



MONASH University

**Investigation of the protein dynamics of the
Plasmodium falciparum M1 and M17 aminopeptidases**

Wei Yang
B.S.c (Biotechnology)
M.S.c (Marine Biology)

A thesis submitted for the degree of *Doctor of Philosophy* at
Monash University in 2017

Department of Microbiology
School of Biomedical Sciences

Copyright notice

© Wei Yang 2017. Except as provided in the Copyright Act 1968, this thesis may not be reproduced in any form without the written permission of the author.

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

Abstract	8
Declaration	10
Publications during enrolment	11
Thesis declaration (including published works)	13
Acknowledgements	15
Abbreviations	18
CHAPTER ONE: INTRODUCTION	19
1.1. Malaria.....	20
1.1.1 Malaria remains a global health problem.....	20
1.1.2 Therapeutics available for the treatment of malaria.....	21
1.2. Aminopeptidases as potential drug targets for the design of new antimalarials.....	22
1.2.1. Proteases as potential drug targets.....	22
1.2.2. Aminopeptidases are involved in Hb digestion in intra-erythrocytic <i>P. falciparum</i> parasites.....	25
1.2.3. The <i>P. falciparum</i> neutral aminopeptidases as potential targets for new antimalarials.....	28
1.3. The M1 aminopeptidase family	29
1.3.1. The role of the M1 aminopeptidases.....	29
1.3.2. The catalytic mechanism of the M1 aminopeptidases.....	30
1.3.3. The structure of the M1 aminopeptidases.....	32
1.3.4. Some members of the M1 aminopeptidase superfamily can exist in an open or closed form.....	34
1.3.5. <i>PfA-M1</i>	35
1.4. The M17 family of Leucine Aminopeptidases.....	38
1.4.1. The role of the M17 aminopeptidases.....	38
1.4.2. The catalytic mechanism of the M17 aminopeptidases.....	39
1.4.3. The structure of the M17 aminopeptidases.....	42
1.4.4. <i>PfA-M17</i>	42

1.5. Molecular dynamics of proteins.....	44
1.5.1. MD simulation of metalloproteins	45
1.6. Aims and Objectives	46
CHAPTER TWO: METHODS	48
2.1. Computational Methods	49
2.1.1. Force field parameterization and evaluation	49
2.1.2. Bestatin parameters	51
2.1.3. Input set up for molecular dynamics simulations	51
2.1.4. Molecular dynamics protocol	51
2.1.5. RMSD calculations.....	52
2.1.6. RMSD clustering analysis	53
2.1.7. RMSF calculations	53
2.1.8. Radial distribution function between zinc ions and water.....	53
2.1.9. Principle Component Analysis (PCA)	54
2.1.10. Calculating free energy landscapes	54
2.1.11. Molecular modelling of <i>PfA</i> -M1-Hb6	55
2.1.12. Steered molecular dynamics (SMD) simulations	56
2.1.13. Calculating the external force and work of SMD	57
2.1.14. MD analysis	57
2.2. Experimental Methods	58
2.2.1. Site-directed mutagenesis.....	58
2.2.2. Protein production	58
2.2.3. Enzyme kinetics	59
2.2.4. Aminopeptidase activity against Hb6 substrate	59
2.2.5. Protein Crystallography.....	60
2.2.6. Structure refinement.....	60
CHAPTER THREE: INVESTIGATING THE PROTEIN DYNAMICS OF <i>PfA</i>-M1.....	61
3.1. Introduction.....	62
3.2. Results.....	63

3.2.1 MD simulations	63
3.2.2. Nucleophilic water behavior throughout simulation	64
3.2.3. Dynamic cross-correlation matrix of <i>PfA</i> -M1 simulations.....	65
3.2.4. Principle component analysis of <i>PfA</i> -M1 simulations.....	68
3.2.5. Mapping the movement and free energy landscape of run 3.....	70
3.2.6. Prediction of residues involved in the hinging motion of <i>PfA</i> -M1.....	74
3.2.7. Testing if residues M462 and A453 contribute to the dynamics of <i>PfA</i> -M1.	75
3.3. Discussion	81
3.4. Conclusions.....	84
CHAPTER FOUR: DETERMINING THE STRUCTURAL FEATURES OF <i>PfA</i>-M1	
BOUND TO A PEPTIDE SUBSTRATE.....	86
4.1. Introduction.....	87
4.2. Results.....	88
4.2.1. Hb6 is a substrate of <i>PfA</i> -M1.....	88
4.2.2. Attempts to produce a co-crystal structure of <i>PfA</i> -M1 bound to Hb6.....	89
4.2.3. Modelling the <i>PfA</i> -M1-Hb6 complex.....	93
4.2.4. Steered MD study suggests N-terminal access point is preferred by Hb6.....	95
4.3. Discussion	98
4.4. Conclusions	100
CHAPTER FIVE: INVESTIGATING THE DYNAMICS OF <i>PfA</i>-M17	102
5.1. Introduction.....	103
5.2. Results.....	103
5.2.1. Nomenclature used to describe the structure of <i>PfA</i> -M17 in this study.....	103
5.2.2. MD simulations of the <i>PfA</i> -M17 monomer and hexamer.....	107
5.2.3. The hexamer protects the active site environment	110
5.2.4. Principle component analysis of the MD simulations	112
5.2.5. The essential dynamics of the <i>PfA</i> -M17 monomer	114
5.2.6. The essential dynamics of the <i>PfA</i> -M17 hexamer.....	117

5.2.7. Hexamer movement is achieved by independent trimer movements.....	120
5.2.8. <i>PfA</i> -M17 has an entrance to the active site cavity that is occluded by cooperative movements of L9 and L3	122
5.2.9. Cooperativity in the conserved metal binding sites of the hexamer	125
5.3. Discussion	126
5.3.1. Why does <i>PfA</i> -M17 form a hexamer?.....	126
5.3.2. Do trimers regulate the activity of <i>PfA</i> -M17?	127
5.3.3. A model for the catalytic mechanism of the M17 aminopeptidase superfamily.	128
5.4. Conclusions	129

CHAPTER SIX: PROTEIN DYNAMICS OF THE CLAN MA, FAMILY M1

AMINOPEPTIDASES	130
6.1. Introduction.....	131
6.2. Results.....	131
6.2.1. Selection of representative M1 aminopeptidase structure for analysis.....	131
6.2.2. MD simulations of selected M1 aminopeptidase structures	137
6.2.3. PCA of the simulated M1 systems.....	141
6.2.3.1. TIFF3.....	141
6.2.3.2. <i>EcAPN</i>	146
6.2.3.3. <i>PfA</i> -M1.....	148
6.2.3.4. <i>hAPA</i>	148
6.2.3.5. <i>hAPN</i>	149
6.2.3.6. <i>hERAP1</i>	151
6.2.3.7. <i>pAPN</i>	155
6.2.4. The mammalian M1 aminopeptidases dimers might possess the same opening movement as were found in monomers	157
6.3. Discussion	159

6.3.1. Six movements can summarize the dynamics that were commonly observed in the M1 aminopeptidase representatives.....	160
6.3.2. The unique movements among the M1 aminopeptidase candidates.	163
6.4. Conclusions	166
CHAPTER SEVEN: MAPPING THE PATHWAY AND DYNAMICS OF BESTATIN INHIBITION OF THE <i>Plasmodium falciparum</i> M1 AMINOPEPTIDASE, PfA-M1.	167
CHAPTER EIGHT: DISCUSSION	203
8.1. The hybrid force field for the M1 and M17 aminopeptidases	204
8.2. The protein dynamics of the clan MA, M1 aminopeptidase superfamily	206
8.3. Insight into the dynamics in PfA-M1 can guide future antimalarial drug design	209
8.4. The protein dynamics of PfA-M17, implications for the M17 aminopeptidase superfamily	212
8.5. Conclusions	214
REFERENCES	215

