNSRRGLAVNEYLVVNGAENVWA
VGDCAVTNYAPTAQVASQEGAFLARLFNTMAKTEAIEKELKKLSEAQAQAKNEEERNQIFDEIRERQKQL
RRTKQIGPFQYSHQGSLAYIGKERAVADISWLSGNIASGGTVTYLFWRSAYLSMCFSTRNRVLVAADWVK
AKLFGRDVSRE

>A.turcosus RLL97513.1
MASPSVMQLSSLSRRSLSTRSPALALRLARPLRSGNFSQQFVQRTFRRSYADAAPAPKPKKRFRFLRWAW
RLTWLSGVGLTGALVYSIYEARHPIEQIEPDPTKKTLVILGTGWGSVSLLKKLDTENYNVVVISPRNYFL
FTPLLPSCTTGQVEHRSIMEPIRNILRLKKAHVKFYEAEATKIDYEKRVVYISDDSEIKGDISHTEVPFD
MLVVGVGAENATFGIKGVKEHSCFLKEVGDAQKIRKRIMDCVETAMFKDQPEEEVKRLLHMVVVGGGPTG
VEFAGELQDFFNEDLKKWIPEIKDNFHVTLVEALPNVLPMFSKQLIDYTESTFKEEAITIRTKTMVKNVT
DKYIEAEVTKPDGTKQLETIPYGLLVWATGNAVRNVVRDLMNQIPAQKDSRRGLAVNEYLVVNGTENVWA
VGDCAVTNYAPTAQVASQEGAFLARLFNTMAKTEAIENELKKLSEAQAQAKNEEERNQIFDEIRERQKQL
RRTKQIGPFQYSHQGSLAYIGKERAVADISWLSGNIASGGTVTYLFWRSAYLSMCFSTRNRVLVTADWLK
AKIFGRDVSRE

>C.coralloides QAT82924.1
MNTTRENLPHVVILGGGFGGLYAARYLRKAGVRVTMVDRHNHHLFQPLLYQVATATLSPSDIAAPLRAML
GREHVQVLLAEVTGVDTARKRVLLADGELAYDFLIVATGATHSYFGNDAWSRHSMGLKTVEDALEIRRRV
LLAFEQAEREPDPERRRALLTFAIIGAGPTGVELAGALAEISRNSLSGDFQNIDPRDARVILIEGMDRVL
