Investigating plasmepsin flexibility as a function of the flap region - a unique structural and dynamic feature of aspartic protease.

Lara McGillewie

214585796

A thesis submitted to the College of Health Sciences, University of KwaZulu-Natal, Westville, in fulfillment of the requirements of the degree of Doctor of Philosophy.

Supervisor

Professor Mahmoud Soliman

KwaZulu-Natal

2015

Investigating plasmepsin flexibility as a function of the flap region - a unique structural and dynamic feature of aspartic protease.

Lara McGillewie

214585796

2015

A thesis submitted to the School of Pharmacy and Pharmacology, Faculty of Health Science, University of KwaZulu-Natal, Westville Campus, for the Doctorate of Medical Sciences (Pharmaceutical Chemistry).

This is a thesis in which the chapters are written as a set of discrete research publications, with an overall introduction and final summary.

This is to certify that the contents of this thesis are the original research work of Miss Lara McGillewie.

As the candidate's supervisor, I have approved this thesis for submission.

Supervisor:

Signed: _____

Name: **Dr. Mahmoud E. Soliman**

Date: _____

Abstract

Malaria is one of the most deadly infectious protozoan diseases known to man. It is spread by the *Plasmodium* parasite through the bite of the female *Anopheles* mosquito. Increasing resistance to currently available antimalarial drugs is a growing concern. Plasmepsins are malarial aspartic proteases, due to their characteristic mechanism of action, the fact that they are found in all *Plasmodium* species and are essential to parasitic survival they represent novel targets in the design of antimalarials. A unique structural feature of aspartic proteases and plasmepsins is the flap region lying perpendicular to the catalytic aspartic acid active, partially covering the active site. The flap region plays an important structural (and kinetic) role in regulating access to the active site, thereby regulating ligand binding.

The present study focused on the flap dynamics of Plm I – V, proposing and validating parameters to accurately quantify the dynamic behaviour of the flap region. The catalytic aspartic acids is highly conserved in the plasmepsin family; sequence analysis revealed that although all plasmepsins are similar in structure, they differ greatly in the residues in the flap region. The heterogeneity in this region gives each plasmepsin unique substrate specificity and response to inhibitors. The parameters proposed in the present study gives a detailed account for the twisting of the flaps which move away from the active site in the absence of an inhibitor. Upon inhibitor binding, residues in the flap region form hydrogen bonds with the inhibitor pulling it inward towards the active site rendering the enzyme inactive. The parameters proposed in the present study will be of great value in the design of novel plasmepsin inhibitors, with increased efficacy and potency.

DECLARATION 1 – PLAGIARISM

I, **Lara McGillewie,** declare that

1. The research reported in this thesis, except where otherwise indicated, is my original work.

2. This thesis has not been submitted for any degree or examination at any other university.

3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a. Their words have been re-written, but the general information attributed to them has been referenced.

b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

5. This thesis does not contain text, graphics or tables copied from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the references section.

A detail contribution to publications that form part and/or include research presented in this thesis is stated (include publications submitted, accepted, in *press* and published).

Signed: _____

DECLARATION 2 - LIST OF PUBLICATIONS

1. Wilson Karubui, Soumendranath Bhakat, Lara McGillewie and Mahmoud E Soliman (2015). Flap dynamics of plasmepsin proteases: insight into proposed parameters and molecular dynamics, *Molecular Biosystems* (***Published***).

Contributions:

Wilson Karubui and Soumendranath Bhakat: Authors contributed to the project by performing all literature reviews, experimental work and data analysis, interpretation of the results as well as manuscript preparation and writing.

Lara McGillewie: Literature review, technical support and editing of the manuscript.

Mahmoud E. Soliman: Supervisor.

2. Lara McGillewie and Mahmoud E Soliman (2015). Flap flexibility amongst plasmepsins I, II, III, IV, and V: Sequence, structural, and molecular dynamics analyses, *Proteins* (***Published***).

Contributions:

Lara McGillewie: Author contributed to the project by performing all literature reviews, experimental work, and data analysis, interpretation of the results as well as manuscript preparation and writing.

Mahmoud E. Soliman: Supervisor.

3. Lara McGillewie and Mahmoud E Soliman (2015). The binding landscape of plasmepsin V and the implications on flap dynamics. (***Submitted***).

   Contributions:

   Lara McGillewie: Author contributed to the project by performing all literature reviews, experimental work, and data analysis, interpretation of the results as well as manuscript preparation and writing.

   Mahmoud E. Soliman: Supervisor.

4. Lara McGillewie and Mahmoud E Soliman (2015). Flap dynamics as unique conformational features among aspartic proteases. (***Submitted***).

   Contributions:

   Lara McGillewie: Author contributed to the project by performing all literature reviews, experimental work, and data analysis, interpretation of the results as well as manuscript preparation and writing.

   Mahmoud E. Soliman: Supervisor.

RESEARCH OUTPUTS

A. PUBLICATIONS

i. Karubui W, Bhakat S, McGillewie L and Soliman M.E (2015). Flap dynamics of plasmepsin proteases: insight into proposed parameters and molecular dynamics. *Molecular Biosystems* (***Published***).

ii. McGillewie L and Soliman M.E (2015). Flap flexibility amongst plasmepsins I, II, III, IV, and V: Sequence, structural, and molecular dynamics analyses. *Proteins* (***Published***).

iii. McGillewie L and Soliman M.E (2015). The binding landscape of plasmepsin V and the implications on flap dynamics (***Submitted***).

iv. McGillewie L and Soliman M.E (2015). Flap dynamics as unique conformational features among aspartic proteases. (***Submitted***).

B. CONFERENCES

i. Oral presentation: 'Flap flexibility amongst plasmepsins I, II, III, IV, and V: Sequence, structural, and molecular dynamics analyses.' UKZN College of Health Sciences Symposium September 2015 (awarded 3rd prize).

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincerest gratitude and appreciation to my supervisor, Prof Soliman, whose guidance, expertise and understanding has mentored me throughout my thesis. His motivation, support and encouragement has made all this possible, thank you Prof.

I would also like to thank the members and friends in the Molecular Modelling and Drug Design research group for all the laughs, all the support and encouragement.

To the Centre of High Performance Computing in Cape Town, as well as the San Diego Computer Centre for computational resources. And to the School of Health Sciences, UKZN Westville Campus.

To the National Research Foundation (NRF) who funded me through my thesis.

And lastly to my family, to my mom and dad, Helena and David, you have made this dream a reality. Your unwavering love and support knows no limits or boundaries. Thank you for being the most incredible parents, I am eternally grateful. To my brother, Dane, for always believing in me no matter the odds. And last but not least to my Daniel, thank you for walking this journey with me, for always encouraging and supporting me. Love you all more than words could ever express.

TABLE OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

LIST OF TABLES

Chapter 5

Chapter 7

List of abbreviations

ACE	Angiotensin converting enzyme
ACT	Artemisinin combination therapy
AD	Alzheimer's disease
AIDS	Acquired immune deficiency syndrome
APP	Amyloid precursor protein
ARB	Angiotensin II receptor blockers
Cat-D	Cathepsin D
DCCM	Dynamic Cross-Correlation Matrices
CDC	Centre for disease control
DPAP1	Dipeptidyl aminopeptidase
ER	Endoplasmic reticulum
FV	Food vacuole
GAFF	General AMBER Force Field
HAP	Histo-aspartic acid
HIV	Human immunodeficiency virus
MD	Molecular dynamics
MM	Molecular mechanics
PCA	Principle component analysis
PES	Potential surface energy
PEXEL	*Plasmodium* export element
Plm	Plasmepsin
PR	Protease
PVM	Parasitophorous vacuolar membrane
QM	Quantum mechanics
RAAS	Renin-angiotensin-aldosterone system
RBC	Red blood cell
SASA	Solvent accessible surface area
WHO	World health organisation

Table of contents

CHAPTER 1

1. Background and rationale for the present study

The current chapter covers the background and novelty of the research project related to malaria and a brief overview of malarial aspartic proteases, plasmepsins, as potential antimalarial targets.

Malaria is one of the most life threatening diseases known to mankind, according to The United Nations Children's Emergency Fund (UNICEF) "malaria kills one child every 30 seconds, about 3000 children every day." [1]. Malaria is caused by the obligate intracellular *Plasmodium* parasite that is passed on from person to person through the bites of infected female mosquitos (*Anopheles*) [2]. The parasite's lifecycle is complex, broadly broken down into a sexual phase (mosquito) and an asexual phase (vertebrate host) [3]. While an infected mosquito feeds, the parasites are released into the host's circulation through the mosquito's saliva, where they move to the liver and infect hepatocytes [4]. In the hepatocytes, the parasites proliferate exponentially eventually causing the cells to rupture upon which thousands of parasites are released into the bloodstream where they invade and infect red blood cells (RBC) [3].

Currently there are several drugs used in clinical practice to treat malaria infection, although increasing resistance of the parasite to antimalarials and of the mosquito vector to insecticides is of growing concern. In July 2015, GlaxoSmithKline' Mosquirix (RTS,S) received positive feedback from the European Medicines Agency, the first antimalarial vaccine to reach this stage [5,6]. In spite of the progress that has been made in recent years in the fight against malaria, there is a dire need for more potent drugs to treat this life threatening infection, with fewer side effects, that are more efficacious, less susceptible to resistance and effective across the *Plasmodium* family. According to the Centre of Disease Control (CDC), drug resistance is one

of the "greatest challenges facing malaria control" [7]. Essentially, drug resistances is responsible for the spread of malaria to new areas and for the re-emergence of the disease in area where it has previously been eradicated. It also has a significant impact on the occurrence and severity of epidemics across the world [7]. We are at the stage where the *Plasmodium* parasite (only *P. falciparum* and *P. vivax*) has shown some degree of resistance to almost every drug used in clinical practice [7]. In the face of increasing resistance, current treatment regimens rely on the combination of two or more actives, therefore no one compound is used as monotherapy with the aim of reducing the spread of antimalarial resistance [8,9].

The first line treatment for uncomplicated malaria is artemisinin combination therapy (ACTs) which is the combination of an artemisinin derivative (artemether, dihydroartemisinin or artesunate) to a partner drug [9]. Artemisinin, also known as qinghao, is a naturally occurring agent isolated from the leaves of the *Artemisia annua* (sweet wormwood) [10,11]. Artemisinin derivatives are semi-synthetic peroxidases synthesised from artemisinin that are effective to all *Plasmodium* species [9]. Partner drugs are usually derived from the natural compound quinine which is from the bark of the cinchona tree. Quinine derivatives can be divided into amino-alcohols (closest chemical resemblance to quinine) such as mefloquine and lumefantrine and 4-aminoquinolines (synthetic mimics) such as chloroquine, piperaquine, pyronaridine and amodiaquine (**Figure 1 and 2**) [9]. In the combination approach, artemisinin is responsible for reducing the parasitic load in the first three days, whereas the partner drug is responsible for eliminating the parasite left over [11]. Resistance to both artemisinin and quinine is on the rise, in February 2015 five countries in the Greater Mekong sub-region of the Asia Pacific has been confirmed resistant to artemisinin and more recently resistance has emerged in sub-Saharan Africa [12,13]. Chloroquine resistance emerged in the 1950s in Southeast Asia, by 1978 chloroquine resistance had spread to South Asia and East Africa and eventually to sub-Saharan

Africa where it drastically increased the mortality in malaria infected children [13,14]. With the resistance to antimalarials and to insecticides increasing, it is of paramount importance to identify essential targets in the search of new antimalarials.

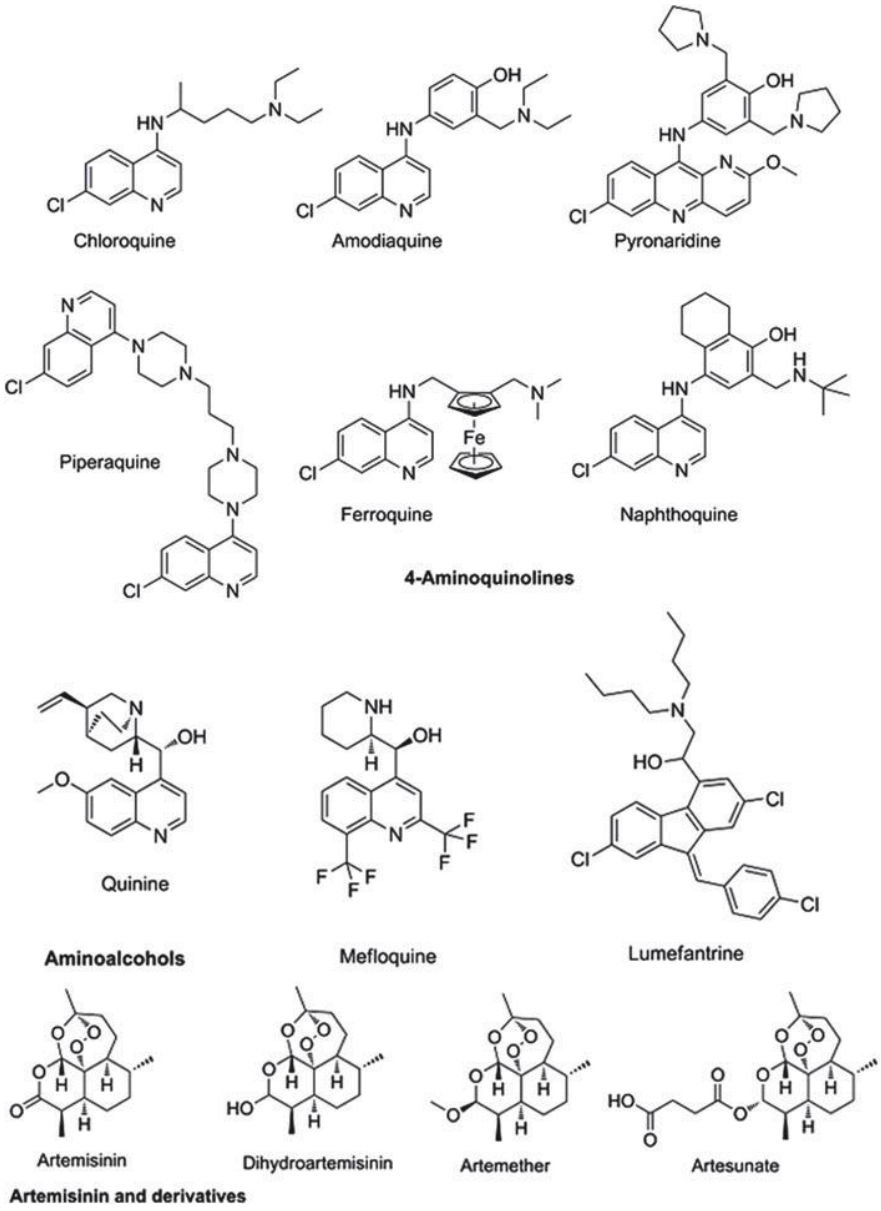


Figure 1. Structures of the antimalarial drugs used to treat the asexual blood stage of the Plasmodium lifecycle [9].

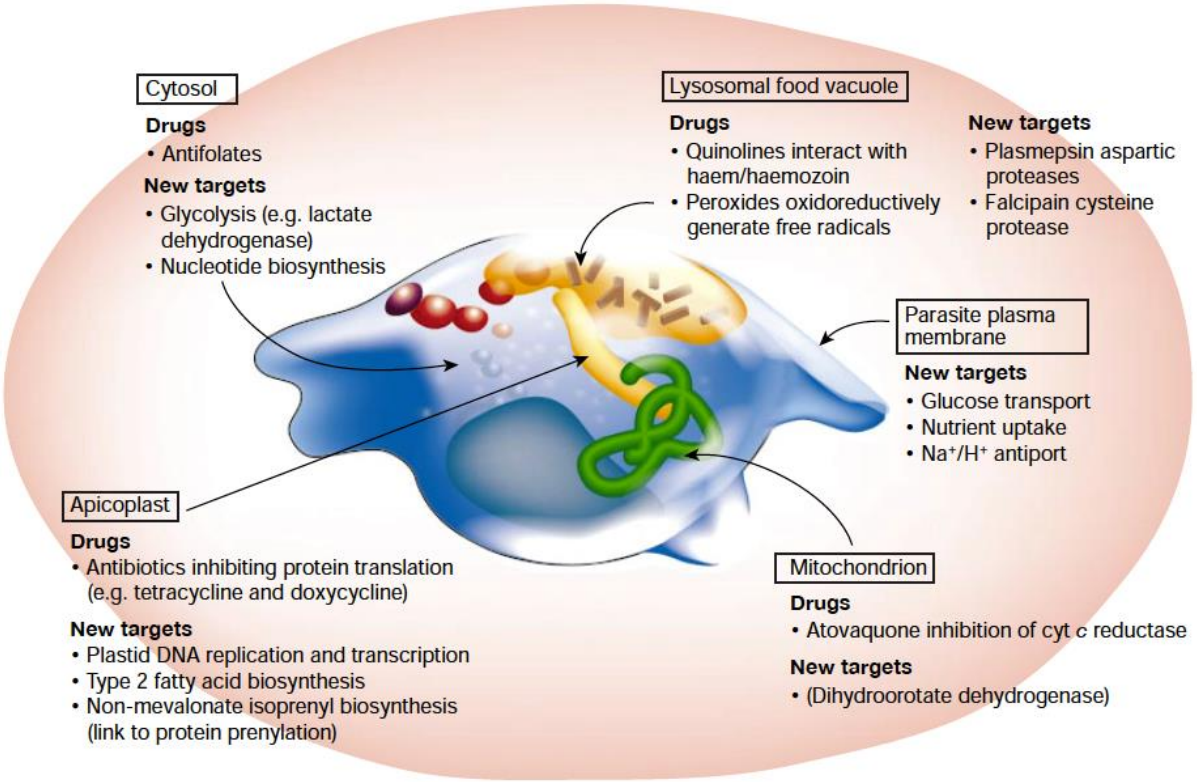


Figure 2. Current antimalarials and their intracellular targets and pathways they disrupt, and potential antimalarial targets [15].

In the infected RBC, hemoglobin degradation is crucial to the survival of the *Plasmodium* parasite and occurs in the acidic food vacuole (FV) [16,17]. Numerous enzymes are involved in the breakdown of hemoglobin, and dual proteinase families with partially redundant overlapping roles are required [17,18]. Plasmepsins (Plm) are aspartic proteases found in malaria, sequencing of the *P. falciparum* genome has led to the identification of ten malarial aspartic proteases, Plm I-X [19,20]. No registered drugs targeting plasmepsin, or other proteases.

2. Aims and objectives

The aim of the present study was to better describe and investigate the dynamic behaviour of the flap and flexible loops covering the active site in plasmepsins, specifically Plm I-V, and

how inhibitor binding alters these dynamic motions. The aims are outlined in more detail hereunder:

1. Define parameters that accurately account for and describe the flexibility and dynamics of the flap region:
    1.1. Extensive review of the literature available on all aspartic proteases, and identify which parameters have previously been used and how they have been used
    1.2. Define additional parameters that we believe would more accurately account for the dynamics of this region in the plasmepsin family
    1.3. Use these parameters in a proof of concept study to test the validity and relevance of the proposed parameters in a continuous 50 ns molecular dynamics (MD) simulation of PlmII as a prototype.
2. To investigate the flap dynamics of apo plasmepsins I-V and how sequence heterogeneity affects the flexibility of the flap region:
    2.1. To ascertain the impact of sequence heterogeneity on flap motions, using sequence and visual analysis.
    2.2. To use available crystal structures and a homology model in continuous 50 ns MD simulations (all structures used were in the apo conformation, thereby establishing a base line for comparison).
    2.3. To use post-MD analysis to understand the flap dynamics and motions of apo Plm I-V.
    2.4. To validate and implement previously proposed parameters to measure the flap dynamics across plasmepsins investigated in the present study.
    2.5. To investigate which plasmepsin is the most flexible is as defined computationally, and correlate findings to flap sequence, biological activity and function.

3. To better understand the binding landscape and impact on flap dynamics of PlmV upon inhibitor binding:
    3.1. To use the crystal structure of PlmV in complex with the inhibitor, WEHI-842, in 50 ns continuous MD simulations.
    3.2. To use post-MD analysis to investigate the dynamics of apo PlmV compared to WEHI-842 bound PlmV.
    3.3. To use the additionally defined parameters to quantify the impact of ligand binding on PlmV as a whole and with respect to the impact ligand binding had on the flap dynamics and flexibility.

3. Novelty and significance of the present study

Unique to the aspartic protease family of enzymes, is a flap region partially covering the active site. The importance of flap dynamics and flexibility in enzyme function is well documented in the literature. The flaps play a crucial role in ligand recognition and in regulating ligand access into the active site. Accurately quantifying the flexibility and dynamics of the flap region is quintessential to the design of novel inhibitors, with improved efficacy and potency. In the case of HIV protease, disrupting the flexibility and mobility of the flap regions leads to an overall inhibition of the enzyme rendering it inactive. The parameter (distance between the flap tip and Leu292 in PlmII) previously used to measure the dynamics of the flap regions in plasmepsins is nothing short of inadequate. To the best of our knowledge, the present study was the first of its kind to accurately define parameters that quantify the dynamic behaviour of the flap region of plasmepsins. Furthermore, how this flexibility varies between plasmepsins and how the dynamics are affected upon ligand binding. The computational parameters proposed and used in the present study could potentially aid in the development of plasmpesin inhibitors as antimalarials. Understanding the dynamic behaviour in this class of enzymes will

assist in the development of potent inhibitors. Additionally, a concise review focused on the flap region, flap dynamics and parameters used to quantify and measure the flexibility of these regions among aspartic proteases is also part of the present study.

4. Overview of thesis

This thesis is divided into seven chapters, including this chapter:

Chapter 2

Gives a general overview on malaria and the current therapies used to fight the malaria endemic. Aspartic proteases, more specifically plasmepsins, are discussed in more detail.

Chapter 3

Provides a general overview of computational chemistry, and focuses on the theories of the molecular modelling methods used in the work presented herein.

Chapter 4 (Published work – this chapter is presented in the required format of the journal and is the final revised accepted version)

This chapter is dedicated to the research paper "Flap dynamics of plasmepsin proteases: insight into proposed parameters and molecular dynamics". In which the parameters to measure flap dynamics was tested and validated.

Chapter 5 (Published work – this chapter is presented in the required format of the journal and is the final revised accepted version)

This chapter is dedicated to the research paper "Flap flexibility amongst plasmepsins I, II, III, IV, and V: Sequence, structural, and molecular dynamics analyses". In which the flap dynamics of apo PlmI-V were analysed. These parameters accurately account for the flexibility of the

active site, of which PlmIV and V were found to be the most dynamic, with the most flexible flap regions.

Chapter 6 (Submitted – this chapter is presented in the required format of the journal and is the submitted version)

This chapter is dedicated to the research paper "The binding landscape of plasmepsin V and the implications on flap dynamics". In which the flap dynamics of unbound PlmV was compared to that of bound PlmV, and to assess the impact of ligand binding on the flexibility of the flap regions; and to analyse the impact of ligand binding on the overall conformation of PlmV.

Chapter 7 (Submitted – this chapter is presented in the required format of the journal and is the submitted version)

This chapter is dedicated to the review "Flap dynamics as unique conformational features among aspartic proteases". Which focuses on aspartic proteases, and the structurally unique flap that partially covers the active site. And presents the parameters currently used to measure the dynamic behaviour of the flap region and its importance in enzyme function and ligand recognition.

5. References

(1) http://www.unicef.org/health/files/health_africamalaria.pdf date accessed 22 November 2015.

(2) White, N. J., Pukrittayakamee, S., Hien, T. T., Faiz, M. A., Mokuolu, O. A., and Dondorp, A. M. (2014) Malaria. *Lancet 383*, 723–35.

(3) Francis, S. E., Sullivan, D. J., and Goldberg, and D. E. (1997) HEMOGLOBIN METABOLISM IN THE MALARIA PARASITE *PLASMODIUM FALCIPARUM*. *Annu. Rev. Microbiol. 51*, 97–123.

(4) http://www.who.int/mediacentre/factsheets/fs094/en/ date accessed 22 November 2015.

(5) https://www.gsk.com/en-gb/media/press-releases/2015/gsk-s-malaria-candidate-vaccine-mosquirix-rtss-receives-positive-opinion-from-european-regulators-for-the-prevention-of-malaria-in-young-children-in-sub-saharan-africa/.

(6) http://www.ema.europa.eu/ema/index.jsp?curl=pages/news_and_events/news/2015/07/news_detail_002376.jsp&mid=WC0b01ac058004d5c1.

(7) Bloland, P. (2001) Drug resistance in malaria.

(8) Duparc, S., Lanza, C., Ubben, D., Borghini-Fuhrer, I., and Kellam, L. (2012) Optimal dose finding for novel antimalarial combination therapy. *Trop. Med. Int. Heal. 17*, 409–413.

(9) Burrows, J. N., Burlot, E., Campo, B., Cherbuin, S., Jeanneret, S., Leroy, D., Spangenberg, T., Waterson, D., Wells, T. N., and Willis, P. (2014) Antimalarial drug discovery - the path towards eradication. *Parasitology 141*, 128–39.

(10) Tu, Y. (2011) The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. *Nat. Med. 17*, 1217–1220.

(11) http://who.int/malaria/media/artemisinin_resistance_qa/en/ accessed 22 November 2015.

(12) Smith Gueye, C., Newby, G., Hwang, J., Phillips, A. a, Whittaker, M., MacArthur, J. R., Gosling, R. D., and Feachem, R. G. (2014) The challenge of artemisinin resistance can only be met by eliminating *Plasmodium falciparum* malaria across the Greater Mekong subregion. *Malar. J. 13*, 286.

(13) Van Hong, N., Amambua-Ngwa, A., Tuan, N. Q., Cuong, D. D., Giang, N. T. H., Van Dung, N., Tinh, T. T., Van Tien, N., Phuc, B. Q., Duong, T. T., Rosanas-Urgell, A., D'Alessandro, U., Van Geertruyden, J. P., and Erhart, A. (2014) Severe malaria not responsive to artemisinin derivatives in man returning from Angola to Vietnam. *Emerg. Infect. Dis. 20*, 1199–1202.

(14) (2013) WHO - World Health Organisation.

(15) Ridley, R. G. (2002) Medical need, scientific opportunity and the drive for antimalarial drugs. *Nature 415*, 686–693.

(16) Olliaro, P. L., and Goldberg, D. E. (1995) The *Plasmodium* digestive vacuole: Metabolic headquarters and choice drug target. *Parasitol. Today 11*, 294–297.

(17) Francis, S. E., Gluzman, I. Y., Oksman, A., Knickerbocker, A., Mueller, R., Bryant, M. L., Sherman, D. R., Russell, D. G., and Goldberg, D. E. (1994) Molecular characterization and inhibition of a *Plasmodium falciparum* aspartic hemoglobinase. *EMBO J. 13*, 306–17.

(18) Liu, J., Istvan, E. S., Gluzman, I. Y., Gross, J., and Goldberg, D. E. (2006) Plasmodium falciparum ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems. *Proc. Natl. Acad. Sci. U. S. A. 103*, 8840–5.

(19) Coombs, G. H., Goldberg, D. E., Klemba, M., Berry, C., Kay, J., and Mottram, J. C. (2001) Aspartic proteases of *Plasmodium falciparum* and other parasitic protozoa as drug targets. *Trends Parasitol 17*, 532–537.

(20) Gardner, M. J., Shallom, S. J., Carlton, J. M., Salzberg, S. L., Nene, V., Shoaibi, A., Ciecko, A., Lynn, J., Rizzo, M., Weaver, B., Jarrahi, B., Brenner, M., Parvizi, B., Tallon, L., Moazzez, A., Granger,

D., Fujii, C., Hansen, C., Pederson, J., Feldblyum, T., Peterson, J., Suh, B., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., White, O., Cummings, L. M., Smith, H. O., Adams, M. D., Venter, J. C., Carucci, D. J., Hoffman, S. L., and Fraser, C. M. (2002) Sequence of *Plasmodium falciparum* chromosomes 2, 10, 11 and 14. *Nature 419*, 531–4.

CHAPTER 2

1. Introduction

Malaria – derived from Italian, meaning 'bad air'

Malaria was first described by the Chinese almost 4000 years ago. It is a mosquito born infection that is completely preventable and treatable. According to the World Health Organisation (WHO), almost half of the world's population (3.2 billion) are at risk of contracting malaria [1]. So far in 2015, approximately 214 million cases of malaria have been reported with an estimated 438 000 deaths [1]. Globally, the mortality rate has decreased by 60% since 2000. Sub-Sharan Africa carries the most significant portion of the malaria burden; in 2015 the region represented 89% of the global malaria cases and 91% of the global malaria deaths (**Figure 1**) [1]. Children under the age of five and pregnant woman are the most vulnerable; of all the malaria deaths reported two thirds (70%) are in children under the age of five. In 2013, South Africa reported 8645 malaria cases (100% *P. falciparum*) and 104 deaths, with approximately 10% (5 million) of the population at risk in certain areas of KwaZulu-Natal, Limpopo and Mpumalanga [2,3]. The need to reduce and eliminate the spread of malaria is so pertinent, the Nobel Prize in Medicine or Physiology has been awarded four times for work related to malaria [4].

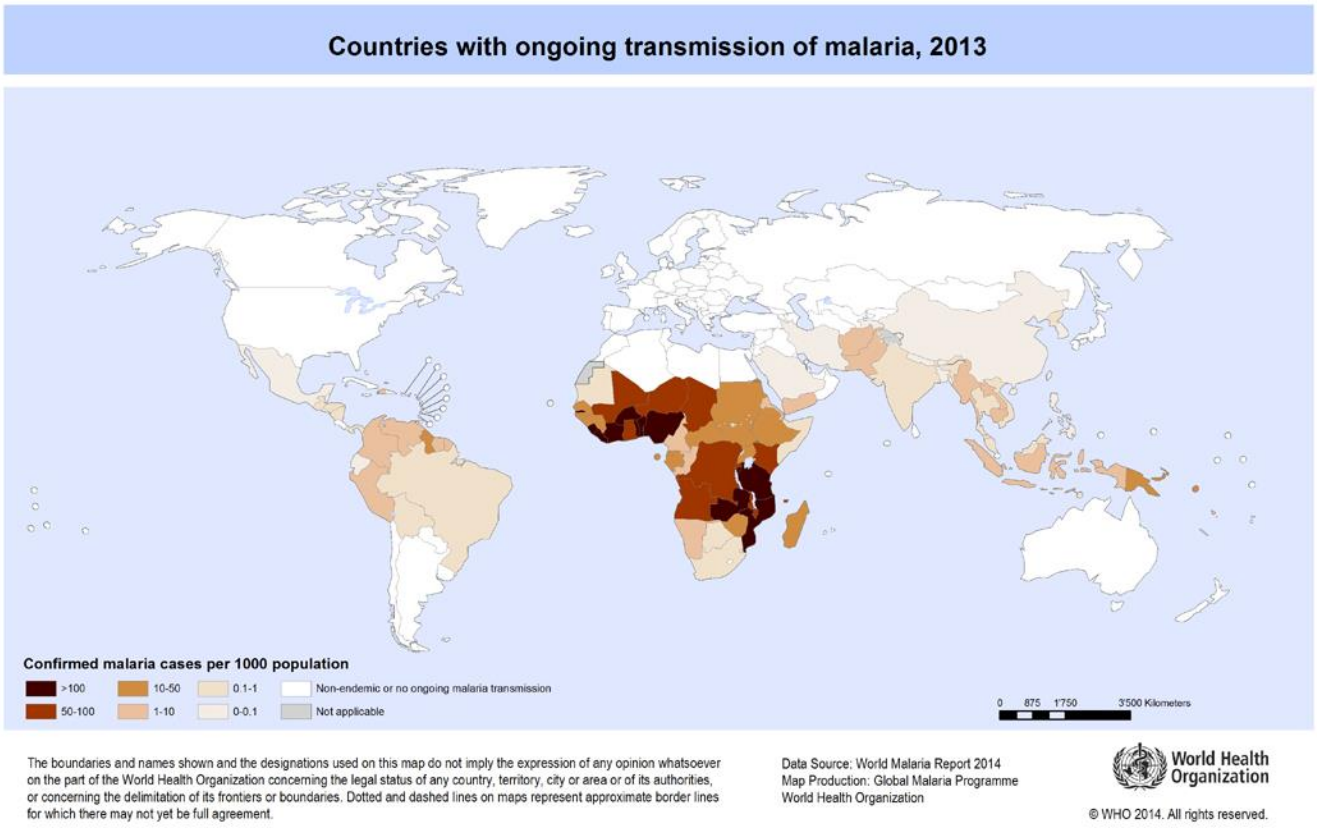


Figure 1. World map representative of the regions with ongoing malarial transmission in 2013 and the severity of each[5].

Malaria is one of the deadliest infectious parasitic diseases known to mankind; caused by the parasitic *Plasmodium* species, and transmitted to humans through the bite of an infected female *Anopheles* mosquito (vector). Throughout the world, there are approximately 400 *Anopheles* species, of which, 60 are natural malaria vectors. There are roughly 200 *Plasmodium* species, of which the five most common species to infect humans are *P. falciparum*, *P. vivax, P. ovale*, *P. malariae* and more recently *P. knowlesi* (Southeast Asia) [6–8]. The deadliest and most virulent strain is *P. falciparum*, causing approximately 90% of all reported deaths [9]. Symptoms usually occur 8 days to 4 weeks after infection, and generally cycle every 48 to 72 hours, depending on the *Plasmodium* strain. Symptoms resemble flu-like symptoms, with characteristic fatigue, high fevers and sudden chills (paroxysm); in severe cases it can cause

convulsions, comas and eventually death [10]. Antimalarials used prophylactically or preventatively delay the onset of symptoms [11].

Strategies to control malaria include vector control such as insecticides, bednets and insect repellents [12]. GlaxoSmithKline is the first company to receive positive results and feedback from regulatory authorities for a malaria vaccine. Antimalarials used in clinical practice can be divided into five classes: artemisinin derivatives, quinolones and arylaminoalcohols, antifolates, antibacterial agents and hydroxynaphthaquinones [12]. No new class has entered the market since 1996 [13]. The most important classes in the fight against malaria is the artemisinins and the quinolones, these classes have shaped and molded the current treatment regimens. The biggest hurdle in the fight against malaria is increasing resistance of the Plasmodium parasite to antimalarial medication. According to the Center for Disease Control (CDC) drug resistance is defined as "the ability of a parasite strain to survive and/or multiply despite the administration and absorption of a drug given in doses equal to or higher than those usually recommended but within tolerance of the subject" [14]. Thus, there is a dire need to design antimalarials with increased efficacy and potency that are less susceptible to resistance.

2. Life cycle

The *Plasmodium* parasite has a complex, multistage lifecycle occurring in two hosts, the female *Anopheles* mosquito vector and a mammalian host (**Figure 2**). During its lifecycle, the malaria parasite undergoes several developmental stages such as sporozoites which are the infectious parasites injected into the host through the saliva of the mosquito, merozoites which invade host red blood cells (RBC), trophozoites which uses the nutrients from hemoglobin catabolism to multiply and gametocytes [6]. The lifecycle can be broken down into a sexual phase

(sporogony) which occurs in the mosquito, and an asexual phase occurring in the host which can further be broken down into an intrahepatic (hepatocytes) and intraerythrocytic (erythrocytes) [7,15]. As a female mosquito feeds on the blood from a malaria infected individual, male and female gametocytes move into the mosquito's gut. Once in the gut, the male and female gametocytes fuse to form zygotes which eventually develop into sporozoites [16]. After 8-15 days the oocytes housing the thousands of sporozoites rupture, releasing the sporozoites into the mosquito's body cavity which then travel to the salivary glands. When the mosquito feeds again, the sporozoites are injected through the saliva into the host's bloodstream. Through the bloodstream the sporozoites travel to the liver, where they infect the hepatocytes. The sporozoites remain in the liver cells for 9 – 16 days, where they asexually replicate and mature into schizonts [16,17]. Each sporozoite develops into a schizont which contain anything from 10 000 to 30 000 merozoites, each merozoite has the ability to infect a RBC [18]. The intrahepatic or pre-erythrocytic phase is a single cycle, at this stage infected individuals are asymptomatic. Merozoite filled merosomes derived in the liver, ensures the parasite bypass the hosts' immune system and delivers the merozoites directly into the bloodstream [19]. Once in the bloodstream, merozoites invade and infect RBCs. Merozoites grow and divide asexually inside RBCs, extensively remodeling the host cell; infected RBC become rigid and deformed which increase their ability to adhere to numerous cell types (such as endothelial cells lining blood vessels) [20]. The various biochemical and structural changes that occur as the parasite develops in the RBC are responsible for the symptoms and pathologies (anemia and sequestering of infected RBC in the vasculature) associated with malaria [21]. Eventually infected RBCs rupture, releasing more merozoites into the circulation to infect more RBCs (**Figure 2**) [22]. At the same time in the infected RBC, a small portion of haploid parasites are stimulated to differentiate into gametocytes (male and female) [23,24]. The gametocytes are non-replicating, do not cause any symptoms and are responsible for the transmission of malaria between hosts.

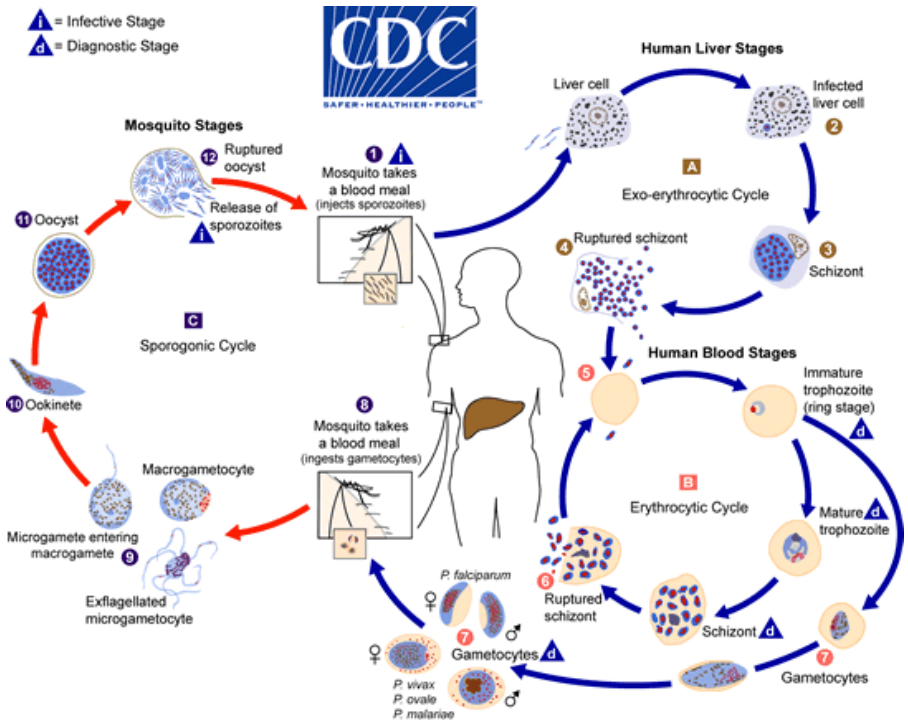


Figure 2. Malaria parasite in both the mosquito vector (sexual phase) and mammalian (human) host (asexual) [25].

3. Hemoglobin degradation

During the intraerythrocytic phase the parasite is encapsulated by three membranes – (i) parasitophorous vacuolar membrane (PVM) which is formed from the RBC plasma membrane as the parasite invades the cell, (ii) the parasite plasma membrane and (iii) the erythrocyte plasma membrane [26]. Inside the RBC, the *Plasmodium* parasite degrades hemoglobin which makes up roughly 33% of a RBC; more than 80% of the RBC hemoglobin is degraded to provide energy and amino acids for parasitic growth as the *Plasmodium* species is incapable of de novo amino acid synthesis and have a limited capacity to exogenously take up amino acids [27]. Hemoglobin degradation is essential to parasitic survival and takes place in the lysosome-like organelle, the acidic food vacuole (FV). Toxic heme, a by-product of hemoglobin hydrolysis, is sequestered into a crystal lattice known as hemozoin or malarial pigment. Hemoglobin hydrolysis is mediated by numerous protease families with functional redundancy

(**Figure 3**) [28–32]. Enzymes responsible for the breakdown of hemoglobin have been grouped together as malarial hemoglobinases; the major hemoglobinolytic enzymes are cysteine and aspartic proteases (**Figure 3**) [32,33]. These proteases work synergistically, in a semi ordered manner and each have their own unique specificity and function in the hemoglobin degradation pathway. The hemoglobin degradation pathway has numerous targets for drug design; proteases involved in the catabolism of hemoglobin include plasmepsins (aspartic proteases), falcipains (cysteine proteases), falcilysins (metalloprotease) and finally dipeptidyl aminopeptidase (DPAP1) (**Figure 3**) [30]. Plamsepsins (I and II) make the initial attack on the native hemoglobin molecule in the conserved hinge region between Phe33 and Leu 34, this unravels the molecule and exposes it to further protease degradation [34,35]. Further along the pathway in the acidic FV, falcipains, HAP and PlmIV degrade the denatured molecule into smaller peptides. Globins are cleaved by a metalloprotease into oligopeptides which are converted to free amino acids by DPAP1 in the parasitic cytosol [7]. Inhibition of both cysteine and aspartic proteases is lethal to the *Plasmodium* parasite (**Figure 3**) [36].

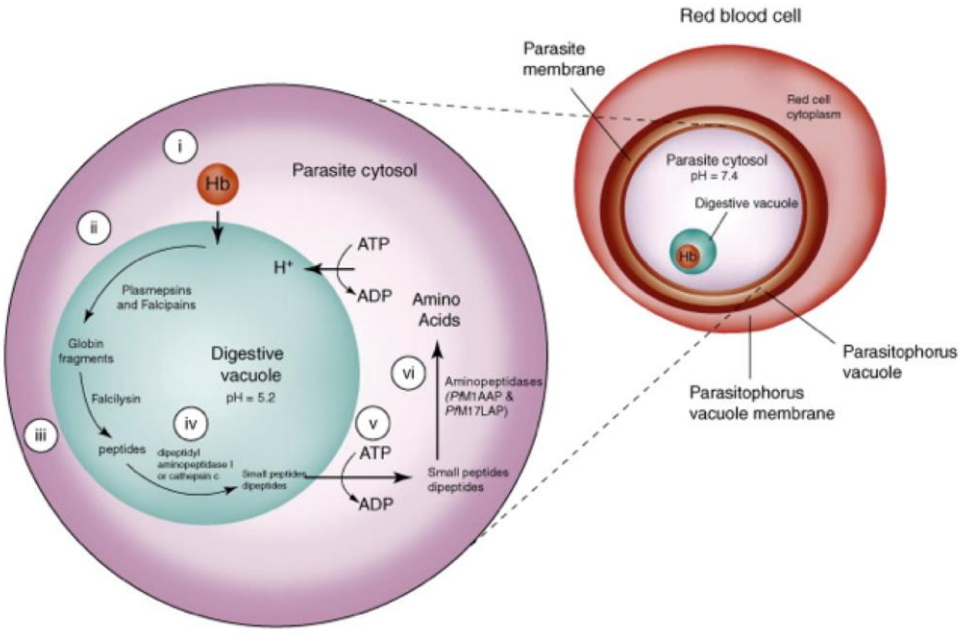


Figure 3. Hemoglobin degradation in the host RBC through numerous parasitic proteases[37].