Drinkwater *et al.*, (2017) *FEBS J.* 28(10):1473-1488 DOI: 10.1111/febs.14009

Yang *et al.*, (2017) *J Biomol Struct Dyn.* doi.org/10.1080/07391102.2017.1364669

Abstract

Proteins of the M1 and M17 aminopeptidase superfamilies play important roles in cellular metabolism. Controlling the activity of these enzymes can modulate chronic and infectious diseases, such as modulation of hypertension, control of tumor growth and treatment of malaria. The ability to produce selective and potent inhibitors for different members of one protein superfamily relies on the knowledge of the mechanism of action, structure of the target and their individual protein dynamics. Each can be exploited for future inhibitor design aimed to produce therapeutic agents for human disease. Two of the representative proteins within M1 and M17 aminopeptidase superfamilies, *PfA-M1* and *PfA-M17* are the focus of this thesis. Both enzymes are metallo-exopeptidases, are essential for the survival of the malaria parasite, *Plasmodium falciparum*, and are attractive new antimalarial drug targets. The dynamics of each protein and how this relates to their enzymatic function remains unknown. Molecular dynamics (MD) simulation is a computational technique used to understand the dynamics of proteins. However, *PfA-M1* and *PfA-M17* are metallo-proteins and have essential zinc ions in their active site. This presents a significant challenge for their simulation by MD techniques due to the lack of appropriate metal parameters for the simulation. In this thesis, new parameters for the MD simulation of M1 and M17 aminopeptidase were designed and implemented. The parameters produced from a hybrid force field (FF) model enabled MD simulations for both M1 and M17 aminopeptidase systems. Classical all-atom MD simulations on seven M1 aminopeptidase enzymes (including *PfA-M1*) were performed in this study, which address the protein dynamics within the M1 aminopeptidase superfamily. Simulation of the inhibitor-bound and substrate bound *PfA-M1* sheds light on the intramolecular pathways that the inhibitor or substrate may take to enter the active site and predicts for the first time, the P2 to P5 substrate pockets of *PfA-M1*. Simulation of the much larger M17 aminopeptidase, *PfA-M17*, identifies reasons as to why the protease is assembled into a homohexamer, showing that the hexamer stabilises the catalytic machinery, including the substrate pockets and catalytic water molecules. Dissection of the hexamer suggests that the dimer of trimers that forms the hexamer is necessary to control substrate movement and that the individual

trimers communicate to allow substrate entry to the interior catalytic cavity of the hexamer. The analysis of the protein dynamics of M1 and M17 aminopeptidase superfamilies presented in this thesis has provided fundamental knowledge on how dynamics affects both catalysis and regulation of activity of the two enzymes and provided key information that will be of broad benefit to future drug discovery programs.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature: Wei Yang 楊威

Print Name: Wei Yang

Date: 30/04/2018

PUBLICATIONS DURING ENROLMENT

* *Related to doctoral research*

- 1) ***Wei Yang**, Blake T. Riley, Xiangyun Lei, Benjamin T. Porebski, Itamar Kass, Ashley M. Buckle, Sheena McGowan. "Generation of AMBER force field parameters for zinc centres of M1 and M17 family aminopeptidases." *Journal of Biomolecular Structure and Dynamics* 2017. DOI: 10.1080/07391102.2017.1364669 (Publication of parts of Chapter 2)
- 2) Drinkwater Nyssa, Lee Jisook, **Yang Wei**, Malcolm Tess R, McGowan Sheena. "M1 Aminopeptidases as drug targets: broad applications or therapeutic niche?" *The FEBS Journal*, 2017. DOI: 10.1111/febs.14009. (Publication of parts of Chapter 1)
- 3) Wang Fangfang, **Wei Yang**, Shi Yonghui, Le Guowei. "Probing the structural requirements for thyroid hormone receptor inhibitory activity of sulfonynitrophenylthiazoles (SNPTs) using 2D-QSAR and 3D-QSAR approaches". *Medicinal Chemistry Research*, 2016, Page 1-17. DOI: 10.1007/s00044-016-1751-3
- 4) Wang Fangfang, **Wei Yang**, Shi Yonghui, Le Guowei. "In silico study on β -aminoketone derivatives as thyroid hormone receptor inhibitors: a combined 3D-QSAR and molecular docking study." *Journal of Biomolecular Structure and Dynamics*, 2016, Volume 34, Issue 12, Pages 2619-2631. DOI: 10.1080/07391102.2015.1124806
- 5) Wang Fangfang, **Wei Yang**, Shi Yonghui, Le Guowei. "Structure-Based Approach for the Study of Thyroid Hormone Receptor Binding Affinity and Subtype Selectivity." *Journal of Biomolecular Structure and Dynamics*, 2016, Volume 34, Issue 10, Pages 2251-2267. DOI: 10.1080/07391102.2015.1113384
- 6) Wang Fangfang, **Wei Yang**, Shi Yonghui, Le Guowei. "Molecular determinants of thyroid hormone receptor selectivity in a series of phosphonic acid derivatives: 3D-QSAR analysis and molecular docking." *Chemico-biological interactions*, Volume 240, Pages 324-335. DOI: 10.1016/j.cbi.2015.09.008
- 7) San Sui Lim, **Wei Yang**, Bankala Krishnarjuna, Komagal Kannan Sivaraman, Indu R. Chandrashekar, Itamar Kass, Christopher A. MacRaid, Shane M. Devine, Cael O. Debono, Robin F. Anders, Martin J. Scanlon, Peter J. Scammells, Raymond S. Norton, Sheena McGowan. "Structure and Dynamics of Apical Membrane Antigen 1 from Plasmodium falciparum FVO." *Biochemistry*, 2014, Vol. 53, Issue 46, 7310-732. DOI: 10.1021/bi5012089
- 8) Wang Fangfang, **Wei Yang**, Shi Yonghui, Le Guowei. "3D-QSAR, molecular docking and molecular dynamics studies of a series of ROR γ t inhibitors." *Journal of Biomolecular Structure and Dynamics*, 2014, Pages 1-12. DOI: 10.1080/07391102.2014.980321 **JOINT FIRST AUTHOR**
- 9) Wang Fangfang, **Wei Yang**, Shi, Yonghui, Le Guowei. "Structural analysis of selective agonists of thyroid hormone receptor β using 3D-QSAR and molecular docking."

Manuscripts under review

Wei Yang, Blake Riley, Ben Porebski, Itamar Kass, Ashley M. Buckle, Sheena McGowan “Mapping the pathway and dynamics of bestatin inhibition of the Plasmodium falciparum M1 aminopeptidase, PfA-M1”. (*Under review, PLOS Computational Biology*, 10 Nov, 2017). (Chapter 7)

Manuscripts in Preparation

Nyssa Drinkwater[#], **Wei Yang**[#], Komagal Kannan Sivaraman, Blake Riley, Itamar Kass, Ashley Buckle, Sheena McGowan. “Cooperation and catalysis: a novel description of the dynamic mechanism of hexameric M17 aminopeptidases. #Joint first authors (Chapter 5)

Wei Yang[#], Nyssa Drinkwater[#], Mauricio Costa, Isaac Kresse, Tess Malcolm, Komagal Kannan Sivaraman, Blake T. Riley, Itamar Kass, David Perahia, Thomas Shaffee, Ganesh Anand, Ashley. M. Buckle, Sheena McGowan. “Aminopeptidase N: domain motions and species complexity in the M1 clan” *Joint first authors. (Chapter 3 and 6)

THESIS INCLUDING PUBLISHED WORKS DECLARATION

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes (2) original papers published in peer reviewed journals and (1) submitted publication. The core theme of the thesis is dynamics, structure and function of the M1 and M17 aminopeptidases. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the Department of Microbiology under the supervision of Sheena McGowan.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

In the case of chapters 1, 2 and 7, my contribution to the work involved the following:

Thesis Chapter	Publication Title	Status (published, in press, accepted or returned for revision, submitted)	Nature and % of student contribution	Co-author name(s) Nature and % of Co-author's contribution*	Co-author(s), Monash student Y/N*
1	M1 aminopeptidases as drug targets: broad applications or therapeutic niche?	Published	10 % Sections of Chapter 1 <i>Introduction</i> were used in this broad review.	1) N Drinkwater 35 % 2) J Lee (10%) 3) W Yang (10%) 4) TR. Malcolm (10%) 5) S McGowan 35%	N Y Y Y N
2	Generation of AMBER force field parameters for zinc centres of M1 and M17 family aminopeptidases.	Published	75%. Concept, experiments and data, wrote manuscript.	1) W Yang (75%) 2) BT Riley, data analysis, (2.5%) 3) X Lei, data analysis, (2.5%) 4) BT Porebski, draft manuscript (2.5%) 5) I Kass, training (2.5%) 6) AM Buckle draft manuscript (5%) 7) S McGowan, concept, data & manuscript (10%)	Y Y N N N N N
7	Mapping the pathway and dynamics of bestatin inhibition of the <i>Plasmodium falciparum</i> M1 aminopeptidase, PfA-M1.	Submitted	77.5%. Concept, experiments and data, wrote manuscript.	1) W Yang (77.5%) 2) BT Riley, data analysis, (2.5%) 3) BT Porebski, draft manuscript (2.5%) 4) I Kass, training (2.5%) 5) AM Buckle draft manuscript (5%) 6) S McGowan, concept, data & manuscript (10%)	Y Y N N N N

I have not renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Student signature: Wei Yang 楊威 **Date:** 30/04/2018

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor signature:  **Date:** 30/04/2018

ACKNOWLEDGEMENTS

Firstly, I would like to express my great thanks to my supervisors, Dr. Sheena McGowan and Associate Professor Ashley Buckle, for their guidance and generous help all along my PhD study. As a student with tremendous interest in bioinformatics without an academic computer science degree, I have been through lots of difficulties in doing this PhD research project. However, it is their excellent mentorship that has helped me through every milestone of my PhD and every chapter of my doctoral thesis. What they have taught me is not only scientific knowledge, but also, more importantly, how to be a professional researcher in the future. This knowledge will be an invaluable treasure for my future career and the rest of my life.

Specially, I would like to give my acknowledgement to Dr. Sheena McGowan for all her great attributes of mentorship. Her approachable attitude resulted in a mentor that you were not afraid to speak to, with whatever result you had in hand, and be it positive or negative. Her office door is always open to discuss scientific ideas. She treats me as every colleague she met, with the same high regard. I feel she wants to inspire more confidence in me by doing so. I thank her for allowing me to perform experiments with ideas that she helped to develop or even mostly produced herself. And by allowing them, take on senior authorship on important papers, even though she could claim it for herself. Although my English skill is limited as an international student, she is always editing my drafts with high level of patience, and providing clear, constructive comments. She is never negative and never belittles me, even though I always committed rudimentary grammar errors. In summary, her wisdom and all her guidance have enlightened me to have the confidence for my future career.

I would like to thank the past and present lab-mates in the McGowan lab and the Buckle Lab: Nyssa Drinkwater, Komagal Kannan Sivaraman, Chiara Ruggeri, Lennart Schnirch, Preethi Jeeva, Tess Malcolm, Sebastian Brøndum, James Andersen, Itamar Kass, David Hoke, Blake Riley, Benjamin Porebski, Emilia Marijanovic, Sarah Le for sharing their

knowledge generously, offering help with my English and helping to make all my presentations friendly to the non-computational in the audience.

Special thanks go to my co-workers, Dr. Itamar Kass, Blake Riley, and Dr. Ben Porebski for their advice, good suggestions and helping to draft my work of my force field parameters and PfA-M1-bestatin dynamics.

I also want to express my thanks to my PhD committee for their help and guidance on my PhD research project. The members of my PhD committee are Associate Professor Anna Roujeinikova, Dr. Michael Kuiper and Associate Professor Martin Stone. I would like to express my appreciation for their great support and valuable suggestions during all of the PhD milestone seminars. I greatly appreciate the helpful and warm discussions with A/Prof. Anna and Martin on my PhD study and life in Melbourne. Just like Michael said, teaching a PhD student from the first day in the University to the last day is like pampering a child from a little baby to a real man, it takes a lot, and thus, I thank them for being good “parents”. I also want to thank Pip Carmen and Sherrie Young for their hard work and time on organizing my transfer from the department of Biochemistry and Molecular biology to the department of Microbiology department and my milestone seminars.

“A good friend is like a four-leaf clover; hard to find and lucky to have.” Yes, as an international student, I had travelled ten thousand miles to make new friends and I was so lucky to have them for sure. I would like to thank Dr. Chen Li, Steve Heaton, Guojie Wu, Adam Quek and Jackie Mao for their help during my hard times. Their words warm my heart during these dark days, and encouraged me continuing my journey onto the ‘Science Mountain’.

I would like to thank my tuition and living allowance scholarship provider, Monash University, for enabling me to pursue my PhD degree here at Monash University in Australia.

Last but not least, I would like to express my great thanks to my family for their support, patience and encouragement during my PhD study. During my PhD study, my great enjoyable time changed from calling my parents on my way back home after work to sharing my baby’s best moments with my parents. Being with my family made me feel delightful and cheerful. They helped me maintained the attitude of being optimistic to

overcome the hurdles. The wisdom they brought to me from our ancient culture is always my motivation for finding the solutions to my difficulties both at work and in life. Therefore, from this moment on, no difficulties in my later life journey will frighten me.

“Our deepest fear is not that we are inadequate. Our deepest fear is that we are powerful beyond measure. It is our light, not our darkness, that most frightens us.”

-‘Our Deepest Fear’

By Marianne Williamson

Abbreviations

ACT	Artemisinin-based combination therapy
ALAP	Adipocyte-derived leucine aminopeptidase
Ang II&III	Angiotensin II & III
APA	Aminopeptidase A
APB	Aminopeptidase B
bILAP	Bovine lens leucine aminopeptidase
DPAP1	Dipeptidyl aminopeptidase 1
DV	Digestive vacuole
<i>EcPepA</i>	<i>Escherichia coli</i> peptidase A
ERAP	Endoplasmic reticulum aminopeptidase
<i>hAPN</i> / CD13	Human aminopeptidase N
Hb	Hemoglobin
IRAP	Insulin-regulated aminopeptidase
LA4HL	Leukotriene A4 hydrolase
LAP	Leucine aminopeptidase
MAPs	Metallo-aminopeptidases
MD	Molecular dynamics
MetAP	Methionine aminopeptidase
MPA1	Meiotic prophase aminopeptidase 1
NPA	N-1-naphthylphthalamic acid
PepA	Cytosol aminopeptidase
<i>PfA-M1</i>	<i>P. falciparum</i> M1 aminopeptidase
<i>PfA-M17</i>	<i>P. falciparum</i> M17 leucyl aminopeptidase
<i>PfAPP</i>	<i>P. falciparum</i> aminopeptidase P
PSAP	Puromycin-sensitive aminopeptidase
<i>SaPepA</i>	<i>Staphylococcus aureus</i> cytosol aminopeptidase
SMD	Steered molecular dynamics
TIF	Tricorn protease-interacting factor
TRH	Pyroglutamyl- -histidyl-prolylamide-NH ₂
TRHDE	Thyrotropin-releasing hormone-degrading ectoenzyme
MSM	Markov State Model
MCPB	Metal center parameters builder
REMD	Replica exchange molecular dynamics simulation

CHAPTER ONE - INTRODUCTION

1.1. Malaria

1.1.1 Malaria remains a global health problem.

The word malaria originally comes from the Italian word *mal'aria*, or the 'bad air', because the belief at the time was that the emitted gas from marshlands was the cause of the disease (8). Today malaria is defined as an intermittent and remittent fever caused by a plasmodium parasite when it invades host red blood cells (9). The disease is transmitted by mosquitoes in many tropical and subtropical regions of the world. Clinical symptoms of malaria include fever and headache (10). If left untreated, malaria can cause severe anemia, reduced local oxygen flow and a block in the immune response (11). If any of these events occur in the brain, cerebral malaria occurs and can cause impaired consciousness, leading to convulsions, coma and death (12).