PTYPENLSAEARRVLVGLGVEVRTGTRVTNIDAAGVDMGPEHLAARTVLWAAGVEASPVARSLGVALDRA
GRVPVTPELTVPGHPDIFVVGDLALVMQEDGSAVPGVAPAAMQEGKHAVLNLRRQLAGQPMQPFHYWDRG
TYAVIGRGHAVGVAFRRVKQSGFVAWLAWLFIHITFLIGFRSKLAVLLNWAYAYLTFGRSARIITGPAPR
LEERVVQPLPASGRSAQVERSPETSIQLGVLVPDKSGTG

>N.crassa Q7S1W8
MASTEPQSTLDAVPEPQSTTAPAPTASTPTTPLSTSIVILGGNLAGLGVLHTLARHTIPALQRLPSQPQY
KITLITPNSHFFFKPASPRALIHPDLLPGGEKQIFRDLSDAVKQYGELVTVVKAYATAVDTSTKEVTIKH
VHTNSPSDSEGTSAVRYDVLVIATGTKARSAAWTMPGQGTDGLHAAEDATKREWAGIRTKLKTLQAGTHP
HHSILIAGGGPVGVETTGEIATLLKAMGKTEGVTITLLSGSDRLLNRNCNESLGRKAEAYLKKNFGSLVE
VKHGGIKVVSVSDSVIEGGEQTAETTVELSDGTTRTVGLYIDATGGSGNADGFLPPTWLDASSRVITKDD
FFRVRGSGSSEETDHQADGIYAVGDVVSGSDNTLISTNFQVPVVCSSIGVDLATKILHPEGGAGGQTSGV
KLPAALKQKSFKNKMAGTLLVPIGPGGGVGQILGWGVPSLMVKTAKAKNFLVDMVEPAVTGSQWK

>S.cerevisiae S288C NP_010198.1
MLPRLGFARTARSIHRFKMTQISKPFFHSTEVGKPGPQQKLSKSYTAVFKKWFVRGLKLTFYTTLAGTLY
VSYELYKESNPPKQVPQSTAFANGLKKKELVILGTGWGAISLLKKLDTSLYNVTVVSPRSFFLFTPLLPS
TPVGTIEMKSIVEPVRSIARRTPGEVHYIEAEALDVDPKAKKVMVQSVSEDEYFVSSLSYDYLVVSVGAK