4. Aspartic proteases

Aspartic proteases belong to the peptidase enzyme family, also known as enzymes of digestion. The peptidase family is not only one largest enzyme families, but also one of the most diverse as they are involved in almost every aspect of biological function [38]. Proteases use water molecules to cleave amide bonds through a nucleophilic attack on the carbonyl carbon. Aspartic proteases share aspartic proteases as their catalytic apparatus, and function optimally in acidic conditions [38]. These enzymes are bilobed (two assymetrical lobes) molecules (N- and C-terminals) with the active site situated at the interface of the two domains, each lobe contributes an aspartic acid to the catalytic dyad of the active site. The two aspartic acids are located in close geometrical proximity to each other, functioning at an optimum pH of roughly 3 which ensures at any given moment one of the aspartic acid residues is protonated (ionised) and the other is deprotonated (unionised) [38–40]. The mechanism of action is through a general acid-base reaction ("push-pull mechanism"), in which a nucleophilic attack occurs through the simultaneous transfer of two protons (between the conserved water molecule, the catalytic dyad and the substrate) leading to the formation of a transient neutral tetrahedral intermediate held non-covalently in the active site until bond cleavage occurs (**Figure 4**) [40–44].

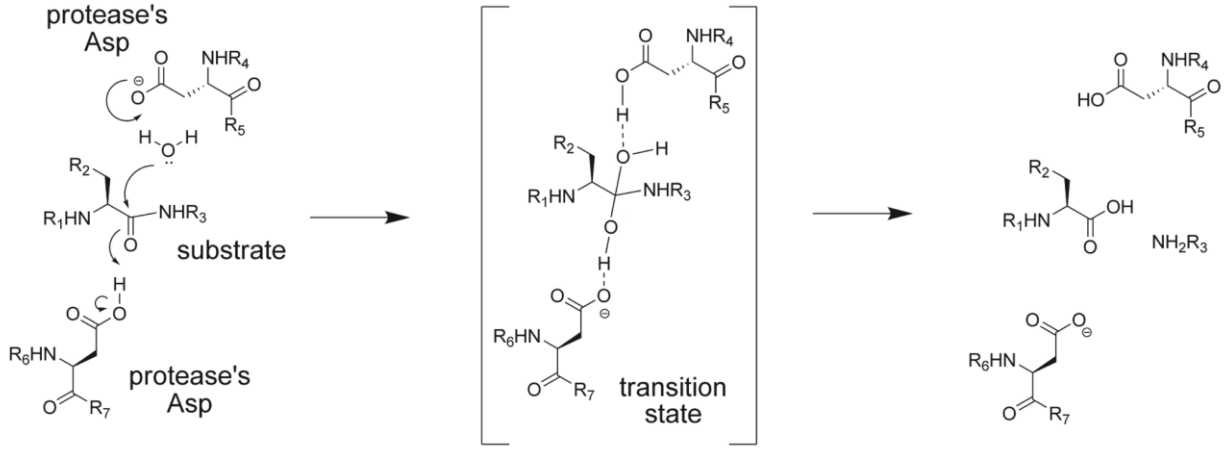


Figure 4. The general acid-base mechanism of proteolytic hydrolysis of aspartic proteases via the formation of the tetrahedral intermediate [45].

Unlike retroviral aspartic proteases eukaryotic aspartic proteases have a single hairpin loop lying perpendicular to the active site, partially covering the active site [39,46]. The importance of the flap region and flap dynamics in regulating protein structure, protein function and ligand binding has been well documented for the HIV aspartic protease. Generally, the flap region moves away from the active site exposing the active site to the ligand, the flaps regulate access to the active site by guiding in the ligand and holing it in close proximity to the catalytic dyad. The flaps play a role in both the structure of the enzyme as well as the kinetics of the enzyme thereby regulating enzyme function and activity [47,48].

5. Plasmepsins

Sequencing of the *P. falciparum* genome has led to the identification of ten (I – X) malarial aspartic proteases, also known as plasmepsins (Plm). The overall structure and general mechanism of action is conserved throughout plasmepsins, although sequence heterogeneity in regions lining the active site such as the flap and other essential regions that regulate ligand access ensures unique substrate specificity and function [49]. Plasmepsin I, II, IV and histo-aspartic protease (HAP) are situated in the FV and are responsible for the degradation of hemoglobin; Plm I, II and IV are classical Asp proteases with a catalytic Asp dyad in the active site whereas the active site of HAP is composed of only one Asp and the other residue of the active site is histidine [29,30,34,36,45,50–52]. *P. falciparum* is the only *Plasmodium* species that actively expresses numerous plasmepsins in the FV, whereas the only aspartic protease in the FV of other *Plasmodium* species is an orthologue of PlmIV [53]. Plasmepsins outside of the FV, Plm V – X, are highly conserved in all *Plasmodium* species [54,55]. Although hemoglobin catabolism is essential to the survival of the malaria parasite, inhibition or knock down of vacuolar plasmepsins is compensated for by the hemoglobin degradation of other proteases such as falcipains albeit less efficacious and to a lesser extent.

Remodeling of the RBC surface ensures that infected cells bypass the host's immune system (specifically the spleen); these changes are mediated through the export of hundreds of parasitic proteins [56]. A staggering 8% of the parasite's proteins are exported via the PEXEL pathway [57,58]. Proteins destined for export are tagged with the five amino acid consensus sequence/motif (RxLxE/Q/D) called the *Plasmodium* export element (PEXEL) [59,60]. These proteins are transported to the endoplasmic reticulum (ER), where the ER resident aspartic protease PlmV cleaves the recognition PEXEL sequence between the highly conserved leucine and glutamic acid residue (RxL ↓ x E/Q/D) [61–65]. Plasmepsin V is distantly related to other plasmepsins and is not inhibited by pepstatin A; deletion and knock-down experiments has shown it is vital to the survival of all *Plasmodium* species [30,53,62,63,66]. Therefore, PlmV is a promising target in the search for potent antimalarials.

6. References

(1) (2015) WHO - World Health Organization.

(2) http://www.who.int/malaria/publications/country-profiles/profile_zaf_en.pdf date accessed 21 November 2015.

(3) http://www.nicd.ac.za/?page=alerts&id=5&rid=330 date accessed 21 November 2015.

(4) http://www.cdc.gov/malaria/about/facts.html accessed 21 November 2015.

(5) (2013) WHO - World Health Organisation.

(6) Tuteja, R. (2007) Malaria - an overview. *FEBS J. 274*, 4670–9.

(7) Ersmark, K., Samuelsson, B., and Hallberg, A. (2006) Plasmepsins as potential targets for new antimalarial therapy. *Med Res Rev 26*, 626–666.

(8) White, N. J., Pukrittayakamee, S., Hien, T. T., Faiz, M. A., Mokuolu, O. A., and Dondorp, A. M. (2014) Malaria. *Lancet 383*, 723–35.

(9) Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y., and Hay, S. I. (2005) The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature 434*, 214–217.

(10) Miller, L. H., Baruch, D. I., Marsh, K., and Doumbo, O. K. (2002) The pathogenic basis of malaria. *Nature 415*, 673–679.

(11) Schwartz, E., Parise, M., Kozarsky, P., and Cetron, M. (2003) Delayed Onset of Malaria — Implications for Chemoprophylaxis in Travelers. *N. Engl. J. Med. 349*, 1510–1516.

(12) Ashley, E., McGready, R., Proux, S., and Nosten, F. Malaria. *Travel Med. Infect. Dis. 4*, 159–73.

(13) Gamo, F.-J., Sanz, L. M., Vidal, J., de Cozar, C., Alvarez, E., Lavandera, J.-L., Vanderwall, D. E., Green, D. V. S., Kumar, V., Hasan, S., Brown, J. R., Peishoff, C. E., Cardon, L. R., and Garcia-Bustos, J. F. (2010) Thousands of chemical starting points for antimalarial lead identification. *Nature 465*, 305–10.

(14) Bloland, P. (2001) Drug resistance in malaria.

(15) Rosenthal, P. J. (2002) Hydrolysis of erythrocyte proteins by proteases of malaria parasites. *Curr Opin Hematol 9*, 140–145.

(16) Mota, M. M., Pradel, G., Vanderberg, J. P., Hafalla, J. C., Frevert, U., Nussenzweig, R. S., Nussenzweig, V., and Rodríguez, A. (2001) Migration of *Plasmodium* sporozoites through cells before infection. *Science 291*, 141–144.

(17) Baldacci, P., and Ménard, R. (2004) The elusive malaria sporozoite in the mammalian host. *Mol. Microbiol. 54*, 298–306.

(18) Prudêncio, M., Rodriguez, A., and Mota, M. M. (2006) The silent path to thousands of merozoites: the *Plasmodium* liver stage. *Nat. Rev. Microbiol. 4*, 849–856.

(19) Sturm, A., Amino, R., van de Sand, C., Regen, T., Retzlaff, S., Rennenberg, A., Krueger, A., Pollok, J.-M., Menard, R., and Heussler, V. T. (2006) Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. *Science 313*, 1287–1290.

(20) Cowman, A. F., Berry, D., and Baum, J. (2012) The cellular and molecular basis for malaria parasite invasion of the human red blood cell. *J. Cell Biol. 198*, 961–71.

(21) Mohandas, N., and An, X. (2012) Malaria and human red blood cells. *Med. Microbiol. Immunol. 201*, 596–598.

(22) Flannery, E. L., Chatterjee, A. K., and Winzeler, E. a. (2013) Antimalarial drug discovery - approaches and progress towards new medicines. *Nat. Rev. Microbiol. 11*, 849–62.

(23) Dixon, M. W. a, Thompson, J., Gardiner, D. L., and Trenholme, K. R. (2008) Sex in *Plasmodium*: a sign of commitment. *Trends Parasitol. 24*, 168–75.

(24) Mantel, P.-Y., Hoang, A. N., Goldowitz, I., Potashnikova, D., Hamza, B., Vorobjev, I., Ghiran, I., Toner, M., Irimia, D., Ivanov, A. R., Barteneva, N., and Marti, M. (2013) Malaria-infected erythrocyte-derived microvesicles mediate cellular communication within the parasite population and with the host immune system. *Cell Host Microbe 13*, 521–34.

(25) http://www.cdc.gov/malaria/about/biology/index.html date accessed 21 November 2015.

(26) Lazarus, M. D., Schneider, T. G., and Taraschi, T. F. (2008) A new model for hemoglobin ingestion and transport by the human malaria parasite *Plasmodium falciparum*. *J. Cell Sci. 121*, 1937–1949.

(27) Sherman, I. W. (1977) Amino acid metabolism and protein synthesis in malarial parasites. *Bull. World Health Organ. 55*, 265–76.

(28) Francis, S. E., Gluzman, I. Y., Oksman, A., Knickerbocker, A., Mueller, R., Bryant, M. L., Sherman, D. R., Russell, D. G., and Goldberg, D. E. (1994) Molecular characterization and inhibition of a *Plasmodium falciparum* aspartic hemoglobinase. *EMBO J 13*, 306–317.

(29) Francis, S. E., Sullivan, D. J., and Goldberg, and D. E. (1997) HEMOGLOBIN METABOLISM IN THE MALARIA PARASITE *PLASMODIUM FALCIPARUM*. *Annu. Rev. Microbiol. 51*, 97–123.

(30) Coombs, G. H., Goldberg, D. E., Klemba, M., Berry, C., Kay, J., and Mottram, J. C. (2001) Aspartic proteases of *Plasmodium falciparum* and other parasitic protozoa as drug targets. *Trends Parasitol. 17*, 532–537.

(31) Liu, J., Gluzman, I. Y., Drew, M. E., and Goldberg, D. E. (2005) The role of *Plasmodium falciparum* food vacuole plasmepsins. *J. Biol. Chem. 280*, 1432–7.

(32) Rosenthal, P. J. (2011) Falcipains and other cysteine proteases of malaria parasites. *Adv. Exp. Med. Biol. 712*, 30–48.

(33) Goldberg, D., and Slater, A. (1990) Hemoglobin degradation in the malaria parasite *Plasmodium falciparum*: an ordered process in a unique organelle. *Proc. … 87*, 2931–5.

(34) Banerjee, R., Francis, S. E., and Goldberg, D. E. (2003) Food vacuole plasmepsins are processed at a conserved site by an acidic convertase activity in *Plasmodium falciparum*. *Mol. Biochem. Parasitol. 129*, 157–165.

(35) Gupta, D., Yedidi, R. S., Varghese, S., Kovari, L. C., and Woster, P. M. (2010) Mechanism-based inhibitors of the aspartyl protease plasmepsin II as potential antimalarial agents. *J Med Chem 53*, 4234–4247.

(36) Rosenthal, P. J. (1998) Proteases of malaria parasites: new targets for chemotherapy. *Emerg. Infect. Dis. 4*, 49–57.

(37) Skinner-Adams, T. S., Stack, C. M., Trenholme, K. R., Brown, C. L., Grembecka, J., Lowther, J., Mucha, A., Drag, M., Kafarski, P., McGowan, S., Whisstock, J. C., Gardiner, D. L., and Dalton, J. P. (2010) *Plasmodium falciparum* neutral aminopeptidases: new targets for anti-malarials. *Trends Biochem. Sci. 35*, 53–61.

(38) Dash, C., Kulkarni, A., Dunn, B., and Rao, M. (2003) Aspartic peptidase inhibitors: implications in drug development. *Crit. Rev. Biochem. Mol. Biol. 38*, 89–119.

(39) Sielecki, A. R., Fujinaga, M., Read, R. J., and James, M. N. G. (1991) Refined structure of porcine pepsinogen at 1·8 Å resolution. *J. Mol. Biol. 219*, 671–692.

(40) Blundell, T. L., Cooper, J. B., Sali, A., and Zhu, Z. Y. (1991) Comparisons of the sequences, 3-D structures and mechanisms of pepsin-like and retroviral aspartic proteinases. *Adv. Exp. Med. Biol. 306*, 443–53.

(41) I Holm, R. O. J. J. P. F. R. (1984) Evolution of aspartyl proteases by gene duplication: the mouse renin gene is organized in two homologous clusters of four exons. *EMBO J. 3*, 557.

(42) Veerapandian, B., Cooper, J. B., Sali, A., Blundell, T. L., Rosati, R. L., Dominy, B. W., Damon, D. B., and Hoover, D. J. (1992) Direct observation by X-ray analysis of the tetrahedral "intermediate" of aspartic proteinases. *Protein Sci. 1*, 322–328.

(43) Dunn, B. M. (2002) Structure and Mechanism of the Pepsin-Like Family of Aspartic Peptidases. *Chem. Rev. 102*, 4431–4458.

(44) Northrop, D. B. (2001) Follow the Protons: A Low-Barrier Hydrogen Bond Unifies the Mechanisms of the Aspartic Proteases. *Acc. Chem. Res. 34*, 790–797.

(45) Nguyen, J.-T., Hamada, Y., Kimura, T., and Kiso, Y. (2008) Design of Potent Aspartic Protease Inhibitors to Treat Various Diseases. *Arch. Pharm. (Weinheim). 341*, 523–535.

(46) Cascella, M., Micheletti, C., Rothlisberger, U., and Carloni, P. (2005) Evolutionary conserved functional mechanics across pepsin-like and retroviral aspartic proteases. *J. Am. Chem. Soc. 127*, 3734–3742.

(47) Hong, L., and Tang, J. (2004) Flap position of free memapsin 2 (beta-secretase), a model for flap opening in aspartic protease catalysis. *Biochemistry 43*, 4689–95.

(48) Patel, S., Vuillard, L., Cleasby, A., Murray, C. W., and Yon, J. (2004) Apo and Inhibitor Complex Structures of BACE (β-secretase). *J. Mol. Biol. 343*, 407–416.

(49) McGillewie, L., and Soliman, M. E. (2015) Flap flexibility amongst plasmepsins I, II, III, IV, and V: Sequence, structural, and molecular dynamics analyses. *Proteins Struct. Funct. Bioinforma. 83*, 1693–1705.

(50) Gluzman, I. Y., Francis, S. E., Oksman, A., Smith, C. E., Duffin, K. L., and Goldberg, D. E. (1994) Order and specificity of the *Plasmodium falciparum* hemoglobin degradation pathway. *J Clin Invest 93*, 1602–1608.

(51) Goldberg, D. E. (1993) Hemoglobin degradation in *Plasmodium*-infected red blood cells. *Semin. Cell Biol. 4*, 355–361.

(52) Bhaumik, P., Xiao, H., Parr, C. L., Kiso, Y., Gustchina, A., Yada, R. Y., and Wlodawer, A. (2009) Crystal Structures of the Histo-Aspartic Protease (HAP) from *Plasmodium falciparum*. *J. Mol. Biol. 388*, 520–540.

(53) Dame, J. B., Yowell, C. a, Omara-Opyene, L., Carlton, J. M., Cooper, R. a, and Li, T. (2003) Plasmepsin 4, the food vacuole aspartic proteinase found in all Plasmodium spp. infecting man. *Mol. Biochem. Parasitol. 130*, 1–12.

(54) Banerjee, R., Liu, J., Beatty, W., Pelosof, L., Klemba, M., and Goldberg, D. E. (2002) Four plasmepsins are active in the *Plasmodium falciparum* food vacuole, including a protease with an active-site histidine. *Proc. Natl. Acad. Sci. 99*, 990–995.

(55) Omara-Opyene, a L., Moura, P. a, Sulsona, C. R., Bonilla, J. A., Yowell, C. a, Fujioka, H., Fidock, D. a, and Dame, J. B. (2004) Genetic disruption of the *Plasmodium falciparum* digestive vacuole plasmepsins demonstrates their functional redundancy. *J. Biol. Chem. 279*, 54088–96.

(56) Spielmann, T., and Gilberger, T.-W. (2015) Critical Steps in Protein Export of *Plasmodium falciparum* Blood Stages. *Trends Parasitol. 31*, 514–525.

(57) Boddey, J. A., Carvalho, T. G., Hodder, A. N., Sargeant, T. J., Sleebs, B. E., Marapana, D., Lopaticki, S., Nebl, T., and Cowman, A. F. (2013) Role of plasmepsin V in export of diverse protein families from the *Plasmodium falciparum* exportome. *Traffic 14*, 532–550.

(58) Sargeant, T. J., Marti, M., Caler, E., Carlton, J. M., Simpson, K., Speed, T. P., and Cowman, A. F. (2006) Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol 7*, R12.

(59) Marti, M. (2004) Targeting Malaria Virulence and Remodeling Proteins to the Host Erythrocyte. *Science (80-.). 306*, 1930–1933.

(60) Hiller, N. L., Bhattacharjee, S., van Ooij, C., Liolios, K., Harrison, T., Lopez-Estrano, C., and Haldar, K. (2004) A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science (80-.). 306*, 1934–1937.

(61) Chang, H. H., Falick, A. M., Carlton, P. M., Sedat, J. W., DeRisi, J. L., and Marletta, M. a. (2008) N-terminal processing of proteins exported by malaria parasites. *Mol. Biochem. Parasitol. 160*, 107–15.

(62) Boddey, J. A., Hodder, A. N., Gunther, S., Gilson, P. R., Patsiouras, H., Kapp, E. A., Pearce, J. A., de Koning-Ward, T. F., Simpson, R. J., Crabb, B. S., and Cowman, A. F. (2010) An aspartyl protease directs malaria effector proteins to the host cell. *Nature 463*, 627–631.

(63) Russo, I., Babbitt, S., Muralidharan, V., Butler, T., Oksman, A., and Goldberg, D. E. (2010) Plasmepsin V licenses *Plasmodium* proteins for export into the host erythrocyte. *Nature 463*, 632–6.

(64) Osborne, A. R., Speicher, K. D., Tamez, P. a., Bhattacharjee, S., Speicher, D. W., and Haldar, K. (2010) The host targeting motif in exported *Plasmodium* proteins is cleaved in the parasite endoplasmic reticulum. *Mol. Biochem. Parasitol. 171*, 25–31.

(65) Tarr, S. J., and Osborne, A. R. (2015) Experimental determination of the membrane topology of the Plasmodium protease Plasmepsin V. *PLoS One 10*, e0121786.

(66) Klemba, M., and Goldberg, D. E. (2005) Characterization of plasmepsin V, a membrane-bound aspartic protease homolog in the endoplasmic reticulum of *Plasmodium falciparum*. *Mol Biochem Parasitol 143*, 183–191.

CHAPTER 3

1. Introduction to Computational chemistry

Computational chemistry, also known as molecular modeling, is a theoretical area of science that merges physics, mathematics, biology and chemistry; and is proving to be an invaluable adjunct to experimental studies. Computational chemistry is a general term, encompassing numerous methods and theories which can be used to solve various biochemical problems. Broadly speaking, it is an application in which computer simulations on an atomistic level are used to generate large volumes of data and information on the behaviour of atoms (molecules) over time. In order to do so, an understanding of the interactions at an atomic level is quintessential. Atomic interactions can be described in two ways; quantum mechanics (QM) and molecular mechanics (MM) [1]. Quantum mechanics solely focuses on electrons, and electron distribution is used to model the interaction between atoms. Whereas, MM uses atoms (not electrons as single particles) and focuses on the classical Newtonian mechanics; modeling atomic interactions as a function of bond angles and bond length, non-bonded forces and dihedral angles [1]. This chapter gives a general introduction to the computational chemistry theories used in MM and molecular dynamics (MD) applicable to the present study such as Schrödinger's equation, the Born-Oppenheimer approximation and potential surface energy (PES).

2. Schrödinger's equation

The behaviour of small particles such as nuclei and electrons cannot be explained or described using classical Newtonian physics. Fundamental to modern physics is Schrödinger's equation, which describes the behaviour of electrons in a molecule as a wave like function and how the molecular system evolves over time (wave mechanics). QM solves Schrödinger's equation and expresses the interaction in terms of a wave like function. This wave like function is a

mathematical function that is used to calculate the electron distribution. From the electron distribution other properties of the molecule can be determined, e.g. which part of the molecule is susceptible to nucleophilic or electrophilic attack [2]. In computational chemistry, the most widely used Schrödinger's equation is the time-dependent equation which is dependent on time and spatial coordinates of a system (time-dependent wave function, ψ). In its simplest, the Schrödinger equation is expressed as a sum of its operators:

$$H\psi = E\psi \tag{1}$$

$$H = T + V \tag{2}$$

Where H is the Hamiltonian operator (total energy of a system), T is the kinetic energy operator of the system and V is the potential energy operator [1]. The Hamiltonian operator can also be defined as:

$$\boldsymbol{H} = \left[-\frac{h^2}{8\pi^2} \sum_i \frac{1}{mj} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \right] + \sum_i \sum < j \left(\frac{eiei}{r_{ij}} \right) \tag{3}$$

3. Born-Oppenheimer approximation

To circumnavigate the limitations of the Schrödinger equations, approximations are introduced for practicality. Quintessential to computational chemistry is the separation of electronic and nuclear motions. From the molecular orbital theory, nuclei move relatively slower compared to the electrons, which travel at the speed of light, – The Born-Oppenheimer approximation [3]. This approximation allows the Schrödinger's equation to be divided and expressed as an 'electronic' equation and a nuclear equation. Once this equation has been solved, it allows for the conceptualisation of the potential energy

surface (PES) which is used to solve for the nuclear motions in the system [1]. A Hamiltonian operator that excludes the kinetic energy from the nuclei can be expressed:

$$\boldsymbol{T}^{elec} = \left[-\frac{h^2}{8\pi^2 m} \sum_i^{electrons} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \right] \quad (4)$$

The electronic Schrödinger's equation with fixed nuclei, is expressed as:

$$H^{elec}\varphi^{elec}(r,R) = E^{eff}(R)\varphi^{elec}(r,R) \quad (5)$$

4. Potential energy surface (PES)

Potential energy surface (PES) is the mathematical (or graphical) relationship between molecular energy and the geometry of a molecule; it plots the energy of a collection of electrons and nuclei (molecular energy) versus the geometric coordinates of the nuclei (molecular geometry), mathematically equating the molecular energy as a function of the nuclear coordinates (**Figure 1**) [2]. Nuclear coordinates are the parameters used to define the geometry of a molecule due to the Born-Oppenheimer approximation. The nuclei on the potential energy surface can be calculated classically via Newtonian motions (molecular mechanics) or through quantum physics using Schrödinger's wave methods.

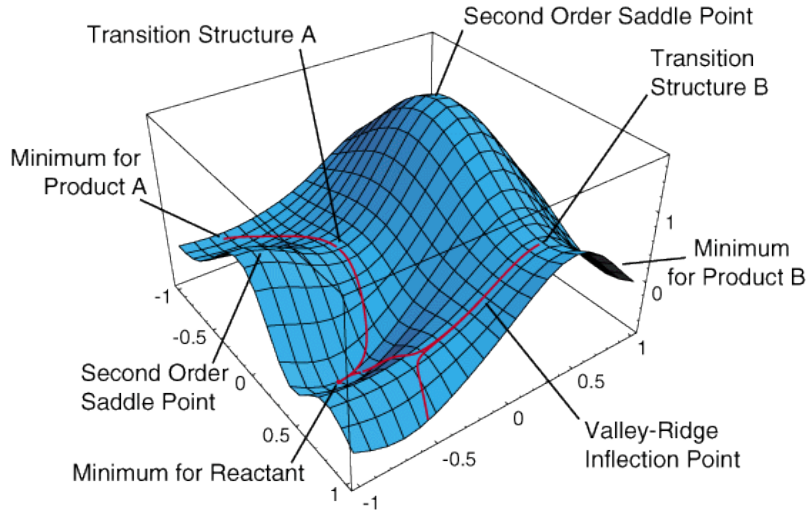


Figure 1. A two-dimensional model of the potential energy surface (PES) [4].

Potential surface energy graphically shows the ratio of the molecule's potential energy to the geometry of the molecules, that is to say the potential energy is directly proportional to the geometry **(Figure 1)**.

5. Molecular mechanics (MM)

From a computational perspective, a researcher has two options when trying to answers questions about a biomolecular system. One is to adhere to the fundamental laws of physics that govern the behaviour of electrons such as Schrödinger's equation, solving the problem using quantum chemistry and physics. The QM approach however, even in its most simplistic form is computationally expensive and time-consuming; and is not feasible for biological systems which contain thousands of atoms. Alternatively, the second approach is to use MM, which is essentially a mechanical model of a molecule, envisioned as a collection of balls (atoms) which are held together by springs (bonds); adhering to Newton's laws of classical mechanics. This mechanical system rotates, vibrates and translates until an energestically favourable conformation is reached while inter- and intra- molecular forces act on the system.

The underlying foundation of MM is that molecules are composed of structural units or functional groups that demonstrate similar behaviour in different molecules; and that mechanical models subjected to forces can be used to calculate the geometry, energy and other molecular properties of the biomolecular system [1]. Therefore, MM expresses the energies of a molecule in relation to its resistance to bond stretching and bending, and atom crowding; and uses this energy equation to find the conformation representative of the minimum energy geometries or PES minima [2]. MM provides an understanding of the mechanical nature of proteins, and provides insights into the changes in the cellular structure, cellular responses and cellular functions of a protein [5].

The Born-Oppenheimer approximation allows for the energy calculation of a system to be represented as a function only of the position of the nuclei, ignoring the motions of electrons which are treated implicitly. Potential energy is a contributor to the total energy of a system and depends on the position of the atoms (expressed as coordinates on a Cartesian plane). With the help of empirical force fields, MM calculates the potential energy of an atom through atomic interactions or terms, each term is described as the energy required to distort a molecule in a particular manner (Equation 6). The position and motion of the nuclei in a molecule on surfaces such as PES will govern the molecule's structure, function and dynamics [6].

5.1. Force Fields

Force fields (interatomic potentials) are the mathematical description that models the interactions between molecules at an atomistic level; it is the functional form of the potential energy function that predicts the molecular energy of a systems in relation to specific particle coordinates. Even though each force field uses different functional forms each equation is

always comprised of bonded terms (bond lengths, bond angles and torsions) and non-bonded terms (electrostatic and van der Waals) (Equation 6 and 7) [1,7].

$$\boldsymbol{E}_{total} = \boldsymbol{E}_{bonded} + \boldsymbol{E}_{non-bonded} \quad (6)$$

$$\boldsymbol{E}_{total} = \boldsymbol{E}_{stretch} + \boldsymbol{E}_{bend} + \boldsymbol{E}_{torsion} + \boldsymbol{E}_{electrostatic} + \boldsymbol{E}_{vdW} \quad (7)$$

Where E_{total} represents the total energy of the systems. $E_{stretch}$ is the energy contribution from bond stretching, E_{bend} is the energy contribution from angle bending, $E_{torsion}$ is the energy contribution from torsional motions around single bonds; these make up the energy contributions from bonded interactions, E_{bonded} (Equation 2.1). $E_{non\text{-}bonded}$, interactions between atoms that are not directly bonded together, is made up of $E_{electrostatic}$ and $E_{van\ der\ Waals}$. Collectively these energies/parameters creates the MM force field, and describes the energy contribution of numerous atomic forces. The energy terms above are parameterized to align to experimental and QM data to ensure systems mimic the behaviour of actual molecules in motion [8]. To date, numerous force fields have been developed and differ in their means of parameterisations; force fields should be selected based on the conditions and the type of system being investigated. The most commonly used biomolecular force fields include CHARMM [9], AMBER [10] and GROMOS [11]. In the present study, the standard AMBER Force Field was used to parameterise the protein (Plasmepsins); and the general AMBER Force Field (GAFF) [12] was used to parameterise the ligands.

6. Molecular dynamics (MD)

Molecular mechanics and MD are related; MD uses MM to generate the forces acting on molecules, which is then used to calculate their motions in time [2]. Structures generated from

NMR and crystallographic studies, represent a static view of biomolecular systems; these structures alone are insufficient in understanding the wide range of biological activity. In essence, MD is a deterministic method that strives to mimic the time dependent behaviour of molecules in space [13–15]. The assumption is that the atoms in a given molecule will interact with each other in accordance to the force field applied; resulting in a MD trajectory. A trajectory is a data set that represents the positions (coordinates) and velocities of particles in the system over time; containing the structural and dynamic properties of the system [16]. Molecular dynamics is based on an integrated approach of Newton's laws of motion (F = m.a), which solves the equation of motion for atoms on an energy surface [1]. Therefore, motion can be simulated for a molecule as it changes conformations over time, or upon ligand binding [2]. The evolution of interacting particles/atoms through time is followed by solving Newton's equation of motions:

$$\boldsymbol{F}_i = m_i \frac{d^2 \boldsymbol{r}_i(t)}{dt^2} \tag{8}$$

Where $\mathbf{F}i$ is the force acting on the particle (i) at t –time and m_i – mass of particle i, and $\mathbf{r}_i(t) = (x_i(t), y_i(t), z_i(t))$ is the location vector of the ith particle. Required for the integration of the second order differential formulas above, is the specification of the immediate forces acting on the particles and their initial velocities and positions. The trajectories generated are defined by both location and velocity vectors which describe the time progression of the biomolecular system through phase space [17]. Numerical integrators, such as the Verlet algorithm, are used to propagate the locations and velocities at specific time intervals to outline the time progression of the MD trajectory. The location (changing through time) for each particle in space is expressed by $\mathbf{r}_i(t)$; and the velocities $\mathbf{v}_i(t)$ are used to determine the thermodynamics

(temperature and kinetic energy) of the biomolecular system. The functional properties of a system can be affected by dynamic events, which can be detected at an atomic level [18]. The advantage of MD, is that by applying kinetic energy or temperature we are able to move along the molecule's PES which enables numerous conformational states to be sampled through the simulation [6].

7. References

(1) Jensen, F. (1999) Introduction to computational chemistry 2nd editio. Wiley and Sons.

(2) Lewars, E. (2004) COMPUTATIONAL CHEMISTRY: Introduction to the Theory and Applications of Molecular and Quantum Mechanics. Kluwer Academic Publishers.

(3) Hehre, W. J. (2003) A Guide to Molecular Mechanics and Quantum Chemical Calculations. Wavefunctiom Inc.

(4) http://www.chem.wayne.edu/~hbs/chm6440/PES accessed 14 November 2015.

(5) Bao, G., and Suresh, S. (2003) Cell and molecular mechanics of biological materials. *Nat. Mater. 2*, 715–725.

(6) Lipkowitz, K. B. (1998) Applications of Computational Chemistry to the Study of Cyclodextrins. *Chem. Rev. 98*, 1829–1874.

(7) Rogers, D. . . (2003) Computational Chemistry Using the PC. Wiley and Sons.

(8) Durrant, J. D., and McCammon, J. A. (2011) Molecular dynamics simulations and drug discovery. *BMC Biol. 9*, 71.

(9) Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., and MacKerell, A. D. (2012) Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ 1 and χ 2 Dihedral Angles. *J. Chem. Theory Comput. 8*, 3257–3273.

(10) Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem. 26*, 1668–1688.

(11) Christen, M., Hünenberger, P. H., Bakowies, D., Baron, R., Bürgi, R., Geerke, D. P., Heinz, T. N., Kastenholz, M. a., Kräutler, V., Oostenbrink, C., Peter, C., Trzesniak, D., and Van Gunsteren, W. F. (2005) The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem. 26*, 1719–1751.

(12) Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field. *J. Comput. Chem. 25*, 1157–1174.

(13) Karplus, M., and Petsko, G. A. (1990) Molecular dynamics simulations in biology. *Nature 347*, 631–639.

(14) Karplus, M., and McCammon, J. A. (2002) Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol. 9*, 646–652.

(15) Kalyaanamoorthy, S., and Chen, Y. P. P. (2014) Modelling and enhanced molecular dynamics to steer structure-based drug discovery. *Prog. Biophys. Mol. Biol. 114*, 123–136.

(16) Cheatham, T. E., and Kollman, P. A. (2000) Molecular dynamics simulation of nucleic acids. *Annu. Rev. Phys. Chem. 51*, 435–71.

(17) González, M. A. (2011) Force fields and molecular dynamics simulations. *École thématique la Société Française la Neutron. 12*, 169–200.

(18) Wang, W., Donini, O., Reyes, C. M., and Kollman, P. a. (2001) Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct. 30*, 211–243.

CHAPTER 4

Flap Dynamics of Plasmepsin Proteases: Proposed Parameters and Molecular Dynamics Insight

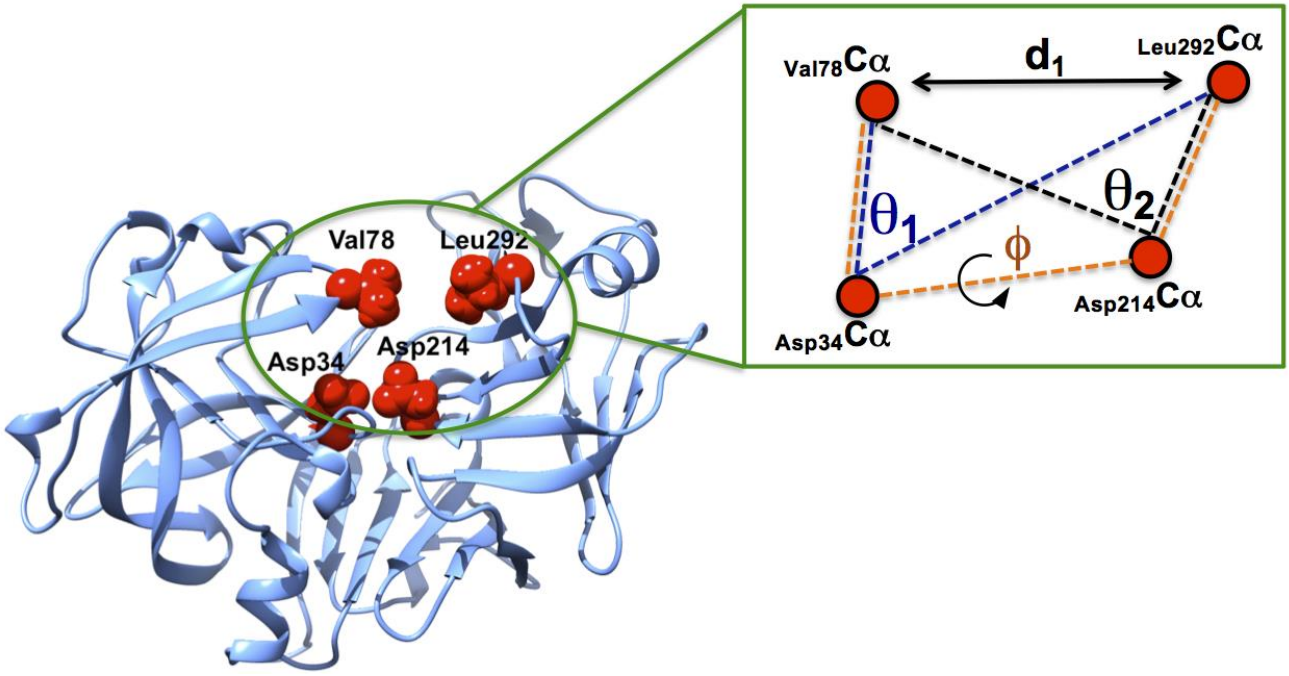
Wilson Karubiu[a,1], Soumendranath Bhakat [a,1], Lara McGillewie[a] and Mahmoud E.S. Soliman[a*]

[a]School of Health Sciences, University of KwaZulu-Natal, Westville, Durban-4000, South Africa

*Corresponding author: Mahmoud E. S. Soliman (email: soliman@ukzn.ac.za)

[1] Authors contributed equally

Graphical Abstract



Abstract

Herein, for the first time, we report the flap opening and closing in Plasmepsin proteases – plasmepsin II (PlmII) was used as a prototype model. We proposed different combined parameters to define the asymmetric flap motion; distance, d_1, between the flaps tips residues (Val78 and Leu292); dihedral angle, ϕ; in addition to TriCα angles Val78-Asp34-Leu292, θ_1, and Val78-Asp214-Leu292, θ_2, are presented. Only three combined parameters, distance, d_1, dihedral angle, ϕ and TriCα angle, θ_1 were found to appropriately define the observed "twisting' motion during the flap opening and closing. The coordinated motions of the proline-rich loop adjacent to the binding cavity rim appeared to exert steric hindrance on the flap residues, driving the flap away from the active site cavity. This loop may also have increased movements around the catalytic dyad residue, Asp214, making the TriCα, θ_2 unreliable in describing the flap motion. The full flap opening at d_1, 23.6 Å corresponded to the largest TriCα angle, θ_1 at 78.6° at ~46800ps time scale. Overall the average θ_1 and θ_2 for the bound was ~46° and ~53°, respectively compared to ~50° and ~59° for the Apo PlmII, indicating a drastic increase in the TriCα as the active site cavity opens. Similar trends in the distance, d_1 and dihedral angle, ϕ were observed during the simulation. The asymmetrical opening of the binding cavity was best described by the large shift in ϕ from -33.90° to +20.99° corresponding to the partial opening of the flap at a range of 22091-31013ps. Though, the dihedral angle described the twisting of the flap, the extent of flap opening can appropriately be defined by combining the d_1 and the θ_1. The results presented here, on the combined parameters, will certainly augment current efforts in designing potent structure-based inhibitors against plasmepsins.

Keywords: Malaria; *Plasmodium falciparum*; Plasmepsin II; Flap motion; Molecular Dynamics.

1. Introduction

The plethora of proteomics data on human pathogens has significantly augmented disease fighting strategies. *Plasmodium falciparum* is responsible for the most lethal form of malaria among the six known *Plasmodium spp*.(*1-3*). Plasmepsins (Plm) are plasmodium encoded proteins, similar to human pepsins, which play critical roles in the erythrocytic stages of plasmodium life cycle (*4*). Currently, only crystal structures of food vacuole plasmepsins (I, II, III, and IV) are available (*5-8*). Plasmepsin II (PlmII), an aspartic protease encoded by *P. falciparum*, has been reported as a virulence factor in malaria where it is involved in haemoglobin degradation. Similar to other aspartic proteases, the PlmII active site contains two aspartic acid residues, a proton donor and acceptor, forming the catalytic dyad when cleaving the peptide bond. General features from experimental studies by Asojo *et al*, show a mature enzyme crystal structure consisting of a single chain, composed of 329 amino acid residues which fold into two topologically similar N and C terminal domains (*6*). The domains make contact along the bottom of the binding cleft, that contains the catalytic dyad Asp34 and Asp214. A single long β hairpin structure (flap, Lys72-Phe85) lies perpendicularly over the binding cleft and lying opposite is a flexible flap-like loop structure (*6, 9*). These two structures are highly flexible where they interact with bound inhibitors and presumably substrates. The amino and carboxyl ends of the polypeptide chain of PlmII are assembled into a characteristic six-stranded inter-domain β-sheet which serves to anchor the domains together (*6, 9*).

Several non-structural proteins display unique and specific motions which are essential in defining their precise function(*10*). Determining the correct parameters which best describe such motions in proteins, is essential in understanding enzyme functions as well as its impact on drug binding and resistance. For instance, flap dynamics is a distinctive motion observed amongst aspartate proteases e.g. HIV protease and cathepsin. Flaps have been shown to

regulate access to the active site of proteases by providing access for substrate and/or inhibitors binding (*10, 11*). Flap opening and closure in HIV protease have been well studied using molecular dynamics (*11-19*). Generally, different parameters have been proposed to describe the HIV PR flap motion (*16*). The distance between Ile50-Ile50' (inter-flap distance) is one of the most commonly used parameter for defining flap motions. However, this parameter does not adequately describe the curling behaviour observed, which lead to the introduction of other parameters such as curling to better define HIV protease flap dynamics (*19, 20*). The flap dynamics in HIV protease is one such example, which highlights the importance of defining the appropriate parameters that best describe specific dynamics associated with inhibitor binding to the binding cavity.

The characteristic flap and flap-like structure have been reported to be highly flexible in plasmepsins I-IV, in both the free and ligand-bound proteases relative to other enzyme structures (*5-8*). These structures are critical determinants in the conformational flexibility of the binding cavity of PlmII in accommodating different inhibitors, where crystallographic studies have demonstrated structural differences between the free and ligand-bound enzymes (*6, 9*). Precise insight into the mechanistic events associated with binding of plasmepsins inhibitors is essential for the design of more potent inhibitors. Crystallographic studies reveal that the binding cavity of PlmII is a highly flexible pocket, thus designing more potent inhibitors will require a better understanding of the plasticity of this cavity in response to different inhibitors. Moreover, different inhibitors have shown a varying scale of binding potential to the binding cavity of PlmII. Experimental parameters previously defined are insufficient in examining and predicting the binding modes and flap dynamics of different inhibitors (*6, 9, 21*). The interaction of the PlmII flap pocket with non-peptidomimetic inhibitors have been documented elsewhere (*22*). Therefore, defining appropriate parameters

to assess flap motions is critical. Furthermore, the exact parameters to precisely describe flap opening and closing are not well defined in literature, neither experimentally nor computationally. This has prompted us to report the first detailed computational study that highlights the flap dynamics amongst plasmepsins and different proposed parameters. We believe that this article will serve as a benchmark for observing this phenomenon in plasmepsins, potentially other proteases.