Malaria is caused by apicomplexan parasites of the genus *Plasmodium*. The five species of *Plasmodium* that can infect humans are *P. falciparum*, *P. malariae*, *P. ovale*, *P. vivax* and *P. knowlesi* (13, 14). Infection by *P. falciparum* has the highest fatality rate with a majority of deaths in children less than 5 years old and women during pregnancy (12, 14-16). The global morbidity and mortality of malaria remain high with an estimated 214 million clinical cases resulting in approximately 438,000 deaths in 2015 (9). This fact makes the treatment of malaria a major health challenge around the world (9).

P. falciparum has a complex life cycle that involves both a mosquito and human host (Fig 1.1) (2). In the human host, the life cycle begins by the injection of sporozoites, the infectious form of the malaria parasite, by a feeding mosquito (Fig 1.1, I). The sporozoites are then carried by the blood circulatory system to the liver where they invade hepatocytes (Fig 1.1, II). Within the hepatocytes (exoerythrocytic phase), the intracellular parasite undergoes asexual replication known as exoerythrocytic schizogony. Exoerythrocytic schizogony culminates in the production of merozoites, the parasite form that is released into the bloodstream and can infect other red blood cells (Fig 1.1, III). Once the merozoites invade the host red-blood cells, the parasite enters the asexual erythrocytic or blood stage, which is the cause of malaria-associated pathology (17). During the blood stage, the

parasites produce a specialized lysosomal-like digestive vacuole (DV) inside the host erythrocyte. During the parasite development from ring stages to trophozoites large amounts of host hemoglobin (Hb) are hydrolyzed in the DV, and parasites grow rapidly to complete their life cycle. The blood stage reproduction is complete within 48 h post-invasion (18). Finally, the parasite can also undergo sexual reproduction where gametocytes are produced. Gametocytes can be taken up by a mosquito, infect the insect host and continue the life cycle (Figure 1.1, IV).

1.1.2 Therapeutics available for the treatment of malaria.

One of the early forms of treatment for malaria was quinine, a natural product that was extracted from the bark of cinchona tree (19). Quinine and chloroquine-substituted quinine were used extensively as malaria medications from the early 1920s until the 1970s (19-21). The 1980s saw the emergence of resistance to chloroquine and derivative

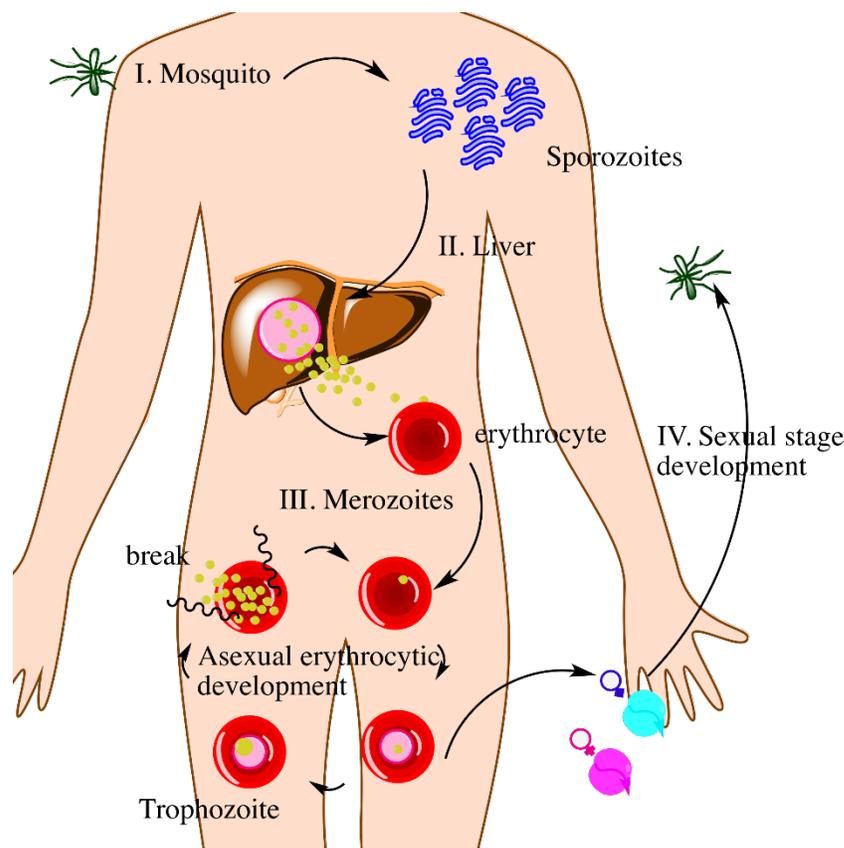


Fig 1.1. The complex life cycle of the malaria parasite (I) Sporozoites are injected into the human body by mosquitos. (II) The exoerythrocytic phase. (III) The asexual erythrocytic development in red blood cells. (IV) The sexual development in mosquito. Reproduced from (2).

medications that has now spread globally (22).

By the 1970s, Tu Youyou and colleagues had discovered the natural product artemisinin, from the plant *Artemisia annua* (23). Artemisinin was shown to be highly efficient for treating *P. falciparum* malaria (24) and the current recommended treatment by the World Health Organization, for uncomplicated malaria, is a Artemisinin-based Combination Therapy (ACT) (25, 26). ACT is highly effective and utilizes multiple artemisinin derivatives that are active against four species of *Plasmodium* (27) and can include amodiaquine, lumefantrine, mefloquine or sulfadoxine/pyrimethamine and the semi-synthetic derivatives: dihydroartemisinin and piperazine (28). ACT is approximately 90 % effective for the treatment of uncomplicated malaria (12, 28, 29). The use of a drug combination therapy is considered advantageous as it decreases the onset of drug resistance to any single drug component (30, 31).

In the early 21st century, malaria parasites with partial resistance to artemisinin-based therapies emerged in South East Asia (32, 33). Recent reports from that region of the world indicate that the efficacy of our “last line of defense”, ACT, is diminishing (34, 35). The discovery of next-generation and completely novel anti-malarial drugs is urgently needed.

1.2. Aminopeptidases as potential drug targets for the design of new antimalarials.

1.2.1 Proteases as potential drug targets.

Proteases (also termed peptidases, proteinases and proteolytic enzymes) are enzymes that hydrolyze peptide bonds (Fig. 1.2) and play important roles in all living cells (36). An early classification of the proteases is from Bergmann & Ross (1936), who suggested that proteases should be categorized by the position by which they cut their peptide substrate(s) (37). In this simple system, a protease can be an endopeptidase (or endoproteinase, Fig. 2), which cuts the peptide bonds of non-terminal amino acids, or an exopeptidase (Fig. 2), which hydrolyzes peptide bonds located at peptide / protein termini. In 1967, in order to distinguish and record the different cleavage positions of peptide substrates, Schechter & Berger (38, 39) proposed numbering amino acid residues of

polypeptide substrates based on the point of cleavage. The cleavage position (bond that was hydrolysed) would be the P1 - P1' and the remaining amino acids numbered from N to C terminal. For example, a six-residue substrate would be labeled as "P2, P1, P1', P2'" (Fig. 1.2). Accordingly, the respective binding pockets for each of the amino acids in the substrate would be defined as "S2, S1, S1', S2'," for the example (Fig.1.2).