TTTFNIPGVYGNANFLKEIEDAQNIRMKLMKTIEQASSFPVNDPERKRLLTFVVVGGGPTGVEFAAELQD
YINQDLRKWMPDLSKEMKVILIEALPNILNMFDKTLIKYAEDLFARDEIDLQVNTAVKVVEPTYIRTLQN
GQTNTDIEYGMLVWATGNEPIDFSKTLMSRIPEQTNRRGLLINDKLELLGSENSIYAIGDCTAHTGFFPT
AQVAHQEGEYLAKILKKLQIEQLEWDMLNSIDEIEVSPLQKEVNLPKSKLPKENYKHMGALAYIGSETA
```

```
>N.crassa Q7S1W8
MASTEPQSTLDAVPEPQSTTAPAPTASTPTTPLSTSIVILGGNLAGLGVLHTLARHTIPALQRLPSQPQY
KITLITPNSHFFFKPASPRALIHPDLLPGGEKQIFRDLSDAVKQYGELVTVVKAYATAVDTSTKEVTIKH
VHTNSPSDSEGTSAVRYDVLVIATGTKARSAAWTMPGQGTDGLHAAEDATKREWAGIRTKLKTLQAGTHP
HHSILIAGGGPVGVETTGEIATLLKAMGKTEGVTITLLSGSDRLLNRNCNESLGRKAEAYLKKNFGSLVE
VKHGGIKVVSVSDSVIEGGEQTAETTVELSDGTTRTVGLYIDATGGSGNADGFLPPTWLDASSRVITKDD
FFRVRGSGSSEETDHQADGIYAVGDVVSGSDNTLISTNFQVPVVCSSIGVDLATKILHPEGGAGGQTSGV
KLPAALKQKSFKNKMAGTLLVPIGPGGGVGQILGWGVPSLMVKTAKAKNFLVDMVEPAVTGSQWK

>S.cerevisiae S288C NP_010198.1
MLPRLGFARTARSIHRFKMTQISKPFFHSTEVGKPGPQQKLSKSYTAVFKKWFVRGLKLTFYTTLAGTLY
VSYELYKESNPPKQVPQSTAFANGLKKKELVILGTGWGAISLLKKLDTSLYNVTVVSPRSFFLFTPLLPS