2. Methods

2.2. System preparation

The apo crystal structure (PDB ID: 1LF4)(*6*) and complex structure of plasmepsin II (PlmII) bound with potent inhibitor EH58 (PDB ID: 1LF3) (*6*) were obtained from the RSCB Protein Data Bank (*23*). The systems were prepared as described in our previous reports (*24*, *25*).

2.3. Molecular dynamic analysis

An all atom explicit solvation molecular dynamics simulation was performed using GPU version of PMEMD engine integrated with Amber14 (*26*). All the systems were set up using the standard methodology discussed in our previous reports (*24*, *27*). Visualisation of enzyme structures was carried out using graphical user interface of UCSF Chimera package (*28*) and data were plotted using the GUI of Microcal Origin data analysis software version 6 (www.originlab.com) and Surfer version 12 (www.goldensoftware.com).

3. Results and Discussion

3.1. Experimentally Determined Parameter and Its Limitations

X-ray diffraction studies on PlmII found that the uncomplexed form (PDB ID: 1LF4, 1.9 Å) (*6*) of the enzyme is more open than the bound, by measuring the distance between Cα of Val78 at the flap tip and Cα Leu292 on the opposite side of the hydrophobic rim of the binding cavity. In the free enzyme crystal structure, this distance was found to be 12.6 Å. However, upon binding to EH58 inhibitor (PDB ID: 1LF3, 2.7 Å) (*6*) this distance slightly reduced to 12.0 Å, indicating the closing of the flap upon ligand entry. In PlmII - Pepstatin A complex, the distance between flaps reduced significantly to 9.9 Å, indicating that the binding cavity closes to embrace the inhibitor (*6*). Thus, as evident from these measurements, the binding cavity of PlmII exhibits a rather flexible conformation which adapts to accommodate inhibitors based on their bulkiness.

However, the defined parameter as described by Asojo *et al.2003*, provide some information on the opening and closing motion of the flap-like structure, this single parameter was based on a rigid crystal structure and only described the motions in one-dimension of space (*6, 9*) (**Figure 1)**. Analysis of the MD trajectories clearly indicated that more parameters need to be defined to appropriately describe the flap motion (**see section 3.2**)

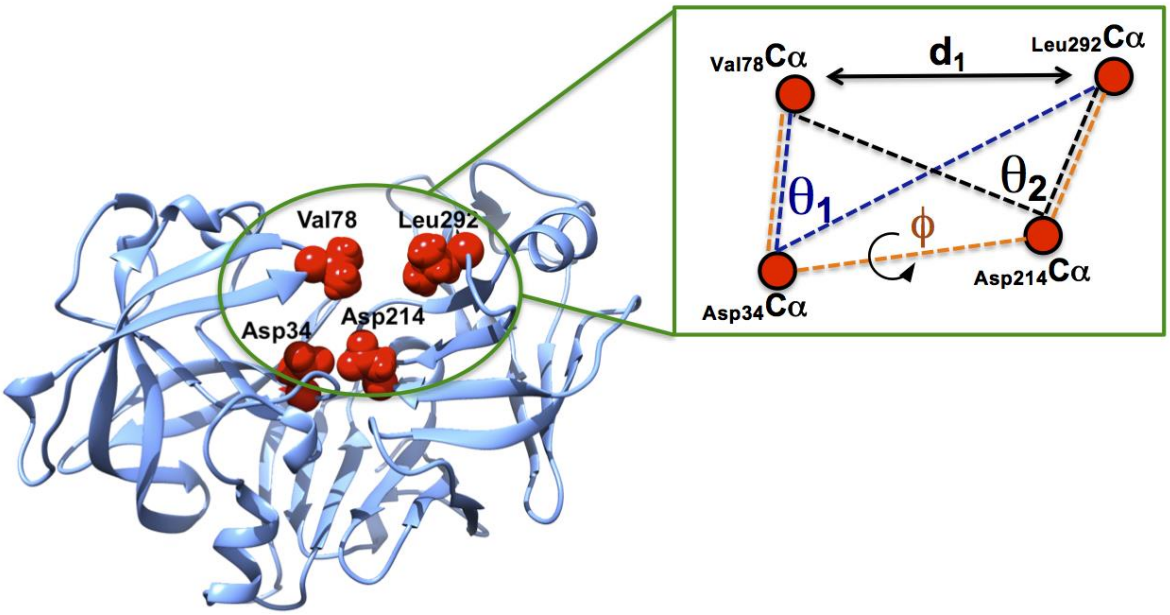


Figure 1. An illustration of the different proposed parameters to describe the flap motion; distance (d1) between the flap tip residues (Val78, Leu292) and TriCα angles, θ1, θ2 and dihedral angle, ϕ.

3.2. Proposed different parameters to describe flap opening and closing

Visual inspection of the molecular dynamic trajectory snapshots shows that the opening and closing of the binding cavity of the free PlmII structure is characterized by an extensive "twisting" of the flap and the concerted recoiling of a proline rich loop (flap-like structure).

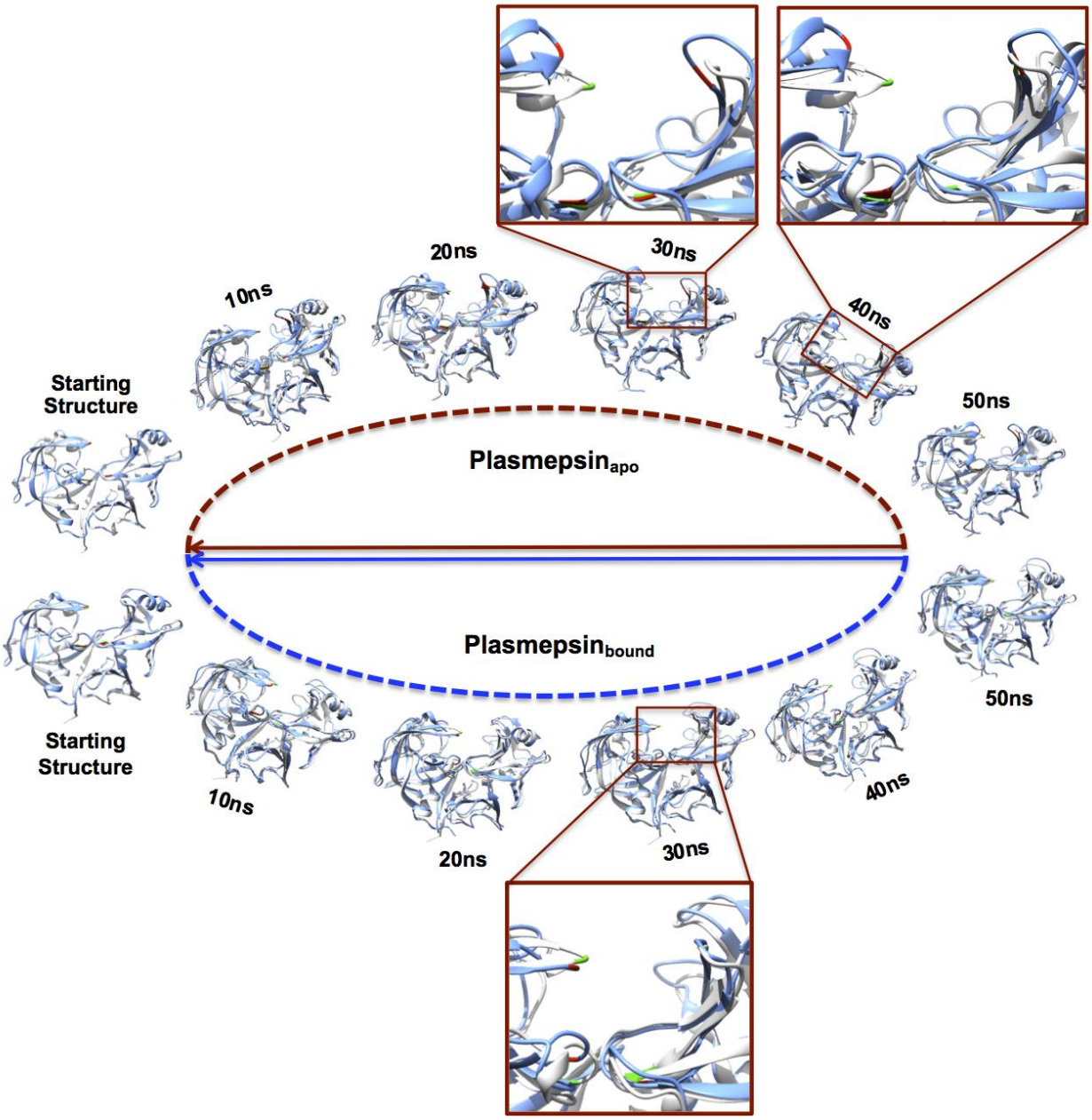


Figure 2. MD trajectory showing the flap dynamics of free and bound (PlmII-EH58) PlmII (blue ribbons) superimposed against the starting structure (grey ribbons). Flap tips for the free and bound are shown in green and red, respectively.

Upon observation of the flap motions in PlmII (**Figure 2**), we propose that these motions can be explained accurately by considering the distance between flap tip residues ($\mathbf{d_1}$) in relation to the TriCα angles, Val78-Asp34-Leu292, $\boldsymbol{\theta_1}$ and Val78-Asp214-Leu292, $\boldsymbol{\theta_2}$ as well as the dihedral angle, $\boldsymbol{\phi}$ (**Figure 1).**

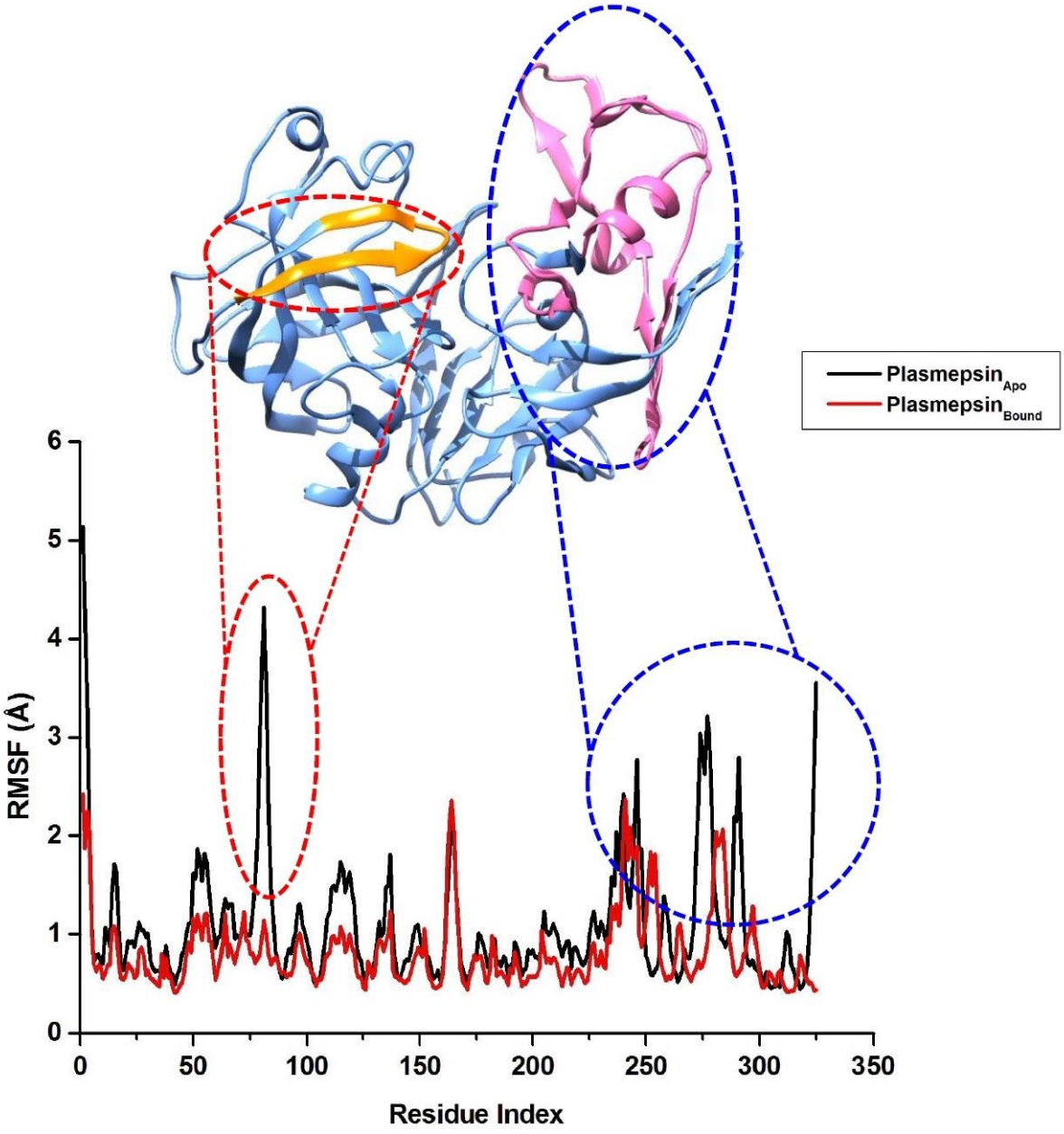


Figure 3. Plot of root mean square fluctuation (RMSF) of apo PlmII (blue) and PlmII-EH58 complex (red). Region showing high residual fluctuation corresponding to the proline-rich region of PlmII is shown in pink whereas the flap region is highlighted in orange.

The "twisting" of the flap and the lateral shift of the highly flexible loop in the open conformation appears to be a coordinated movement, involving the highly flexible loop a proline-rich region (223-295) (**Figure 3**). As evident from the root mean square fluctuation (RMSF) plot of the free and bound PlmII, the apo PlmII shows high residual fluctuation (average fluctuation ~1.37 Å) around this proline-region loop compared to the bound PlmII

(average fluctuation ~1.07 Å), indicating less residual interaction in the bound state of the enzyme. The average RMSF of the apo was 1.11 Å, whereas it was 0.82 Å for the bound enzyme, indicating a more compact bound enzyme with less residual fluctuations and a more flexible apo enzyme with high residual atomic fluctuations. Higher residual fluctuations were observed in the proline rich area, as well as the flap residues. The flap residues (residue 76-86) displayed a significant increase (~4 Å) in fluctuation in case of apo conformation as compared to bound Plm-II complex. Interestingly the RMSF of flap tip residue of β-hairpin region (otherwise known as flap region), Val78 found to be 1.97 Å in case of apo whereas in case of bound the RMSF significantly decreased and found to be 0.75 Å. Similarly, the fluctuation of Leu292 found to be 1.97 Å in case of apo as compared to 0.64 Å in bound conformation. Higher RMSF of flap tips (Val78 and Leu292) further confirmed the flexible nature of flaps whose opening leads to incorporation of inhibitor in the protease active site. The bound PlmII adopts a compact closed conformation all through the simulation, whereas in the free enzyme, the flexible regions adopt different conformations with the opening and closing of the binding cavity. This loop, recoils inside the binding cavity towards the catalytic dyad (Asp34, Asp214) as it adopts a lateral shift away from the main enzyme structure. Interestingly, the residue at the tip (Leu292) of this loop moves relatively less compared to the entire loop. Notably, the intense motion at the highly flexible proline-rich loop seems not to orient Leu292 away from the binding cavity rather, the loop movement seems to cause steric hindrance on the flap tip residues resulting in the movement of the flap away from the binding cavity. The average radius of gyration (R_g) for the free PlmII was found to be 12.31 Å at 46819 ps which corresponds to the fully opened conformation of apo PlmII. Whereas, the initial value of Rg was 4.92 Å and closing was observed at 49999ps with an Rg value of 6.96 Å. The trend of Rg in case of apo conformation of PlmII was found to be similar that of d_1 as well as θ_1. Whereas the bound conformation of Plm II observed a stable Rg fluctuation with an average fluctuation of 6.02 Å

(Figure S3, Supplementary Information). This significant fluctuation in case of apo Plm II highlighting the rigorous asymmetrical opening and closing of the flap which may have led to changes in moment of inertia.

As such, this intensive and coordinated flap motion may certainly not be defined by distance (d_1) alone, but rather a combination of parameters including TriCα angles (θ_1, θ_2) and dihedral angle, ϕ, as have been proposed in this paper. The distance, d_1, between flap tip residues was 10.69Å at the starting structure of the free enzyme. Partial flap opening was observed at ~30 ns, with the full flap opening occurring at 46820 ps (d_1 = 23.61 Å). Towards the end of the 50 ns MD simulation, the closing of the flap was observed at 49999 ps (d_1 = 12.69Å). Further, for the first time we observed an asymmetrical opening of the binding cavity evident by the "twisting" of the flap and a maximum "twisting" at the time range of 22091-31013 ps corresponding to a large shift in the dihedral angle, ϕ from -33.90° to +20.99°. Although "twisting" alone may not solely describe the extent of flap opening, it however reveals the asymmetrical nature of flap opening as defined by the dihedral angle, ϕ. This demonstrates the intensive motion which involves the movement of adjacent residues around the highly flexible loop which seems to generate steric hindrance that drives away the flap lying at the active site cavity.

3.3. Defining the appropriate "Combined" Parameters

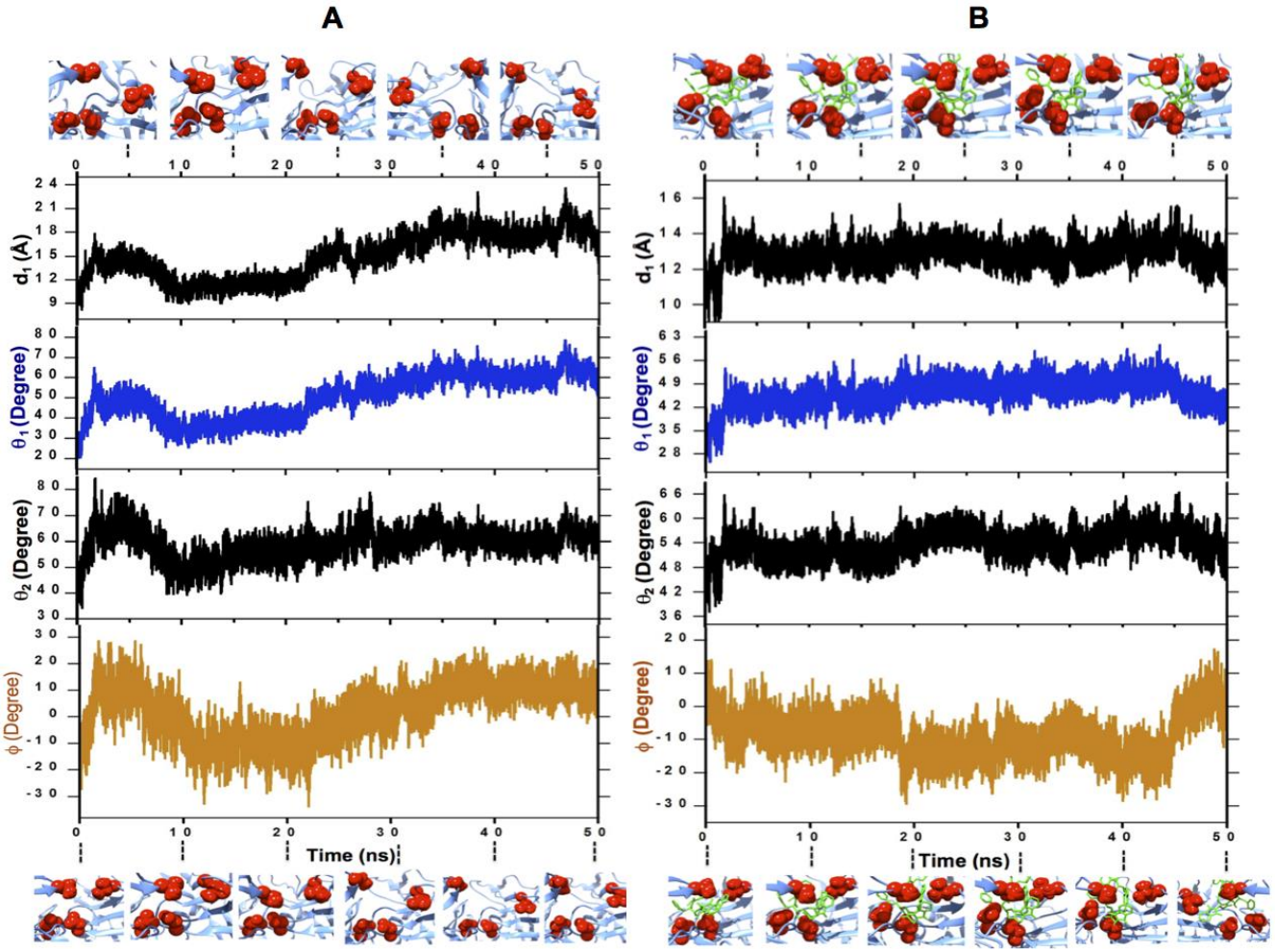


Figure 4. Combined plots of distance, d1, between flap tip residues (Val78, Leu292), dihedral angle, ϕ, and TriCα angles, θ_1, θ_2 against time (ns) of the Apo PlmII [A] and PlmII-EH58 complex [B].

3.3.1. Distance, d_1, between flap tip residues and dihedral angle, phi (ϕ)

Plotting the dihedral angle, ϕ and distance, d_1 against time (ns) for the free PlmII (**Figure 4A**), it is evident that d_1 increases with an increase in the angle ϕ. A sharp increase in d_1 and ϕ corresponding to the partial opening of the active site cavity is observed at 30 ns. This is consistent with the "twisting" motion of the flap as it moves away from the active site pocket. The partial opening of the active site cavity at 30 ns is accompanied by an increase in the dihedral angle, ϕ which demonstrates that though movement is observed at the flaps, there is

an intensive orientation of the peptide bond, evident by the large shift in the dihedral angle, ϕ, which results in a concerted movement of the flap away from the active site pocket.

In the PlmII-EH58 complex, there was no significant movement of the flap as the ligand (EH58) remained embraced by the flap at the active site cavity. The average distance between Val78 at the flap tip and Leu292 remained at 13 Å across the 50 ns simulation of the bound PlmII (**Figure 4B**).

3.3.2. Distance, d_1 between flap tips and TriCα Angles, θ_1 and θ_2

Measuring the distance as well as the angle between the flap tip residues and the catalytic residues may better explain the flap dynamics in this protease. The aspartic residues sitting in the active site are observed to show minimal movement relative to the highly dynamic flap tip residues and the highly flexible loop. Interestingly, the trend in distance of Cα of Val78 and Leu292 highly correlates with the angle between Val78-Asp34-Leu292. However, the angle between Val78-Asp214-Leu292 did not accurately describe the flap movement observed in the MD simulation as well as the observed "twisting" of the flap structure to allow the opening of the binding cavity. Thus, the opening and closing of the binding cavity can better be explained by measuring the angle, θ_1, between Val78-Asp34-Leu292 and correlating it with the distance between Cα of the flap tip residues (**Figure 4**). The apo conformation of plasmepsin II displayed a θ_1 of 78.64 Å at 46840 ps which highly corresponds to fully opened conformation of PlmII. The θ_1 displayed a value of 44.61 Å at 4999 ps which suggests a tendency of flap closing which is similar to the trend in d_1. The active site residue, Asp214, has a high atomic fluctuation compared to Asp34. This may be attributed to its closeness to the highly flexible proline-rich loop region located at the binding cavity rim. Thus, Asp214 is highly mobile and

may not give an accurate representation of the "twisting" as well as the opening and closing of the flap.

In the PlmII-EH58 complex, there was no observed drastic fluctuation in the TriCα angles, θ_1 and θ_2. Average angle (θ_1) during the 50 ns simulation was ~46°. This trend is similar to the observed minimal changes in the d_1 during the 50 ns simulation average distance between Val78 at the flap tip and Leu292 remained at 13Å across the 50 ns simulation of the bound PlmII (**Figure 4B**).

4. Concluding Remarks

The binding cavity of plasmepsin has been observed to adapt to different inhibitors demonstrating its plasticity. Precise parameters to define flap motion in plasmepsin will certainly assist in the design of potent inhibitors to bind to the flexible binding cavity. In addition, recent identification of flap pockets for potent non-peptidomimetic inhibitors further raises the important of this study.

In conclusion, the asymmetrical flap opening and closing was evident, characterized by a "twisting" motion. Only three parameters we postulated namely distance, dihedral angle and TriCα angles which appropriately defined the observed "twisting' motion during the flap opening and closing. However, the extensive motion of a proline-rich loop adjacent to the binding cavity rim appeared to exert steric hindrances on the flap residues, driving the flap away from the active site cavity. This loop may also have increased movements around the catalytic dyad residue, Asp214, resulting in the observed distortion of the TriCα angle, θ_2. Overall the average θ_1 and θ_2 for the bound was ~46° and ~53°, respectively compared to ~50°

and ~59° for the Apo PlmII, indicating intensive changes in the TriCα as the active site cavity opens. Similar trends in the distance, d_1 and dihedral angle, ϕ were observed during the simulation. The asymmetrical opening of the binding cavity was best described by the large shift of -33.90° to +20.99° in the ϕ, corresponding to the partial opening of the flap at a range of 22091-31013 ps. Though, the dihedral angle describes the twisting of the flap, the extent of flap opening can be defined by combining the distance,d_1 and the Val78-Asp34-Leu292 TriCα angle, θ_1. In contrast to the apo conformation the bound conformation of Plm II observed a stable trend in d1, θ_1, Rg, as well as dihedral angle. This is due to the fact that EH58 bound conformation pf Plm II displayed crucial hydrogen bond interactions between the ligand (Eh58) and flap tip residue, Val78; catalytic residues Asp34 and Asp214 (**Figure S4**, Supplementary Informations) which played a key role in retaining the closed conformation of Plm II. Hence designing novel inhibitors which can interact with flap tip residue, Val78 as well as catalytic aspartic residues, Asp34 and Asp214 will be crucial to maintain a tight closed conformation of Plm II which will help in keeping the inhibitor tightly within the active site pocket of Plm II. Thus interactions with flap tip residue and catalytic dyads will play an important role in designing selective and potent Plm II inhibitors and pharmacophores of EH58 might act as a template in this regard. The results presented here on the combined parameters will certainly augment current efforts in designing potent structure-based inhibitors against plasmepsins.

Acknowledgements

The authors like to acknowledge the School of Health Sciences, University of KwaZulu-Natal, and Westville for financial support.

Conflicts of Interests

Authors declare no conflicts of interest.

5. References

1. Gil L, A., Valiente, P. A., Pascutti, P. G., and Pons, T. (2011) Computational perspectives into plasmepsins structure-function relationship: implications to inhibitors design, *Journal of tropical medicine 2011*, 657483-657483.
2. Greenwood, B., Bojang, K., Whitty, CJ., Targett, GA. . (2005) Malaria, *Lancet 365*, 1487-1498.
3. Greenwood, B. (2014) Treatment of Malaria - A Continuing Challenge, *New England Journal of Medicine 371*, 474-475.
4. Bhaumik, P., Gustchina, A., and Wlodawer, A. (2012) Structural studies of vacuolar plasmepsins, *BBA-Proteins Proteomics 1824*, 207-223.
5. Banerjee, R., Liu, J., Beatty, W., Pelosof, L., Klemba, M., and Goldberg, D. E. (2002) Four plasmepsins are active in the *Plasmodium falciparum* food vacuole, including a protease with an active-site histidine, *Proceedings of the National Academy of Sciences of the United States of America 99*, 990-995.
6. Asojo, O. A., Gulnik, S. V., Afonina, E., Yu, B., Ellman, J. A., Haque, T. S., and Silva, A. M. (2003) Novel uncomplexed and complexed structures of plasmepsin II, an aspartic protease from *Plasmodium falciparum*, *Journal of Molecular Biology 327*, 173-181.
7. Bhaumik, P., Horimoto, Y., Xiao, H., Miura, T., Hidaka, K., Kiso, Y., Wlodawer, A., Yada, R. Y., and Gustchina, A. (2011) Crystal structures of the free and inhibited forms of plasmepsin I (PMI) from *Plasmodium falciparum*, *Journal of Structural Biology 175*, 73-84.
8. Bhaumik, P., Xiao, H., Parr, C. L., Kiso, Y., Gustchina, A., Yada, R. Y., and Wlodawer, A. (2009) Crystal Structures of the Histo-Aspartic Protease (HAP) from *Plasmodium falciparum*, *Journal of Molecular Biology 388*, 520-540.
9. Asojo, O. A., Afonina, E., Gulnik, S. V., Yu, B., Erickson, J. W., Randad, R., Medjahed, D., and Silva, A. M. (2002) Structures of Ser205 mutant plasmepsin II from *Plasmodium falciparum* at 1.8 angstrom in complex with the inhibitors rs367 and rs370, *Acta Crystallographica Section D-Biological Crystallography 58*, 2001-2008.
10. Ishima, R., and Louis, J. M. (2008) A diverse view of protein dynamics from NMR studies of HIV-1 protease flaps, *Proteins-Structure Function and Bioinformatics 70*, 1408-1415.
11. Cai, Y., Yilmaz, N. K., Myint, W., Ishima, R., and Schiffer, C. A. (2012) Differential Flap Dynamics in Wild-Type and a Drug Resistant Variant of HIV-1 Protease Revealed by Molecular Dynamics and NMR Relaxation, *Journal of Chemical Theory and Computation 8*, 3452-3462.
12. Hornak, V., Okur, A., Rizzo, R. C., and Simmerling, C. (2006) HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations, *Proceedings of the National Academy of Sciences of the United States of America 103*, 915-920.
13. Scott, W. R. P., and Schiffer, C. A. (2000) Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance, *Structure 8*, 1259-1265.
14. Ishima, R., Freedberg, D. I., Wang, Y. X., Louis, J. M., and Torchia, D. A. (1999) Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function, *Structure with Folding & Design 7*, 1047-1055.
15. Toth, G., and Borics, A. (2006) Flap opening mechanism of HIV-1 protease, *Journal of Molecular Graphics & Modelling 24*, 465-474.
16. Galiano, L., Bonora, M., and Fanucci, G. E. (2007) Interflap distances in HIV-1 protease determined by pulsed EPR measurements, *Journal of the American Chemical Society 129*, 11004-+.

17. Hornak, V., Okur, A., Rizzo, R. C., and Simmerling, C. (2006) HIV-1 protease flaps spontaneously close to the correct structure in simulations following manual placement of an inhibitor into the open state, *Journal of the American Chemical Society 128*, 2812-2813.
18. Galiano, L., Ding, F., Veloro, A. M., Blackburn, M. E., Simmerling, C., and Fanucci, G. E. (2009) Drug Pressure Selected Mutations in HIV-1 Protease Alter Flap Conformations, *Journal of the American Chemical Society 131*, 430-+.
19. Heaslet, H., Rosenfeld, R., Giffin, M., Lin, Y.-C., Tam, K., Torbett, B. E., Elder, J. H., McRee, D. E., and Stout, C. D. (2007) Conformational flexibility in the flap domains of ligand-free HIV protease, *Acta Crystallographica Section D-Biological Crystallography 63*, 866-875.
20. Kovalskyy, D., Dubyna, V., Mark, A. E., and Kornelyuk, A. (2005) A molecular dynamics study of the structural stability of HIV-1 protease under physiological conditions: The role of Na+ ions in stabilizing the active site, *Proteins-Structure Function and Bioinformatics 58*, 450-458.
21. Prade, L., Jones, A. F., Boss, C., Bildstein, S. R., Meyer, S., Binkert, C., and Bur, D. (2005) X-ray structure of plasmepsin II complexed with a potent achiral inhibitor, *Journal of Biological Chemistry 280*, 23837-23843.
22. Zuercher, M., Gottschalk, T., Meyer, S., Bur, D., and Diederich, F. (2008) Exploring the flap pocket of the antimalarial target plasmepsin II: The "55% rule" applied to enzymes, *Chemmedchem 3*, 237-240.
23. H.M. Berman, J. W., Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne (2000) The Protein Data Bank, *Nucleic Acids Research 28.*, 235-242.
24. Karubiu, W., Bhakat, S., and Soliman, M. E. S. (2014) Compensatory Role of Double Mutation N348I/M184V on Nevirapine Binding Landscape: Insight from Molecular Dynamics Simulation, *The protein journal 33*, 432-446.
25. Bhakat, S., Martin, A. J. M., and Soliman, M. E. S. (2014) An integrated molecular dynamics, principal component analysis and residue interaction network approach reveals the impact of M184V mutation on HIV reverse transcriptase resistance to lamivudine, *Molecular BioSystems 10*, 2215-2228.
26. D.A. Case, V. B., J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu and P.A. Kollman. (2014) AMBER 14, University of California, San Francisco.
27. Bhakat, S., Martin, A. J. M., and Soliman, M. E. S. (2014) An Integrated Molecular Dynamics, Principal Component Analysis and Residue Interaction Network Approach Reveals the Impact of M184V Mutation on HIV Reverse Transcriptase Resistance to Lamivudine, *Molecular BioSystems*.
28. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF chimera - A visualization system for exploratory research and analysis, *Journal of Computational Chemistry 25*, 1605-1612.

Supplementary Information

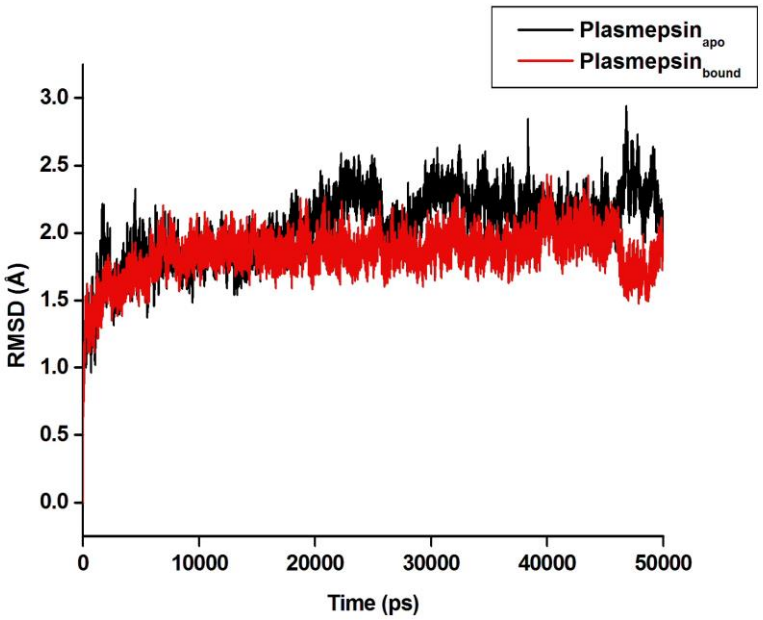


Figure S1. The RMSD fluctuation during the simulation time for both apo (black) and bound (red) conformations. The RMSD was calculated taking in account C-α atoms of the systems.

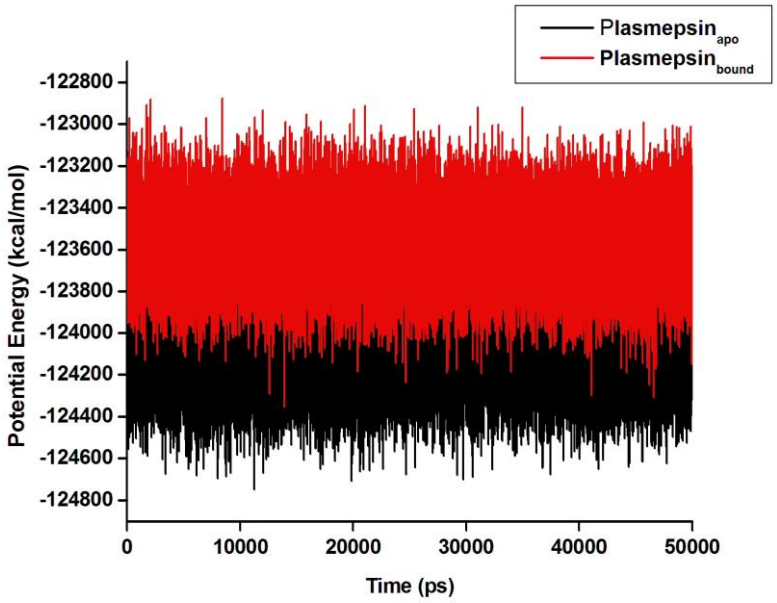


Figure S2. The potential energy fluctuations for both apo (black) and bound (red) conformations during simulation time.

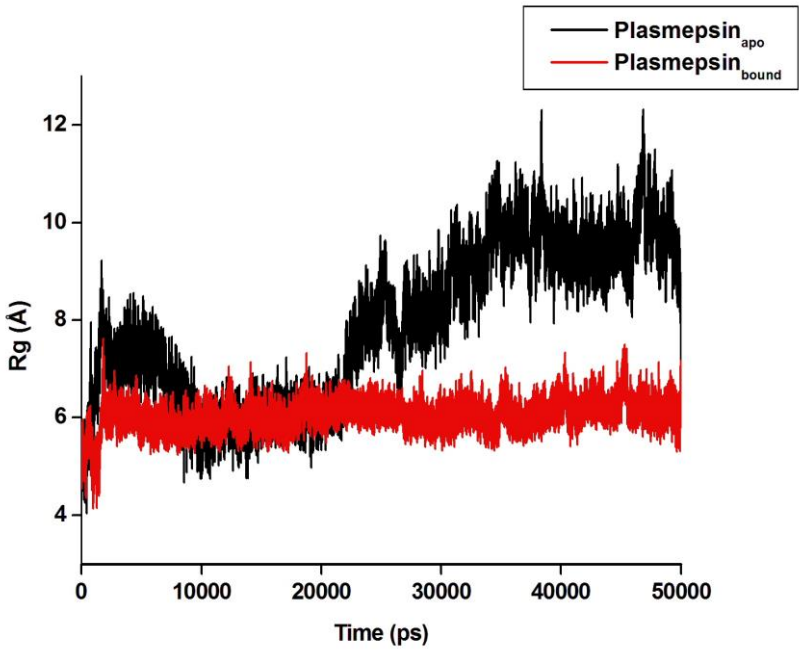


Figure S3. The fluctuation in radius of gyration (R_g) for apo and bound conformations of plasmepsin taking in account

CHAPTER 5

Flap flexibility amongst plasmepsins I, II, III, IV and V: Sequence, structural and molecular dynamics analyses.

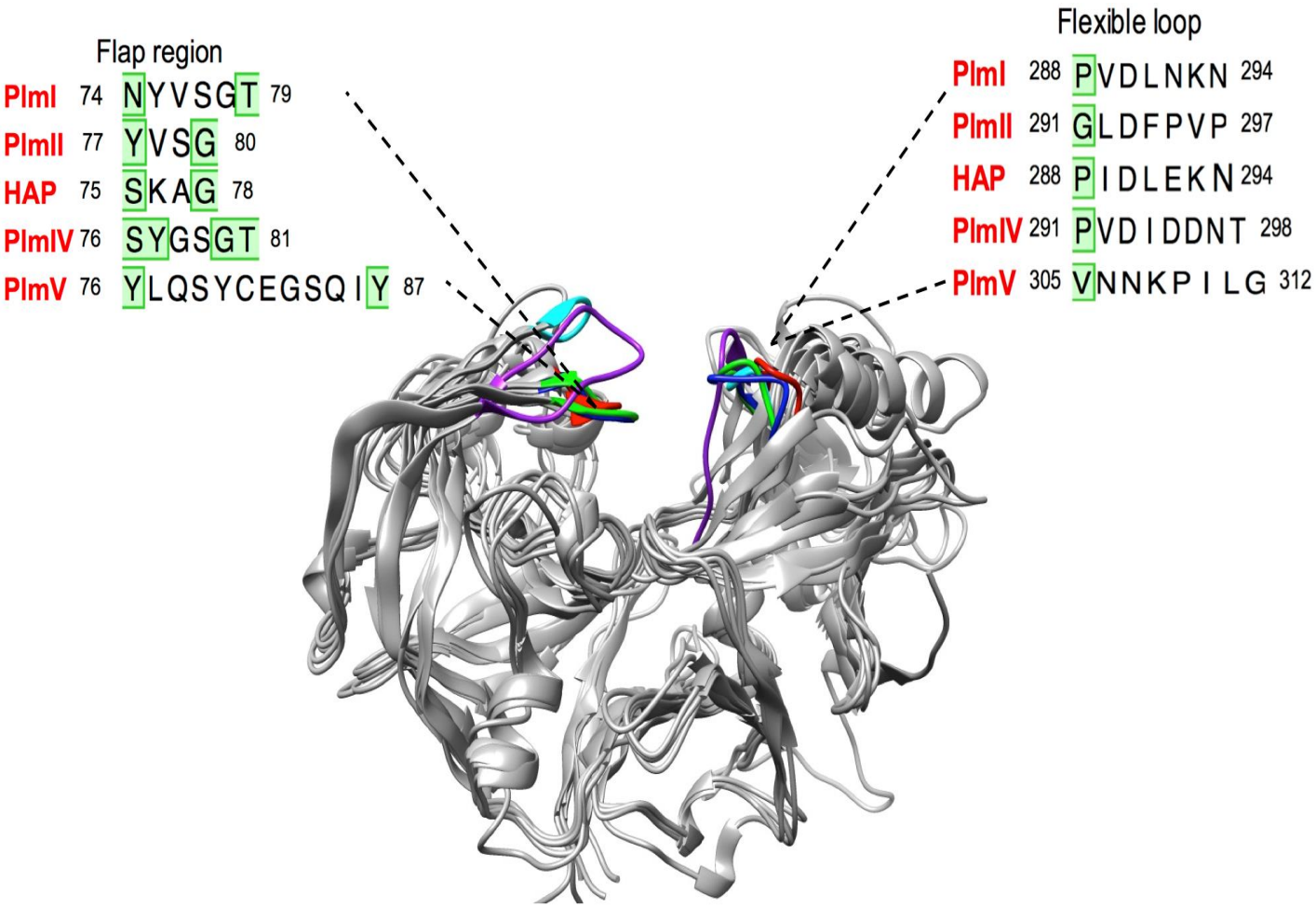
Lara McGillewie and Mahmoud E. Soliman*

Molecular Modelling & Drug Design Research Group, School of Health Sciences, University of KwaZulu-Natal, Westville, Durban 4001, South Africa

*Corresponding author: Mahmoud E.S. Soliman, email: soliman@ukzn.ac.za

Telephone: +27 031 260 7413, Fax: +27 031 260 7792

Graphical Abstract



Abstract

Herein, for the first time, we comparatively report the opening and closing of apo plasmepsin I – V. Plasmepsins belong the aspartic protease family of enzymes, and are expressed during the various stages of the *P. falciparum* lifecycle, the species responsible for the most lethal and virulent malaria to infect humans. Plasmepsin I, II, IV and HAP degrade hemoglobin from infected red blood cells, whereas plasmepsin V transport proteins crucial to the survival of the malaria parasite across the endoplasmic reticulum. Flap structures covering the active site of aspartic proteases (such as HIV protease) are crucial to the conformational flexibility and dynamics of the protein, and ultimately control the binding landscape. The flap-structure in plasmepsins is made up of a flip tip in the N-terminal lying perpendicular to the active site, adjacent to the flexible loop region in the C-terminal. Using molecular dynamics, we propose three parameters to better describe the opening and closing of the flap-structure in apo plasmepsins. Namely, the distance, d_1, between the flap tip and the flexible region; the dihedral angle, ϕ, to account for the twisting motion; and the TriCα angle, θ_1. Simulations have shown that as the flap-structure twists, the flap and flexible region move apart opening the active site, or move towards each other closing the active site. The data from our study indicate that of all the plasmepsins investigated in the present study, Plm IV and V display the highest conformational flexibility and are more dynamic structures versus Plm I, II and HAP.