The simplistic classification of *endo* or *exo* peptidase was complicated by the discovery that some peptidases show both endopeptidase and exopeptidase activity, e.g., cathepsin B (40-42). This led to other methods of classifying proteases that included their optimal pH for proteolytic activity (such as acid proteases, neutral proteases, basic or alkaline proteases) (43) as well as a more detailed catalog based on their catalytic residue (such as serine proteases, cysteine proteases, threonine proteases, aspartic proteases, glutamic proteases, metalloproteases, asparagine peptide lyases) (44, 45). An up-to-date detailed classification of protease evolutionary superfamilies can be found in the MEROPS database (46), in which, proteases are firstly classified by 'clan' (superfamily) based on basic structure and mechanism before being assigned into families based on sequence similarity within each clan (46).

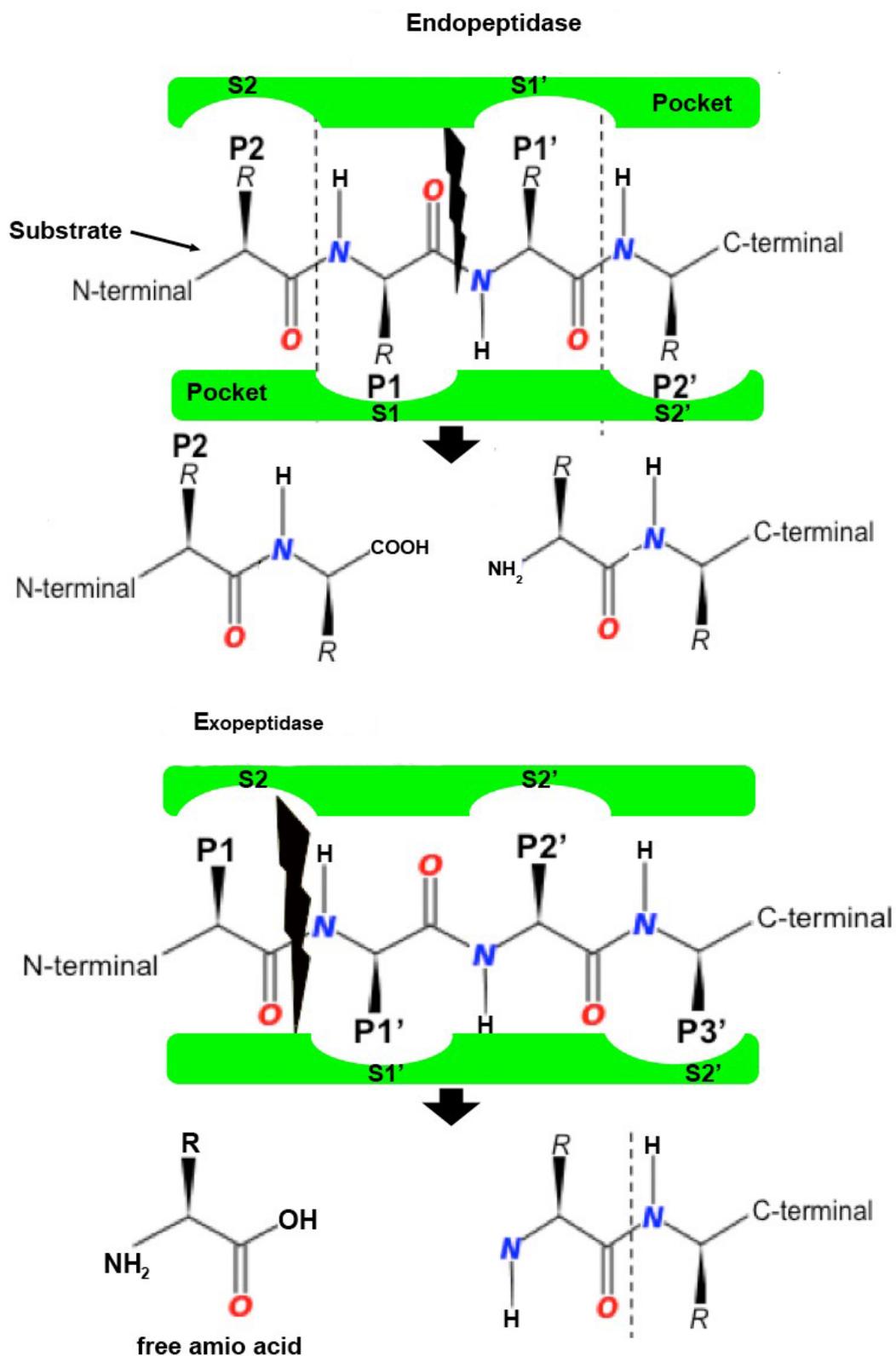


Fig 1.2 Protease activity and nomenclature. The action of an (A) endopeptidase and (B) an exopeptidase is shown diagrammatically with a four amino acid substrate labelled according to standard nomenclature of Schechter & Berger (38, 39).

For cysteine and serine proteases, the side chain that is eponymous residue (Cys or Ser) is the catalytic residue that acts as a proton-withdrawing group to initiate the nucleophilic attack to the peptide bond (47, 48), whereas the aspartyl and metalloproteases trigger a nucleophilic water molecule to attack the scissile bond (49-52). For threonine proteases, the secondary alcohol of their N-terminal threonine is used as a nucleophilic agent to perform the reaction (53, 54). The glutamic protease contains glutamic acid and glutamine catalytic dyad, which act as a nucleophile, with the glutamic acid serving as a general acid to donate a proton to protonate the carbonyl oxygen in the peptide bond of the substrate. One to two water molecule(s) involved in the reaction supply the hydroxyl group to pass the proton to the amide nitrogen on the scissile peptide bond (55).

Protease activity is increasingly seen as a regulatory step in various cellular processes, including development, organelle shaping, metabolism, pathogenicity and degenerative disease (56). Aberrations and / or malfunction of protease activity are often related to a disease state (56-60). Thus proteases are considered attractive therapeutic targets for a range of disease states (4, 60-82).

One important type of exopeptidase is the aminopeptidase. Aminopeptidases can be found in many classes of proteases in MEROPS (46) and are the most abundant in the form of metalloproteases (M1-98) (83). The M1 family of the metalloprotease class contains the majority of the aminopeptidases (46). The function of an aminopeptidase is to digest the N-terminal peptide bond of a protein or peptide substrate, releasing a single (or multiple) amino acids (46, 83). The activity of aminopeptidases is essential to all cellular functions in both eukaryotes and prokaryotes and, along with other proteases, can regulate diverse proteolytic pathways (84, 85). Aminopeptidases can be found in many and multiple biological processes including nutrient acquisition, protein maturation and hormone regulation (84, 86). Recently, they have been of interest as potential new drug targets to treat both chronic and infectious diseases of humans (4, 87-93).

1.2.2 Aminopeptidases are involved in hemoglobin digestion in intra-erythrocytic *P. falciparum* parasites.

The blood stage of the *P. falciparum* life cycle is the cause of the clinical symptoms of malaria. During the blood stage, human hemoglobin (Hb) is the main source of nutrition or food for malaria parasites (94). Digestion of Hb occurs via a multi-step proteolytic cascade that involves many different *P. falciparum* proteases (Fig. 1.3) (2). Initially, Hb is endocytosed and transported to the acidic DV inside the infected red blood cells (Fig. 1.3). Within this specialized organelle, proteases including plasmepsins I, II, IV, falcipains 2, 2', 3 and a histo-aspartic protease, falcilysin, and aminopeptidases digest Hb to globin peptide fragments (Fig. 1.3) (94). Once the Hb peptides are approximately 5-10 amino acids in length, a dipeptidyl aminopeptidase I (DPAP1) and an M1 aminopeptidase (*PfA-M1*) take over digestion (95, 96). When the globin peptides are further reduced in length, they are actively transported out of the DV into parasite cytosol (Fig. 1.3) (97). Finally, within the cytosol, free amino acids are thought to be liberated by the action of various clans of metallo-aminopeptidases (MAPs) (2). The amino acids generated from the catabolism of Hb are not only necessary for protein synthesis but also function to maintain an osmotically stable environment, and can create a gradient by which amino acids that are rare or not present in Hb are transported into the parasite from host serum (98).