TPVGTIEMKSIVEPVRSIARRTPGEVHYIEAEALDVDPKAKKVMVQSVSEDEYFVSSLSYDYLVVSVGAK
TTTFNIPGVYGNANFLKEIEDAQNIRMKLMKTIEQASSFPVNDPERKRLLTFVVVGGGPTGVEFAAELQD
YINQDLRKWMPDLSKEMKVILIEALPNILNMFDKTLIKYAEDLFARDEIDLQVNTAVKVVEPTYIRTLQN
GQTNTDIEYGMLVWATGNEPIDFSKTLMSRIPEQTNRRGLLINDKLELLGSENSIYAIGDCTAHTGFFPT
AQVAHQEGEYLAKILDKKLQIEQLEWDMLNSTDETEVSRLQKEVNLRKSKLDKFNYKHMGALAYIGSETA
IADLHMGDSSYQLKGMFAFLFWKSAYLAMCLSIRNRILIAMDWTKVYFLGRDSSV

>Y.lipolytica VBB82463.1
MLRLRPAVRAVSVARSVALTRSLHVSVAKFNKIEGTAPAGLPKEVKQTAGHQGHHQEIPKPDENHPRRKK
FHFWRSLWRLTYLSAIASLGYIGYRIYVIRNPSDQLPADPSKKTLVVLGSGWGSVSFLKKLDTSNYNVIV
VSPRNYFLFTPLLPSCPTGTIEHRSIMEPIRGIIRHKQAECQYLEADATKIDHEKRIVTIRSAVSENSKE
EVIKEIPFDYLVVGVGAMSSTFGIPGVQENACFLKEIPDAQQIRRTLMDCIEKAQFEKDPEVRKRLLHTV
VVGGGPTGVEFAAELQDFFEDDLRKWIPDIRDDFKVTLVEALPNVLPSFSKKLIDYTEKTFSDEKISILT
KTMVKSVDENVIRAEQTKGDGTKETLEMPYGTLVWATGNTVRPVVRELMSKIPAQKGSRRGLLVNEYLVV
EGTEGIWALGDCSATKYAPTAQVASQEGSYLANLLNGIAKTEDLNNEITNLEKQSEHTFDEQERKNIFAQ
LESKSRKLRRSRAMLPFEYSHQGSLAYIGSDRAVADLSFNFWGIMNWSSGGTMTYYFWRSAYVSMCFSMR
NKILVCIDWMKVRVFGRDISRE