Keywords: Malaria; *Plasmodium falciparum*; Plasmepsin; Flap motion; Molecular Dynamics

1. Introduction

Malaria is one of the most infectious and deadly protozoan diseases known to man. In 2012, malaria caused an estimated 207 million cases and approximately 627 00 deaths, of which 90% occurred in sub-Saharan Africa mostly in young children under the age of 5 [(1)]. The spread of the malaria parasite, *Plasmodium*, in humans is dependent on the female *Anopheles* mosquito vector [(2)]. Humans are infected by the following *Plasmodium* species - *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae* and *Plasmodium falciparum* [(2, 3)]. *P. falciparum*, which predominates in Africa, is the most virulent human parasite, causing almost all human deaths [(4, 5)]. Worldwide the malaria pandemic is not improving [(6)] as there are no effective vaccines against malaria, and increasing resistance to both insecticides (to control mosquito vectors) and antimalarial drugs is of growing concern [(2, 5)].

The *Plasmodium* parasite has a complex life cycle, during which the plasmodium parasite undergoes numerous developmental stages [(5)]. The erythrocytic phase is responsible for the clinical symptoms of malaria, typically characterized by high fevers due to the lysis of infected red blood cells (RBCs) [(2, 5)]. Hemoglobin degradation occurs in the acidic food vacuole (FV), with a pH of approximately 5 [(8, 9)]. Hemoglobin degradation provides energy for the parasite, and generates amino acids used in parasitic protein biosynthesis [(10)]. In *P. falciparum*, several enzymes have been implicated in hemoglobin proteolysis and represent potential targets for drug design: three cysteine proteases – falcipains; four aspartic proteases – plasmepsins (Plm) and histo-aspartic protease (HAP); one metalloproteases – falcilysins and one dipeptidyl aminopeptidase (DPAP1) [(12)]. Among all the *Plasmodium* species that infect humans, *P. falciparum* is the only species that has numerous plasmepsins active within the FV [(13)]. Hemoglobin degradation occurs in an semi ordered manner in the acidic food vacuole [(12)], in which plasmepsins (specifically Plm I and II) initially cleave the native hemoglobin tetramer

between αPhe33-αLeu34 in the highly conserved hinge region, unravelling and exposing the protein to further degradation [14, 15]. Smaller proteins (globins) are generated by further plasmepsin (HAP and Plm IV) and falcipain degradation; metalloproteases then cleave globins into oligopeptides. Hydrolysis to free amino acids occurs via aminopeptidases in the parasitic cytosol [2]. Protease inhibitors have shown that hemoglobin degradation is critical for the survival of the *Plasmodium* parasite [16], although dual protease families with overlapping function and partial redundant degradation pathways are utilised in hemoglobin degradation [17]. Aspartic and cysteine protease inhibitors show synergistic activity in inhibiting the growth of cultured *Plasmodium*, a combination of aspartic and cysteine protease inhibitors may offer the most effective treatment, potentially limiting resistance to protease inhibitors [18].

Aspartic proteases are found in most eukaryotes, generally functioning in a similar manner [12]. In this class of enzymes, a water molecule acts as the active nucleophile that breaks the scissile bond of the substrate. Different *p*Ki values of the aspartic residues within the catalytic dyad, ensures that one acts as an acid and the other a base, a 'push and pull' reaction forming a transient tetrahedral oxyanion intermediate. Once the peptide bond is protonated, the substrate is cleaved into two shorter peptides [12]. Aspartic proteases found in *Plasmodium*, previously known as aspartic hemoglobinases, are functionally classed together as plasmepsins. Sequencing of the *P. falciparum* genome has revealed 10 genes encoding aspartic proteases I-X, including the closely related histidine-aspartic protease (HAP and/or Plm III) [12, 19]. Despite a high degree of sequence similarity, each plasmepsin has different responses to inhibitors and unique substrate specificity due to the heterogeneous nature of the residues lining the active site, including the flap and flexible loop region (flap-structure) [20]. The present study will focus on plasmepsins I-V.

In *P. falciparum*, vacuolar plasmepsins (Plm I,II, IV and HAP) are expressed during the intraerythrocytic stages and reside within the acidic FV where they are responsible for hemoglobin degradation [(14)]. The four genes are clustered together on chromosome 14, and are more closely related than those encoding Plm V-X [(12)]. Vacoular plasmepsins I, II and HAP are not found in the genomes of other *Plasmodium* species, only *P. falciparum* actively expresses all four FV plasmepsins throughout the asexual phase [(13)]. Whereas, plasmepsins located outside the FV, Plm V-X, can be found in all other *Plasmodium* species [(14, 21)]. Sequence comparison has revealed a single FV plasmepsin in other *Plasmodium* species, orthologous to PlmIV, responsible for hemoglobin degradation. It has been postulated that Plm I, II and HAP were created by gene duplications and have evolved unique, although redundant functions in *P. falciparum* [(13)]. Plasmepsin I and II (75% identity) are expressed in the early stages of RBC infection, PlmI readily cleaves native hemoglobin whereas PlmII prefers acid denatured hemoglobin [(12, 22)]. Plasmepsin IV and HAP are expressed during the later stages of the erythrocytic phase, where they further degrade hemoglobin peptides (globins) generated from Plm I and II degradation. Plasmepsin II and IV have also been implicated in the degradation of the membrane skeleton proteins of RBCs [(14)]. Histo-aspartic protease has a 60% identity to PlmI, II and IV, preferentially cleaves denatured globins at unknown sites downstream of Plm I and II [(14, 23)]. The functional redundancy of FV plasmepsins in *P. falciparum* confers a selective growth advantage, whereby multiple FV plasmepsins (with different cleavage specificity and complementary roles) synergistically cleave hemoglobin to increase catalytic efficiency and degradation of hemoglobin [(14, 24)]. Individually, vacuolar plasmepsins are not essential during the asexual phase and knock down or elimination of all four FV plasmepsins shows that other proteases are capable of degrading hemoglobin sufficient for survival [(25)]. Due to the functional redundancy within the FV plasmepsins, the use of selective inhibitors targeting a single plasmepsin will not effectively or efficiently kill the parasite [(24)].

Malaria parasites develop in the PV, and export over 450 proteins into the infected RBC cytoplasm that are crucial to cellular remodelling and virulence [(26)]. Proteins destined for export contain an N-terminal signal sequence (RxLxE/Q/D) known as the plasmodium export element (PEXEL) [(27, 28)]. Upon export, this motif is cleaved by PlmV in the endoplasmic reticulum (ER). Plasmepsin V is located on chromosome 13, and localized to the parasitic ER [(29)]. Deletion of the PlmV gene has revealed that it is crucial to the survival of the malaria parasite [(29-31)]. Analyses have shown that PlmV is only distantly related (17% to Plm II) to the other plasmespsins and is conserved in all *Plasmodium* species [(12, 13)]. Plasmepsin V cleaves PEXEL-containing proteins at the carboxy terminal, after the conserved Leucine residue (RxL↓) [(32, 33)]. Approximately 8% of *P. falciparum*'s proteins are exported via the PEXEL exportome pathway [(33, 34)]. Plasmepsin V is distinguishable from most aspartic proteases, including vacuolar plasmepsins, based on its affinity and binding to pepstatin (potent inhibitor of aspartic proteases) [(29)]. Unlike vacuolar plasmepsins, where pepstatin acts as a potent inhibitor, pepstatin does not bind to PlmV [(29)]. It is well known that *P. falciparum* is more virulent compared to other *Plasmodium* species, this is due to the fact that infected RBCs have the ability to adhere to the walls of blood vessels via exported molecules known as adhesins. Thus, bypassing the spleen, where splenic macrophages detect and eliminate deformed RBCs with altered antigenicity [(35)].

Plasmepsins have a varying degree of identity and similarity with human aspartic proteases, with the highest degree of similarity with Cathepsin D (catD) a lysosomal enzyme [(36)]. Thus, catD is often used as a marker for cross inhibition when designing inhibitors of plasmepsins [(36, 37)]. Similar to most proteases, plasmepsins are synthesized as inactive zymogens, pro-plasmepsins, enzymatically activated upon cleavage of N-terminal pro-segment [(38, 39)].

Activation of pro-plasmepsin I, II, HAP and IV requires the acidic conditions present in the FV, removal of the pro-domains releases mature active enzymes [12]. The pro-domain of PlmV is non-inhibitory, thus removal is not necessary for the activation of PlmV [6, 29]. The mature form of the enzymes fold into two topologically related N- and C-terminal domains. These two domains are connected towards the bottom, forming the binding cleft housing the active site with two catalytic aspartic acid (Asp/D) residues [12, 14], with the exception of HAP in which one catalytic Asp is replaced by histidine (His/H) [40, 41]. The His residue within the active site of HAP is protonated, thus HAP functions in an aspartic-protease fashion with lower affinity for hemoglobin relative to other plasmepsins, functioning at an optimum pH near 6 [14, 23]. The amino and carboxyl ends of each domain, are assembled into a characteristic six-stranded inter-domain β-sheets, which anchors the two domains together [42]. The highly flexible domains are conserved within the aspartic protease class of enzymes, where substrate interaction occurs. Similar to other aspartic acid proteases, the N-terminal domain of plasmepsins contains a single long β-hairpin loop (flap), which lies perpendicularly over the binding pocket adjacent to the flexible loop region found in the C-terminal domain [42]. Crystallographic studies have shown a high degree of enzyme flexibility, this substantial flexibility is essential for the recognition and binding to different sequences within the hemoglobin molecule [2, 43]. Conformational changes in the flap and flexible loop region regulate substrate access, assuming a more closed conformation upon substrate binding.

Flap dynamics is a fundamental aspect in understanding the ligand binding landscape, as well as the overall conformational flexibility of enzymes. In HIV protease, flaps covering the active site play a major role in substrate binding, where flap motions dictate the movement of the incoming ligand and tightly bind the ligand in the active site [44]. Two main flap conformations

have been observed experimentally; (i) a ligand-bound closed conformation, and (2) a apo semi-open conformation [44]. The most commonly used metric to measure various flap confirmations is based on the distance between the Cα of Ile50 and Ile149 [44-46]. This parameter alone does not adequately describe the curling or twisting motion observed in HIV protease, which has led to the introduction of additional parameters to more accurately describe flap motions and dynamics [46, 47]. The flap dynamics of HIV protease highlights the significance of defining more suitable parameters that best describe flap dynamics for plasmepsins.

It is well known that the flap and loop region is highly flexible for Plm I-IV, both in the free and ligand-bound conformations [14, 48-50]. Crystallographic studies of PlmII has shown that residues in the active site move to accommodate substrate binding, while others fold inwards to form hydrogen bonds. Residues in the flap (Val78) and loop region (Ile290, Leu292, Phe294) undergo majority of the relative changes between free and bound PlmII conformations, not only in relative positions of the side chains but also movements within the main chains [42]. Flap opening and closing for PlmII was measured by the distance between the flap tip, Val78, and the residue opposite the tip in the hydrophobic rim of the active site, Leu292 (hinge residue) [42]. These two residues close the active site upon substrate binding, as can been seen upon pepstatin A binding [43]. Previous parameters to measure flap motion were defined using a rigid crystal structure for PlmII, and only described motions in a one-dimensional space [42, 51]. The parameters used to accurately describe flap motions are poorly defined in the literature, both experimentally and computationally. Recently, we have proposed additional parameters to measure flap dynamics, which we believe best describes not only flap and flexible region motions, but also the characteristic 'twisting' [52] motion during the opening and closing of plasmepsins [51]. The aim of the present study is to investigate and compare the flap-structure movements of free, unbound Plm I-V.

2. Computational Methodology

2.1. System Preparation

The simulated systems, crystal structures (PDB codes) and corresponding abbreviations are listed in **Table 1**. All crystal structures were obtained from the protein data bank (PDB) [53] and prepared as detailed in our previous reports [54, 55]. Free un-liganded structures were generated by manually deleting bound inhibitors from the crystal structures complexes where applicable. For PlmV, where to date no crystal structure has yet been published, a homology model was used in the simulations as previously described [26, 56]. All enzymes were prepared and visualised using Chimera [57]. In total, 5 systems (**Table 1**) were subjected to 50 ns continuous molecular dynamic simulations as described below in section 2.2. To verify and validate the results from the single continuous MD approach adopted herein, we ran multiple MD simulations (using 3 replicas) for one system and compared the results.

Table 1. Crystal structures of the simulated systems, PDB codes and abbreviations

System	PDB code	Abbreviation*
Apo Plm I	3QRV	PlmI
Apo Plm II	1LF4	PlmII
Apo HAP	3FNS	HAP
Apo Plm IV	1LS5	PlmIV
Apo Plm V		PlmV

* Abbreviations as used throughout the manuscript

2.2. Molecular Dynamics and Post dynamics analyses

To investigate the flap opening and closing of apo Plm I-V, a continuous 50 ns MD approach was utilised in the present study. All-atom, explicit solvation unrestrained molecular dynamic simulations was performed using the GPU version of the PMEMD engine incorporated with

the Amber 14 package [52]. The ff99sb Amber force field was used to describe the protein systems [53]. The LEAP module of Amber 14 was used to generate the topologies for the system, by adding missing hydrogen atoms and counter ions for neutralization. Before simulations, all systems were solvated in a orthorhombic box of TIP3P [54] water molecules at a distance of 10 Å from all protein atoms. Simulations were performed using periodic boundary conditions. The particle-mesh Ewald [55] (PME) method was used to treat long-range electrostatic interactions, with a direct space and van der Waals cut-off distance of 12 Å. All systems were minimised in two steps – partial minimisation followed by full minimisation. Initially all systems underwent minimisation with 750 steps of steepest decent, thereafter 1750 steps of conjugate gradients and harmonic restraints with a constant force of 500 kcal/mol $Å^2$ was applied to the solute. Following this, a further 1000 a step of unrestrained conjugate gradient energy minimisation was performed on all systems. All systems were gradually heated from 0 to 300 K for 50 ps in the canonical ensembles (NVT), by the application of harmonic restraints of 10 kcal/mol $Å^2$ and collision frequency of 1.0 p.s-1 to all solutes in the system. The simulation temperature was controlled using the Langevin thermostat [56]. All systems were equilibrated at 300 K in an NPT ensemble for 500 ps with no restraints and a constant pressure (1 bar) was maintained using the Berendsen [57] thermostat. The SHAKE algorithm [58] was used to constrict the bonds of all hydrogen atoms. A 2 fs time step and the SPFP precision model [59] was used for all simulations. Continuous MD was performed on all systems for 50 ns in an NPT ensemble with a constant pressure of 1 bar, constant temperature of 300 K and a pressure coupling constant of 2 ps. Trajectories were saved every 1 ps and further analysed (e.g. RMSD, RMSF, d_1, angles) using the CPPTRAJ module incorporated in Amber 14. As an additional validation of the continuous MD approach utilised in the current study, a multiple MD approach was performed for one system. In which three 50 ns MD runs with different initial velocities

were performed and the average trajectory analysed. The trends for both approaches are similar, the results are reported in the supplementary information (graphs S4 – S11).

The graphical interface of UCSF chimera [50] was used to visualise all enzyme structures, and data was analysed and plotted using the GUI of Microcal Origin data analysis software version 6 [60].

2.3. Parameters used in the present study

The present study utilised parameters previously proposed by our group to more accurately describe flap dynamics for aspartic proteases, in which Plm II was used as a prototype [44]. These parameters are explained in more detail and illustrated below (Section 3.1, Figure 2, Table 1 and 2).

3. Results and discussion

3.1. Sequence analysis and experimentally defined parameters

Although the full sequence analysis of Plm I-V has previously been reported [12, 65], the present study provides the first in-depth analysis of the two regions that could impact on the flap dynamics and motion of plasmepsins, more specifically, the flap region including the flap tip, and the flexible loop region including the hinge residue – collectively called the flap-structure (**Table 1**, **Figure 1 and 3**). Previously the flexible loop region has been referred to as a proline-rich area in homologous aspartic proteases renin and catD, where three consecutive proline residues make up the loop that forms part of the binding site [43]. In PlmII, the β-strand that makes up the loop is shorter and has only two proline residues separated by a valine [42]. Upon

sequence analysis, we found that the proline-rich area is not present in the other plasmepsin variants; henceforth the previously defined proline-rich area will be referred to as the flexible loop region (**Table 2, Figure 1**). From the sequence analysis it can be seen that the flap-structure is heterogeneous, where the flap tip and hinge region reportedly responsible for the motion of the flap-structure is not conserved between plasmepsins (**Table 2, Figure 1**). In Plm I, II, IV and V a highly conserved tyrosine residue [(52)] lies adjacent to the flap tip in the β-hairpin loop, with the exception of HAP where it is replaced by a serine residue [(23)]. In the flexible loop region, the hinge residue is followed by an aspartic acid with the exception of PlmV where the hinge residue is followed by an asparagine residue (**Table 2, Figure 1**). The changes observed in the flap-structure sequence will ultimately affect the motion of the flap-structure and the binding landscape.

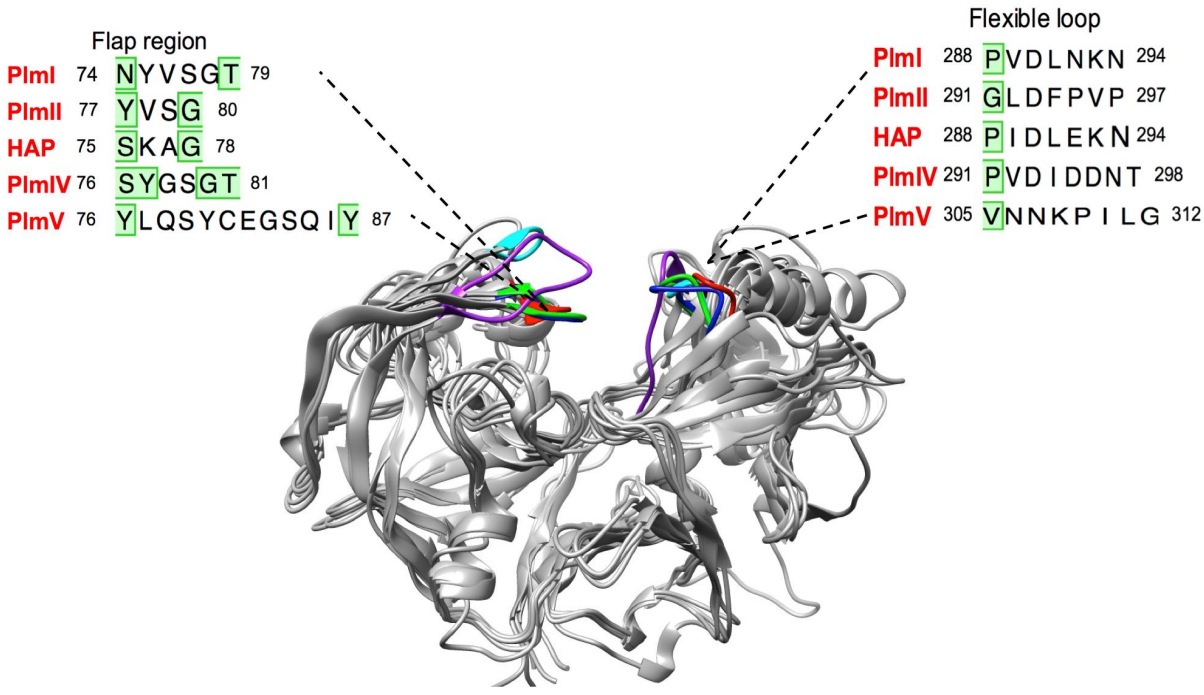


Figure 5. Schematic representation of the flap-structure and corresponding sequences for Plm I (blue), II (red), IV (cyan) and V (purple) and HAP (green).

Table 2. Plasmepsin I–V - Length, active sites, flap and flexible loop regions.

	Length (amino acids)	Resolution (Å)	R-value	Active sites	Flap region	Flexible region
PlmI *(66)*	336	2.40	0.211	Asp32, Asp215	Asn74,Tyr75,***Val76***,Gly78, Thr79	Pro288,***Val289***,Asp290,Leu291, Asn292,Lys296,Asn297
PlmII *(42)*	331	1.90	0.217	Asp34, Asp214	Tyr77,***Val78*****,**Ser79,Gly80	Gly291,***Leu292***,Asp293,Phe294, Pro295,Val296,Pro297
HAP *(40)*	332	2.50	0.226	His32, Asp215	Ser75,***Lys76*****,**Ala77,Gly78	Pro288,***Ile289***,Asp290,Leu291**,** Glu292,Lys296,Asn297
PlmIV *(42)*	328	2.80	0.220	Asp34, Asp214	Ser76,Tyr77,***Gly78***,Ser79, Gly80,Thr81	Pro291,***Val292***,Asp293,Ile294, Asp295,Asp296,Asn297,Thr298
PlmV *(26, 56)*	357			Asp36, Asp220	Tyr76,Leu77,Gln78,Ser79, Tyr80,***Cys81***,Glu82,Gly83, Ser83,Gln85,Ile86,Tyr87	Val305,***Asn306***,Asn307,Lys308, Pro309,Ile310,Leu311,Gly312

Blue – hydrophilic residues; purple – hydrophobic residues. Flap tips are bolded and italicised. The flexible loop is taken from one beta sheet to the following, outlined by the residues above. The residue responsible for the lateral 'pivoting' motion upon substrate binding (hinge residue) is in bold and italicised. +

To date, no computational studies have comparatively investigated the opening and closing of the flap-structure in plasmepsins. From previous molecular dynamic simulations of free and bound PlmII, we have shown a characteristic 'twisting' motion of the flap region and the coordinated recoiling of the flexible loop region (flap-structure) *(51)*. In this study we have defined more accurate parameters to describe the opening and closing of both free and bound PlmII, by calculating the distance between the flap tip and the hinge residue of the flexible loop (d_1), in relation to the dihedral angle ϕ, as well as the TriCα angles, θ_1 and θ_2 (**Figure 2, Table 2 and 3**) *(51)*. These parameters, in addition to other post MD analyses, were used to compare the flap-structure motions and dynamics between Plm I-V.

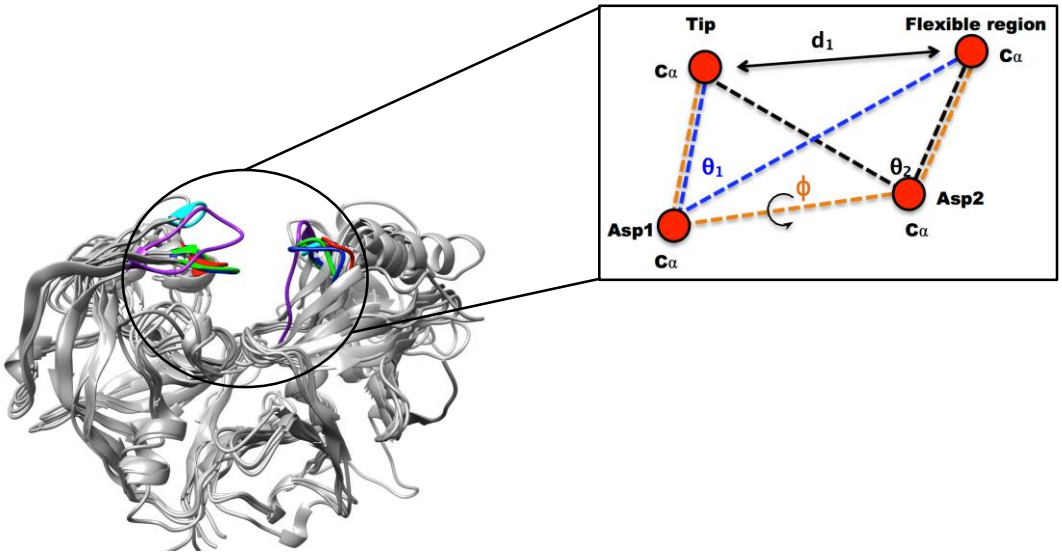


Figure 6. Schematic representation of the parameters used to define the flap-structure motion: d_1 the distance between the flap tip and flexible region hinge residue, the dihedral angle ϕ and the TriCα angles, θ_1 and θ_2. Asp1 denotes the first aspartic acid residue for Plm I (blue), II (red), IV (cyan) and V (purple), in HAP (green) the first catalytic Asp is replaced with by His. Asp2 denotes the second Asp making up the active site for Plm I-V (refer to Table 1).

Table 3. Residues used to calculate the distance (d1), dihedral angle (ϕ), and TriCα θ1 and θ2.

	PlmI	**PlmII**	**HAP**	**PlmIV**	**PlmV**
Distance (d_1)	Val76-Val289	Val78-Leu292	Lys76-Ile286	Gly78-Val292	Cys81-Asn306
Dihedral ϕ	Val76-Asp32-Asp215-Val289	Val78-Asp34-Asp214-Leu292	Lys76-His32-Asp215-Ile286	Gly78-Asp34-Asp214-Val292	Cys81-Asp36-Asp220-Asn306
TriCα θ_1	Val76-Asp32-Val289	Val78-Asp34-Leu292	Lys76-His32-Ile286	Gly78-Asp34-Val292	Cys81-Asp36-Asn306
TriCα θ_2	Val76-Asp215-Val289	Val78-Asp214-Leu292	Lys76-Asp214-Ile286	Gly78-Asp214-Val292	Cys81-Asp220-Asn306

3.2. Molecular dynamic simulations and post-dynamic analysis

Snapshots of the molecular dynamics trajectories of Plm I-V were analysed throughout the 50 ns simulation (**Figure 3**). From visual analysis of the snapshots, it can be seen that all plasmepsins transition between an open and semi-open conformations throughout the duration of the simulation. As previously reported for apo PlmII [(51)], the flap and flexible loop region of

Plm I-V recoils inwards towards the active site as the enzymes closes. Between the enzymes, a varying degree of motion and flexibility was observed. This in part can be attributed to sequence heterogeneity within the flap-structures (**Table 2**, **Figure 1**). The differences in motion from the trajectories are described in more detail below.

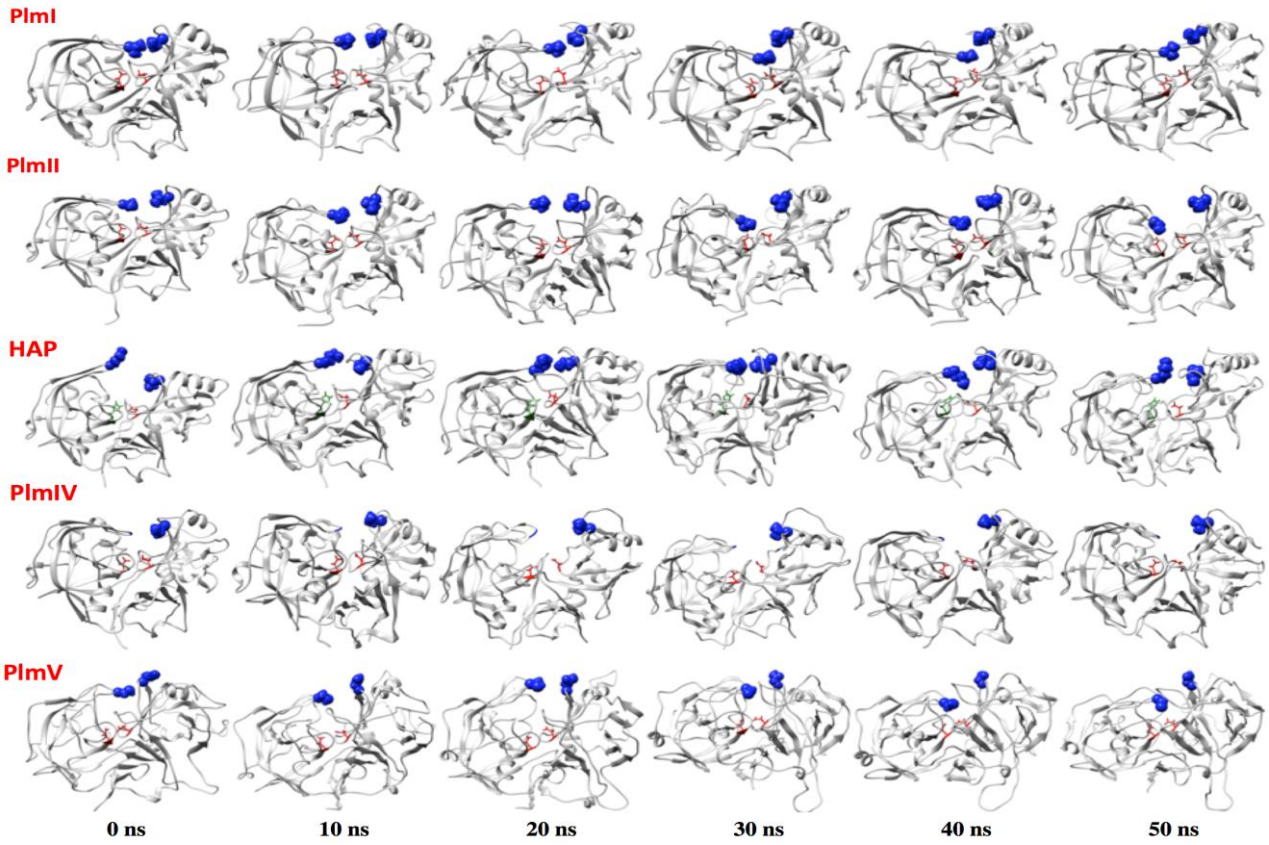


Figure 7. Flap and flexible loop region movements of apo Plm I-V throughout a 50 ns molecular dynamics simulation.The blue regions denote the flap tip and the hinge region within the flexible loop. Red highlights the catalytic aspartic acid residues, with the exception of HAP in which the green highlights the catalytic histidine forming the active site.

3.2.1. System stability

Prior to MD analysis, the root mean square deviation (RMSD) of the Cα backbone and potential energy fluctuations of the trajectories were monitored during the 50 ns simulation to ensure stability within the systems and accuracy of successive post-dynamic analyses. Convergence and stabilisation in the potential energies was observed, with no fluctuations greater than 1000

kcal/mol for all systems (Figure S1). Convergence in the RMSD trajectories occurred after approximately 10ns (Figure S2).

3.2.2. Root mean square fluctuation (RMSF)

The per-residue fluctuation (RMSF) of the 5 different apo plasmepsins, show the conformational flexibility and dynamics of the proteins (**Figure 4**). The amino acids residues between 225 and 300 show a higher fluctuation relative to other residues. This is where the flexible loop region is situated, and the fluctuation can be attributed to the 'twisting' motion of the flexible region as a whole and not to a single residue. Previous studies have identified four loops that display large structural changes upon ligand binding in vacuolar plasmepsins – L1 residues 12-14, L2 residues 158-165, L3 residues 231-244 and L4 residues 277-283 (PlmII numbering) [(67, 68)]. The L3 loop was also shown to be highly flexible, and entropic analysis revealed that the L3 loop inherently loses entropy upon ligation in order to attain stability. Residues within these loops have negligible nonpolar and electrostatic interactions with bound inhibitors, but play a critical role in the openness of the binding pocket [(67, 68)]. High residual fluctuations can be seen for all plasmepsins in the L3 loop region, indicative of conformational flexibility (**Figure 4**). Higher fluctuations are also observed between residue 50 and 80 where the flap region is situated (**Figure 4**).

Overall, all plasmepsins show a high conformational flexibility with a similar trend in fluctuations, PlmIV and PlmV show the most erratic and volatile fluctuations relative to PlmI, II and HAP (**Figure 4**). This can also been seen in the higher average fluctuations for Plm IV and V, 1.33 Å and 1.58 Å respectively (**Table 4**). Higher residual fluctuations were observed in the flap tip residues for all plasmepsins, although Plm IV and V displayed the highest

fluctuations in this region, 2.85 Å and 4.51 Å respectively (**Table 4**). Flap tip residues of all plasmepsins showed higher fluctuations relative to the hinge residue in the flexible loop, indicative of more movement and flexibility in this region (**Table 4**). As previously reported, the hinge residue in the flexible loop (Leu292, PlmII numbering) shows less fluctuation and moves relatively less compared to the loop region as a whole for all plasmepsins investigated in the present study (**Figure 4**). The fluctuations around the two residues within the active sites are relatively low, and all follow a similar trend. With the exception of PlmIV, where the second Asp residue of the catalytic dyad (Asp214) shows a higher residual fluctuation compared to the first (Asp34), and compared to the active site residues in other plasmepsins (**Table 4**). These results indicate that Plm I-V are all highly flexible enzymes, most of the flexibility observed is within the flap and flexible loop region. Plamsmepsin IV and V appear to be the most flexible and dynamic enzymes. The fluctuations observed in the RMSF values, are supported by a similar trend in the fluctuations in the b-factor values (Figure S3).

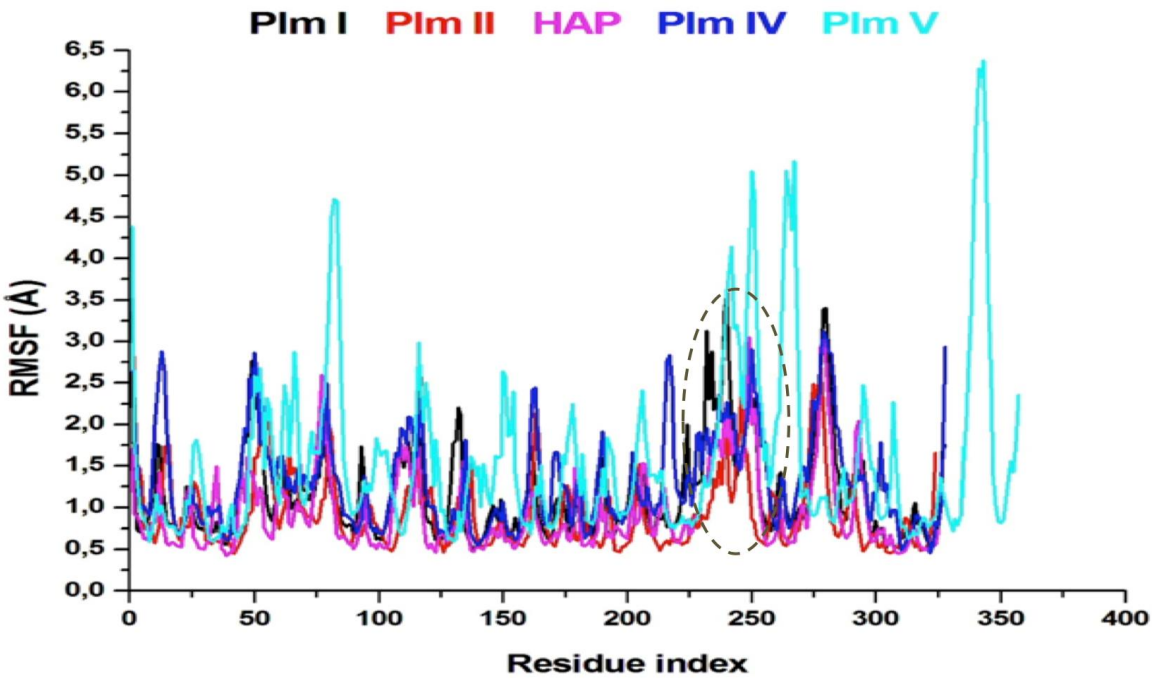


Figure 8. Plot comparing the C-α root mean square fluctuations (RMSF) of apo Plm I (black), II (red), HAP (magenta), IV (blue) and V (cyan). Dotted region highlights highly flexible L3 loop.

Table 4. Root mean square fluctuation (RMSF) values of apo Plm I-V for the binding site residues, tip and hinge residue and total average fluctuations.

	PlmI	PlmII	HAP	PlmIV	PlmV
Asp1*	0.75	0.59	0.79	0.96	0.64
Asp2	0.87	0.61	0.94	1.20	0.80
Flap tip	1.75	2.00	2.23	2.85	4.51
Hinge	1.31	0.93	0.95	1.30	1.30
Average	1.16	0.94	0.94	1.33	1.58

*Active site residue 1 for Plm I, II, IV and V = Asp, HAP = His, distances measured in Å.

3.2.3. Distance (d_1) and dihedral angle (ϕ)

The distance, d_1, calculated between the Cα atom of the flap tip and the Cα atom of the opposite hinge residue in the flexible loop; and the dihedral angle, ϕ, calculated between the Cα atoms of the tip - catalytic dyad – hinge residue, has previously been used by our group to best describe flap motion for several aspartic proteases [(51)] (**Figure 2, Table 3**). For plasmepsins, d_1 accurately describes the opening and closing of the flap structure, whereas ϕ is used to gain insight into the twisting motion of the flap structure.

Throughout the simulation, all plasmepsins move between open and semi-open conformations (**Figure 5A**). Plasmepsin I has two opening conformations, at ~8-12 ns and again at ~33-36 ns. Assuming a more closed conformation from ~19-30 ns and a semi-open configuration towards the end of the simulation (**Figure 5A**). Plasmepsin II assumes a more semi-open conformation throughout the simulation, opening briefly around ~20-21 ns and again toward the end of the simulation at ~43.5 ns (**Figure 5A**). Of all the plasmepsins, HAP fluctuated the most between semi-open conformations throughout the simulation, as reflected by erratic changes in d_1. Starting the simulation in a more open configuration till ~8 ns where it remains in a semi-open conformation opening briefly at ~25.3 ns, 36.8 ns and 41.2 ns (**Figure 5A**). Plasmepsin IV starts the simulation in a more closed conformation until ~5.3 ns, where after it briefly (~6.1 ns) assumes a semi-open conformation returning to a more closed conformation at ~8.5 ns,

fluctuating between a semi-open and more closed conformation. For the remainder of the simulation Plm IV assumes an open conformation (**Figure 5A**). Plasmepsin V starts the simulation in a more closed conformation up until 8 ns, briefly moving to an open conformation from ~10.2-12.2 ns. Thereafter it fluctuates between semi-open conformations till ~31.5 ns, assuming a more open conformation until the end of the simulation (**Figure 5A**).

The highest average distance was observed for Plm IV and V, 13.97 Å and 13.56 Å respectively (**Table 5**). The flap structure for Plm IV and V seem to move apart the most, as reflected by the maximum distance of 19.63 Å and 20.75 Å respectively and wider range of motion (Δ**)** (**Table 5**). Whereas PlmII with a maximum d_1 of 16.78 Å, has the lowest range movement in the flap structure (**Table 5**). Thus, Plm IV and V are more flexible and dynamic compared to PlmII.

Table 5. Distance by which the flap structure moves, measured in angstroms (Å).

	PlmI	**PlmII**	**HAP**	**PlmIV**	**PlmV**
Average	9.67	13.24	11.91	13.97	13.56
Max	15.15 @34.8ns	16.78 @33.3ns	17.36 @7.8ns	19.63 @22.6ns	20.75 @10.6ns
Min	5.94 @16.4ns	8.92 @12ns	8.62 @17ns	7.47 @18ns	6.46 @21.5ns
Δ*	9.22	7.87	8.74	12.16	14.29

* Change between the maximum and minimum distance.

The dihedral angle (ϕ) calculated in the post MD analysis gives insight into the twisting motion of the Cα atoms between the flap tip, active site and hinge residue in the flexible loop region. Overall, all plasmepsins display a twisting motion (change in ϕ: from negative to positive, and/or positive to negative) which is reflected by a change in d_1, albeit small (**Figure 5B**). Two major twisting motions can be seen for PlmI, which coincide with an increase in d_1 (**Figure**

5B). Between ~3.5-12 ns PlmI twists from -6.31° to 24.07°, and again at ~32.7-38.5 ns from 24.6° to -14.39° roughly correlating with the more open conformations of the flap-structure at these time points. Plasmepsin II showed less definitive twisting events in the flap-structure, the twisting observed is more gradual and less erratic. Most of the twisting observed in PlmII occurred in the beginning of the simulation until ~19 ns, -33.36° to 27°, which coincides with the semi-open conformation. Thereafter, no twisting is observed up until the end of the simulation at ~41.7-44.2 ns, where PlmII twists from -12.9 to 27.17° as it transitions from a semi-open conformation to a more open conformation (**Figure 5B**). Only one major twisting motion can be seen in HAP, at ~4.3-9.3 ns, where the flap-structure twists from 8.39° to -36.06° (7.1 ns) and then to 3.5°. This twisting movement occurs as HAP assumes a more open conformation, with the highest distance between the flap and flexible loop region between 6.6-7.8 ns (**Figure 5B**). As PlmIV moves into a more open conformation, a major twisting motion can be seen at ~14.6-20.5 ns as the flap-structure moves from -22.43° to 22.98°. Thereafter, as PlmIV remains in a more open conformation and no significant changes in ϕ are observed (**Figure 5B**). Plasmepsin V starts off the simulation in a semi-open conformation, as the flap-structure moves towards a more open conformation a twisting motion at ~4.5-9 ns can be seen, as ϕ changes from 19.7° to -24.53°. As PlmV transitions into an open conformation towards the end of the simulation, a twisting motion between ~27.2-29.4 ns is observed as ϕ moves from -7° to -39.78° (**Figure 5B**). Plasmepsin IV and V twist more compared to other plasmepsins in the present study, as reflected by the magnitude and wide range of ϕ values. Overall, an increase in d_1 and shift in ϕ corresponds to the opening of the flap structure as the flap twists away from the binding cavity.

3.2.4. TriCα angles, θ_1 and θ_2

In addition to calculating the distance and the dihedral angle fluctuations in the flap-structure, we have previously reported that calculating the TriCα angles (θ_1 and θ_2) better explain the flap dynamics of PlmII [(51)]. In unpublished work by our group, we have also shown the importance of TriCα angles, more specifically θ_2, in understanding the flap opening and closing of apo and bound beta-secretase (an aspartic protease with distant homology to PlmIV [(31)]).