There are nine aminopeptidases in *P. falciparum* (97). There are four methionine aminopeptidases (*PfAMetAP1*_{a,b,c} & *PfAMetAP2*) and a single leucine aminopeptidase (*PfA-M17*), alanyl aminopeptidase (*PfA-M1*), aspartyl aminopeptidase (*PfA-M18*), aminopeptidase P (*PfAAP*) and dipeptidyl aminopeptidase (DPAP1). The methionine aminopeptidases perform essential housekeeping role(s) in catalyzing the removal of the N-terminal initiator methionine during protein synthesis (Fig. 1.3) (99). DPAP1 bridges the gap between endopeptidase and aminopeptidase activity as it further hydrolyzes the oligopeptides to dipeptides (Fig. 1.3) (95). *PfA-M1* hydrolyzes most amino acids from the N-terminus of the peptides (the P1 position), except for the acidic acids like Glu or Asp (100). *PfA-M17* can process hydrophobic amino acids with a preference for Leu (100). *PfA-M18* will only hydrolyse acidic amino acids, including both Asp and Glu (101). *PfAAP* removes the N-terminal residues from the peptide substrates with a proline at the P1' position (102, 103).

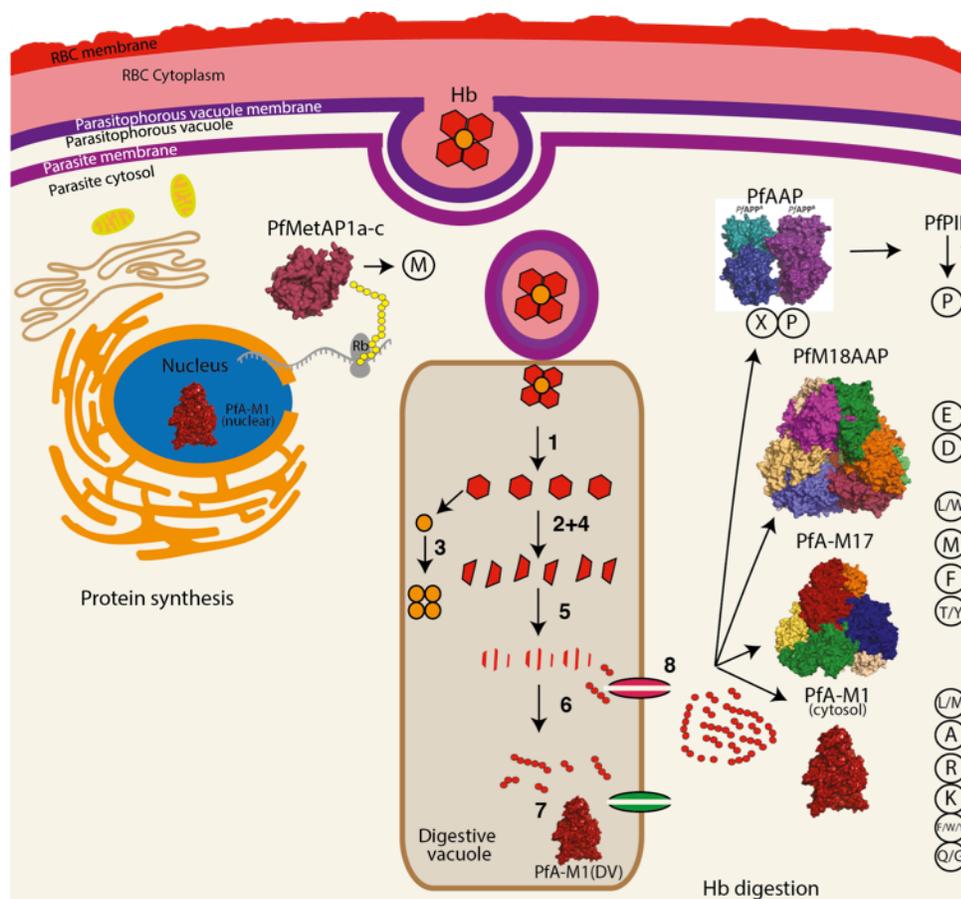


Fig 1.3. The overview of Hb digestion in the malaria parasite *P. falciparum*. 1. Plasmepsin I and II initiate Hb digestion, which releases the monomers from the Hb tetramers. 2. Plasmepsin I, II, IV continue to cleave the Hb with divergent specificities. 3. Heme is released and hemozoin formed, as can be observed as malaria pigment hemozoin in the ring-stage of parasite development. 4. Falcipains 2, 2', 3 will then take over the digestion process and catalyze the products generated from plasmepsins to globular peptides of 10-20 residues in length. 5. Falcilysin then further processes the products into oligopeptides of approximately 5-10 residues. 6. Aminopeptidases are involved in the final process of digestion and release free amino acids. 7. *PfA-M1* has been shown to be active in the DV, however, can also be found in the cytosol and the nucleus. 8. Protein transporters move the Hb peptides from DV to parasite cytosol. (Figure reproduced with permission from McGowan laboratory).

1.2.3 The *P. falciparum* neutral aminopeptidases as potential targets for new antimalarials.

The two neutral aminopeptidases, *PfA-M1* and *PfA-M17*, have been validated as drug targets for the development of new antimalarials. Both aminopeptidases are encoded by a single copy of their respective gene, and attempts to genetically disrupt these genes in *P. falciparum* have been unsuccessful (4, 95). This indicates that the genes or gene products are essential for *P. falciparum* parasites. Inhibition of the activity of the two enzymes prevents the growth of *P. falciparum* parasites in culture (4, 98, 104). In the *P. chaubaudi chaubaudi* murine model of malaria, inhibitors of *PfA-M1* and *PfA-M17* were able to control parasitemia indicating the potential of these enzymes as new antimalarial drug targets (104, 105). These early inhibitors lacked optimal pharmacokinetic properties for further development and therefore, new drug-like inhibitors possessing both potency and selectivity are still required.

One idea for the inhibitor design for any potential new drug target is to mimic the natural substrate. To use this approach for *PfA-M1* and *PfA-M17*, small molecules that are a similar size to a dipeptide would be needed. These compounds would need to be able to bind to the active site of the enzymes and lock the enzymes in a transition state. Various studies have aimed to produce inhibitors of both *PfA-M1* and / or *PfA-M17*. These include the development of peptidomimetic scaffolds including bestatin analogues (106), phosphonic acids (107), and phosphinopeptides (4, 81, 98, 107, 108). Virtual screens of compound libraries have also been performed and in some cases experimentally validated (109-112). Screening of the Medicines for Malaria Venture (MMV) / SCYNEXIS anti-malarial compound library (MMV400) suggested that neither *PfA-M1* nor *PfA-M17* were the primary targets of these compounds (108). Drug re-purposing studies also showed the antimalarial efficacy of Tosedostat, a chemotherapeutic agent that is currently in phase II trials for various human cancers (105). The study of phospho-arginine mimetics to probe the specificity of the S1 pockets of *PfA-M1* and *PfA-M17* provided a route to a small molecule scaffold that could target both enzymes (107). Potent dual inhibitors have been produced,

using a hydroxamic acid scaffold, which shows both nanomolar potencies in anti-aminopeptidase and anti-parasitic activity (113, 114).

To facilitate the design of new inhibitors and discover potential new pharmacophores within *PfA*-M1 and -M17, it is important to understand the relationship between the structures and function of both of the two enzymes. Therefore, in this thesis, we will introduce both the M1 and M17 aminopeptidase families in detail and discuss current literature relating to the family and specifically to *PfA*-M1 and *PfA*-M17.