```

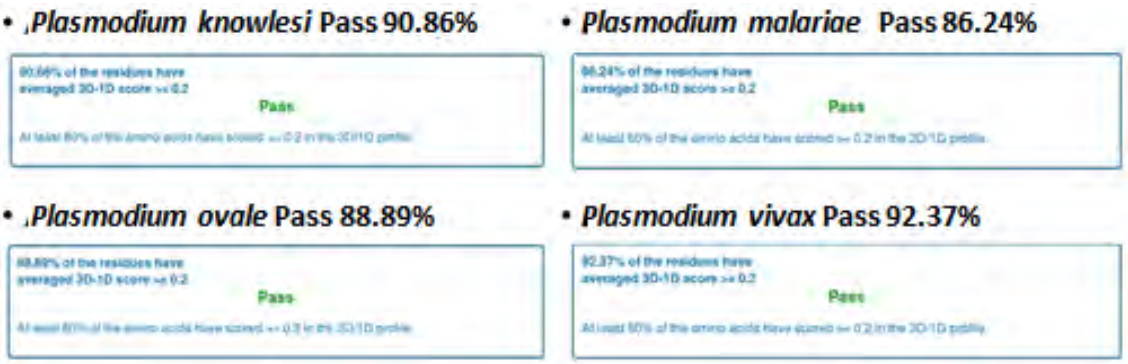

Figure S1 Verify3-D output results for the homology modelled structures

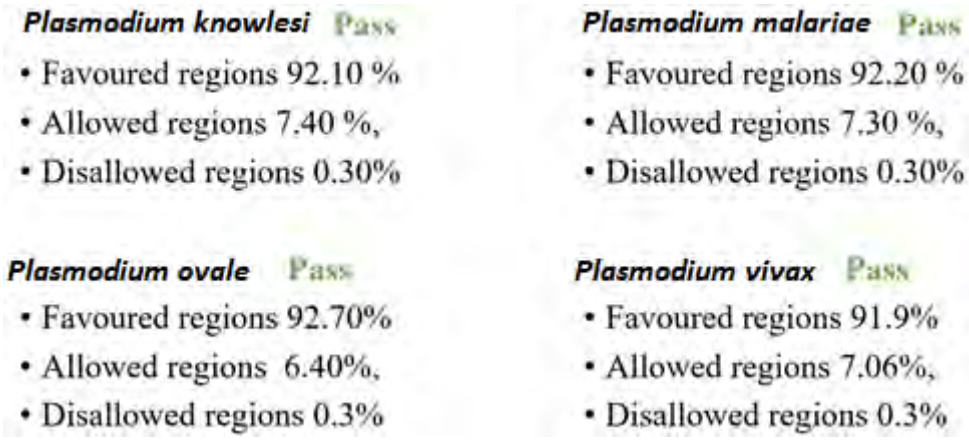

Figure S2 PROCHECK output results for the homology modelled structures

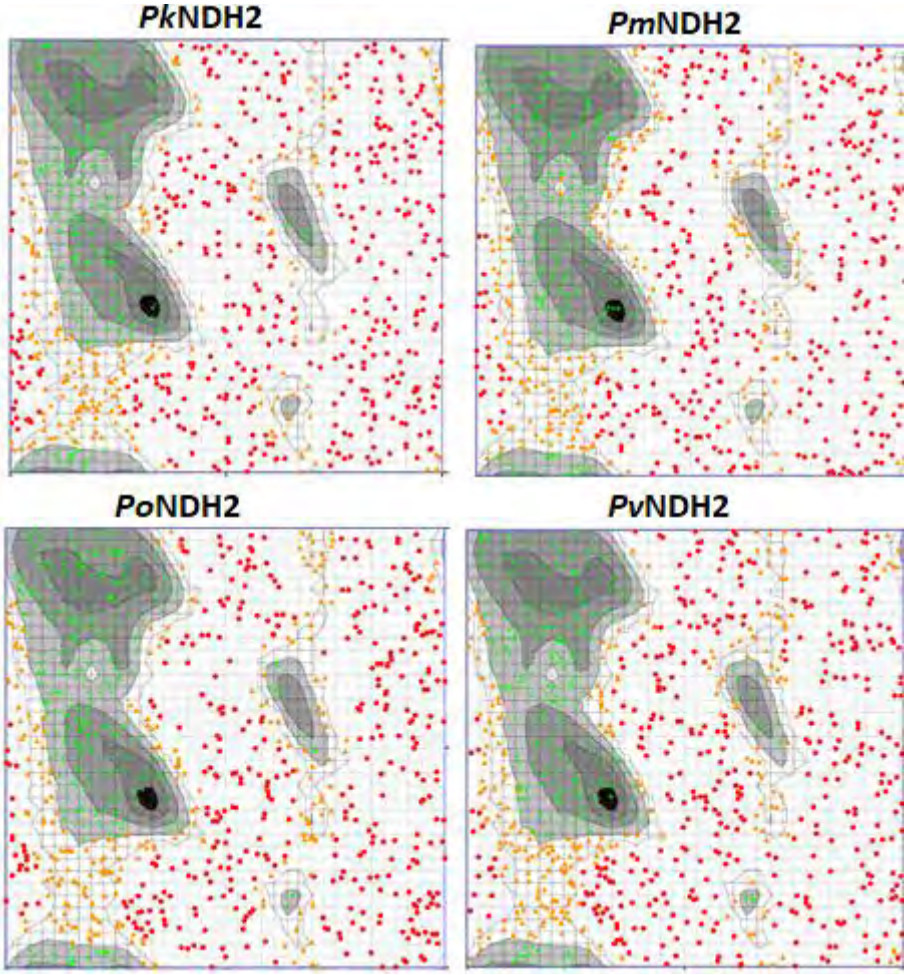

Figure S3 Ramachandran Plots of the Homology models, black, dark grey, grey, light grey represents highly preferred conformations, Delta >= -2. White with black grid represents preferred conformations -2> Delta >= -4. White with grey grid represents questionable conformations Delta < -4. Highly preferred observations shown as green crosses. Preferred observations shown as brown triangles and questionable observations shown as red circles.

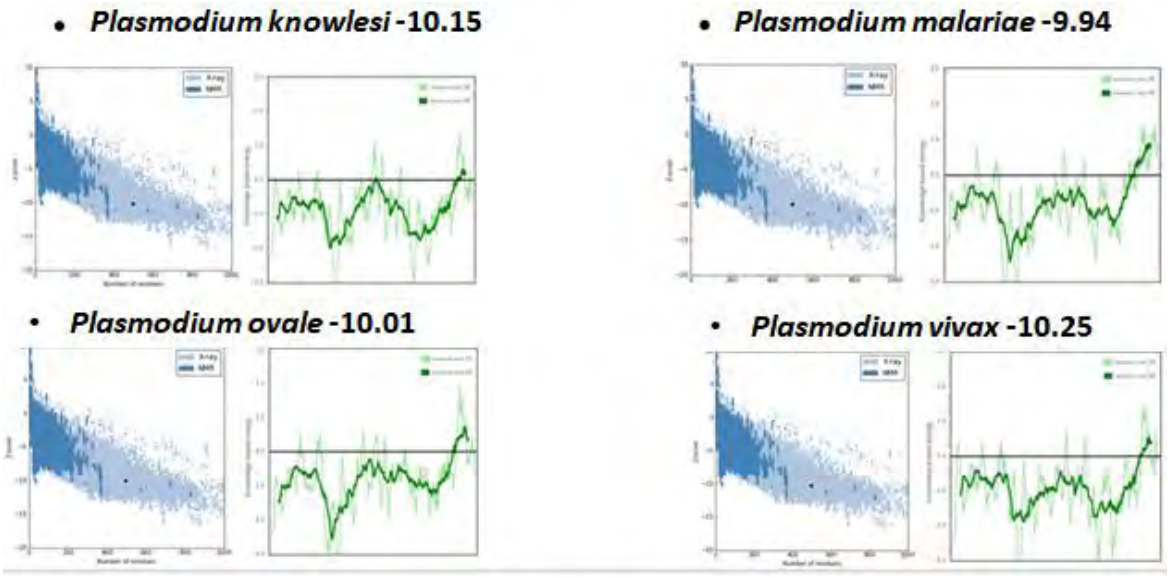

Figure S4 PROSA output results for the homology modeled structures

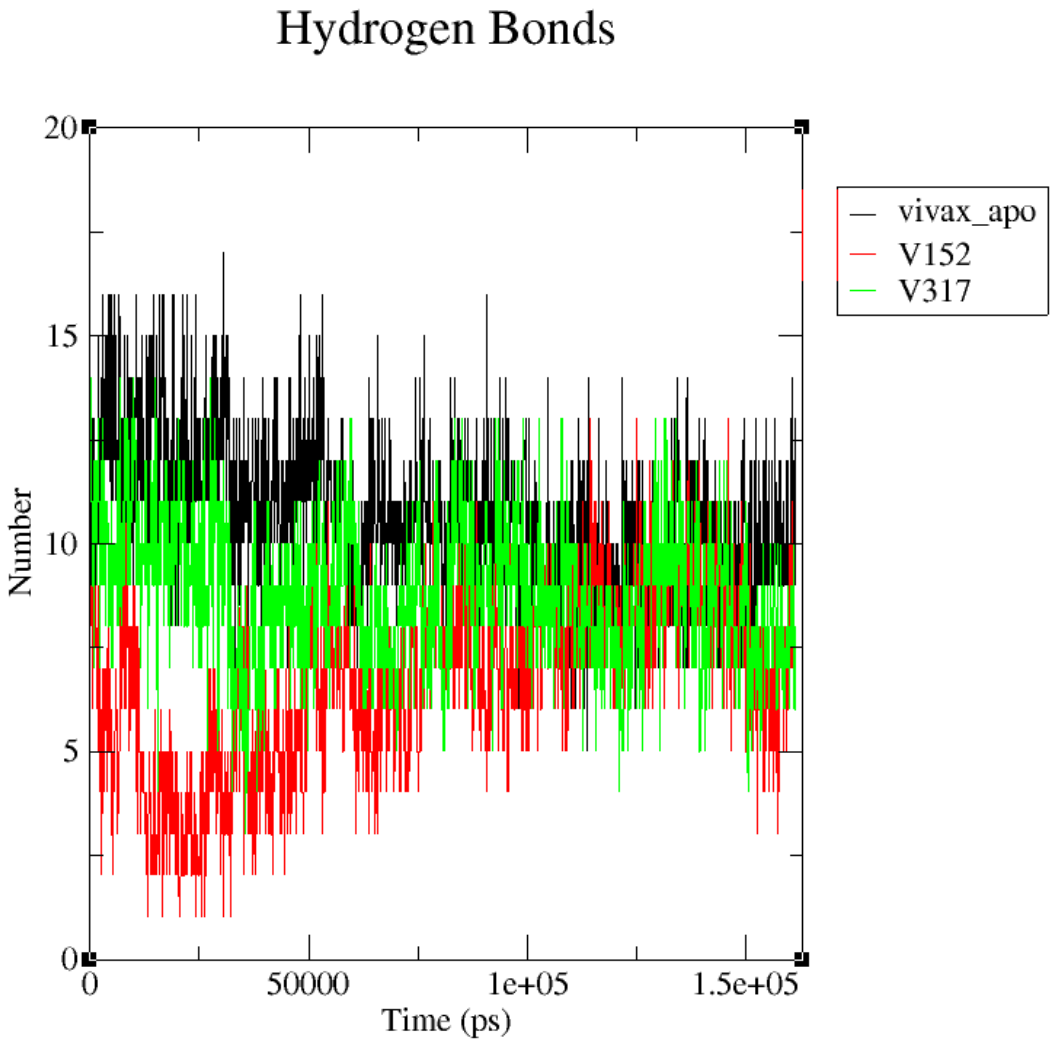

Figure S5 The total number of hydrogen bonds in each time step for the *Pv*NDH2_*apo*, *Pv*NDH2_152 (V152) and *Pv*NDH2_317 (V317)