In all plasmepsins investigated in the present study, θ_1 and θ_2 follow similar trends to each other (**Figure 5C and 5D**). The trend observed in θ_1 and θ_2 corresponds to the fluctuations observed in d_1 for Plm I-V. The maximum θ_1 and θ_2 values coincide with opening of the flap-structure, as the flap and flexible loop move away from each other exposing the active site (**Table 6**). The minimum θ_1 and θ_2 values correlate with a more closed flap-structure for Plm I-V, as the flap folds inwards towards the active and the flexible loop recoils closing the active site. Overall, as the flap-structure transitions into more open conformations (increase in d_1), both θ_1 and θ_2 increase, whereas as the flap-structure moves towards a more closed conformations (decrease in d_1) θ_1 and θ_2 decrease accordingly (**Figure 5C and 5D**).

Plasmepsin II has the highest average values for both θ_1 and θ_2, although the minimum values for θ_1 and θ_2 (27.93° and 40.69° respectively) are significantly higher compared to other plasmepsins (**Table 6**). Plasmepsin I, II and HAP have similar ranges in motion reflected by the difference between the maximum and minimum θ_1 and θ_2 values, denoted as Δ in Table 6. Plasmepsin IV and V show the widest range of motion, as reflected by the significant differences between the maximum and minimum θ_1 and θ_2 values (**Table 6**). Thus, Plm IV and V appear to be more flexible and dynamic as the flap-structure is capable of a wider range of

motion, moving the flap and flexible loop closer into the binding cavity as it closes and further away as it opens.

TriCα θ_2 more accurately describes the opening and closing of the flap-structure, although θ_1 and θ_2 follow a similar trend to the changes observed in d_1, the magnitude of θ_2 more accurately describes the extent of flap opening and closing in all plasmepsins investigated in the present study (**Figure 5D**). For PlmII, the maximum d_1 of 16.78 Å at 33.3 ns corresponds to the maximum θ_2 of 84.12° at 33.3 ns (**Table 6**). The wider range of motion around θ_2 could be due to the fact that the second residue of the active site (Asp2) lies in close proximity to the highly flexible L3 loop (residues 231-244, PlmII numbering) [67, 68].

Table 6. TriCα angles, θ_1 and θ_2.

	PlmI		**PlmII**		**HAP**		**PlmIV**		**PlmV**	
	θ_1	**θ_2**	**θ_1**	**θ_2**	**θ_1**	**θ_2**	**θ_1**	**θ_2**	**θ_1**	**θ_2**
Average	27.46°	40.45°	41.79°	60.62°	34.97°	49.76°	39.24°	58.41°	34.64°	49.93°
Max	48.22° @11.8ns	64.5° @33.1ns	57.47° @45.8ns	84.12° @33.3ns	52.03° @2.7ns	65.07° @43.5ns	61.89° @5.6ns	86.32° @5.6ns	55.36° @10.6ns	83.36° @41.4ns
Min	10.8° @2.3ns	22.76° @16.4ns	27.93° @11.9ns	40.69° @1.1ns	21.21° @21.4ns	36.08° @22.2ns	12.26° @2.5ns	26.90° @2.7ns	4.42° @41.5ns	14.84° @2.1ns
Δ*	37.42°	41.75°	29.54°	43.42°	30.82°	28.99°	49.63°	59.42°	50.94°	68.53°

* Change between the maximum and minimum distance.

3.2.5. Radius of gyration (R_g)

The radius of gyration is indicative to the compactness of the tertiary structure of a protein i.e. how folded or unfolded a protein is, and gives insight into the stability of biological molecules during the MD simulation. The R_g for all plasmepsins investigated in the present study, follow similar trends to those reported for d_1 and θ_2 (**Figure 5E**). With the exception of Plm II and V. Plasmepsin II follows a similar trend to d_1 and θ_2 for most of the simulation. However, towards

the end of the simulation as the flap-structure moves to a more open conformation the R_g values do not increase accordingly and remain fairly constant with a brief increase at ~47.5 ns (**Figure 5E**). The R_g for PlmV does not follow a similar trend to that seen in d_1 and θ_2 for most of the simulation, although there are periods during the simulation where R_g increases with an increase in d_1 and θ_2 (**Figure 5E**). This discrepancy could be due to the fact that a homology model was used for PlmV in the MD simulations. In addition, apo PlmV displayed an average R_g of 21.44 Å comparatively higher as compared to PlmI (20.5 Å), PlmII (20.48 Å), HAP (20.79 Å) and PlmIV (20.66 Å). Indicating PlmV is a more open, dynamic and flexible enzyme.

4. Conclusion

As previously reported, the present study indicates that differences in plasmepsin substrate specificity depend on the conformational differences at distant sites rather than variations within the active site itself [(69, 70)]. Molecular dynamic simulations has shown that the activation energy of other aspartic proteases is highly sensitive to the distance between the substrate and catalytic aspartic acid residues in the active, which is controlled by the motion and dynamics of the flaps covering the binding site [(71)]. In the present study we comparatively show the difference in flap motion and dynamics between apo Plm I-V. We also show how the flap opening and closing of plasmepsins can be more accurately describe by combining parameters namely – distance (d_1), the dihedral angle (ϕ) and TriCα angle (θ_2). These three parameters accurately account for the twisting motion responsible for opening the binding cavity, as the flap and flexible loop region move away from each other, and closing of the active site as the flap and flexible loop region move inwards towards each other and the active site.

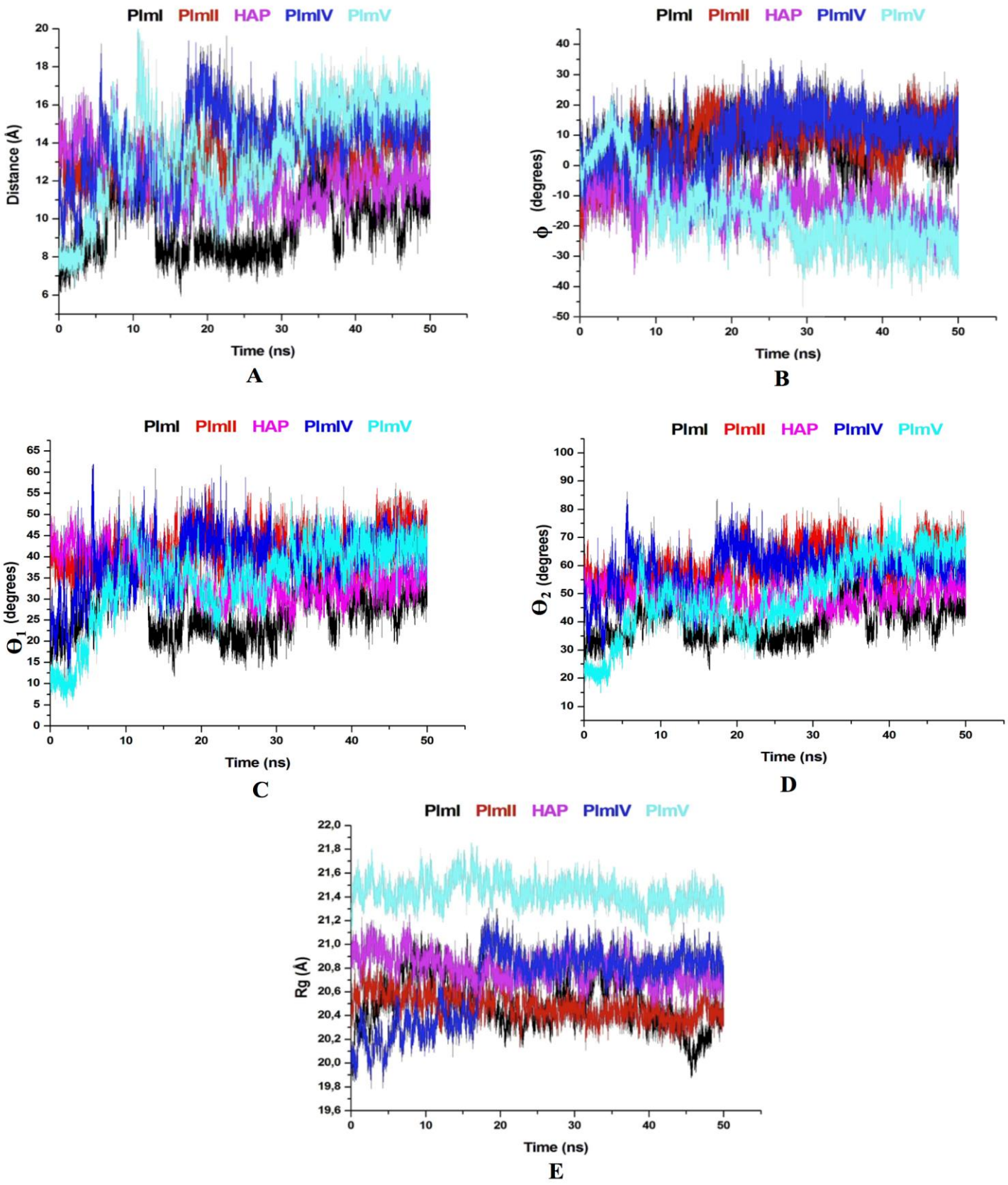


Figure 9. Graphs showing the fluctuation in distance, d1 (A); dihedral angle ϕ (B), Tri Cα angles Θ1 (C) and Θ2 (D); radius of gyration, R_g (E) throughout the 50 ns simulation time.

The data presented above indicate that Plm IV and V are the most flexible and dynamic, capable of wider ranges of motion. The flexibility observed in both these enzymes correlate to their function, as well as their conservation throughout all *Plasmodium* species. From the sequence analysis of the flap-structure, it can be seen the PlmIV has a glycine residue situated at the flap tip (Gly78). The inherent flexibility of glycine accounts for the mobility and increased flexibility observed in PlmIV [72, 73]. Glycine is the smallest amino acid, due to it being achiral, with two hydrogen atoms composing the side chain. This offers less to no steric hindrance of the flap to the opposite flexible loop in PlmIV. Thus the flap-structure of PlmIV is highly mobile and flexible. In *P. falciparum* vacuolar plasmepsins have evolved from PlmIV, whereas in other species PlmIV is the only plasmepsin responsible for hemoglobin degradation. As observed in the present study, mobility and flexibility is less in Plm I, II and HAP possibly indicating that they have evolved more specific functions and unique substrate specificity versus PlmIV. Thereby, ensuring efficient and total degradation of all forms of hemoglobin in the virulent *P. falciparum*.

Higher fluctuations and increased flexibility was also observed for endoplasmic reticulum resident PlmV. The instability reported, could be due to the fact that a homology model was used for the simulations, results need to be validated and verified once a crystal structure becomes available. The wide range of motion and high conformational flexibility seen in PlmV could be to accommodate the wide variety of PEXEL containing proteins destined for export. Unlike other plasmepsins investigated in the present study, PlmV has a range of intracellular ligands. Thus, the high flexibility of the flap-structure in PlmV ensures that all naturally occurring substrates can be accommodated within the active site.

The parameters used in the present will aid in a better understanding of not only the flap dynamics across plasmepsins but also across the aspartic protease family, for both apo and bound enzyme conformations. The results obtained from the present study, further highlight the importance of flap dynamics and how they affect the binding landscape. Which will ultimately aid in the design and development of novel antimalarial drugs.

Acknowledgments

The authors would like to acknowledge The School of Health Sciences at the University of KwaZulu Natal, Westville campus for financial support. Further we would like to acknowledge the CHPC, Cape Town, RSA and the cBio cluster at MSKCC for high performance computational resources.

Conflicts of interest

Authors declare no financial or intellectual conflicts of interest.

5. References

1. http://www.who.int/malaria/media/world_malaria_report_2013/en/.
2. Ersmark, K., Samuelsson, B., and Hallberg, A. (2006) Plasmepsins as potential targets for new antimalarial therapy, *Med Res Rev 26*, 626-666.
3. White, N. J. (1993) Malaria parasites go ape, *Lancet 341*, 793.
4. Miller, L. H., Ackerman, H. C., Su, X. Z., and Wellems, T. E. (2013) Malaria biology and disease pathogenesis: insights for new treatments, *Nat Med 19*, 156-167.
5. Rosenthal, P. J. (2002) Hydrolysis of erythrocyte proteins by proteases of malaria parasites, *Curr Opin Hematol 9*, 140-145.
6. Xiao, H., Bryksa, B. C., Bhaumik, P., Gustchina, A., Kiso, Y., Yao, S. Q., Wlodawer, A., and Yada, R. Y. (2014) The zymogen of plasmepsin V from *Plasmodium falciparum* is enzymatically active, *Molecular and biochemical parasitology 197*, 56-63.
7. Aurrecoechea, C., Brestelli, J., Brunk, B. P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Innamorato, F., Iodice, J., Kissinger, J. C., Kraemer, E., Li, W., Miller, J. A., Nayak, V., Pennington, C., Pinney, D. F., Roos, D. S., Ross,

C., Stoeckert, C. J., Jr., Treatman, C., and Wang, H. (2009) PlasmoDB: a functional genomic database for malaria parasites, *Nucleic Acids Res 37*, D539-543.

8. Goldberg, D. E. (1993) Hemoglobin degradation in *Plasmodium*-infected red blood cells, *Semin Cell Biol 4*, 355-361.

9. Francis, S. E., Sullivan, D. J., Jr., and Goldberg, D. E. (1997) Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum*, *Annu Rev Microbiol 51*, 97-123.

10. Sherman, I. W. (1977) Amino acid metabolism and protein synthesis in malarial parasites, *Bull World Health Organ 55*, 265-276.

11. Lew, V. L., Macdonald, L., Ginsburg, H., Krugliak, M., and Tiffert, T. (2004) Excess haemoglobin digestion by malaria parasites: a strategy to prevent premature host cell lysis, *Blood Cells Mol Dis 32*, 353-359.

12. Coombs, G. H., Goldberg, D. E., Klemba, M., Berry, C., Kay, J., and Mottram, J. C. (2001) Aspartic proteases of *Plasmodium falciparum* and other parasitic protozoa as drug targets, *Trends Parasitol 17*, 532-537.

13. Dame, J. B., Yowell, C. A., Omara-Opyene, L., Carlton, J. M., Cooper, R. A., and Li, T. (2003) Plasmepsin 4, the food vacuole aspartic proteinase found in all *Plasmodium* spp. infecting man, *Molecular and biochemical parasitology 130*, 1-12.

14. Banerjee, R., Liu, J., Beatty, W., Pelosof, L., Klemba, M., and Goldberg, D. E. (2002) Four plasmepsins are active in the *Plasmodium falciparum* food vacuole, including a protease with an active-site histidine, *Proceedings of the National Academy of Sciences of the United States of America 99*, 990-995.

15. Gupta, D., Yedidi, R. S., Varghese, S., Kovari, L. C., and Woster, P. M. (2010) Mechanism-based inhibitors of the aspartyl protease plasmepsin II as potential antimalarial agents, *J Med Chem 53*, 4234-4247.

16. Francis, S. E., Gluzman, I. Y., Oksman, A., Knickerbocker, A., Mueller, R., Bryant, M. L., Sherman, D. R., Russell, D. G., and Goldberg, D. E. (1994) Molecular characterization and inhibition of a *Plasmodium falciparum* aspartic hemoglobinase, *EMBO J 13*, 306-317.

17. Liu, J., Istvan, E. S., Gluzman, I. Y., Gross, J., and Goldberg, D. E. (2006) *Plasmodium falciparum* ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems, *Proceedings of the National Academy of Sciences of the United States of America 103*, 8840-8845.

18. Rosenthal, P. J. (1998) Proteases of malaria parasites: new targets for chemotherapy, *Emerging infectious diseases 4*, 49-57.

19. Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M. S., Nene, V., Shallom, S. J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, L. M., Subramanian, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M., and Barrell, B. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*, *Nature 419*, 498-511.

20. Westling, J., Cipullo, P., Hung, S. H., Saft, H., Dame, J. B., and Dunn, B. M. (1999) Active site specificity of plasmepsin II, *Protein Sci 8*, 2001-2009.

21. Omara-Opyene, A. L., Moura, P. A., Sulsona, C. R., Bonilla, J. A., Yowell, C. A., Fujioka, H., Fidock, D. A., and Dame, J. B. (2004) Genetic disruption of the *Plasmodium falciparum*

digestive vacuole plasmepsins demonstrates their functional redundancy, *The Journal of biological chemistry 279*, 54088-54096.

22. Gluzman, I. Y., Francis, S. E., Oksman, A., Smith, C. E., Duffin, K. L., and Goldberg, D. E. (1994) Order and specificity of the *Plasmodium falciparum* hemoglobin degradation pathway, *J Clin Invest 93*, 1602-1608.

23. Bjelic, S., and Aqvist, J. (2004) Computational prediction of structure, substrate binding mode, mechanism, and rate for a malaria protease with a novel type of active site, *Biochemistry 43*, 14521-14528.

24. Liu, J., Gluzman, I. Y., Drew, M. E., and Goldberg, D. E. (2005) The role of *Plasmodium falciparum* food vacuole plasmepsins, *The Journal of biological chemistry 280*, 1432-1437.

25. Bonilla, J. A., Bonilla, T. D., Yowell, C. A., Fujioka, H., and Dame, J. B. (2007) Critical roles for the digestive vacuole plasmepsins of *Plasmodium falciparum* in vacuolar function, *Molecular microbiology 65*, 64-75.

26. Sleebs, B. E., Lopaticki, S., Marapana, D. S., O'Neill, M. T., Rajasekaran, P., Gazdik, M., Gunther, S., Whitehead, L. W., Lowes, K. N., Barfod, L., Hviid, L., Shaw, P. J., Hodder, A. N., Smith, B. J., Cowman, A. F., and Boddey, J. A. (2014) Inhibition of Plasmepsin V activity demonstrates its essential role in protein export, PfEMP1 display, and survival of malaria parasites, *PLoS Biol 12*, e1001897.

27. Marti, M., Good, R. T., Rug, M., Knuepfer, E., and Cowman, A. F. (2004) Targeting malaria virulence and remodeling proteins to the host erythrocyte, *Science (New York, N.Y.) 306*, 1930-1933.

28. Hiller, N. L., Bhattacharjee, S., van Ooij, C., Liolios, K., Harrison, T., Lopez-Estrano, C., and Haldar, K. (2004) A host-targeting signal in virulence proteins reveals a secretome in malarial infection, *Science (New York, N.Y.) 306*, 1934-1937.

29. Klemba, M., and Goldberg, D. E. (2005) Characterization of plasmepsin V, a membrane-bound aspartic protease homolog in the endoplasmic reticulum of *Plasmodium falciparum*, *Molecular and biochemical parasitology 143*, 183-191.

30. Boddey, J. A., Hodder, A. N., Gunther, S., Gilson, P. R., Patsiouras, H., Kapp, E. A., Pearce, J. A., de Koning-Ward, T. F., Simpson, R. J., Crabb, B. S., and Cowman, A. F. (2010) An aspartyl protease directs malaria effector proteins to the host cell, *Nature 463*, 627-631.

31. Russo, I., Babbitt, S., Muralidharan, V., Butler, T., Oksman, A., and Goldberg, D. E. (2010) Plasmepsin V licenses *Plasmodium* proteins for export into the host erythrocyte, *Nature 463*, 632-636.

32. Chang, H. H., Falick, A. M., Carlton, P. M., Sedat, J. W., DeRisi, J. L., and Marletta, M. A. (2008) N-terminal processing of proteins exported by malaria parasites, *Molecular and biochemical parasitology 160*, 107-115.

33. Boddey, J. A., Carvalho, T. G., Hodder, A. N., Sargeant, T. J., Sleebs, B. E., Marapana, D., Lopaticki, S., Nebl, T., and Cowman, A. F. (2013) Role of plasmepsin V in export of diverse protein families from the *Plasmodium falciparum* exportome, *Traffic 14*, 532-550.

34. Sargeant, T. J., Marti, M., Caler, E., Carlton, J. M., Simpson, K., Speed, T. P., and Cowman, A. F. (2006) Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites, *Genome biology 7*, R12.

35. Baruch, D. I., Pasloske, B. L., Singh, H. B., Bi, X., Ma, X. C., Feldman, M., Taraschi, T. F., and Howard, R. J. (1995) Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes, *Cell 82*, 77-87.

36. Silva, A. M., Lee, A. Y., Erickson, J. W., and Goldberg, D. E. (1998) Structural analysis of plasmepsin II. A comparison with human aspartic proteases, *Adv Exp Med Biol 436*, 363-373.

37. McKay, P. B., Peters, M. B., Carta, G., Flood, C. T., Dempsey, E., Bell, A., Berry, C., Lloyd, D. G., and Fayne, D. (2011) Identification of plasmepsin inhibitors as selective anti-malarial agents using ligand based drug design, *Bioorg Med Chem Lett 21*, 3335-3341.

38. Bhaumik, P., Gustchina, A., and Wlodawer, A. (2012) Structural studies of vacuolar plasmepsins, *Biochim Biophys Acta 1824*, 207-223.

39. Khan, A. R., and James, M. N. (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes, *Protein Sci 7*, 815-836.

40. Bhaumik, P., Xiao, H., Parr, C. L., Kiso, Y., Gustchina, A., Yada, R. Y., and Wlodawer, A. (2009) Crystal structures of the histo-aspartic protease (HAP) from *Plasmodium falciparum*, *Journal of molecular biology 388*, 520-540.

41. Berry, C., Humphreys, M. J., Matharu, P., Granger, R., Horrocks, P., Moon, R. P., Certa, U., Ridley, R. G., Bur, D., and Kay, J. (1999) A distinct member of the aspartic proteinase gene family from the human malaria parasite *Plasmodium falciparum*, *FEBS Lett 447*, 149-154.

42. Asojo, O. A., Gulnik, S. V., Afonina, E., Yu, B., Ellman, J. A., Haque, T. S., and Silva, A. M. (2003) Novel uncomplexed and complexed structures of plasmepsin II, an aspartic protease from *Plasmodium falciparum*, *Journal of molecular biology 327*, 173-181.

43. Silva, A. M., Lee, A. Y., Gulnik, S. V., Maier, P., Collins, J., Bhat, T. N., Collins, P. J., Cachau, R. E., Luker, K. E., Gluzman, I. Y., Francis, S. E., Oksman, A., Goldberg, D. E., and Erickson, J. W. (1996) Structure and inhibition of plasmepsin II, a hemoglobin-degrading enzyme from *Plasmodium falciparum*, *Proc Natl Acad Sci U S A 93*, 10034-10039.

44. Hornak, V., Okur, A., Rizzo, R. C., and Simmerling, C. (2006) HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations, *Proc Natl Acad Sci U S A 103*, 915-920.

45. Zhu, Z. W., Schuster, D. I., and Tuckerman, M. E. (2003) Molecular dynamics study of the connection between flap closing and binding of fullerene-based inhibitors of the HIV-1 protease, *Biochemistry 42*, 1326-1333.

46. Heaslet, H., Rosenfeld, R., Giffin, M., Lin, Y.-C., Tam, K., Torbett, B. E., Elder, J. H., McRee, D. E., and Stout, C. D. (2007) Conformational flexibility in the flap domains of ligand-free HIV protease, *Acta Crystallographica Section D-Biological Crystallography 63*, 866-875.

47. Kovalskyy, D., Dubyna, V., Mark, A. E., and Kornelyuk, A. (2005) A molecular dynamics study of the structural stability of HIV-1 protease under physiological conditions: The role of Na+ ions in stabilizing the active site, *Proteins-Structure Function and Bioinformatics 58*, 450-458.

48. Asojo, O. A., Gulnik, S. V., Afonina, E., Yu, B., Ellman, J. A., Haque, T. S., and Silva, A. M. (2003) Novel uncomplexed and complexed structures of plasmepsin II, an aspartic protease from *Plasmodium falciparum*, *Journal of Molecular Biology 327*, 173-181.

49. Bhaumik, P., Horimoto, Y., Xiao, H., Miura, T., Hidaka, K., Kiso, Y., Wlodawer, A., Yada, R. Y., and Gustchina, A. (2011) Crystal structures of the free and inhibited forms of plasmepsin I (PMI) from *Plasmodium falciparum*, *Journal of Structural Biology 175*, 73-84.

50. Bhaumik, P., Xiao, H., Parr, C. L., Kiso, Y., Gustchina, A., Yada, R. Y., and Wlodawer, A. (2009) Crystal Structures of the Histo-Aspartic Protease (HAP) from *Plasmodium falciparum*, *Journal of Molecular Biology 388*, 520-540.

51. Karubiu, W., Bhakat, S., McGillewie, L., and Soliman, M. E. (2015) Flap dynamics of plasmepsin proteases: insight into proposed parameters and molecular dynamics, *Molecular bioSystems 11*, 1061-1066.

52. Prade, L., Jones, A. F., Boss, C., Richard-Bildstein, S., Meyer, S., Binkert, C., and Bur, D. (2005) X-ray structure of plasmepsin II complexed with a potent achiral inhibitor, *J Biol Chem 280*, 23837-23843.

53. http://www.rcsb.org/pdb/home/home.do.

54. Bhakat, S., Martin, A. J., and Soliman, M. E. (2014) An integrated molecular dynamics, principal component analysis and residue interaction network approach reveals the impact of M184V mutation on HIV reverse transcriptase resistance to lamivudine, *Molecular bioSystems 10*, 2215-2228.

55. Karubiu, W., Bhakat, S., and Soliman, M. E. (2014) Compensatory role of double mutation N348I/M184V on nevirapine binding landscape: insight from molecular dynamics simulation, *Protein J 33*, 432-446.

56. Sleebs, B. E., Gazdik, M., O'Neill, M. T., Rajasekaran, P., Lopaticki, S., Lackovic, K., Lowes, K., Smith, B. J., Cowman, A. F., and Boddey, J. A. (2014) Transition State Mimetics of the Plasmodium Export Element Are Potent Inhibitors of Plasmepsin V from P. falciparum and P. vivax, *J Med Chem*.

57. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *Journal of computational chemistry 25*, 1605-1612.

58. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs, *Journal of Computational Chemistry 26*, 1668-1688.

59. Jorgensen, W. L. C., J; Madura, J. D; Impey, R. W and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *The Journal of Chemical Physics 79*, 926-935.

60. Bhakat, S., Martin, A. J. M., and Soliman, M. E. S. (2014) An integrated molecular dynamics, principal component analysis and residue interaction network approach reveals the impact of M184V mutation on HIV reverse transcriptase resistance to lamivudine, *Molecular Biosystems 10*, 2215-2228.

61. Karubiu, W., Bhakat, S., and Soliman, M. S. (2014) Compensatory Role of Double Mutation N348I/M184V on Nevirapine Binding Landscape: Insight from Molecular Dynamics Simulation, *Protein J*, 1-15.

62. Ahmed, S. M., Kruger, H. G., Govender, T., Maguire, G. E. M., Sayed, Y., Ibrahim, M. A. A., Naicker, P., and Soliman, M. E. S. (2013) Comparison of the Molecular Dynamics and Calculated Binding Free Energies for Nine FDA-Approved HIV-1 PR Drugs Against Subtype B and C-SA HIV PR, *Chemical Biology & Drug Design 81*, 208-218.

63. Naicker, P., Achilonu, I., Fanucchi, S., Fernandes, M., Ibrahim, M. A. A., Dirr, H. W., Soliman, M. E. S., and Sayed, Y. (2013) Structural insights into the South African HIV-1 subtype C protease: impact of hinge region dynamics and flap flexibility in drug resistance, *J Biomol Struct Dyn 31*, 1370-1380.

64. http://www.originlab.com/

65. Nezami, A., and Freire, E. (2002) The integration of genomic and structural information in the development of high affinity plasmepsin inhibitors, *International journal for parasitology 32*, 1669-1676.

66. Bhaumik, P., Horimoto, Y., Xiao, H., Miura, T., Hidaka, K., Kiso, Y., Wlodawer, A., Yada, R. Y., and Gustchina, A. (2011) Crystal structures of the free and inhibited forms of plasmepsin I (PMI) from *Plasmodium falciparum*, *J Struct Biol 175*, 73-84.

67. Gil, L. A., Valiente, P. A., Pascutti, P. G., and Pons, T. (2011) Computational perspectives into plasmepsins structure-function relationship: implications to inhibitors design, *Journal of tropical medicine 2011*, 657483.

68. Bhargavi, R., Sastry, G. M., Murty, U. S., and Sastry, G. N. (2005) Structural and active site analysis of plasmepsins of *Plasmodium falciparum*: potential anti-malarial targets, *International journal of biological macromolecules 37*, 73-84.

69. Siripurkpong, P., Yuvaniyama, J., Wilairat, P., and Goldberg, D. E. (2002) Active site contribution to specificity of the aspartic proteases plasmepsins I and II, *The Journal of biological chemistry 277*, 41009-41013.

70. Valiente, P. A., Batista, P. R., Pupo, A., Pons, T., Valencia, A., and Pascutti, P. G. (2008) Predicting functional residues in *Plasmodium falciparum* plasmepsins by combining sequence and structural analysis with molecular dynamics simulations, *Proteins 73*, 440-457.

71. Piana, S., Carloni, P., and Parrinello, M. (2002) Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease, *Journal of molecular biology 319*, 567-583.

72. Okoniewska, M., Tanaka, T., and Yada, R. Y. (2000) The pepsin residue glycine-76 contributes to active-site loop flexibility and participates in catalysis, *The Biochemical journal 349*, 169-177.

73. Scott, W. R., and Schiffer, C. A. (2000) Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance, *Structure (London, England : 1993) 8*, 1259-1265.

Supplementary Materials

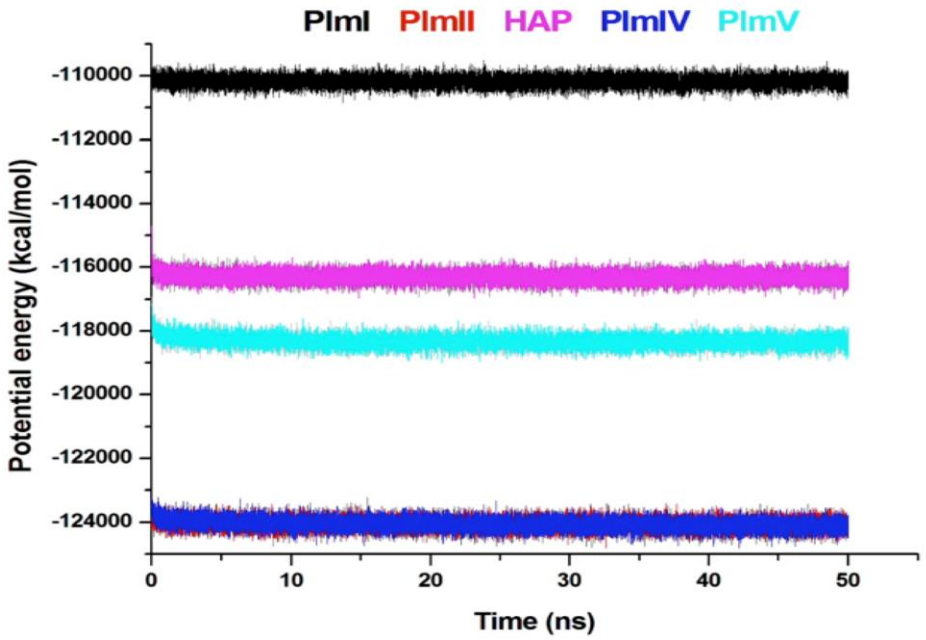


Figure S1. Potential energy fluctuations throughout the 50 ns continuous simulation for apo Plm I (black), II (red), HAP (magenta), IV (blue) and V (cyan).

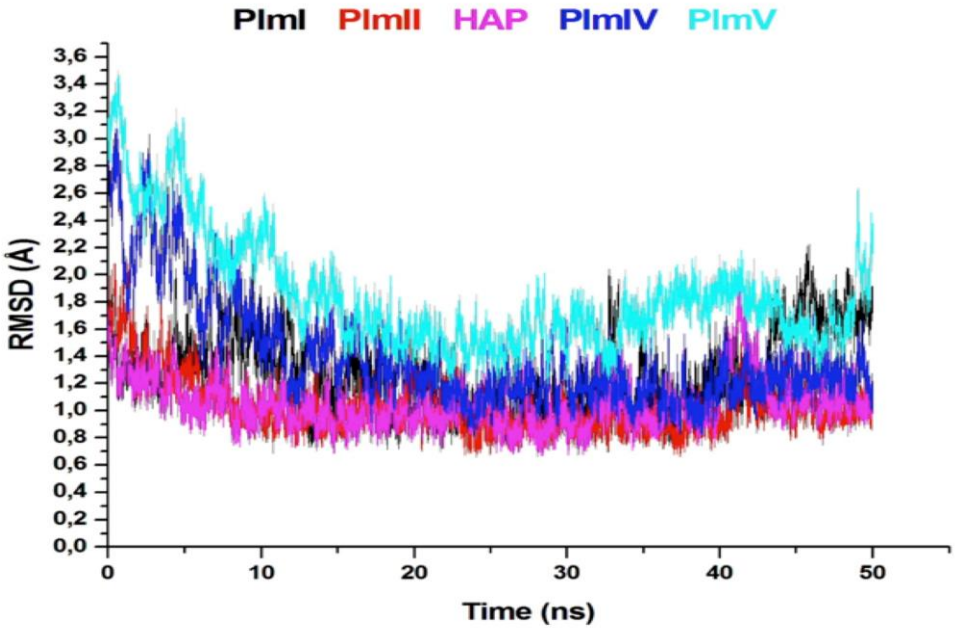


Figure S2. RMSD fluctuations throughout the 50 ns continuous simulation for apo Plm I (black), II (red), HAP (magenta), IV (blue) and V (cyan).

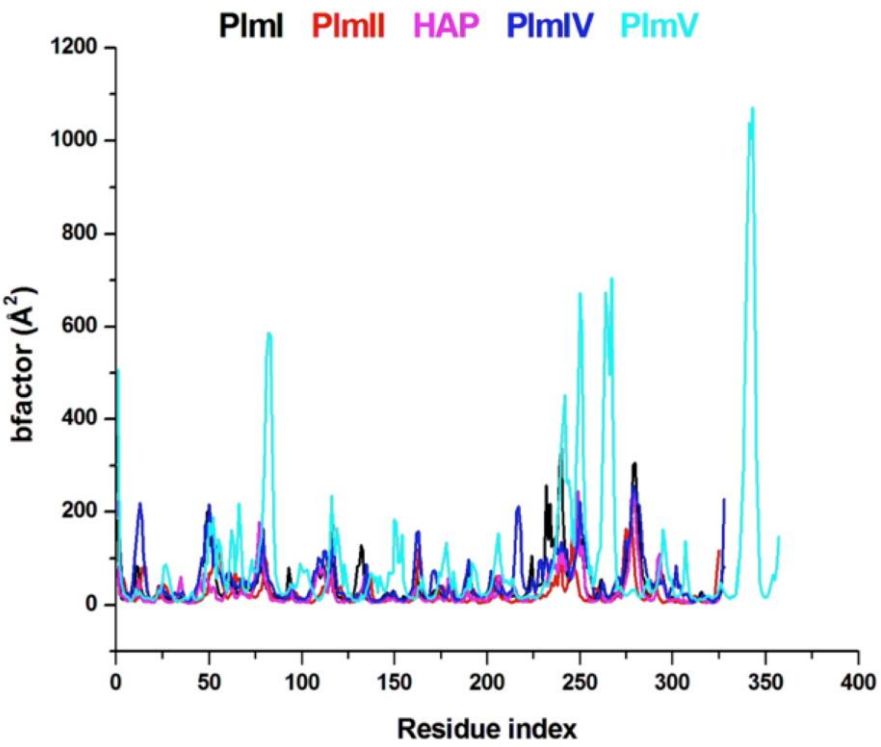


Figure S3. Per-residue fluctuations of the b-factors throughout simulation for apo Plm I (black), II (red), HAP (magenta), IV (blue) and V (cyan).

CHAPTER 6

The binding landscape of plasmepsin V and the implications on flap dynamics

McGillewie L and Mahmoud E. Soliman*

Molecular Modelling & Drug Design Research Group, School of Health Sciences, University of KwaZulu-Natal, Westville, Durban 4001, South Africa

* Corresponding author: Mahmoud E.S. Soliman, email: soliman@ukzn.ac.za

Telephone: +27 031 260 7413, Fax: +27 031 260 7792

Abstract

Plasmepsin V belongs to the plasmepsin family of aspartic proteases. PlmV is unique compared to other plasmepsins, as this membrane bound aspartic protease resides in the endoplasmic reticulum and is responsible for cleavage of PEXEL tagged proteins destined for export outside of the host red blood cell. Plasmepsin V is highly conserved throughout the *Plasmodium* species, and is essential to the survival of the parasite. Recently, two potent inhibitors of Plmv have been identified, WEHI-916 and WEHI-842. Of these inhibitors, WEHI-842 has a higher binding affinity for *P.vivax* PlmV and a crystal structure of PlmV in complex with WEHI-842 has recently been resolved (4ZL4). The structure of PlmV is unique compared to other plasmepsins, it is stablised internally by seven disulphide bonds, a NAP1 insert/fold is associated with the movement of the flap covering the active site and a highly conserved helix-turn-helix is situated towards the C-terminal. Flap motion and dynamics play an important role in enzyme selectivity and function. To better understand the impact of ligand binding on the flap dynamics; molecular dynamic simulations and post dynamic analysis was employed in the present study on PlmV in complex with WEHI-842. Previously defined parameters, which accurately account for the opening and closing of the active site were used to assess the conformational changes induced in the absence and presence of WEHI-842. From the simulations it can be seen that inhibitor binding significantly reduces the flexibility and mobility of not only the flap and flexible loop but areas outside of the active site. Ligand binding leads to the formation of a more stable compact structure. This being said, there is the

possibility of reducing the flexibility even further with potentially more lethal effects on the plasmodium parasite. We believe that results presented herein, would assist researches in the discovery of potent PlmV inhibitors as potential antimalarial therapies.

Keywords: aspartic proteases; flap dynamics; plasmepsin; flexibility; plasmodium; malaria; molecular dynamics

1. Introduction

Malaria is a highly infectious parasitic disease caused by the *Plasmodium* protozoa, of which the most virulent species to infect man is *P. falciparum* and *P. vivax* [1,2]. In 2014, 97 countries had ongoing malarial transmissions with an estimated 3.3 billion people at risk globally and in 2013 198 million cases were reported worldwide, with an estimated 584 000 deaths [3]. To date, no malaria vaccine has been developed and drug resistance is widespread, with no new antimalarial classes entering clinical practice since 1996 [1].

The malaria parasite invades host erythrocytes, where it resides and replicates within a parasitophorous vacuole. To ensure survival in the host cell, the parasite degrades hemoglobin via a plethora of protein degrading enzymes with functional redundancy and exports numerous parasitic proteins into the red blood cell (RBC) cytoplasm and onto the surface of infected RBCs, remodeling the host cell [4]. Export relies mostly on a vacuolar transport signal (VTS) or the *Plasmodium* export element (PEXEL), an N-terminal motif found in over 450 *P. falciparum* proteins [5]. In *P. falciparum*, this is approximately 9% [6,7] of all proteins (exportome), of which 20% or more are predicted to be essential for survival [8]. The conserved pentameric PEXEL motif contains a consensus sequence, RxLxE/Q/D, and PEXEL-proteins destined for export are cleaved in the parasitic endoplasmic reticulum (ER) at the C-terminal of the leucine residue (RxL) [9,10]. Cleavage requires the highly conserved arginine x leucine, and is performed by the ER aspartic protease, plasmepsin V (PlmV) (**Figure 1**) [6,11,12]. Plasmepsin V and PEXEL tagged proteins are conserved throughout all *Plasmodium* species [7,13,14], deletion of the *PMV* gene has suggested it is essential to the survival of the parasite [11,12,14]. The ER-resident PlmV is situated on chromosome 13 consisting of 590 amino acids (68.4kDa) [14], is only distantly related to other plasmepsins (17% to PlmII); and is phylogenetically unique as it shares distant homology with human aspartic proteases [11,12,14]. Unlike most aspartic proteases, PlmV does not bind to the

aspartic protease inhibitor pepstatin A [14]. These findings indicate that PlmV could be a very promising target for antimalarial treatment.

Plasmepsin V belongs to the family of aspartic proteases, a class of proteolytic enzymes that typically house two catalytic aspartic acids in the active site partially covered by a β-hairpin loop (flap) (**Figure 1**). Traditionally, aspartic proteases are inhibited by transition state peptidomimetics which mimic and outcompete the biological substrate. In these peptidomimetic inhibitors, a non-cleavable moiety mimics the amide bond preventing proteolysis (e.g. statine) [5]. Due to the hydroxyl groups in the transition state analogs, hydrogen bonds are formed between the inhibitor and the catalytic aspartic acids holding the inhibitor in the active site [19]. Peptides transformed into therapeutic drugs/agents tend to be problematic due to their metabolic instability and poor bioavailability [16]. It is well known that plasmepsin substrate specificity is dictated by conformational deviations in residues surrounding the actives site rather than the active site itself [11,12]. Flap dynamics and motions play a big role in the overall conformational flexibility of a protein and is crucial in understanding the ligand binding landscape. Previously we have shown the flexibility of free unbound plasmepsins, and how the flap region covering the active site is the most mobile and flexible region in the active site [19]. In an unbound conformation, the flap region moves freely opening and closing the active site. Using previously defined parameters to account for the dynamic behaviour of the active site, it was observed that Plm IV and V were the most flexible [19].

Recently, a potent inhibitor (PEXEL-mimetic) of PlmV has been identified, WEHI-916, with an IC_{50} (nM) of 0.020 and 0.024 for *P. falciparum* and *P. vivax* respectively with good selectivity over BACE1 and CatD [20]. Overexpression of PlmV led to an increase in resistance to WEHI-916, whereas knockdown of PlmV increased susceptibility and sensitivity to WEHI-

916 [20]. Due to the polarised guanidinium group present on the P_3 Arg, WEHI-916 was poorly membrane permeable thus had a high affinity for endogenous PlmV and only modestly affected *P. falciparum* growth [21]. A new compound, WEHI-842, was synthesised in which the P_3 arganine was replaced by a non-proteinogenic amino acid (cavanine) and the sulfonamide of WEHI-916 was replaced by an N-terminal carbamate [21]. WEHI-842 has been shown to be more potent than WEHI-916 against recombinant PlmV, with a binding affinity of 13.8 nM compared to 42.0 nM for WEHI-916. A structure of *P. vivax* PlmV in complex with WEHI-842 was subsequently crystallised (PDB code 4ZL4, resolution 2.37 Å) (**Figure 1**) [21].