1.3. The M1 aminopeptidase family

The M1 aminopeptidase family is the largest of the metalloprotease clans (115). The M1 family is defined by the presence of two conserved sequence motifs that are required for catalytic function; a consensus zinc-binding motif (HEXXH-(X₁₈)-E) and the 'GXMEN' exopeptidase motif (116). M1 family members are distributed throughout the bacteria, archaea, protozoa, fungi, plants and animals (116). They can be found in the cytoplasm, many subcellular organelles and / or as integral membrane proteins (85, 117, 118).

1.3.1 The role of the M1 aminopeptidases

There are nine M1-type aminopeptidases in humans (119) that perform a variety of essential cellular functions (120). The leukotriene A4 hydrolase (LTA4H) has dual activities and can process LTA4 to LTB4 to make chemotactic agents for human neutrophils, eosinophils, monocytes, and T cells (121) as well as possessing epoxide hydrolase activity (122). LTA4 was the first M1 to be structurally characterised via X-ray crystallography (6). The aminopeptidase A (APA) processes angiotensin II (Ang II) to angiotensin (Ang III) in the renin-angiotensin system (123). This step is essential in regulating blood pressure in mammals (124). The aminopeptidase N (hAPN / CD13) regulates blood pressure, cyclin accumulation, mitotic cell division and antigen processing for MHC molecules by degrading various substrates. Dysregulation of APN function is associated with various cancer-like states (125, 126). The insulin-regulated aminopeptidase (IRAP), also known as cystinyl aminopeptidase (CAP), human placental leucine aminopeptidase (PLAP), oxytocinase, and

vasopressinase, functions as an oxytocinase by degrading oxytocin during human pregnancy (127) and is active against N-terminal cysteine residues in vasopressin and oxytocin (128-130). Endoplasmic Reticulum Aminopeptidase(s) (ERAP1 / ERAP2) were initially identified as aminopeptidases also involved in the processing of the Ang II and Ang III as well as kallidin (131). They take part in post-proteasome processing in the immune response and have a role in pathogenesis of ankylosing spondylitis (131). Unlike many M1 enzymes, Thyrotropin Releasing Hormone Degrading Ectoenzyme (TRHDE) hydrolyses only one substrate, pyroglutamyl-histidyl-prolylamide-NH₂ (TRH), which acts as a neurotransmitter within both the central and peripheral nervous systems and promotes the yield of thyrotropin and prolactin (132-136). Aminopeptidase B (APB) cleaves basic N-terminal amino acids and can process glucagon into miniglucagon (137, 138). APB is more closely related to LTA4H (139).

In the microbial world, M1 aminopeptidases are used primarily for nutrient acquisition and metabolic pathways. The *Escherichia coli* aminopeptidase N, *EcAPN*, has been studied in detail and like most bacterial M1s, is not essential for bacterial survival (140). *EcAPN* is involved in ATP-dependent downstream processing during cytosolic protein degradation, however, there appears to be redundant peptidases, which likely accounts for its non-essential nature (141). Similarly lysyl aminopeptidase in *Lactococcus lactis* helps in the degradation of casein into peptides and amino acids, which can then be taken up by the organism as the external source of amino acids (142). In the archaeon *Thermoplasma acidophilum*, the tricorn interacting factor (TIF) F2 and F3 are both members of the M1 superfamily (143). Together with TIF1, these enzymes degrade the tricorn protease products and generate free amino acids. In protozoa, the best studied example of an M1 aminopeptidase is the enzyme from *P. falciparum*, *PfA-M1*, that is discussed in Section 1.3.4 of this chapter.

1.3.2. The catalytic mechanism of the M1 aminopeptidases.

A catalytic mechanism, based on structural comparisons to the prototypical zinc-peptidase, thermolysin, has been proposed for the M1 aminopeptidases (144). A peptide substrate enters the buried active site and the carbonyl oxygen and nitrogen of the scissile peptide bond of the substrate are activated by interactions with three atoms: 1) the zinc ion, 2) a highly conserved tyrosine residue and 3) the carboxyl oxygen of the E residue in the GXMEN motif (Fig 1.4A). The catalytic water is coordinated by a conserved glutamate (that is not part of the zinc-coordinating triad) and is activated by the zinc ion. The activated nucleophilic water molecule attacks the carbonyl carbon of the scissile peptide bond, allowing the traffic of a proton from the activated catalytic water to the leaving nitrogen group (144). The events surrounding the actual hydrolysis has been proposed to occur through two sequential transition states (Fig 1.4 B, C), where state I is the interaction of the nucleophilic proton with a conserved glutamate and state II is the subsequent transfer to the P1' amide of the scissile bond (Fig 1.4B).