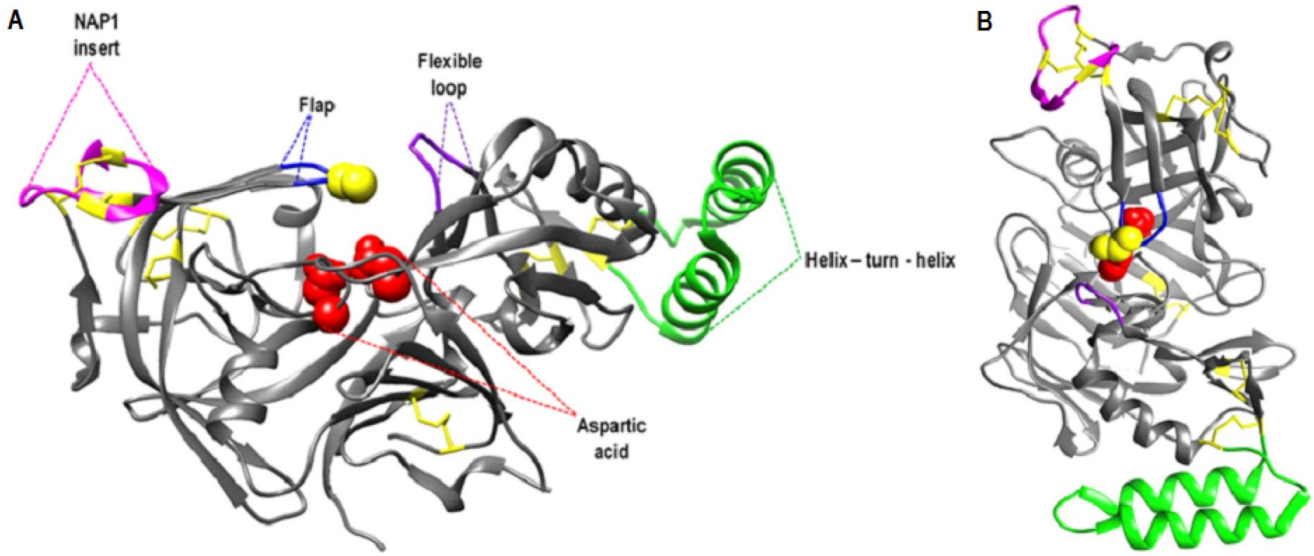


Figure 1. Crystal structure of plasmepsin V (PDB 4ZL4) 22 showing the key regions and residues. Cysteine residues are highlighted in yellow, and disulfide bonds shown in yellow linkages, nepenthesin 1-type aspartyl protease (NAP1) fold (magenta), catalytic aspartic acid residues are shown in red, the flap region (blue), unpaired cysteine flap tip (yellow sphere), flexible loop (purple) and the helix-turn-helix motif (green). A. Side on view, B. Top view.

Herein, we use previously defined parameters that we believe accurately depicts the opening and closing of the PlmV active site to ascertain how ligand binding affects protein dynamics and flexibility. The data presented herein will aid in the development of more potent inhibitors of PlmV.

2. Computational methodology

2.1. System preparation

Plasmepsin V in complex with WEHI-842 (PDB code 4ZL4 [23]) was prepared using the GUI interface of the UCSF Chimera software package (www.cgl.ucsf.edu) [24]. The structure was crystalised as a dimer (chain A and B), although only chain A was used in the simulations. It has been reported that PlmV is composed of 590 amino acids, including the peptide signal region, the pro-segment, mature PlmV and the transmembrane region [25]. The crystal structure started at residue number 33, and the structure of residues in the loop region between Arg241 and Glu272 could not be determined due to poor electron density [22]. The authors noted that these residues are not of importance in the binding of WEHI-842 or effector proteins [22].

2.2. Molecular dynamics simulations

Simulations were carried out using the GPU version of PMEMD engine incorporated with the Sander module of Amber14 [26]. Protein systems were modelled using the ff99sb force field in Amber14 [27], the LEAP module of Amber14 was used to add missing hydrogen and heavy atoms for system stabilization [26]. The systems were neutralized by the addition of Na^+ counter ions. Ligands were parameterised using gasteiger charges in Avogadro [28], the Antechamber module by applying the GAFF (generalized Amber force field) [29]. All systems were immersed within a orthorhombic box of TIP3P [30] water molecules such that protein atoms were within 10 Å of any box edge throughout the simulations. Periodic boundary conditions were used on all systems, long range electrostatic interactions were treated with the particle mesh Ewald (PME) method [31] in Amber14 with direct space and van der Waals interactions restricted to 12 Å. Systems were subjected to two minimisation steps, partial minimisation followed by full minimisation. Initial minimisation (1000 steps) was performed on all systems with restrained harmonic constraints (constant force of 500 kcal/mol^{-1} Å2) using the steepest descent

algorithm. Thereafter, all atom energy minimisations without any restraints were conducted for 1000 steps using the conjugate gradient method. Minimised systems were gradually heated from 0 to 300 K in the NVT ensemble using harmonic constraints of 5 kcal/mol $Å^2$ (all solvent molecules) and a Langevin thermostat (collision frequency of 1 ps-1) [32] regulated and maintained temperatures throughout the simulations. Systems were equilibrated at 300 K in the non-restrained NPT ensemble for 500 ps prior to production runs, restraints were removed and constant pressure (1 bar) was maintained using a Barendson barostat [33]. The SHAKE [34] algorithm was used throughout runs to constrain all bonds involving hydrogen atoms. All simulations were run using a 2fs time step and the SPFP precision model [35].

Continuous 50 ns MD simulations were run using the NPT ensembles (isothermal and isobaric) at a constant target pressure of 1 bar and a pressure coupling constant of 2 ps. Resulting coordinates were saved every 1 ps. Trajectories were analysed using the CPPTRAJ modules in Amber14. Trajectories were visualised using the GUIs Chimera [24]. Results were analysed and plotted using Origin [36]. In previous reports we have shown that a continuous approach for 50 ns is sufficient to observe conformational changes and flap motions of aspartic proteases and ensures sample convergence and system stability [19,37].

2.3. Post-dynamic analysis

2.3.1. Principle component analysis (PCA)

Principle component analysis (PCA) also known as essential dynamics (ED) [38], is a technique that aids in the understanding of the dynamic behavior of biological systems. Essentially, PCA defines atomic displacement and identifies conformational changes by extracting various conformational modes of the protein complex during the MD simulation. Principle component analysis defines the eigenvectors (direction of motion) and eigenvalues (magnitude of motion)

for the protein systems [39]. The analysis was carried out by constructing a covariance matrix of the Cα atom displacements. After stripping counter ions (Na^+) and solvent molecules from the 50 ns trajectories, PCA was performed on all Cα backbone atoms. Data was averaged over 1000 snapshots taken at 100 ps time intervals using the CPPTRAJ module of AMBER14 to generate the first two principal components, PCA1 and PC2 (corresponds to the first two eigenvectors of the matrix). Corresponding scatter plots were generated and analysed using Origin software (http://www.originlab.com/).

2.3.2. Dynamic Cross-Correlation Matrices (DCCM)

The cross-correlation of dynamic trajectories is useful in understanding correlated motions of residual-based fluctuations throughout a simulation. The matrix is a 3D representation which graphically depicts time-correlated information among the residues of the protein systems [40] and residue-base time correlated data can be analysed using visual pattern recognition [41]. To better understand the dynamics of apo versus bound PlmV, a DCCM was generated to depict cross-correlated displacements of backbone Cα atoms in the trajectories. Dynamic cross-correlation matrices were generated using the equation:

$$C_{ij} = \langle \Delta r_i * \Delta r_j \rangle / \left(\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle \right)^{1/2} \quad (1)$$

Where, *i* and *j* corresponds to the *i*th and *j*th residue and $\Delta \boldsymbol{r}_i$ and $\Delta \boldsymbol{r}_j$ represents the displacement of the *i*th and *j*th atom from the mean respectively. The cross-correlation coefficient $C_{ij,}$ varies between the range of -1 to +1, where the upper and lower limits correspond to a fully correlated (+1) and anti-correlated (-1) motion throughout the simulation. The analysis was executed using the CPPTRJ module of AMBER 14. Matrices were generated and analysed using Origin software (http://www.originlab.com/).

3. Results and discussion

3.1. Molecular dynamics simulations

3.1.1. System stability

Before the MD trajectories were analysed, root mean square deviations (RMSD) and potential energy fluctuations were monitored throughout the 50 ns simulation. This was to ensure stability and reliability of the post-dynamic analyses to follow. Stabilisation and convergence was observed for both systems, and no major energy fluctuations were observed (Figure S1). After approximately 10 ns, the RMSD trajectories converged and fluctuations did not exceed 2 Å for either system for the duration of the simulation (Figure S2). On average apo and bound PlmV show a similar trend in RMSD fluctuations, with an average RMSD of 1.2 Å and 1 Å respectively, although towards the end of the simulation apo PlmV fluctuates slightly higher than the bound conformation. In the apo conformation, PlmV reaches a maximum fluctuation of 2.7 Å compared to bound PlmV which only reaches a maximum of 1.8 Å. These results indicate that albeit slightly, apo PlmV displays more flexibility and deviates more compared to bound PlmV.

3.1.2. Visual tajectory analysis

Snapshots of the trajectories were taken at 10 ns intervals for both apo and bound PlmV, and analysed for the 50 ns duration (**Figure 2 and 3**). From the visual inspection of the snapshots it can be seen that in an unbound apo state, PlmV is much more flexible as the flap tip moves away from the flexible region opening up the active site (reflected by an increase in d_1). Throughout the simulation it can be seen that apo PlmV deviates significantly more than bound PlmV from their initial starting structures (cornflower blue). It can be seen that the tip (Cys140) and the opposite flexible loop of apo PlmV moves a significant distance from the initial starting structure (**Figure 2**). The core of both apo and bound PlmV appears rigid, whereas significant

changes can be seen in the extremities more specifically in the nepenthesin 1-type aspartyl protease (NAP1) fold, regions adjacent to the NAP1 insert (N-terminal side) and the helix-turn-helix towards the C-terminal. These regions appear to be flexible and mobile in both apo and bound PlmV, although ligand binding reduces mobility, decreases flexibility and increases the compactness of PlmV. In the presence of the inhibitor WEHI-842, the flexibility and dynamic motion of the flap tip and flexible loop is drastically reduced. The inhibitor stabilizes and reduces the flexibility of the active site; as the flap tip and flexible region move closer together (decrease in d_1), folding inwards towards the two catalytic aspartic residues holding the ligand firmly in the active site (**Figure 2 and 3**). From the snapshots it can also be seen that the catalytic aspartic acids remain relatively constant throughout the simulation, and that ligand binding increases rigidity by binding to other key residues lining the binding cavity (**Figure 2 and 3**). Interestingly, from the snapshots it seems as if ligand binding stabalises the flap region by reducing mobility and flexibility in the NAP1 insert region which moves significantly in the absence of WEHI-842. The helix-turn-helix flexibility is also hindered and reduced upon binding of WEHI-842, suggesting that WEHI-842 inhibits PlmV by inducing conformational changes outside of the active site. It is worth noting, that in the apo conformation the orientation of the flap tip (Cys140) is erratic, disordered and highly flexible; whereas in the bound conformation the orientation Cys140's side chain is more stable and tends to orientate inwards towards the active site (**Figure 3**).

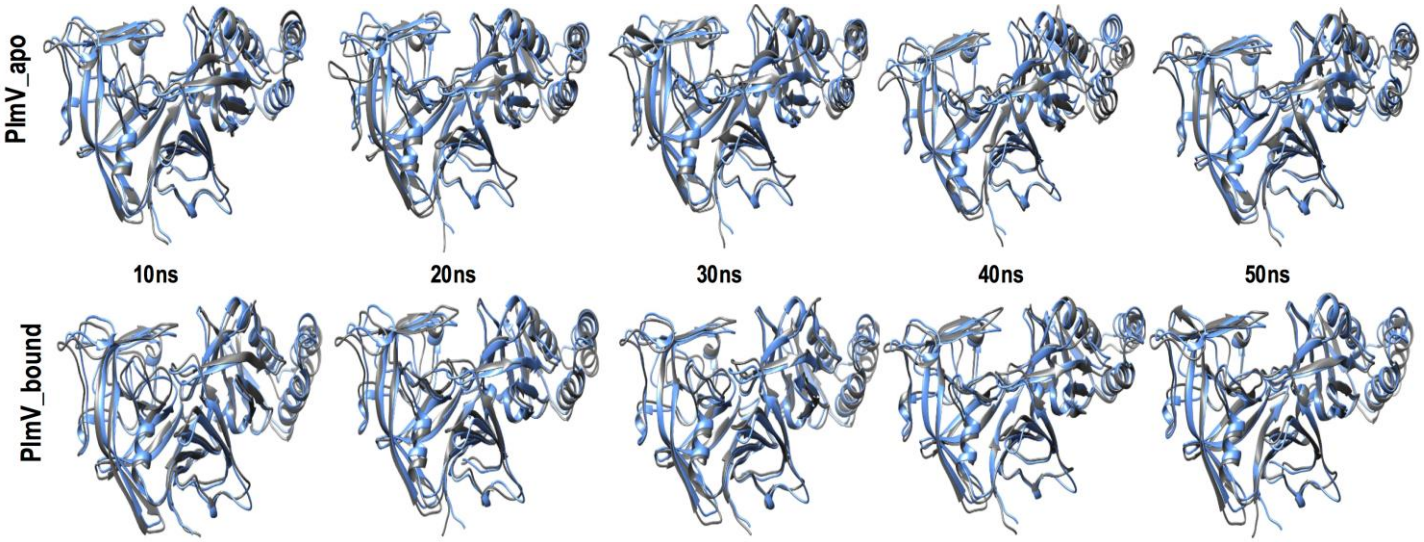


Figure 2. Visual analysis of the MD trajectories throughout the 50 ns simulation.

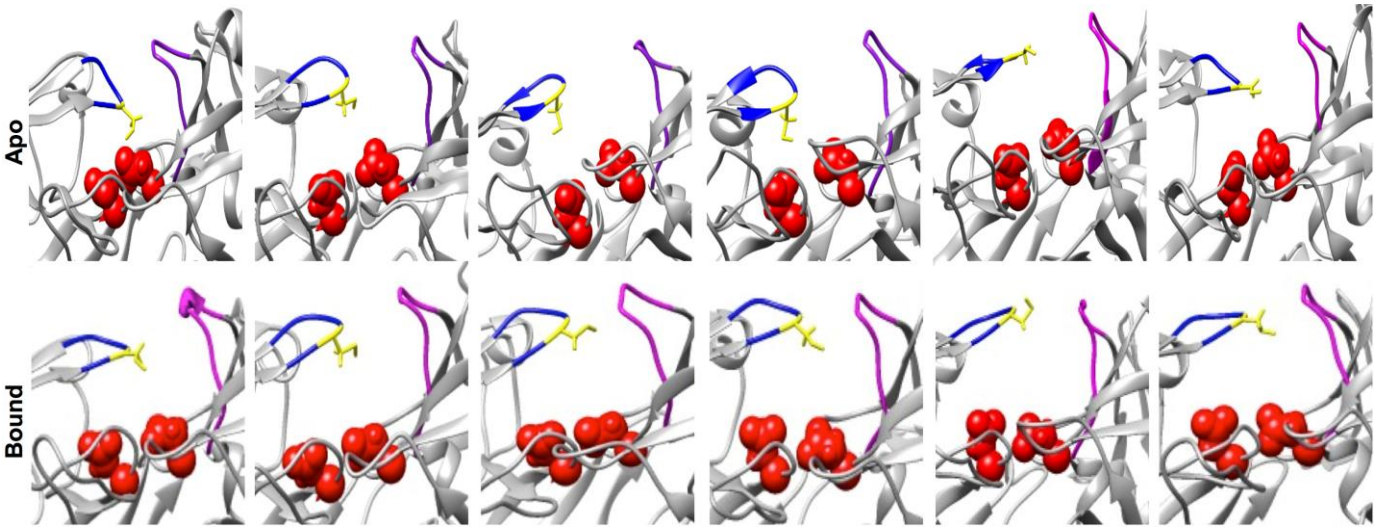


Figure 3. Zoomed analysis of the MD trajectories throughout the 50 ns simulation, for both apo and WEHI-842 bound PlmV. Flap (blue), flap tip (yellow), catalytic aspartic acids (red) and flexible loop (magenta).

3.2. Post MD analysis

3.2.1. Root mean square fluctuation (RMSF)

Protein function can be altered directly through interactions with the active site, or indirectly by disrupting motions essential to PlmV function. More specifically, the essential motions induced by the conformational changes that accompany ligand binding and dissociation [42,43]. To better understand the structural fluctuations that occur upon ligand binding, the RMSF of

both apo and bound PlmV was calculated i.e. standard deviation of Cα atom from average structure (**Figure 4**).

The protein core appears to be more rigid compared to the solvent exposed loops which are more flexible as indicated by the fluctuations in the RMSF values (**Figure 4**). Overall, the average RMSF does not differ much with a slightly lower RMSF for bound PlmV, 0.9 Å, compared to 1.1 Å for the apo conformation. The catalytic aspartic acid residues (80,313) showed no significant fluctuations upon ligand binding, thus binding of WEHI-842 inhibits PlmV activity due to conformational changes induced in residues lining the active site. The most significant changes in fluctuation can be seen in residues in the flap, notably Cys140 and Glu141. This is due to the formation of a salt bridge between the guanidinium ion of WEHI-842 and the carboxylic acid of Glu141, inhibitor binding reduces the flexibility of Glu141 from 2.4 Å to 1.6 Å. The formation of the salt bridge, further reduces the flexibility of the flap by indirectly lowering the fluctuation of the flap tip, Cys140, from 2 Å to 1.6 Å. Both the NAP1 fold and helix-turn-helix are less flexible in the bound conformation. In the apo conformation, Asn95 has the highest fluctuation, 4.3 Å, upon ligand binding it becomes more rigid and fluctuates significantly less (2.7 Å). Asn95 is situated between Cys93 and Cys96 which form the second disulfide bond (C2) in the PlmV structure. The highest fluctuations were observed between residue 93-97, with an average RMSF of 3.5 Å in the apo conformation, and 2.5 Å in the bound conformation. Ligand binding lowers the fluctuation in this region, although even after ligand binding this region still fluctuates more than the average for the total bound conformation. The trend and conformational fluctuations observed in the b-factor values (Figure S3), support the fluctuations observed in the RMSF values.

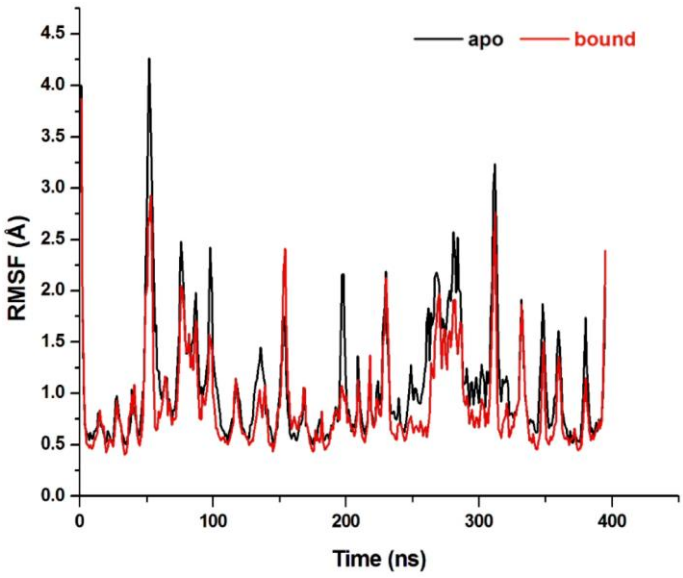


Figure 4. RMSF of apo PlmV (black) and WEHI-842 bound PlmV (red).

3.2.2. Flexibility of the active site

Previously defined and reported parameters [19,37] were used to investigate the flap opening and closing of PlmV in the absence and presence of WEHI-842. These parameters accurately account for the twisting motion of the flap tip and the recoiling motion of the flexible loop. The parameters measured in the present study were the distance (d_1) between the flap tip (Cys140) and the hinge residue of the flexible loop (Gln433), in relation to TriCα angles θ_1 (Cys140-Asp80-Gln433) and θ_2 (Cys140-Asp313-Gln433), and to the dihedral angle ϕ (Cys140-Asp80-Asp313-Gln433) (Figure S4).

As previously reported using a homology model for PlmV [20], the results presented in the present study show that in the apo conformation PlmV transitions between semi-open and more open conformations. Throughout the simulation, the flap and flexible regions of PlmV move significantly from each other; opening the active site exposing the catalytic aspartic acids. In comparison to apo PlmV, the flexibility and mobility of the flap region and the flexible loop of WEHI-842 bound PlmV is significantly reduced. Overall, the average distance of the flap tip

and flexible loop for apo PlmV was 13.4 Å which was reduced to 10.1 Å upon binding of WEHI-842. The major difference between the two conformations was in the range of motion, in a bound state PlmV showed more restricted movements between the flap tip and the flexible region (7.7 Å – 13.7 Å); whereas in the unliganded conformation PlmV reaches a maximum distance of 17.1 Å (**Figure 5**). It is worthwhile noting, that the distance between the flap and the flexible region in the bound conformation is not constant throughout the simulation; this in part could be attributed to the fact that WEHI-842 forms a salt bridge with Glu141 adjacent to the Cys140 flap tip, indirectly reducing the range of motion and flexibility of Cys140. In both the apo and bound conformations, the Tri Cα angles, θ_1 and θ_2, follow a similar trend to that observed in d_1 (**Figure 5**). As the flap and flexible loop move away from each other ($d_1 \uparrow$), both the Tri Cα angles, θ_1 and θ_2, increase in the apo conformation. The wide range of d_1 in the apo conformation corresponds to the wide angular motion of the active site; as θ_1 transitions between 17.4° and 56.6°, whereas θ_2 reaches a staggering 84.8° from a base of 33.3°. Enabling the tip and flexible loop to move apart, reaching a maximal d_1 of 17.1 Å exposing the active site. Ligand binding to PlmV induces rigidity and compactness of the active site as represented by the decreased range of motion of the Tri Cα angles, θ_1 and θ_2, 24.9° - 48.4° and 33.9° - 59.6° respectively and corresponding decrease in d_1. As previously reported for plasmepsins, θ_2 more accurately accounts for the magnitude of flap opening of PlmV. Ligand binding significantly reduces the angle between the flap tip, the second aspartic acid and the flexible loop; in an apo conformation the average value for θ_2 is 65.7° which is reduced to 44.8° upon binding of WEHI-842 (**Figure 5**). Ligand binding also stabalises and reduces the twisting motion of the active site, as reflected by the conserved range of ϕ values (**Figure 5**). In the bound conformation, there is less twisting of the active site as bonds formed between WEHI-842 and residues in the active site prevent large torsional shifts. The apo conformation of PlmV twists more aggressively to open the active site, which can be seen in the large shift in the dihedral

angles in the beginning of the simulation as PlmV moves from a semi-open configuration to a more open configuration. As to be expected, the further the active site opens the more active site twists to accommodate the opening of the active site.

The radius of gyration (R_g) measures the thermodynamics and kinetics of protein folding, and is an indication of the compactness and stability of protein complexes [44]. From **Figure 6**, it can be seen that the R_g for both the apo and bound conformations is fairly similar to one another, with an average R_g of 23.7 Å and 23.4 Å respectively. Ligand binding induces a shift to a more compact structure, albeit only a 0.3 Å change, a relatively more stable structure is formed (**Figure 6**).

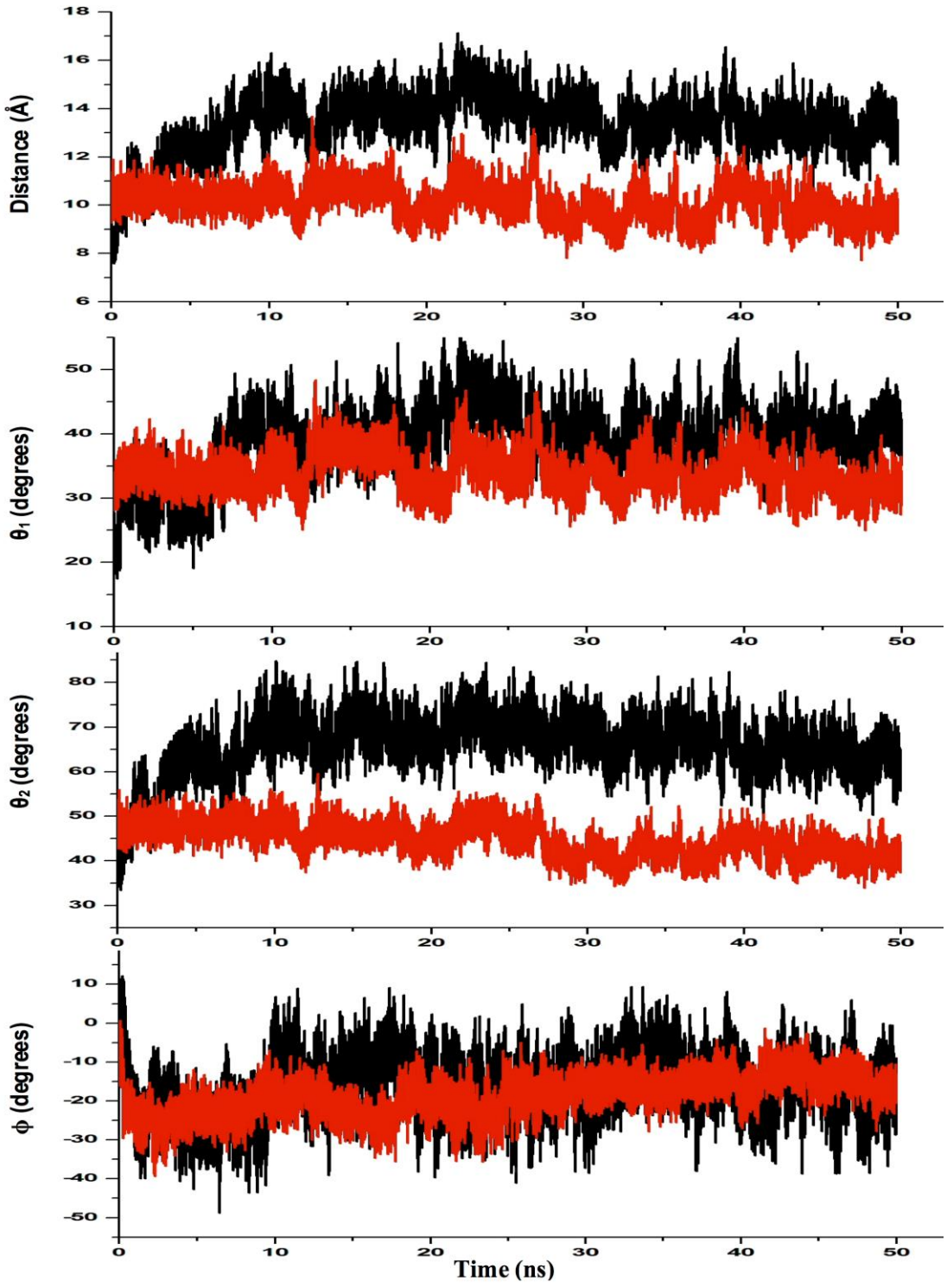


Figure 5. Plots showing the opening and closing of the active site of PlmV in two conformations, Apo PlmV (black) and WEHI-842 bound PlmV (red) as represented by the changes in distance in relation to the TriCα angles, θ1 and θ2, and the dihedral angle (ϕ).

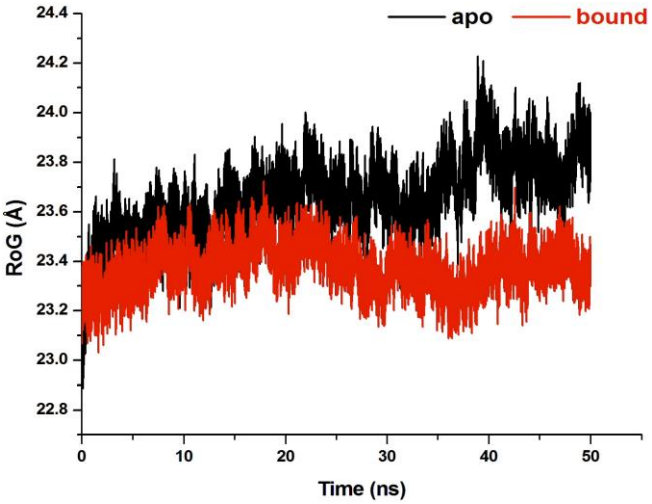


Figure 6. Radius of gyration of Apo PlmV (black) and WEHI-842 bound PlmV (red).

To further elucidate the impact of ligand binding on the structure of PlmV, the solvent accessible surface area (SASA) was measured. Solvent accessibility measures how much of a protein is exposed to or accessible by the solvent molecules [45]. Overall, the apo conformation of PlmV has a larger surface area exposed to solvent compared to bound PlmV, average SASA of 17784.9 Å and 17254.8 Å respectively (**Figure 7**). These findings suggest that as WEHI-842 binds to PlmV the structure becomes more compact as the active site folds inwards encompassing the inhibitor.

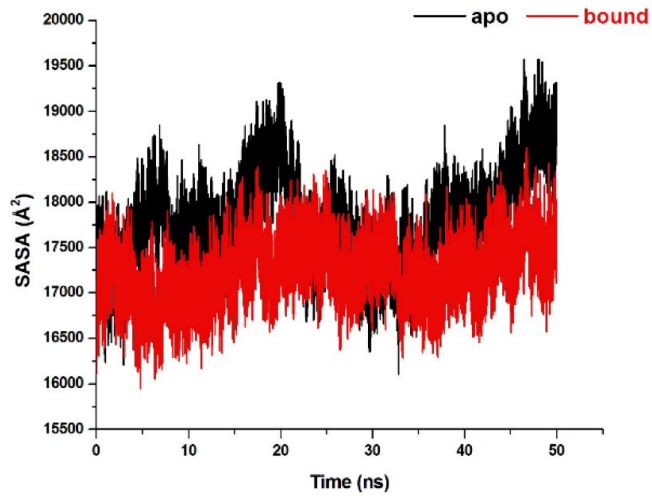


Figure 7. Solvent accessible surface area of Apo PlmV (black) and WEHI-842 bound PlmV (red).

3.3. Dynamic cross-correlation matrix (DCCM) analysis

To further investigate conformational changes upon WEHI-842 binding, DCCM analysis was performed on the positions of the Cα backbone atoms throughout each of the independent simulations, apo and bound, to determine the presence of correlated motions (**Figure 8**). Highly positive or correlated motions of specific residues are represented as red – yellow regions, whereas, highly negative or anti-correlated movements of specific residues are represented as blue – black regions. From the DCCM analysis it can be seen that WEHI-842 binding changes the structure of PlmV as reflected by the changes in correlated motions and dynamics. In the apo conformation (**Figure 8A**), there are small areas of correlated motions (red regions) which appear to be between regions where the disulfide bonds (between cysteine residues) are formed. This correlation, strongly increases upon ligand binding (**Figure 8B**). In the apo conformation, Cys141 is slightly correlated to Cys117 and Cys122 where the NAP1 insert is located. Upon binding of WEHI-842, this correlation strongly increases suggesting that flap movements and dynamics in part depends on the flexibility and dynamics of the NAP1 fold and surrounding regions (**Figure 8B**). The correlation observed for the flap tip can also be seen for Glu141, upon ligand binding and the formation of the salt bridge; correlated movements between the flap and NAP1 and surrounding regions increase. Suggesting, that upon ligand binding the correlation in movement between the flap region and the NAP1 insert and surrounding regions increase. Furthermore, ligand binding induces a shift to anti-correlated movement between the flap region and the helix-turn-helix (blue region) (**Figure 8B**). The flap region movement is more correlated to Asp80, this correlation increases upon ligand binding. Whereas, anti-correlated movements increase between the flap region and Asp313 in the bound conformation of PlmV.

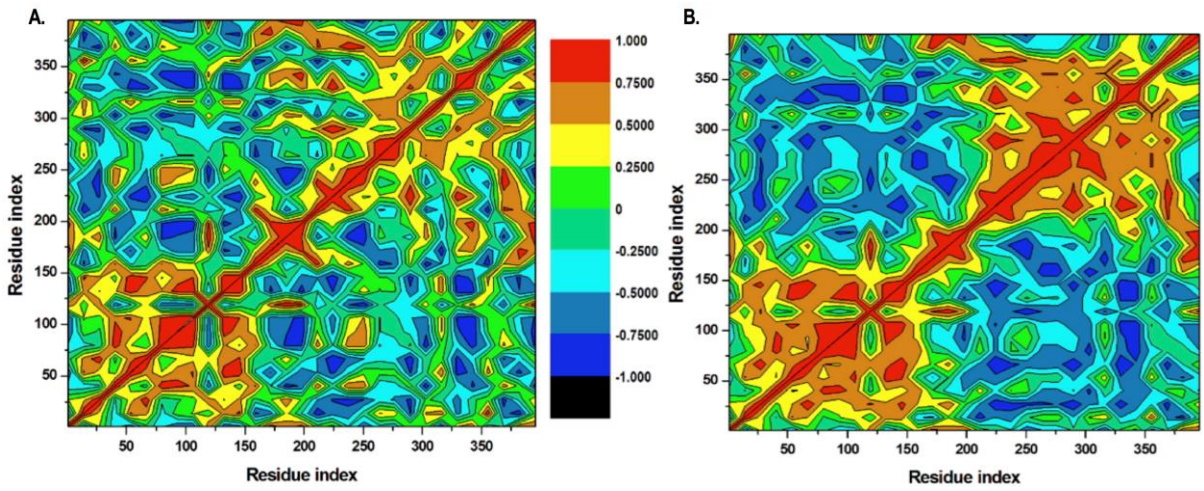


Figure 8. Cross-correlation matrices of the Cα fluctuations in both A. Apo PlmV and B. WEHI-842 bound PlmV throughout the 50 ns simulation.

3.4. Principle component analysis (PCA)

PCA is a technique which reduces the size or dimensionality of a given set of data, while retaining as much variation as possible and has been used extensively to account for variations detected in experimental simulations. PCA or essential dynamics provides insight into conformational changes and dynamics in proteins by detecting large and concerted conformational dynamics and fluctuations. For the purpose of the present study, a restricted analysis for the first two modes was implemented. **Figure 9** shows the essential dynamics for apo PlmV and WEHI-842 bound PlmV during the 50 ns MD simulation, where PC1 (X axis) and PC2 (Y axis) are constructs from a covariance matrix with the eigenvectors removed (rational motion). Each eigenvector is representative of the motion the Cα backbone atoms in a single direction through a multidimensional space; with a corresponding eigenvalue representative of amplitude. After collation of both the apo and bound systems, it can be seen that ligand binding to PlmV induces a conformation with restricted mobility and flexibility relative to its Cα backbone. Compared to apo PlmV, bound PlmV shows reduced spatial occupancy at any given moment (**Figure 9**). In the apo state two dominant conformations are observed, which is condensed to one main conformation upon ligand binding.

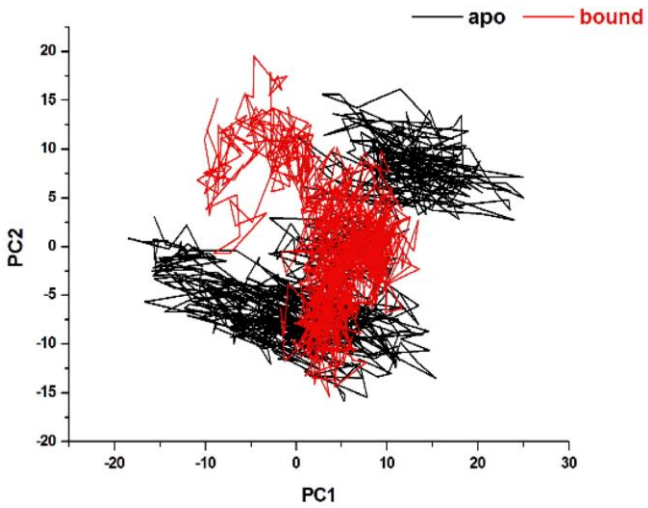


Figure 9. PCA scatter plots of 1000 frames along two planes, PC1 and PC2 for apo PlmV and WEHI-842 bound PlmV showing the difference in the eigenvectors of the 50 ns trajectories.

4. Conclusion

The aspartic protease, ER resident membrane bound plasmepsin V is responsible for the cleavage of PEXEL tagged proteins, and is vital for *Plasmodium* survivial[11,20,46]. Plasmepsin V represents a novel target in the search for antimalarials, as it is highly conserved throughout all *Plasmodium* species; and is structurally and functionally different to other members of the plasmepsin family. Recently, two potent inhibitors of PlmV has been identified WEHI-916 (IC_{50} *P. falciparum* 3.5nM; *P. vivax* 0.2 nM) and WEHI-842 (IC_{50} *P. falciparum* 2.1 nM; *P. vivax* 0.4 nM) [20,22]. We previously documented the flap dynamics of free unliganded plasmepsins I-V, using a homology model for PlmV as no crystal structure was available. Subsequently, a crystal structure of PlmV (*P. vivax*) in complex with the peptidomimetic inhibitor WEHI-842 (PDB 4ZL4) has been resolved [22]. In the present study, we comparatively show the differences in flap dynamics upon binding of WEHI-842 to PlmV using the crystal structure, MD simulations and previously defined parameters to measure flap opening and closing. Plasmepsin V has an array of intracellular target proteins, of varying shapes and sizes. The opening and closing of the active site ensures that target proteins are bought into the active site and held in close proximity to the catalytic aspartic dyad until amide bond cleavage occurs

releasing substrates for export. The sequence of the flap covering the active site of PlmV is highly conserved in *Plasmodium* species, this flap houses the unpaired cysteine residue at the flap tip (Cys140) [22]. Unlike other plasmepsins, PlmV is stabalised internally through the formation of 7 disulfide bonds between 14 cysteine residues [22]. As reported by Hodder and colleagues, we observed an association between the flap, the NAP1 insert and surrounding regions; from our simulations we observed flap opening as the NAP1 insert and surrounding regions become more disordered pulling the tip away from the flexible loop opening up the active site. Binding of WEHI-842 reduced flexibility and mobility in this region and in the flap region itself, via a salt bridge formed between Glu141 and WEHI-842. Inhibitor binding significantly reduced the flexibility and dynamics of the active site, although there is still movement between the flap and the flexible loop in the presence of WEHI-842. More specifically, the flexible loop region adjacent to the flap region is more flexible than expected. This could be due to the presence of the helix-turn-helix, a motif unique to PlmV which is highly conserved throughout *Plasmodium* species [22] and is not present in other members of the plasmepsin family. Lining the outside of the helix-turn-helix edges are highly conserved hydrophobic residues, which hold the amphipathic helices in strong antiparallel orientations. Upon visual inspection, ligand binding reduces flexibility in this region although visible movements throughout the simulation can be seen. Ligand binding induces a more compact rigid conformation, although regions located at the extremities are not as rigid as the core of PlmV. Thus ligand binding to PlmV, induces conformational changes outside of the active site to inhibit enzymatic activity. Such changes include reduced flexibility in the flap region, as the flap folds down towards the active site holding the inhibitor firmly in place thereby preventing activity. The current scaffold of WEHI-842 could potentially be further refined to allow the formation of a disulphide bond with the unpaired Cys141 residue in the flap; which could

potentially reduce flexibility and dynamics of the active site even more. The insights presented in the present study will aid in the design of more potent selective inhibitors of PlmV.

Acknowledgements

The authors would like to acknowledge the National Research Foundation (NRF) and The School of Health Sciences at the University of KwaZulu Natal, Westville campus for financial support. Further we would like to acknowledge the CHPC, Cape Town, RSA and the cBio cluster at MSKCC (San Diego) for high performance computational resources. We would also like to thank Dr Boddey and his team for sending us the PlmV homology model and the 4ZL4 crystal structure research article.

Conflicts of interest

Authors declare that there are no conflicts of interest.

5. References

(1) Gamo, F.-J., Sanz, L. M., Vidal, J., de Cozar, C., Alvarez, E., Lavandera, J.-L., Vanderwall, D. E., Green, D. V. S., Kumar, V., Hasan, S., Brown, J. R., Peishoff, C. E., Cardon, L. R., and Garcia-Bustos, J. F. (2010) Thousands of chemical starting points for antimalarial lead identification. *Nature 465*, 305–10.

(2) Rosenthal, P. J. (2002) Hydrolysis of erythrocyte proteins by proteases of malaria parasites. *Curr. Opin. Hematol. 9*, 140–5.

(3) (2015) WHO - World Health Organization.

(4) Sleebs, B. E., Gazdik, M., O'Neill, M. T., Rajasekaran, P., Lopaticki, S., Lackovic, K., Lowes, K., Smith, B. J., Cowman, A. F., and Boddey, J. A. (2014) Transition state mimetics of the Plasmodium export element are potent inhibitors of Plasmepsin V from *P. falciparum* and *P. vivax*. *J. Med. Chem. 57*, 7644–62.

(5) Sleebs, B. E., Gazdik, M., O'Neill, M. T., Rajasekaran, P., Lopaticki, S., Lackovic, K., Lowes, K., Smith, B. J., Cowman, A. F., and Boddey, J. A. (2014) Transition State Mimetics of the Plasmodium Export Element Are Potent Inhibitors of Plasmepsin V from *P. falciparum* and *P. vivax*. *J Med Chem*.

(6) Boddey, J. a, Carvalho, T. G., Hodder, A. N., Sargeant, T. J., Sleebs, B. E., Marapana, D., Lopaticki, S., Nebl, T., and Cowman, A. F. (2013) Role of plasmepsin V in export of diverse protein families from the *Plasmodium falciparum* exportome. *Traffic 14*, 532–50.

(7) Sargeant, T. J., Marti, M., Caler, E., Carlton, J. M., Simpson, K., Speed, T. P., and Cowman, A. F. (2006) Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol 7*, R12.

(8) Maier, A. G., Rug, M., O'Neill, M. T., Brown, M., Chakravorty, S., Szestak, T., Chesson, J., Wu, Y., Hughes, K., Coppel, R. L., Newbold, C., Beeson, J. G., Craig, A., Crabb, B. S., and Cowman, A. F. (2008) Exported Proteins Required for Virulence and Rigidity of *Plasmodium falciparum*-Infected Human Erythrocytes. *Cell 134*, 48–61.

(9) Chang, H. H., Falick, A. M., Carlton, P. M., Sedat, J. W., DeRisi, J. L., and Marletta, M. a. (2008) N-terminal processing of proteins exported by malaria parasites. *Mol. Biochem. Parasitol. 160*, 107–15.

(10) Hiller, N. L., Bhattacharjee, S., van Ooij, C., Liolios, K., Harrison, T., Lopez-Estrano, C., and Haldar, K. (2004) A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science (80-.). 306*, 1934–1937.

(11) Boddey, J. A., Hodder, A. N., Gunther, S., Gilson, P. R., Patsiouras, H., Kapp, E. A., Pearce, J. A., de Koning-Ward, T. F., Simpson, R. J., Crabb, B. S., and Cowman, A. F. (2010) An aspartyl protease directs malaria effector proteins to the host cell. *Nature 463*, 627–631.

(12) Russo, I., Babbitt, S., Muralidharan, V., Butler, T., Oksman, A., and Goldberg, D. E. (2010) Plasmepsin V licenses *Plasmodium* proteins for export into the host erythrocyte. *Nature 463*, 632–636.

(13) Marti, M. (2004) Targeting Malaria Virulence and Remodeling Proteins to the Host Erythrocyte. *Science (80-.). 306*, 1930–1933.

(14) Klemba, M., and Goldberg, D. E. (2005) Characterization of plasmepsin V, a membrane-bound aspartic protease homolog in the endoplasmic reticulum of *Plasmodium falciparum*. *Mol Biochem Parasitol 143*, 183–191.

(15) Park, H., and Lee, S. (2003) Determination of the active site protonation state of beta-secretase from molecular dynamics simulation and docking experiment: implications for structure-based inhibitor design. *J. Am. Chem. Soc. 125*, 16416–22.

(16) Ripka, A. S., Satyshur, K. A., Bohacek, R. S., and Rich, D. H. (2001) Aspartic protease inhibitors designed from computer-generated templates bind as predicted. *Org. Lett. 3*, 2309–12.

(17) Siripurkpong, P., Yuvaniyama, J., Wilairat, P., and Goldberg, D. E. (2002) Active site contribution to specificity of the aspartic proteases plasmepsins I and II. *J Biol Chem 277*, 41009–41013.

(18) Bhargavi, R., Sastry, G. M., Murty, U. S., and Sastry, G. N. (2005) Structural and active site analysis of plasmepsins of *Plasmodium falciparum*: potential anti-malarial targets. *Int J Biol Macromol 37*, 73–84.

(19) McGillewie, L., and Soliman, M. E. (2015) Flap flexibility amongst plasmepsins I, II, III, IV, and V: Sequence, structural, and molecular dynamics analyses. *Proteins Struct. Funct. Bioinforma. 83*, 1693–1705.

(20) Sleebs, B. E., Lopaticki, S., Marapana, D. S., O'Neill, M. T., Rajasekaran, P., Gazdik, M., Günther, S., Whitehead, L. W., Lowes, K. N., Barfod, L., Hviid, L., Shaw, P. J., Hodder, A. N.,

Smith, B. J., Cowman, A. F., and Boddey, J. a. (2014) Inhibition of Plasmepsin V activity demonstrates its essential role in protein export, PfEMP1 display, and survival of malaria parasites. *PLoS Biol. 12*, e1001897.

(21) Hodder, A. N., Sleebs, B. E., Czabotar, P. E., Gazdik, M., Xu, Y., O'Neill, M. T., Lopaticki, S., Nebl, T., Triglia, T., Smith, B. J., Lowes, K., Boddey, J. A., and Cowman, A. F. (2015) Structural basis for plasmepsin V inhibition that blocks export of malaria proteins to human erythrocytes. *Nat. Struct. Mol. Biol. 22*, 590–596.

(22) Hodder, A. N., Sleebs, B. E., Czabotar, P. E., Gazdik, M., Xu, Y., O'Neill, M. T., Lopaticki, S., Nebl, T., Triglia, T., Smith, B. J., Lowes, K., Boddey, J. A., and Cowman, A. F. (2015) Structural basis for plasmepsin V inhibition that blocks export of malaria proteins to human erythrocytes. *Nat. Struct. Mol. Biol. 22*, 590–596.

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

(24) Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem 25*, 1605–1612.

(25) Boonyalai, N., Sittikul, P., and Yuvaniyama, J. (2015) *Plasmodium falciparum* Plasmepsin V (PfPMV): Insights into recombinant expression, substrate specificity and active site structure. *Mol. Biochem. Parasitol. 201*, 5–15.

(26) Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem. 26*, 1668–1688.

(27) Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins 78*, 1950–8.

(28) Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., and Hutchison, G. R. (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform. 4*, 17.

(29) Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field. *J. Comput. Chem. 25*, 1157–1174.

(30) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926.

(31) Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys. 103*, 8577.

(32) Grest, G., and Kremer, K. (1986) Molecular dynamics simulation for polymers in the presence of a heat bath. *Phys. Rev. A 33*, 3628–3631.

(33) Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys. 81*, 3684.

(34) Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. . (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys. 23*, 327–341.

(35) Le Grand, S., Götz, A. W., and Walker, R. C. (2013) SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun. 184*, 374–380.

(36) http://www.originlab.com/ .

(37) Karubiu, W., Bhakat, S., McGillewie, L., and Soliman, M. E. (2015) Flap dynamics of plasmepsin proteases: insight into proposed parameters and molecular dynamics. *Mol Biosyst 11*, 1061–1066.

(38) Amadei, A., Linssen, A. B. M., and Berendsen, H. J. C. (1993) Essential dynamics of proteins. *Proteins Struct. Funct. Genet. 17*, 412–425.

(39) Cocco, S., Monasson, R., and Weigt, M. (2013) From Principal Component to Direct Coupling Analysis of Coevolution in Proteins: Low-Eigenvalue Modes are Needed for Structure Prediction. *PLoS Comput. Biol.* (Wallner, B., Ed.) *9*, e1003176.

(40) Arnold, G. E., and Ornstein, R. L. (1997) Molecular dynamics study of time-correlated protein domain motions and molecular flexibility: cytochrome P450BM-3. *Biophys. J. 73*, 1147–1159.

(41) Swaminathan, S., Harte, W. E., and Beveridge, D. L. (1991) Investigation of domain structure in proteins via molecular dynamics simulation: application to HIV-1 protease dimer. *J. Am. Chem. Soc. 113*, 2717–2721.

(42) Mao, Y. (2011) Dynamical Basis for Drug Resistance of HIV-1 Protease. *BMC Struct. Biol. 11*, 31.

(43) Szarecka, A., Xu, Y., and Tang, P. (2007) Dynamics of firefly luciferase inhibition by general anesthetics: Gaussian and anisotropic network analyses. *Biophys. J. 93*, 1895–1905.

(44) Lobanov, M. Y., Bogatyreva, N. S., and Galzitskaya, O. V. (2008) Radius of gyration as an indicator of protein structure compactness. *Mol. Biol. 42*, 623–628.

(45) Richmond, T. J. (1984) Solvent accessible surface area and excluded volume in proteins. *J. Mol. Biol. 178*, 63–89.

(46) Russo, I., Babbitt, S., Muralidharan, V., Butler, T., Oksman, A., and Goldberg, D. E. (2010) Plasmepsin V licenses Plasmodium proteins for export into the host erythrocyte. *Nature 463*, 632–6.

Supplementary Materials

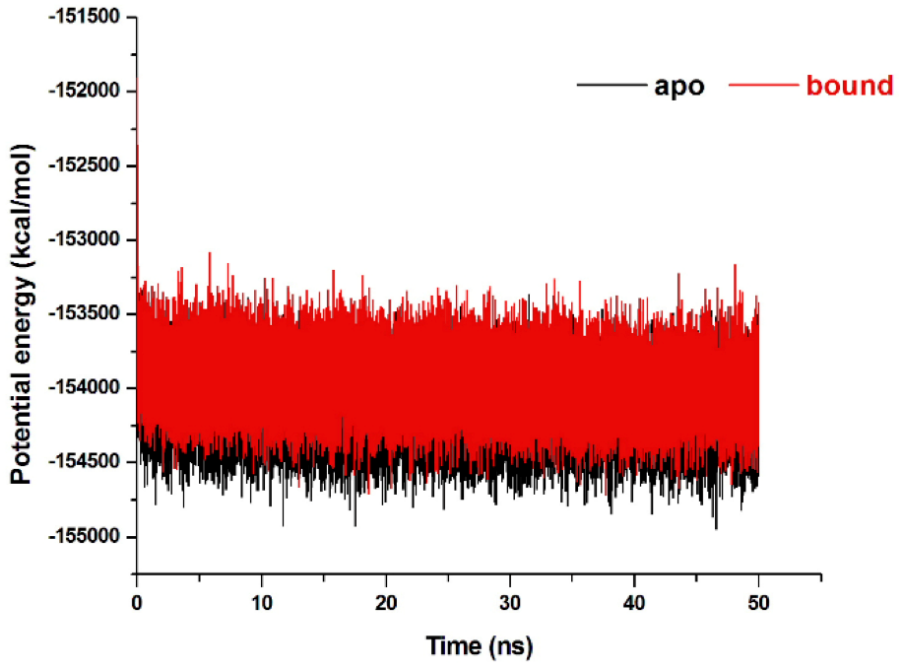


Figure S1. Potential energy fluctuations of apo PlmV (black) and WEHI-842 bound PlmV (red).

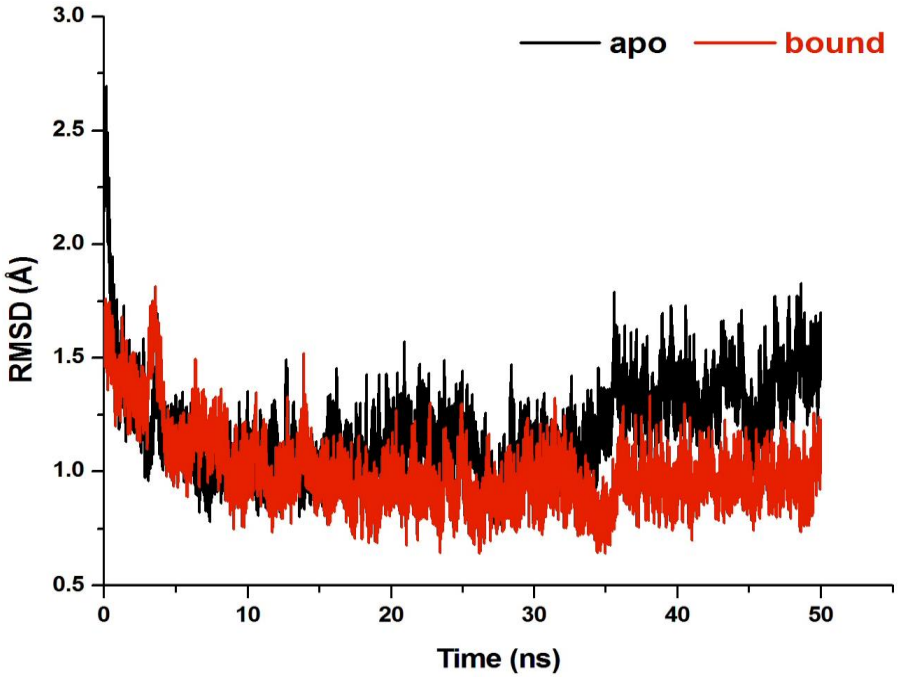


Figure S2. The Cα root mean square deviations (RMSD) for apo PlmV (black) and WEHI-842 bound PlmV (red).

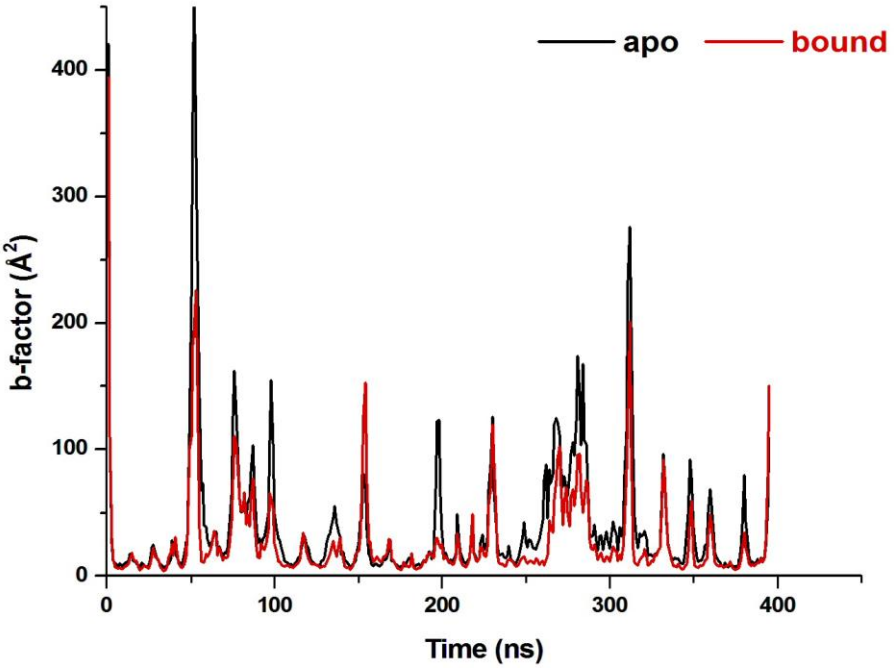


Figure S3. Plot of the b-factor values for apo (black) and bound (red) PlmV.

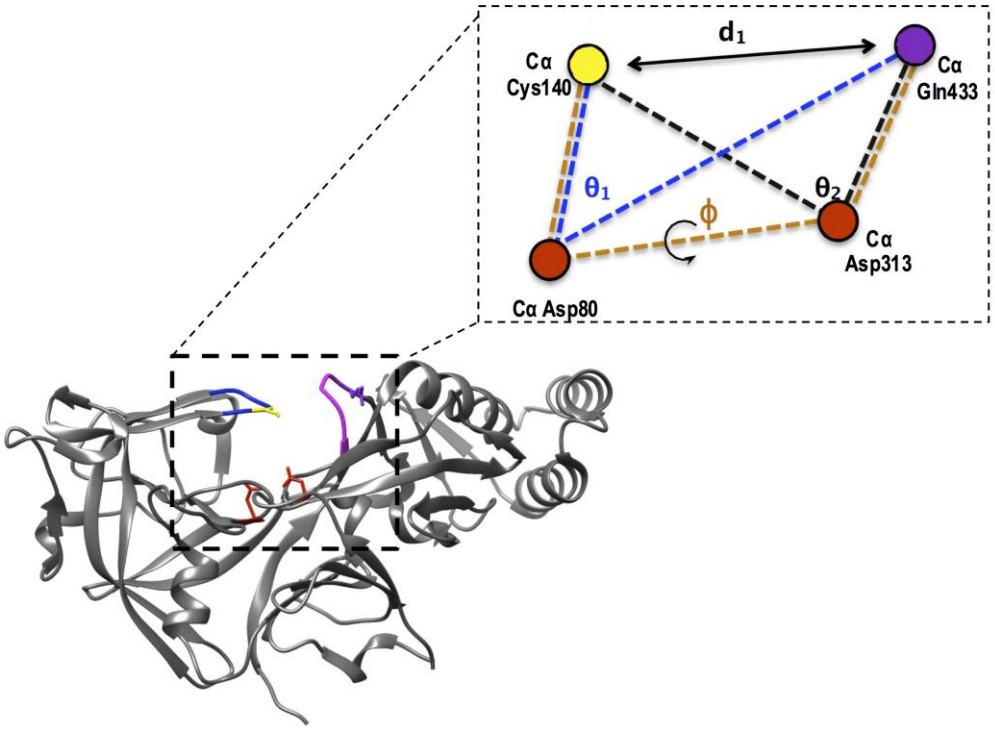


Figure S4. Schematic of the parameters used to measure flap opening and closing. Flap region 138-144(blue), flap tip Cys140 (yellow), catalytic aspartic acids 80, 313 (red), flexible loop region 433-439 (magenta) and hinge residue Gln433 (purple).

CHAPTER 7

Flap dynamics as unique conformational features among aspartic proteases

McGillewie L and Mahmoud E. Soliman*

Abstract

Aspartic proteases are hydrolytic enzymes that have been implicated in a number of diseases such as HIV, malaria, cancer and Alzheimer's. Generally speaking, their structure is relatively conserved in which the active site (aspartic acids) is housed at the interface of the N- and C-terminal lobes. Partially covering the active site is a flexible hairpin flap, lying perpendicular to the active site. The flap region in this class of enzyme is a characteristic unique to aspartic proteases; and has a profound impact on protein function. Dynamics in the flap region substantially affects the overall dynamics and conformation of the protein. Flap dynamics play a crucial role in drug binding, regulating access to the active site. Mutations in this region has a significant impact on the conformation of the protein and on drug binding; mutations that alter flap dynamics can lead to the development of resistance. Therefore understanding the dynamics and behaviour of the flap regions is crucial in the design of potent, selective inhibitors of aspartic proteases with increased efficacy and potency, and potentially reduce the incidence of resistance. The current review focuses on the flap region of aspartic proteases and the parameters used to measure the dynamics of the active site. Clear, concise and accurate parameters will aid in a better understanding of the dynamic nature of this enzyme family.

Keywords: parameters, flap dynamics, HIV, malaria, Alzheimer's, hypertension, cancer, aspartic protease, flexibility

1. Introduction to protein dynamics

Enzyme catalysis in its very nature is a dynamic process, where the binding of ligands often induce subtle and/or dramatic conformational changes [1]. Ultimately, atomic fluctuations result in bonds being broken and new ones being formed. It is well known that there is a correlation between the structure of a protein, the dynamics of a protein and the function of a protein – dynamics has a profound effect on function, therefore understanding the dynamics of a given system is critical to the success of ligand and protein design alike [1]. Protein flexibility is paramount to ligand recognition, predicting flexibility on static structural data alone is problematic. Which has been circumvented by the introduction of molecular modelling studies, which seek to understand the dynamic behaviour of biological systems on an atomistic level.

1.1. Aspartic proteases

Aspartic proteases are a family of enzymes involved in numerous important biological processes, found in plants, viruses, protozoa and vertebrates [2,3]. They are important hydrolytic enzymes that have become targets for diseases such as AIDS, malaria, hypertension and Alzheimers. Initially synthesised as proenzymes, aspartic protease are activated upon cleavage of the pro-segment. Aspartic proteases belong to a class of protease enzymes which use a catalytic aspartic acid dyad to cleave peptide bonds, through the activation and polarization of a water molecule and the formation of a transient tetrahedral intermediate [4–6]. The catalytic aspartic acid residues take on different roles in a general acid-base or 'push and pull' reaction[7]. In spite of the heterogeneity among aspartic proteases, the active site (motif Asp-Ser/Thr-Gly) and mechanism of action is conserved throughout the family [7–9]. The mechanism of action is highly conserved between members of the aspartic protease family, so much so that in 1991 Sharma and colleagues demonstrated that HIV-1 protease (viral) has the ability to produce angiotensin I at levels comparable to that of human renin [10].

The family is divided broadly into two groups based on their characteristic folds [3,11]. Retroviral proteases are β homodimers where the catalytic aspartic acid dyad is located at the monomer interface (1 Asp from each monomer) and the active site is covered by two identical hairpin loops [12]. Whereas, eukaryotic aspartic proteases are either α or β monomers made up of two asymmetrical lobes where the catalytic aspartic acid dyad is situated at the interface of the two lobes covered by a single hairpin loop [12,13]. Flaps covering the active site have a dual role in protease function: (i) structural, as interactions formed between a ligand and the flaps stabilise the ligand-protease complex, and (ii) kinetic, as flap closing induces ligand binding and flap opening induces ligand release [14,15]. The mechanism of action or catalytic cycle, for aspartic proteases is generally a gating mechanism where binding of the ligand forms a loose complex with the enzyme, flap closure inwards towards the active site orientating the geometry of ligand for catalysis (reactive conformation), bond cleavage, opening of flaps releasing the products and active site reformation [16–18]. By transitioning between different geometrical conformations of their active sites, aspartic proteases have the ability to modulate their catalytic activity (enhance or suppress). The residues making up the flap region and the flap tip of some of the aspartic proteases is listed in **Table 1**.

Table 1. Flap regions and flap tip of aspartic proteases.

Protein		**Flap region**
HIV protease		K43-P44-K45-M46-I47-G48-G49-***I50***-G51-G52-F53-I54-K55-V56-R57-Q58
Plasmepsin*	**Plasmepsin I**	N74-Y75-***V76***-S77-G78-T79
	Plasmepsin II	Y77-***V78***-S79-G80
	Plasmepsin III	S75-***K76***-A77-G78
	Plasmepsin IV	S76-Y77-***G78***-S79-G80-T81
	Plasmepsin V	S138-Y139-***C140***-E141-G142-S143-Q144
Renin		L73-R74-Y75-***S76***-T77-G78-T79-V80
Cathepsin D		S73-F74-D75-I76-H77-Y78-***G79***-S80-G81-S82-L83-S84-G85-Y86-L87
β- secretase 1		V67-Y68-V69-P70-Y71-***T72***-Q73-G74-A75

*Only Plm I-V, as the is not much data in the literature for VI-X

The opening and closing of the flap structures of aspartic proteases is crucial to their catalytic activity, better understanding flap dynamics will aid in the design of more efficient inhibitors. Additionally, disrupting flap movements could potentially inhibit this class of enzymes. The current review serves as a compilation of information on the flap regions of aspartic proteases, and the parameters used to measure the dynamics of these regions. Members of the aspartic protease are discussed in more detail below.

2. HIV protease

The human immunodeficiency virus (HIV) protease (PR) plays a role in the development of acquired immune deficiency syndrome (AIDS), and is an important target in the development of drugs targeted against AIDS. HIV-PR is responsible for cleaving the peptide bond in *pol* and *gag* polyprotein (between Phe and Pro, or Tyr and Pro) precursors into more mature structural proteins [19,20]. Inhibition prevents viral maturation and prevents the virus from taking its infectious form [21]. This aspartic protease is a symmetric 22 kDa homodimer, each chain consisting of 99 amino acid residues (Chain A 1-99, Chain B 100-198) (**Figure 1**) [1,22]. The base of the binding pocket, situated at the interface between the two dimers is composed of a conserved catalytic triad, Asp25-Thr26-Gly27 (Asp124-Thr125-Gly126), where the aspartic acid dyad (Asp25 Asp124) forms the active site which is firmly held in place through a network of strong hydrogen bonds forming the fireman's grip (residues 24 – 29, and 123 - 128) [22]. Covering the active site is a pair of overlapping flexible flaps (residues 43 – 58, and 142 - 157), with the flap tips, Ile50 and Ile149, being the outmost residue of the flap [16,23]. Other important structural features, named after their hypothesized function, include the fulcrum (residues 11 – 22, and 110 - 121), the flap elbow (residues 35 – 42, and 134 - 141) and the cantilever (residues 59 – 75, and 158 - 174) (**Figure 1**). Unique to HIV-PR is the conserved water molecule in the flap region, 'flap water', which is hydrogen bonded to the flap region via Ile50 [22]. Generally,

two distinct conformations is adopted by HIV PR; semi-open in apo HIV PR and a more closed conformation in the ligand bound HIV PR [24,25]. Numerous crystallographic, NMR and computational studies have revealed that in an unbound conformation the flaps loosely transition between numerous conformations, most of the time a free HIV PR is in a semi-open conformation sterically regulating ligand access into the active site [16,26–28]. The glycine rich flaps (Met-Ile-Gly-Gly-Ile-Gly-Gly-Phe-Ile) are highly mobile and flexible; and transitions between semi-open and open conformations in the absence of a ligand [23,29–32]. The flexibility and curling of the flap regions, is in part due to the high degree of flexibility of glycine residues either side of the centrally located isoleucine [23]. During simulations it was observed that in the ligand free conformation HIV PR assumes a semi-open conformation, as the flaps curl upwards away from the active site due to a concentrated downward movement in the fulcrum, flap elbows and cantilever. Ligand binding induces large conformational changes allowing the flaps to move from a semi-open conformation to a closed conformation; adopting a more stable compact structure, as the flaps are pulled inward [16,24,25]. Stability conferred upon ligand binding is due to interactions between the ligand and the flaps, rather than the partially stable bonds that temporarily form between the flaps themselves [16].

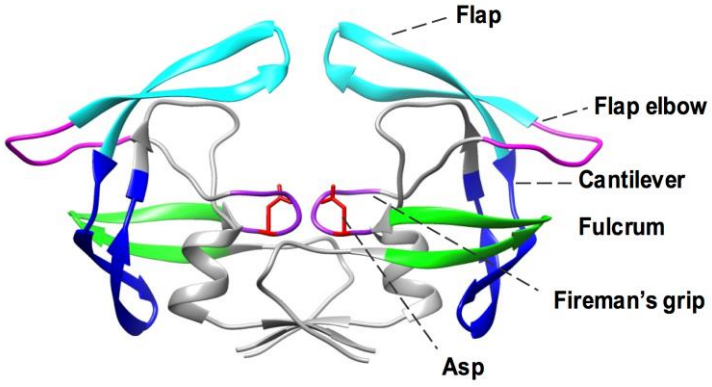
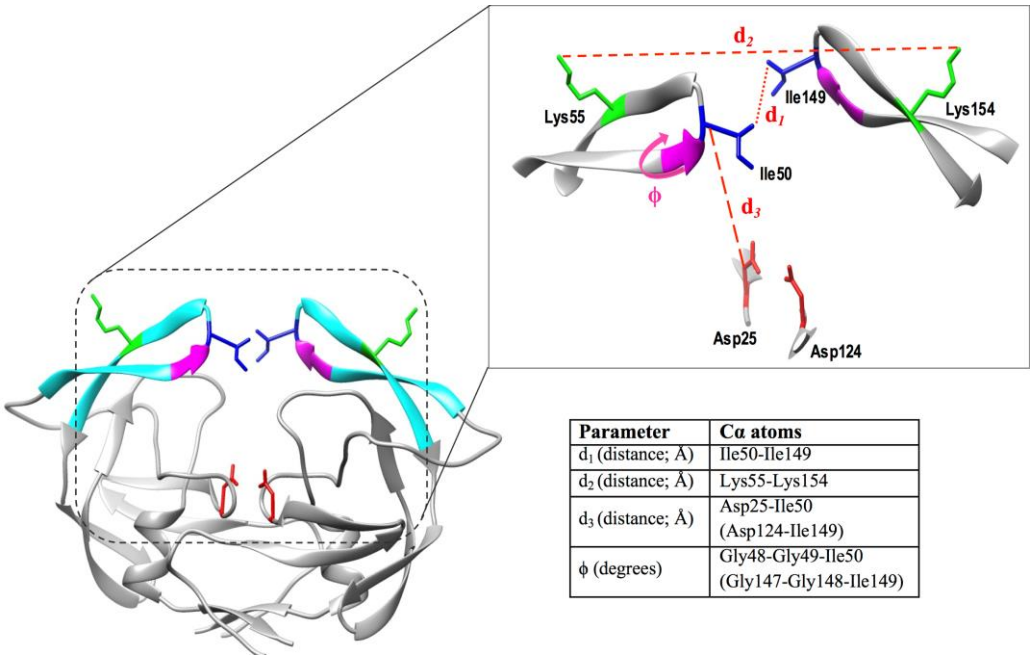


Figure 1. Schematic of HIV PR dimer (PDB 2UYO [33]) denoting the active site aspartic acid residues (red, 25), the fulcrum (green; 11–22), the fireman's grip (purple; 24-29), the flap elbow (magenta; 35-42), the flap (cyan; 43-58) and the cantilever (blue; 59-75).

Currently there are several protease inhibitors used in clinical practice such as saquinavir, lopinavir and darunavir to name a few, unfortunately the benefits of these drugs is compromised by the high rate of HIV mutation and increasing drug resistance [7,34–36]. Structurally, mutations in HIV PR are either located in the active site or outside the active site (non-active site mutations). Most mutations are situated in the active site, and tend to be conservative as they preserve the charge and polarity yet change the geometry and conformation of the active site. Mutations outside of the active site, cause structural geometrical rearrangements that distort the active site and alter the dynamics of HIV PR, such mutations can occur in the flap regions [7,36–38]. It has been reported that not only do mutations alter the binding affinity of HIV PR inhibitors, some also have the ability to affect the dynamics of the flap. Mutations in the flap regions also have the ability to alter the flexibility of the flap region [39]. In the face of increasing resistance, it is of paramount importance to design novel inhibitors that fits tightly in the active site that has the ability to inhibit different strains and variants of the target protein. Such adaptive inhibitors will slow down resistance, be more efficacious, suppress the viral load quicker and in the long run will be more affordable and economically viable.

The importance of flap dynamics and protein flexibility is highlighted in HIV strains in which mutations in regions regulating flap movement, confer increased resistance and decreased susceptibility to HIV PR inhibitors. The flap dynamics and flap motions of HIV PR has been well characterised in the literature, although the exact parameters which accurately account for the opening and closing is more vague. The most commonly used metric to measure flap opening and closing, is to measure the distance (d_1) between the Cα of Ile50 and the Cα of Ile149 (**Figure 2**). In the fully open conformation, the distance between the two flap tips is ~30 Å [16,24,25,40–42]. Measuring the distance between the Cα of Lys55 and the Cα of Lys154 (d_2)

seem to give a more representative value for the magnitude and extent of opening of the active site, where a fully open active site reaches ~43 Å [43–45]. More recently, it has been observed that the distance between Asp25-Ile50 and/or Asp124-Ile149 (d_3) more accurately define flap movements, as it is accounts for both the asymmetry of the flaps as well as the curling of the flaps [46,47]. Chibi and colleagues used all three distance parameters mentioned above and the dihedral angle (ϕ) of Gly48-Gly49-Ile50 (A) and Gly147-Gly148-Ile149 (B) to monitor the curling behaviour and opening and closing of HIV PR [48]. From this study they suggest the most accurate parameter to use is d_2, the distance between Cα Lys55 – Cα Lys154 as it shows complete flap opening and closing. Whereas the dihedral angle is representative of the tightness and compactness of the active site [48]. All parameters in the literature to date is captured in Figure 2 and Table 1 delineates the parameters to their corresponding residues and atoms. In the case of HIV PR, Scott and colleagues suggested that an inhibitor that locks the flaps of HIV PR in an open position preventing the flaps from closing, would be less susceptible to drug resistance. Therefore, understanding the flap behaviour and impact on protein dynamics and function is quintessential to the development of potent HIV PR inhibitors.



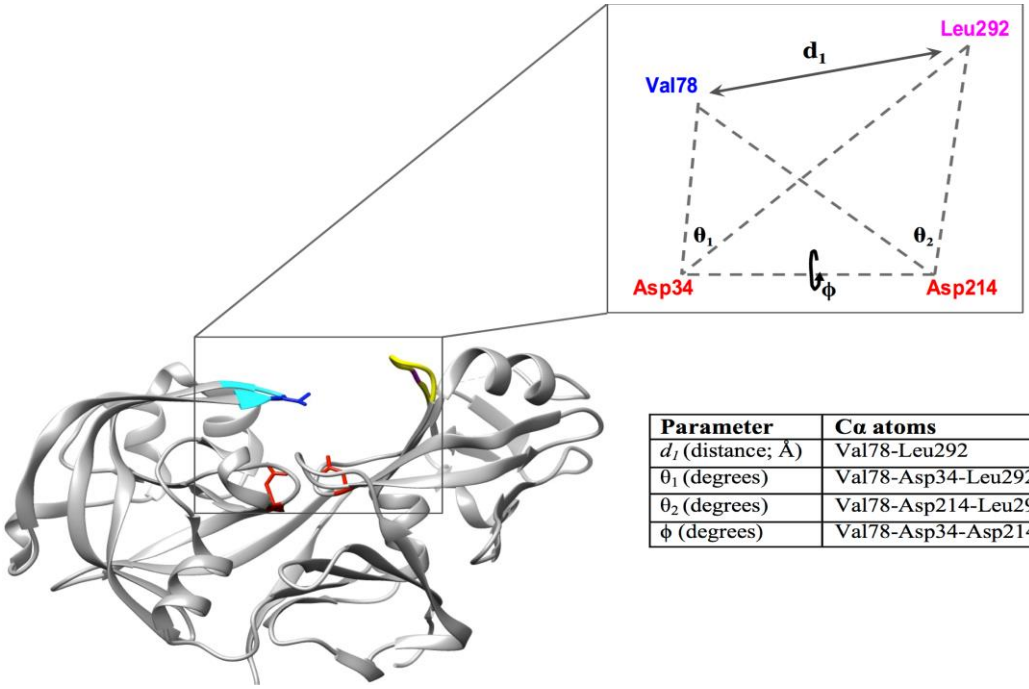
Parameter	Cα atoms
d_1 (distance; Å)	Ile50-Ile149
d_2 (distance; Å)	Lys55-Lys154
d_3 (distance; Å)	Asp25-Ile50 (Asp124-Ile149)
ϕ (degrees)	Gly48-Gly49-Ile50 (Gly147-Gly148-Ile149)

Figure 2. HIV PR dimer (PDB 2UYO [33]) showing the different parameters used to account for the flexibility of the active site and flap structure.

3. Plasmepsins

Malaria is one of the deadliest diseases known to mankind; in 2013 alone, 198 million cases were reported with an estimated 584 000 deaths worldwide [49]. Malaria is an infection caused by the *Plasmodium* parasite [50,51]. Of all the species to infect man, *Plasmodium falciparum* is the most lethal and resistance of this strain to current antimalarials is increasing [52]. Plasmepsins (Plm), aspartic proteases unique to the *Plasmodium* species, are attractive candidates in the development of novel potent antimalarials [52,53]. Ten genes, encoding I-X plasmepsins have been identified through the sequencing of the *P. falciparum* genome [54,55]. Plasmepsin I, II, IV and the closely related histo-aspartic protease (HAP) (vacuolar plasmepsins) are situated in the acidic food vacuole (FV) of infected red blood cells (RBC), where they are responsible for the degradation of hemoglobin [52,54,56–58]. Hemoglobin catabolism is crucial to the survival of the malaria parasite, although the overlapping function of dual proteases and redundant catabolic hemoglobin pathways ensures optimum degradation; and necessitates the need for dual inhibition of aspartic and cysteine proteases [53,59,60]. *P. falciparum* is the only *Plasmodium* species that actively expresses four Plms in the FV, other *Plasmodium* species only express PlmIV [61]. Plasmepsins expressed outside of the FV, Plm V-X, are expressed in all other *Plasmodium* species [52,62]. Vacuolar plasmepsins have roughly 60 – 75% identity with each other; HAP is structurally and functionally similar to vacuolar plasmepsins with the exception that the conserved catalytic aspartic dyad active site is composed of an Asp-His active site [54,59,62]. Plasmepsin V is conserved throughout all *Plasmodium* species and distantly related to other plasmodium aspartic proteases (17% to PlmII); deletion and knock down of PlmV has shown it is crucial to the survival of all *Plasmodium* species [63–66]. Localised to the parasitic endoplasmic reticulum, this membrane bound aspartic protease is responsible for the export of PEXEL (plasmodium export element) tagged proteins into the cytoplasm and surface of infected RBCs [63,67,68]. Similar to other aspartic proteases, mature plasmepsins fold into two

topologically similar N- and C- terminal lobes, anchored through a six stranded inter domain towards the bottom where the active site is housed [52,61]. The active site is composed of two catalytic aspartic acids, which is partially covered by a β-hairpin flap [69]. Numerous studies have shown that residues lining the active site such as the flap and flexible loop, undergo large conformational changes upon ligand binding. Essentially, the flap moves to a more closed conformation in the presence of an inhibitor or substrate, closing the active site [52,69–72]. Despite the high degree of similarity in the overall structure and conservation of the active site, each plasmepsin has unique specificity towards substrates. Substrate specificity is due to the sequence heterogeneity of residues lining the active site, such as the flap and flexile loop [73]. Using a set of parameters to quantify flap dynamics, we have previously reported how changes in the sequence of the flap and flexible loop region affect and alter the flexibility of the active site [70,74]. These parameters measures the distance (d_1) between Cα carbon atoms of the flap tip and the hinge residue of the flexible loop; in relation to the TriCα angles, θ_1 and θ_2, and the dihedral angle (ϕ) between the flap tip, the catalytic dyad and the hinge residue of the flexible loop (**Figure 3**). For the purpose of the review, the schematic is in accordance to PlmII and residues pertaining to the active site of PlmII; although these parameters were also used to measure flap dynamics and flexibility of Plm I, II, HAP, IV and V (**Figure 3**). To date, there are no approved plasmepsin inhibitors; thus better understanding the dynamics is crucial to the design of novel inhibitors more resilient to resistance.



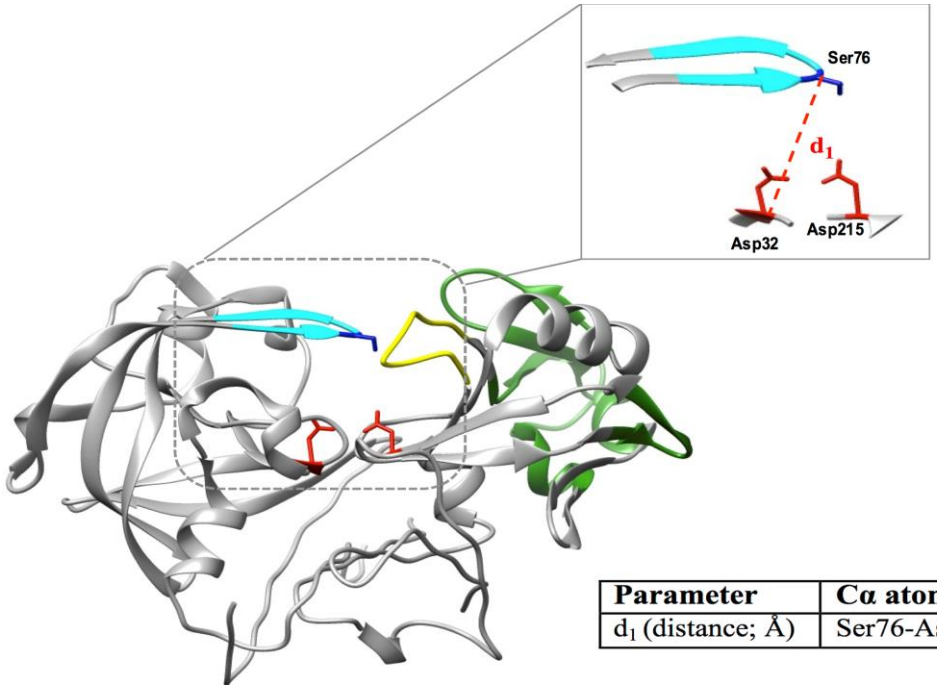
Parameter	Cα atoms
d_1 (distance; Å)	Val78-Leu292
θ_1 (degrees)	Val78-Asp34-Leu292
θ_2 (degrees)	Val78-Asp214-Leu292
ϕ (degrees)	Val78-Asp34-Asp214-Leu292

Figure 3. Plasmepsin II structure (PDB 1LF4 [69]) showing the different parameters used to account for the flexibility of the active site and flap region.

4. Renin

Renin, also known as angiotensinogenase, is an aspartic protease involved in the renin-angiotensin-aldosterone system (RAAS). This system is responsible for regulating the body's water and electrolyte homeostasis; and is responsible for regulating arterial blood pressure [75]. Renin is secreted from specialized granular cells found in the juxtaglomerular apparatus of the kidneys; and is secreted in response to decreased arterial blood pressure, decreased sodium chloride levels and sympathetic nervous system activation. Renin is classically termed a hormone, although renin is an enzyme responsible for the hydrolysis of the amide bond @ L10 – V11 of angiotensinogen (from the liver) to angiotensin I (rate limiting step), which is converted to the potent vasoconstrictor angiotensin II [76,77]. Angiotensin II leads to the constriction of blood vessels, increased secretion of antidiuretic hormone and aldosterone; all causing an increase in blood pressure. Defects in the RAAS, such as an over reactive RAAS causes vasoconstriction, retention of water and Na^+ leading to the development of

hypertension. Since renin determines the rate limiting step in the formation of angiotensin I, inhibition would be useful in the management of hypertension [77,78]. In clinical practice hypertension is treated using angiotensin converting enzyme (ACE) inhibitors or angiotensin II receptor blockers (ARB) [79,80]. In 2007 the first direct renin inhibitor was approved by the FDA, aliskiren, which binds in the active pocket inhibiting the formation of angiotensin I [81–84]. Synthesised as a precursor or pro-enzyme, renin is made up of ~406 amino acids and upon cleavage of the pre and pro-segments mature renin consists of 340 amino acids (37kDa). The active site, Asp32-Thr33-Gly34 and Asp215-Thr216-Gly217, is situated deep inside the binding pocket, covered by a single β-hairpin loop Leu73-Val80 (**Figure 4**) [85]. A proline rich flexible loop lies across from the flap tip, Pro292, Pro293, Pro294 and Pro297. Similarly to other aspartic proteases, in the apo conformation renin is highly flexible as flap motions regulate ligand access to the active site. Ligand binding induces a more stable structure as the flap assume a closed position [86,87]. The open and closed conformations differ in their orientations of the C-terminal loops 198-204, 236-254 and 270-286 (**Figure 4**) [88,89]. The importance of flap movements and dynamics in regulating ligand access into the active site is clearly highlighted in the literature; although clear, concise and consistent parameters to measure the opening and closing of the renin active site has not yet been reported. One parameter that has been reported, is the distance between the flap tip (Ser76) and Asp32 in the active site [85]; a distance parameter also used to account for the flap dynamics of HIV PR (**Figure 4**).



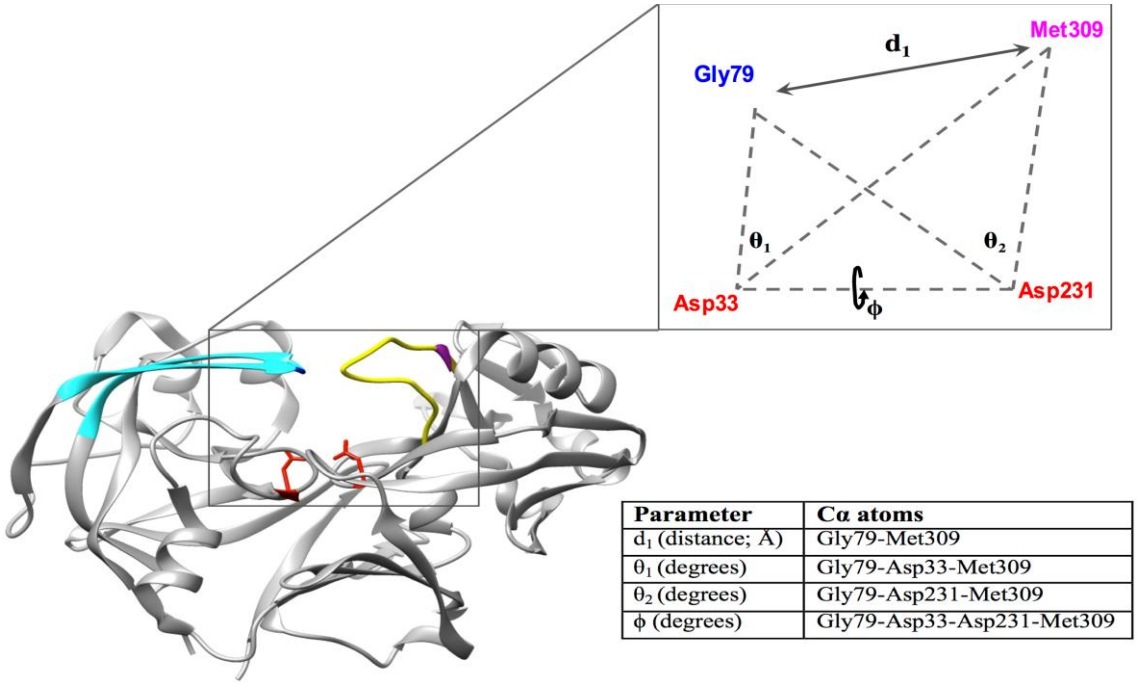
Parameter	Cα atoms
d_1 (distance; Å)	Ser76-Asp32

Figure 4. Renin structure (PDB 2V0Z [89]) showing the different parameters used to account for the flexibility of the active site and flap region.

5. CathepsinD

Cathepsins are an intracellular lysosomal enzymes, found in eukaryotes and are responsible for the degradation of proteins in lysosomes at acidic pH's [90,91]. Cathepsins are subdivided according to the amino acids present in their active sites – cysteine (B, C, H, F, L, O, S, V, W), serine (G) or aspartic (D and E) [91,92]. Cathepsin D (cat-D) is the major protein degrading enzyme in phagosomes and lysosomes (lysosomal hydrolase) [91]. Over expression and hyper secretion of cat-D has been linked to an increased risk of cancer, and stimulates metastasis via effects on cell proliferation [93–96]. Mature cat-D is mainly found in a two chain structure and phosphorylated N-linked oligosaccharides in cat-D target it to lysosomes, two characteristic features of lysosomal hydrolases. Typical of aspartic proteases cat-D consists of three distinct regions namely, the N-terminal domain, an antiparallel interdomain β-sheet and the C-terminal domain [90]. The interdomain links the N and C terminal, each contributing an aspartic acid forming the active site of cat-D (Asp33 and Asp231) [90]. The flexible β-hairpin loop (flap)

covering the active site is composed of residues 72-87, with the Gly79 situated at the flap tip. Upon binding of pepstatin, the flexibility of the flap region is reduced [90]. Situated opposite the flap region, on the C-terminal side of cat-D, is a proline rich loop (Pro312-Pro313-Pro314-Ser315-Gly316-Pro17) analogous to the proline rich loop found in renin [90]. Although flap motions, and conformational changes induced upon ligand have been reported for cat-D, no concise or accurate parameters have yet been reported. In unpublished work from our group, we have proposed parameters adopted from other aspartic proteases (BACE and plasmepsins) to more accurately account for the flexibility of the cat-D active site (**Figure 5**). These parameters measure the distance (d_1) between Cα carbon atoms of the flap tip (Gly79) and the hinge residue of the proline rich loop (Met309); in relation to the TriCα angles, θ_1 (Gly79-Asp33-Met309) and θ_2 (Gly79-Asp231-Met309), and the dihedral angle (ϕ) between Gly79-Asp33-Asp231-Met309 (**Figure 5**).



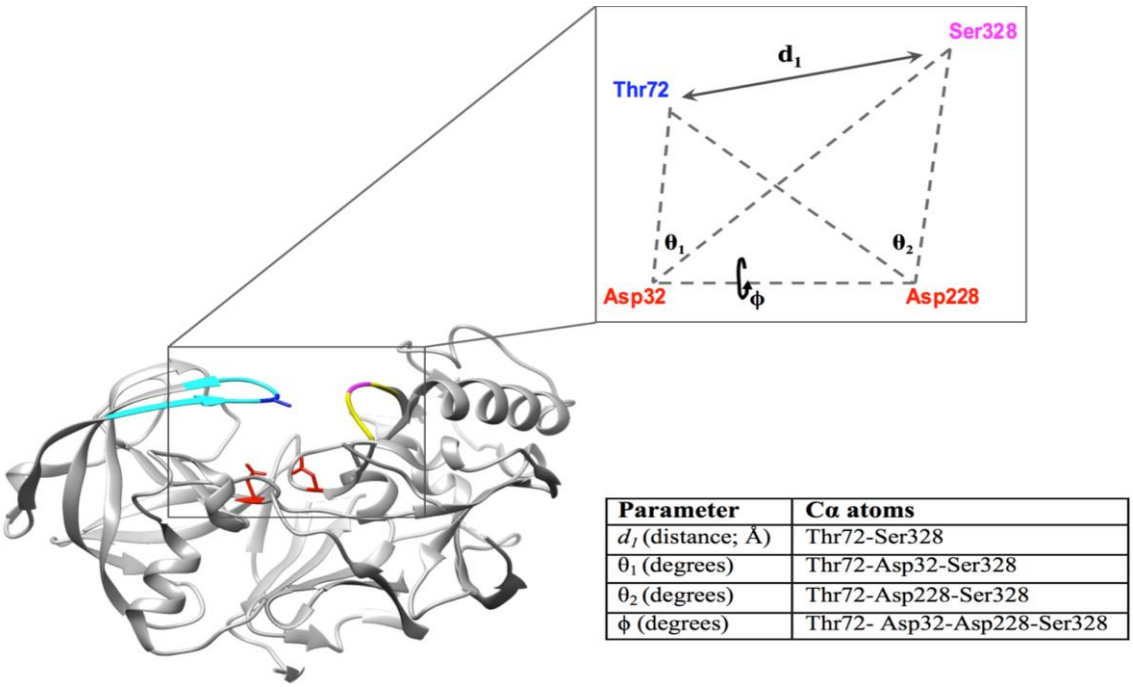
Parameter	Cα atoms
d_1 (distance; Å)	Gly79-Met309
θ_1 (degrees)	Gly79-Asp33-Met309
θ_2 (degrees)	Gly79-Asp231-Met309
ϕ (degrees)	Gly79-Asp33-Asp231-Met309

Figure 5. Cathepsin D structure (PDB 1LYA [90]) showing the different parameters used to account for the flexibility of the active site and flap region.

6. *β*-secretase (BACE)

Alzheimer's disease (AD) is a fatal progressive neurodegenerative disorder that affects more than 36 million people worldwide; and is the leading cause of dementia in the elderly [97]. The hallmark of AD is the formation amyloid plaques (extracellular) and neurofibrillary tangles (intracellular) in the cortical grey areas of the brain and in the hippocampus [17]. The amyloid plaques are formed due to the accumulation and deposition of small peptides (40 or 40 amino acids) called amyloid β (A*β*) peptides; generated from the proteolytic cleavage of amyloid precursor protein (APP), a transmembrane protein, by the two aspartic proteases β- and γ-secretase [98–100]. Amyloid peptides are generated by the sequential cleavage of APP; first by β-secretase (BACE) which represents the rate limiting step of the amylogenic pathway, followed by γ- secretase degradation [101–104]. As BACE plays a crucial role in the rate limiting step of APP hydrolysis, inhibition of BACE would regulate the formation of amyloid plaques and potentially the progression of AD [105]. There are two forms of the BACE enzyme, BACE1 (501 amino acids) and BACE2 (518 amino acids), which are 45% identical and share ~75% homology [106,107]. Both BACE1 and BACE2 have the ability to cleave APP at the β-site, although BACE2 has a higher preference for within the Aβ peptide [108]. BACE1 is highly expressed throughout the brain, whereas BACE2 is expressed at low levels in the brain but expressed at varying levels in peripheral tissue [100,106]. Due to the cleavage specificity and localization of BACE1 activity, BACE1 is believed to be the secretase responsible for the formation of plaques in the brain [109]. The membrane anchored BACE1 structure can broadly be broken down into three regions: the N-terminal 'ectodomain', a transmembrane domain and a cytosolic C-terminal domain [110]. The protease domain of BACE1 is located in the N-terminal ectodomain; a conserved aspartic protease fold houses the active site at the interface between the N and C terminal lobes [111]. The conserved catalytic aspartic acid dyad, Asp32 and Asp228, is partially covered by a hairpin loop (flap). The flap is made up of residues 67-75 and lies

perpendicular to the active site [112]. In the apo conformations of BACE1, the flap region adopts more open configurations; whereas ligand binding causes the flap to move inwards towards the active site residues closing the binding cleft [14,111,113,114]. Parameters that accurately account for the flexibility, opening and closing; as well as the twisting of the BACE1 active site have been reported and are depicted below (**Figure 6**) [115]. These parameters measure the distance (d_1) between Cα carbon atoms of the flap tip (Thr72) and the hinge residue of the proline rich loop (Ser328); in relation to the TriCα angles, θ_1 (Thr72-Asp32-Ser328) and θ_2 (Thr72-Asp228-Ser328), and the dihedral angle (ϕ) between Thr72-Asp32-Asp228-Ser328 (**Figure 6**) [115].



Parameter	Cα atoms
d_1 (distance; Å)	Thr72-Ser328
θ_1 (degrees)	Thr72-Asp32-Ser328
θ_2 (degrees)	Thr72-Asp228-Ser328
ϕ (degrees)	Thr72- Asp32-Asp228-Ser328

Figure 6. BACE1 structure (PDB 3TPL [116]) showing the different parameters used to account for the flexibility of the active site and flap region

7. Conclusion

Out of the family of aspartic proteases that cause disease, only two have inhibitors that are currently used in clinical practice. These inhibitors are targeted towards HIV PR (Saquinavir, Ritonavir, Lopinavir to name a few)[117] and renin (Aliskiren). To date, there are no approved inhibitors targeted towards plasmepsins, BACE1 and cat-D. In the case of HIV PR, even though

drugs targeted against HIV PR are currently used, increasing resistance is proving problematic. Numerous studies have shown the importance of protein dynamics and flexibility in protein function [118–121]. Conformational changes in the flap region control ligand access into the active site, align the ligand in the correct geometry and assists in the removal of the byproducts of substrate hydrolysis [114]. Therefore, understanding and quantifying the flexibility and dynamic behaviour of the flap region is advantageous to the design of novel inhibitors. Immobilising the flap region and the active site, could potentially render the aspartic protease inactive.

Acknowledgements

The authors would like to acknowledge the National Research Foundation (NRF) for financial assistance and the School of Health Science, University of KwaZulu-Natal, Westville Campus.

Conflict of Interest

Authors declare no financial and intellectual conflict of interests.

8. References

(1) Boehr, D. D., Dyson, H. J., and Wright, P. E. (2006) An NMR Perspective on Enzyme Dynamics. *Chem. Rev. 106*, 3055–3079.

(2) Cooper, J. B. (2002) Aspartic proteinases in disease: a structural perspective. *Curr. Drug Targets 3*, 155–73.

(3) Dunn, B. M. (2002) Structure and Mechanism of the Pepsin-Like Family of Aspartic Peptidases. *Chem. Rev. 102*, 4431–4458.

(4) Northrop, D. B. (2001) Follow the Protons: A Low-Barrier Hydrogen Bond Unifies the Mechanisms of the Aspartic Proteases. *Acc. Chem. Res. 34*, 790–797.

(5) Davies, D. (1990) The Structure And Function Of The Aspartic Proteinases. *Annu. Rev. Biophys. Biomol. Struct. 19*, 189–215.

(6) Simões, I., and Faro, C. (2004) Structure and function of plant aspartic proteinases. *Eur. J. Biochem. 271*, 2067–2075.

(7) Castro, H. C., Abreu, P. A., Geraldo, R. B., Martins, R. C. A., dos Santos, R., Loureiro, N. I. V., Cabral, L. M., and Rodrigues, C. R. (2011) Looking at the proteases from a simple perspective. *J. Mol. Recognit. 24*, 165–181.

(8) Tang, J., James, M. N. G., Hsu, I. N., Jenkins, J. A., and Blundell, T. L. (1978) Structural evidence for gene duplication in the evolution of the acid proteases. *Nature 271*, 618–621.

(9) Swanstrom, R. (2000) Human immunodeficiency virus type-1 protease inhibitors therapeutic successes and failures, suppression and resistance. *Pharmacol. Ther. 86*, 145–170.

(10) Sharma, S. K., Evans, D. B., Hui, J. O., and Heinrikson, R. L. (1991) Could angiotensin I be produced from a renin substrate by the HIV-1 protease? *Anal. Biochem. 198*, 363–367.

(11) Rawlings, N. D., and Barrett, A. J. (1995) Familes of aspartic peptidases, and those of unknown catalytic mechanism. *Methods 248*, 1995.

(12) Cascella, M., Micheletti, C., Rothlisberger, U., and Carloni, P. (2005) Evolutionary conserved functional mechanics across pepsin-like and retroviral aspartic proteases. *J. Am. Chem. Soc. 127*, 3734–3742.

(13) Sielecki, A. R., Fujinaga, M., Read, R. J., and James, M. N. G. (1991) Refined structure of porcine pepsinogen at 1·8 Å resolution. *J. Mol. Biol. 219*, 671–692.

(14) Hong, L., and Tang, J. (2004) Flap Position of Free Memapsin 2 (β-Secretase), a Model for Flap Opening in Aspartic Protease Catalysis † , ‡. *Biochemistry 43*, 4689–4695.

(15) Patel, S., Vuillard, L., Cleasby, A., Murray, C. W., and Yon, J. (2004) Apo and Inhibitor Complex Structures of BACE (β-secretase). *J. Mol. Biol. 343*, 407–416.

(16) Tóth, G., and Borics, A. (2006) Closing of the Flaps of HIV-1 Protease Induced by Substrate Binding: A Model of a Flap Closing Mechanism in Retroviral Aspartic Proteases. *Biochemistry 45*, 6606–6614.

(17) Barman, A., and Prabhakar, R. (2014) Computational Insights into Substrate and Site Specificities, Catalytic Mechanism, and Protonation States of the Catalytic Asp Dyad of β -Secretase. *Scientifica (Cairo). 2014*, 598728.

(18) Dunn, B. M. (2002) Structure and Mechanism of the Pepsin-Like Family of Aspartic Peptidases. *Chem. Rev. 102*, 4431–4458.

(19) Kohl, N. E., Emini, E. a, Schleif, W. a, Davis, L. J., Heimbach, J. C., Dixon, R. a, Scolnick, E. M., and Sigal, I. S. (1988) Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci. 85*, 4686–4690.

(20) Seelmeier, S., Schmidt, H., Turk, V., and von der Helm, K. (1988) Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A. *Proc. Natl. Acad. Sci. 85*, 6612–6616.

(21) Katz, R. (1994) The Retroviral Enzymes. *Annu. Rev. Biochem. 63*, 133–173.

(22) Czodrowski, P., Sotriffer, C. A., and Klebe, G. (2007) Atypical Protonation States in the Active Site of HIV-1 Protease: A Computational Study. *J. Chem. Inf. Model. 47*, 1590–1598.

(23) Scott, W. R., and Schiffer, C. A. (2000) Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. *Structure 8*, 1259–1265.

(24) Hornak, V., Okur, A., Rizzo, R. C., and Simmerling, C. (2006) HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A. 103*, 915–20.

(25) Hornak, V., and Simmerling, C. (2007) Targeting structural flexibility in HIV-1 protease inhibitor binding. *Drug Discov. Today 12*, 132–138.

(26) SPINELLI, S., LIU, Q., ALZARI, P., HIREL, P., and POLJAK, R. (1991) The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. *Biochimie 73*, 1391–1396.

(27) Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., Merson, J. R., Whittle, P. J., Danley, D. E., Geoghegan, K. F., Hawrylik, S. J., Lee, S. E., Scheld, K. G., and Hobart, P. M. (1989) X-ray analysis of HIV-1 proteinase at 2.7 Å resolution confirms structural homology among retroviral enzymes. *Nature 342*, 299–302.

(28) Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B., Baldwin, E., Weber, I., Selk, L., Clawson, L., Schneider, J., and Kent, S. (1989) Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science (80-.). 245*, 616–621.

(29) Ishima, R., Freedberg, D. I., Wang, Y.-X., Louis, J. M., and Torchia, D. a. (1999) Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function. *Structure 7*, 1047–S12.

(30) Freedberg, D. I., Ishima, R., Jacob, J., Wang, Y.-X., Kustanovich, I., Louis, J. M., and Torchia, D. a. (2009) Rapid structural fluctuations of the free HIV protease flaps in solution: Relationship to crystal structures and comparison with predictions of dynamics calculations. *Protein Sci. 11*, 221–232.

(31) Katoh, E., Louis, J. M., Yamazaki, T., Gronenborn, A. M., Torchia, D. a, and Ishima, R. (2003) A solution NMR study of the binding kinetics and the internal dynamics of an HIV-1 protease-substrate complex. *Protein Sci. 12*, 1376–1385.

(32) Nicholson, L. K., Yamazaki, T., Torchia, D. a, Grzesiek, S., Bax, A., Stahl, S. J., Kaufman, J. D., Wingfield, P. T., Lam, P. Y. ., Jadhav, P. K., Hodge, C. N., Domaille, P. J., and Chang, C.-H. (1995) Flexibility and function in HIV-1 protease. *Nat. Struct. Biol. 2*, 274–280.

(33) Wu, X., Öhrngren, P., Ekegren, J. K., Unge, J., Unge, T., Wallberg, H., Samuelsson, B., Hallberg, A., and Larhed, M. (2008) Two-Carbon-Elongated HIV-1 Protease Inhibitors with a Tertiary-Alcohol-Containing Transition-State Mimic ⊥. *J. Med. Chem. 51*, 1053–1057.

(34) Clavel, F., and Hance, A. J. (2004) HIV Drug Resistance. *N. Engl. J. Med. 350*, 1023–1035.

(35) Velazquez-Campoy, A., Muzammil, S., Ohtaka, H., Schon, A., Vega, S., and Freire, E. (2003) Structural and Thermodynamic Basis of Resistance to HIV-1 Protease Inhibition: Implications for Inhibitor Design. *Curr. Drug Target -Infectious Disord. 3*, 311–328.

(36) Muzammil, S., Ross, P., and Freire, E. (2003) A Major Role for a Set of Non-Active Site Mutations in the Development of HIV-1 Protease Drug Resistance †. *Biochemistry 42*, 631–638.

(37) Dam, E., Quercia, R., Glass, B., Descamps, D., Launay, O., Duval, X., Kräusslich, H.-G., Hance, A. J., and Clavel, F. (2009) Gag Mutations Strongly Contribute to HIV-1 Resistance to Protease Inhibitors in Highly Drug-Experienced Patients besides Compensating for Fitness Loss. *PLoS Pathog.* (Luban, J., Ed.) *5*, e1000345.

(38) Ohtaka, H., Schön, A., and Freire, E. (2003) Multidrug Resistance to HIV-1 Protease Inhibition Requires Cooperative Coupling between Distal Mutations †. *Biochemistry 42*, 13659–13666.

(39) Ahmed, S. M., Maguire, G. E. M., Kruger, H. G., and Govender, T. (2014) The impact of active site mutations of South African HIV PR on drug resistance: Insight from molecular dynamics simulations, binding free energy and per-residue footprints. *Chem. Biol. Drug Des. 83*, 472–81.

(40) Tóth, G., and Borics, A. (2006) Flap opening mechanism of HIV-1 protease. *J. Mol. Graph. Model. 24*, 465–474.

(41) Zhu, Z. W., Schuster, D. I., and Tuckerman, M. E. (2003) Molecular dynamics study of the connection between flap closing and binding of fullerene-based inhibitors of the HIV-1 protease. *Biochemistry 42*, 1326–1333.

(42) Heaslet, H., Rosenfeld, R., Giffin, M., Lin, Y.-C., Tam, K., Torbett, B. E., Elder, J. H., McRee, D. E., and Stout, C. D. (2007) Conformational flexibility in the flap domains of ligand-free HIV protease. *Acta Crystallogr. Sect. D-Biological Crystallogr. 63*, 866–875.

(43) Torbeev, V. Y., Raghuraman, H., Mandal, K., Senapati, S., Perozo, E., and Kent, S. B. H. (2009) Dynamics of "Flap" Structures in Three HIV-1 Protease/Inhibitor Complexes Probed by Total Chemical Synthesis and Pulse-EPR Spectroscopy. *J. Am. Chem. Soc. 131*, 884–885.

(44) Galiano, L., Bonora, M., and Fanucci, G. E. (2007) Interflap distances in HIV-1 protease determined by pulsed EPR measurements. *J. Am. Chem. Soc. 129*, 11004–+.

(45) Karthik, S., and Senapati, S. (2011) Dynamic flaps in HIV-1 protease adopt unique ordering at different stages in the catalytic cycle. *Proteins Struct. Funct. Bioinforma. 79*, 1830–1840.

(46) Meher, B. R., and Wang, Y. (2012) Interaction of I50V mutant and I50L/A71V double mutant HIV-protease with inhibitor TMC114 (darunavir): molecular dynamics simulation and binding free energy studies. *J. Phys. Chem. B 116*, 1884–900.

(47) Seibold, S. A., and Cukier, R. I. (2007) A molecular dynamics study comparing a wild-type with a multiple drug resistant HIV protease: Differences in flap and aspartate 25 cavity dimensions. *Proteins Struct. Funct. Bioinforma. 69*, 551–565.

(48) Chibi, B. (2012) COMPUTATIONAL STUDIES OF PENTACYCLOUNDECANE PEPTIDE BASED HIV-1 PROTEASE INHIBITORS. University of KwaZulu-Natal.

(49) (2015) WHO - World Health Organization.

(50) Ersmark, K., Samuelsson, B., and Hallberg, A. (2006) Plasmepsins as potential targets for new antimalarial therapy. *Med. Res. Rev. 26*, 626–66.

(51) White, N. J. (1993) Malaria parasites go ape. *Lancet 341*, 793.

(52) Banerjee, R., Liu, J., Beatty, W., Pelosof, L., Klemba, M., and Goldberg, D. E. (2002) Four plasmepsins are active in the *Plasmodium falciparum* food vacuole, including a protease with an active-site histidine. *Proc. Natl. Acad. Sci. 99*, 990–995.

(53) Rosenthal, P. J. (1998) Proteases of malaria parasites: new targets for chemotherapy. *Emerg Infect Dis 4*, 49–57.

(54) Coombs, G. H., Goldberg, D. E., Klemba, M., Berry, C., Kay, J., and Mottram, J. C. (2001) Aspartic proteases of *Plasmodium falciparum* and other parasitic protozoa as drug targets. *Trends Parasitol. 17*, 532–537.

(55) Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M., Nene, V., Shallom, S. J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M. A., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, L. M., Subramanian, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M., and Barrell, B. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature 419*, 498–511.

(56) Goldberg, D. E. (1993) Hemoglobin degradation in *Plasmodium*-infected red blood cells. *Semin Cell Biol 4*, 355–361.

(57) Francis, S. E., Sullivan Jr., D. J., and Goldberg, D. E. (1997) Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum*. *Annu Rev Microbiol 51*, 97–123.

(58) Sherman, I. W. (1977) Amino acid metabolism and protein synthesis in malarial parasites. *Bull World Heal. Organ 55*, 265–276.

(59) Gluzman, I. Y., Francis, S. E., Oksman, A., Smith, C. E., Duffin, K. L., and Goldberg, D. E. (1994) Order and specificity of the *Plasmodium falciparum* hemoglobin degradation pathway. *J Clin Invest 93*, 1602–1608.

(60) Liu, J., Istvan, E. S., Gluzman, I. Y., Gross, J., and Goldberg, D. E. (2006) *Plasmodium falciparum* ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems. *Proc Natl Acad Sci U S A 103*, 8840–8845.

(61) Dame, J. B., Yowell, C. a, Omara-Opyene, L., Carlton, J. M., Cooper, R. a, and Li, T. (2003) Plasmepsin 4, the food vacuole aspartic proteinase found in all Plasmodium spp. infecting man. *Mol. Biochem. Parasitol. 130*, 1–12.

(62) Omara-Opyene, A. L., Moura, P. A., Sulsona, C. R., Bonilla, J. A., Yowell, C. A., Fujioka, H., Fidock, D. A., and Dame, J. B. (2004) Genetic disruption of the *Plasmodium falciparum* digestive vacuole plasmepsins demonstrates their functional redundancy. *J Biol Chem 279*, 54088–54096.

(63) Klemba, M., and Goldberg, D. E. (2005) Characterization of plasmepsin V, a membrane-bound aspartic protease homolog in the endoplasmic reticulum of *Plasmodium falciparum*. *Mol Biochem Parasitol 143*, 183–191.

(64) Boddey, J. A., Hodder, A. N., Gunther, S., Gilson, P. R., Patsiouras, H., Kapp, E. A., Pearce, J. A., de Koning-Ward, T. F., Simpson, R. J., Crabb, B. S., and Cowman, A. F. (2010) An aspartyl protease directs malaria effector proteins to the host cell. *Nature 463*, 627–631.

(65) Russo, I., Babbitt, S., Muralidharan, V., Butler, T., Oksman, A., and Goldberg, D. E. (2010) Plasmepsin V licenses *Plasmodium* proteins for export into the host erythrocyte. *Nature 463*, 632–6.

(66) Hodder, A. N., Sleebs, B. E., Czabotar, P. E., Gazdik, M., Xu, Y., O'Neill, M. T., Lopaticki, S., Nebl, T., Triglia, T., Smith, B. J., Lowes, K., Boddey, J. A., and Cowman, A. F. (2015) Structural basis for plasmepsin V inhibition that blocks export of malaria proteins to human erythrocytes. *Nat. Struct. Mol. Biol. 22*, 590–596.

(67) Marti, M., Good, R. T., Rug, M., Knuepfer, E., and Cowman, A. F. (2004) Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science (80-.). 306*, 1930–1933.

(68) Hiller, N. L., Bhattacharjee, S., van Ooij, C., Liolios, K., Harrison, T., Lopez-Estrano, C., and Haldar, K. (2004) A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science (80-.). 306*, 1934–1937.

(69) Asojo, O. a., Gulnik, S. V., Afonina, E., Yu, B., Ellman, J. a., Haque, T. S., and Silva, A. M. (2003) Novel Uncomplexed and Complexed Structures of Plasmepsin II, an Aspartic Protease from *Plasmodium falciparum*. *J. Mol. Biol. 327*, 173–181.

(70) Karubiu, W., Bhakat, S., McGillewie, L., and Soliman, M. E. (2015) Flap dynamics of plasmepsin proteases: insight into proposed parameters and molecular dynamics. *Mol Biosyst 11*, 1061–1066.

(71) Asojo, O. A., Afonina, E., Gulnik, S. V, Yu, B., Erickson, J. W., Randad, R., Medjahed, D., and Silva, A. M. (2002) Structures of Ser205 mutant plasmepsin II from *Plasmodium falciparum* at 1.8 Å in complex with the inhibitors rs367 and rs370. *Acta Crystallogr. Sect. D Biol. Crystallogr. 58*, 2001–2008.

(72) Bhaumik, P., Horimoto, Y., Xiao, H., Miura, T., Hidaka, K., Kiso, Y., Wlodawer, A., Yada, R. Y., and Gustchina, A. (2011) Crystal structures of the free and inhibited forms of plasmepsin I (PMI) from *Plasmodium falciparum*. *J. Struct. Biol. 175*, 73–84.

(73) Westling, J., Cipullo, P., Hung, S. H., Saft, H., Dame, J. B., and Dunn, B. M. (1999) Active site specificity of plasmepsin II. *Protein Sci. 8*, 2001–9.

(74) McGillewie, L., and Soliman, M. E. (2015) Flap flexibility amongst plasmepsins I, II, III, IV, and V: Sequence, structural, and molecular dynamics analyses. *Proteins Struct. Funct. Bioinforma. 83*, 1693–1705.

(75) Tice, C. M. (2006) Renin Inhibitors, pp 155–167.

(76) Paul, M. (2006) Physiology of Local Renin-Angiotensin Systems. *Physiol. Rev. 86*, 747–803.

(77) Wood, J. M., Stanton, J. L., and Hofbauer, K. G. (1987) Inhibitors of renin as potential therapeutic agents. *J Enzym. Inhib 1*, 169–185.

(78) Powell, N. a., Ciske, F. L., Cai, C., Holsworth, D. D., Mennen, K., Van Huis, C. a., Jalaie, M., Day, J., Mastronardi, M., and McConnell, P. (2007) Rational design of 6-(2,4-diaminopyrimidinyl)-1,4-benzoxazin-3-ones as small molecule renin inhibitors. *Bioorg. Med. Chem. 15*, 5912–5949.

(79) Hedner, T., Sun, X., Junggren, I. L., Pettersson, A., and Edvinsson, L. (1992) Peptides as targets for antihypertensive drug development. *J Hypertens Suppl 10*, S121–32.

(80) Cody, R. J. (1994) The clinical potential of renin inhibitors and angiotensin antagonists. *Drugs 47*, 586–598.

(81) Staessen, J. a., Li, Y., and Richart, T. (2006) Oral renin inhibitors. *Lancet 368*, 1449–1456.

(82) Fisher, N. D. L., and Hollenberg, N. K. (2005) Renin inhibition: what are the therapeutic opportunities? *J. Am. Soc. Nephrol. 16*, 592–9.

(83) Azizi, M., Webb, R., Nussberger, J., and Hollenberg, N. K. (2006) Renin inhibition with aliskiren: where are we now, and where are we going? *J. Hypertens. 24*, 243–56.

(84) Brown, M. J. (2008) Aliskiren. *Circulation 118*, 773–84.

(85) Tzoupis, H., Leonis, G., Megariotis, G., Supuran, C. T., Mavromoustakos, T., and Papadopoulos, M. G. (2012) Dual Inhibitors for Aspartic Proteases HIV-1 PR and Renin: Advancements in AIDS–Hypertension–Diabetes Linkage via Molecular Dynamics, Inhibition Assays, and Binding Free Energy Calculations. *J. Med. Chem. 55*, 5784–5796.

(86) Scheiper, B., Matter, H., Steinhagen, H., Stilz, U., Böcskei, Z., Fleury, V., and McCort, G. (2010) Discovery and optimization of a new class of potent and non-chiral indole-3-carboxamide-based renin inhibitors. *Bioorg. Med. Chem. Lett. 20*, 6268–6272.

(87) Politi, A., Leonis, G., Tzoupis, H., Ntountaniotis, D., Papadopoulos, M. G., Grdadolnik, S. G., and Mavromoustakos, T. (2011) Conformational Properties and Energetic Analysis of Aliskiren in Solution and Receptor Site. *Mol. Inform. 30*, 973–985.

(88) Rahuel, J., Priestle, J. P., and Grütter, M. G. (1991) The crystal structures of recombinant glycosylated human renin alone and in complex with a transition state analog inhibitor. *J. Struct. Biol. 107*, 227–36.

(89) Rahuel, J., Rasetti, V., Maibaum, J., Rüeger, H., Göschke, R., Cohen, N.-C., Stutz, S., Cumin, F., Fuhrer, W., Wood, J., and Grütter, M. (2000) Structure-based drug design: the discovery of novel nonpeptide orally active inhibitors of human renin. *Chem. Biol. 7*, 493–504.

(90) Baldwin, E. T., Bhat, T. N., Gulnik, S., Hosur, M. V, Sowder, R. C., Cachau, R. E., Collins, J., Silva, A. M., and Erickson, J. W. (1993) Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design. *Proc. Natl. Acad. Sci. U. S. A. 90*, 6796–800.

(91) Liaudet-Coopman, E., Beaujouin, M., Derocq, D., Garcia, M., Glondu-Lassis, M., Laurent-Matha, V., Prébois, C., Rochefort, H., and Vignon, F. (2006) Cathepsin D: newly discovered functions of a long-standing aspartic protease in cancer and apoptosis. *Cancer Lett. 237*, 167–179.

(92) De Duve, C. (1983) Lysosomes revisited. *Eur. J. Biochem. 137*, 391–7.

(93) Garcia, M., Derocq, D., Pujol, P., and Rochefort, H. (1990) Overexpression of transfected cathepsin D in transformed cells increases their malignant phenotype and metastatic potency. *Oncogene 5*, 1809–1814.

(94) Liaudet, E., Garcia, M., and Rochefort, H. (1994) Cathepsin D maturation and its stimulatory effect on metastasis are prevented by addition of KDEL retention signal. *Oncogene 9*, 1145–1154.

(95) Liaudet, E., Derocq, D., Rochefort, H., and Garcia, M. (1995) Transfected cathepsin D stimulates high density cancer cell growth by inactivating secreted growth inhibitors. *Cell growth Differ. Mol. Biol. J. Am. Assoc. Cancer Res. 6*, 1045–1052.

(96) Glondu, M., Coopman, P., Laurent-Matha, V., Garcia, M., Rochefort, H., and Liaudet-Coopman, E. (2001) A mutated cathepsin-D devoid of its catalytic activity stimulates the growth of cancer cells. *Oncogene 20*, 6920–9.

(97) Thies, W., and Bleiler, L. (2013) 2013 Alzheimer's disease facts and figures. *Alzheimers. Dement. 9*, 208–45.

(98) Suo, Z., Humphrey, J., Kundtz, A., Sethi, F., Placzek, A., Crawford, F., and Mullan, M. (1998) Soluble Alzheimers ??-amyloid constricts the cerebral vasculature in vivo. *Neurosci. Lett. 257*, 77–80.

(99) Pillot, T., Drouet, B., Queillé, S., Labeur, C., Vandekerckhove, J., Rosseneu, M., Pinçon-Raymond, M., and Chambaz, J. (1999) The nonfibrillar amyloid β-peptide induces apoptotic neuronal cell death: Involvement of its C-terminal fusogenic domain. *J. Neurochem. 73*, 1626–1634.

(100) Vassar, R., Bennett, B. D., Babu-Khan, S., Kahn, S., Mendiaz, E. a, Denis, P., Teplow, D. B., Ross, S., Amarante, P., Loeloff, R., Luo, Y., Fisher, S., Fuller, J., Edenson, S., Lile, J., Jarosinski, M. a, Biere, a L., Curran, E., Burgess, T., Louis, J. C., Collins, F., Treanor, J., Rogers, G., and Citron, M. (1999) Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science 286*, 735–741.

(101) Cai, H., Wang, Y., McCarthy, D., Wen, H., Borchelt, D. R., Price, D. L., and Wong, P. C. (2001) BACE1 is the major beta-secretase for generation of Abeta peptides by neurons. *Nat. Neurosci. 4*, 233–234.

(102) Sinha, S., Anderson, J. P., Barbour, R., Basi, G. S., Caccavello, R., Davis, D., Doan, M., Dovey, H. F., Frigon, N., Hong, J., Jacobson-Croak, K., Jewett, N., Keim, P., Knops, J., Lieberburg, I., Power, M., Tan, H., Tatsuno, G., Tung, J., Schenk, D., Seubert, P., Suomensaari, S. M., Wang, S., Walker, D., Zhao, J., McConlogue, L., and John, V. (1999) Purification and cloning of amyloid precursor protein beta-secretase from human brain. *Nature 402*, 537–540.

(103) Yan, R., Bienkowski, M. J., Shuck, M. E., Miao, H., Tory, M. C., Pauley, a M., Brashier, J. R., Stratman, N. C., Mathews, W. R., Buhl, a E., Carter, D. B., Tomasselli, a G., Parodi, L. a, Heinrikson, R. L., and Gurney, M. E. (1999) Membrane-anchored aspartyl protease with Alzheimer's disease beta-secretase activity. *Nature 402*, 533–537.

(104) Zhang, Y., Thompson, R., Zhang, H., and Xu, H. (2011) APP processing in Alzheimer's disease. *Mol. Brain 4*, 3.

(105) Luo, Y., Bolon, B., Kahn, S., Bennett, B. D., Babu-Khan, S., Denis, P., Fan, W., Kha, H., Zhang, J., Gong, Y., Martin, L., Louis, J.-C., Yan, Q., Richards, W. G., Citron, M., and Vassar, R. (2001) Mice deficient in BACE1, the Alzheimer's β-secretase, have normal phenotype and abolished β-amyloid generation. *Nat. Neurosci. 4*, 231–232.

(106) Bennett, B. D., Babu-Khan, S., Loeloff, R., Louis, J. C., Curran, E., Citron, M., and Vassar, R. (2000) Expression analysis of BACE2 in brain and peripheral tissues. *J. Biol. Chem. 275*, 20647–20651.

(107) Sun, X., Wang, Y., Qing, H., Christensen, M. a, Liu, Y., Zhou, W., Tong, Y., Xiao, C., Huang, Y., Zhang, S., Liu, X., and Song, W. (2005) Distinct transcriptional regulation and function of the human BACE2 and BACE1 genes. *FASEB J. 19*, 739–749.

(108) Yan, R., Munzner, J. B., Shuck, M. E., and Bienkowski, M. J. (2001) BACE2 Functions as an Alternative -Secretase in Cells. *J. Biol. Chem. 276*, 34019–34027.

(109) Ahmed, R. R., Holler, C. J., Webb, R. L., Li, F., Beckett, T. L., and Murphy, M. P. (2010) BACE1 and BACE2 enzymatic activities in Alzheimer's disease. *J. Neurochem. 112*, 1045–1053.

(110) Lin, X., Koelsch, G., Wu, S., Downs, D., Dashti, A., and Tang, J. (2000) Human aspartic protease memapsin 2 cleaves the beta-secretase site of beta-amyloid precursor protein. *Proc. Natl. Acad. Sci. U. S. A. 97*, 1456–1460.

(111) Hong, L., Turner, R. T., Koelsch, G., Shin, D., Ghosh, A. K., and Tang, J. (2002) Crystal structure of memapsin 2 (β-secretase) in complex with an inhibitor OM00-3. *Biochemistry 41*, 10963–10967.

(112) Hong, L., and Tang, J. (2004) Flap position of free memapsin 2 (beta-secretase), a model for flap opening in aspartic protease catalysis. *Biochemistry 43*, 4689–95.

(113) Hong, L., Koelsch, G., Lin, X., Wu, S., Terzyan, S., Ghosh, a K., Zhang, X. C., and Tang, J. (2000) Structure of the protease domain of memapsin 2 (beta-secretase) complexed with inhibitor. *Science 290*, 150–153.

(114) Shimizu, H., Tosaki, A., Kaneko, K., Hisano, T., Sakurai, T., and Nukina, N. (2008) Crystal structure of an active form of BACE1, an enzyme responsible for amyloid beta protein production. *Mol. Cell. Biol. 28*, 3663–3671.

(115) Kumalo, H. M., Bhakat, S., and Soliman, M. E. (2015) Investigation of flap flexibility of β-secretase using molecular dynamic simulations. *J. Biomol. Struct. Dyn.* 1–12.

(116) Xu, Y., Li, M., Greenblatt, H., Chen, W., Paz, A., Dym, O., Peleg, Y., Chen, T., Shen, X., He, J., Jiang, H., Silman, I., and Sussman, J. L. (2012) Flexibility of the flap in the active site of BACE1 as revealed by crystal structures and molecular dynamics simulations. *Acta Crystallogr. Sect. D Biol. Crystallogr. 68*, 13–25.

(117) Lv, Z., Chu, Y., and Wang, Y. (2015) HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV. AIDS. (Auckl). 7*, 95–104.

(118) Mesecar, a D., Stoddard, B. L., and Koshland, D. E. (1997) Orbital steering in the catalytic power of enzymes: small structural changes with large catalytic consequences. *Science 277*, 202–206.

(119) Falke, J. J. (2002) Enzymology. A moving story. *Science 295*, 1480–1.

(120) Benkovic, S. J., and Hammes-Schiffer, S. (2003) A perspective on enzyme catalysis. *Science (80-.). 301*, 1196–202.

(121) Bhabha, G., Lee, J., Ekiert, D. C., Gam, J., Wilson, I. a, Dyson, H. J., Benkovic, S. J., and Wright, P. E. (2011) A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science 332*, 234–238.

CHAPTER 8

Conclusion

The present study sought to understand the dynamic behaviour of the flap region of plasmepsins, and how these regions not only affect the overall structure but regulate the binding landscape. The flap region is a structural feature unique to aspartic proteases; in retroviral proteases two identical flaps partially cover the active site, whereas eukaryotic proteases have only a single hairpin loop which lies perpendicularly over the active site. The role flap motions and dynamics play in enzyme function is well documented for HIV protease; these regions are so important that alterations in the in the sequence of residues found in the flaps induced by mutations significantly alter the sensitivity to HIV protease inhibitors. Mutations in the flap region potentially leads to the development of resistance. On the flip side, inhibiting the dynamic motions and flexibility of the flap region leads to the inactivation of HIV protease.

Plasmepsins have been identified as targets to develop novel drugs to treat malaria. In the design of potent plasmepsin inhibitors it is crucial to understand and quantify flap dynamics, due to the unique flap region and its essential role in enzyme function. The catalytic aspartic acid dyad is highly conserved in all plasmepsins (with the exception of HAP), therefore residues outside of the active site regulate each plasmepsin's unique substrate specificity and individual response to inhibitors. The impact of sequence heterogeneity is highlighted in vacuolar plasmepsins, *P. falciparum* has four plasmepsins active in the food vacuole with varying affinities for hemoglobin. The differences in affinity is regulated, in part, by the residues making up the flap and flexible loop region. Between the vacuolar plasmepsins, the sequence of the flap region and flexible loop region vary significantly. These changes alter the dynamics and flexibility of not only the flap region, but the enzyme as a whole. It is believed that the vacuolar plasmepsin of *P. falciparum* resulted from gene duplication of an orthologue

of PlmIV found in all *Plasmodium* species. In the comparison of apo vacuolar plasmepsin, we observed PlmIV to be the most flexible and dynamic of all the vacuolar plasmepsins, capable of the widest range of motion, reaching a maximum opening distance 19.63 Å. This was accompanied by an aggressive twisting motion of the active site as reflected by changes in the dihedral angles, suggesting that flap movements not only opens the active site as the flap tip moves away from the active site but also induces large structural changes in the dynamics of the enzyme. The increased flexibility of PlmIV can be attributed to the glycine residue situated at the flap tip, which offers little to no steric hindrance to the movement of the flap. This correlates well to PlmIV's biological function, and bearing in mind it is the only aspartic protease in other *Plasmodium* species therefore flexibility in the flap region ensures hemoglobin irrespective of shape or size is recognized and accommodated by the active site.

Plasmepsin V is a promising target, although it is structurally similar to other plasmepsins it only shares a mere 17% homology with PlmII and is very unique. The membrane bound aspartic protease is highly conserved throughout all malaria species and is essential to the survival of the parasite. As with other plasmepsins and aspartic proteases, PlmV has a flap region partially covering the active site. Due to the biological function of PlmV, the active site is highly flexible as to accommodate numerous different PEXEL tagged proteins in the active site. The mobility of the flap region enable PlmV to accommodate proteins of various shapes and sizes, while the disulfide bonds between adjacent cysteine residues stabalises the enzyme. In an apo conformation the active site of PlmV transitions between open and semi open conformations; as the flap moves freely away from the flexible loop and active site. Upon inhibitor binding, WEHI-842, the flap folds down into the active site allowing the formation of hydrogen bonds between the catalytic dyad holding the inhibitor tightly in the active site.

Although the inhibitor does not directly bind to the flap tip, it binds to the adjacent glutamic acid which stabilizes and reduces the flexibility of the flap region and the active site.

The parameters proposed in the present study accurately accounts for and quantifies the dynamic behaviour of the flap region. And will provide valuable insights into the development of plasmepsin inhibitors.

Future studies

The computational studies and parameters used in the present study is a cost effective and efficient tool that can be implemented in drug design and discovery. Going forward, implementing QM or hybrid QM/MM methodologies would provide a better understanding of the dynamic motions of these regions, give a more detailed account of the molecular interactions and bond formations with respect to the distribution of motion. Further refinement of the WEHI-842 scaffold, or a scaffold that reduces the flap flexibility to a greater extent is a possibility. To computationally search for inhibitor with dual activity against aspartic and cysteine proteases using virtual screening protocols, thereby reducing the number of compounds. Using key interacting residues to build a pharmacophore model based on the binding fee energy contribution of each residue to search for potential inhibitors of the plasmepsin family, more specifically targeted towards PlmV. Although 50 ns was sufficient for the present study, it would be worthwhile running simulations for longer time frames e.g. 200ns and to further assess and validate the parameters proposed.