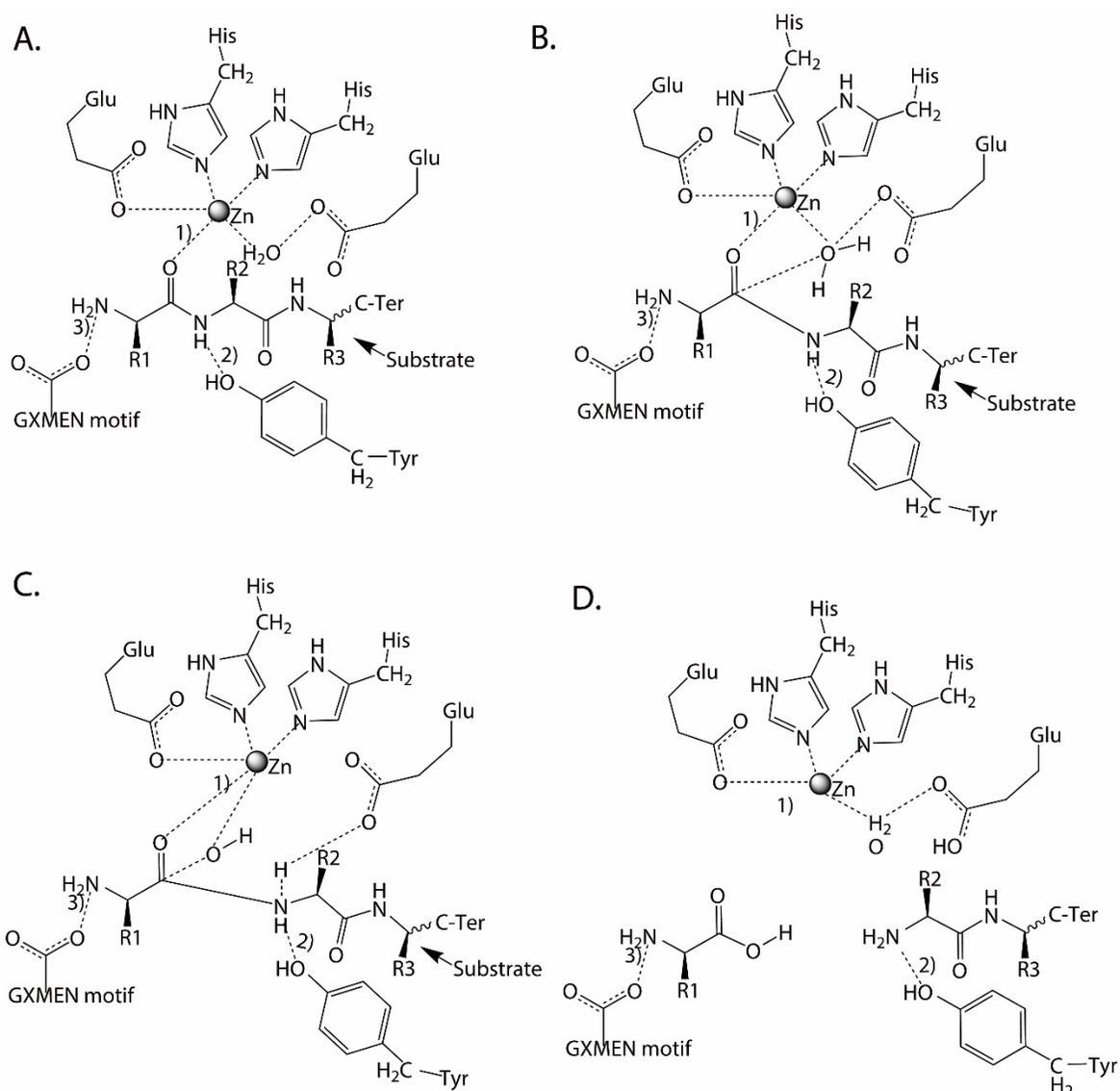


Fig 1.4 The catalytic mechanism of M1 aminopeptidases. (A) Active site arrangement in M1 aminopeptidases showing the zinc-coordinating residues and atoms, and the coordination of the peptide substrate with the zinc ion. Transition state I **(B)** and II **(C)** are also indicated and the interactions between catalytic residues of the M1 aminopeptidase and scissile peptide bond of the substrate peptide are shown. **(D)** The production of the catalytic process. The figure is based on (144)

1.3.3 The structure of the M1 aminopeptidases.

As of September 2016, the International Protein Databank (PDB) contained 17 unique structures of M1 aminopeptidases from archaea, fungi, bacteria, and animals, which provides a wealth of structural information to aid our examination of the family. M1 aminopeptidases can exist as both monomers and dimers (Fig 1.5A and 4B), are generally

membrane bound and can be found surface-exposed or within the cytoplasm (145, 146). The general structure comprises a short tail from the transmembrane anchor/stalk and a large ectodomain with a conserved 3–4 domain fold (Fig 1.5) (144).

The N-terminal domain I has a β -sheet core which, while largely solvent-exposed, contains a hydrophobic region that links to a membrane-spanning anchor (144). The thermolysin-like catalytic domain II contains the active site flanked by a mixed β -sheet and α -helical structure that is highly conserved throughout the family. Within the active site, the catalytic zinc ion is coordinated by N ϵ 2 atoms of two conserved histidine residues, the carboxyl O ϵ atom of the glutamate and a nucleophilic water molecule (144). Domain III is composed of an immunoglobulin-like fold and is completely absent in some family members, such as leukotriene A4 hydrolase (Fig 1.5E) (6). The C-terminal domain IV is the most variable region of the protein family and is completely helical in nature, arranged to cap the active site and sequester it from bulk solvent. Domain IV is also involved in dimerization, primarily in mammalian forms (1, 144, 147, 148).

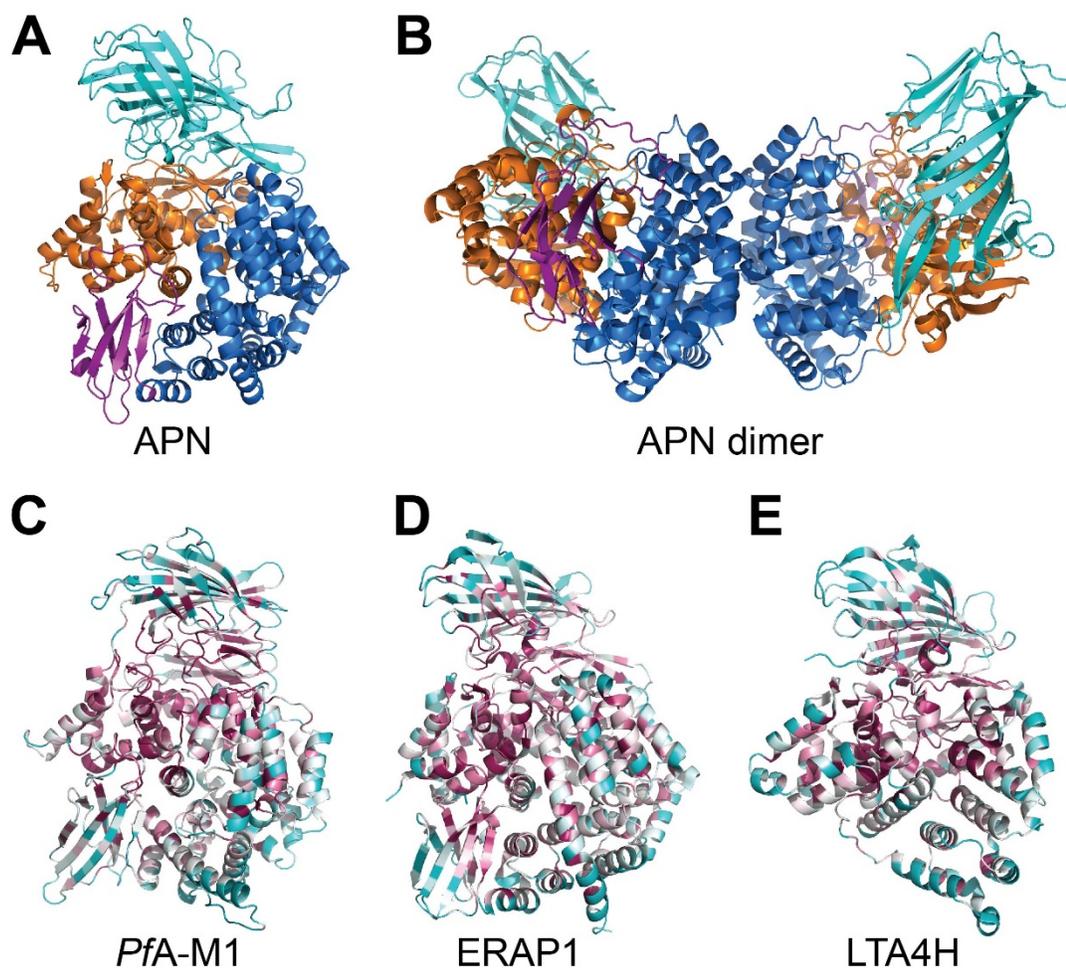


Fig 1.5: Structure and conservation of M1 aminopeptidases. (A) X-ray crystal structure of human aminopeptidase N (APN) (1) colored according to domains with domain I in teal, domain II in orange, domain III in magenta, and domain IV in blue. (B) Crystallographically observed dimer of APN, that is also proposed to occur on the cell surface membrane (1) to which APN is attached via domain I (teal). Domains colored as in A. (C–E) Characteristic structures of M1 aminopeptidases colored according to sequence conservation (high degree of conservation in purple, average in white, and low in cyan). Sequence alignments and conservation calculations performed using ConSurf (3). (C) the M1 aminopeptidase from *P. falciparum* (*PfA-M1*) (4), (D) human endoplasmic reticulum aminopeptidase 1 (ERAP1) (5), (E) human leukotriene A4 hydrolase (LTA4H) (6). This figure is reproduced with permission from (7).

1.3.4. Some members of the M1 aminopeptidase superfamily can exist in an open or closed form.

The determination of the TIFF3 structure identified an open form of the M1 aminopeptidase where the C-terminal domain IV is positioned away from the rest of the molecule (Fig. 1.6) (149). Interestingly, this study captured both conformations (open and closed) of TIFF3 in a single crystal (PDBID: 1Z5H, 2.3 Å) as well as a semi-open conformation in a different crystal (PDBID: 1Z1W, 2.7 Å) (149). Comparison of the structures shows that the open form has clear access to the active site in domain II, whilst in the closed form, domain IV is in close contact with domains I + II (Fig. 1.6) (149). Otto *et al* proposed that the open form is in an inactive conformation, since Tyr³⁵¹, a residue proposed to be involved in the transfer a proton donor to the zinc ion, was shifted 2.7 Å away from the zinc ion in the open form (149).

Open and closed forms of ERAP1 have also been captured using protein crystallography (5). Kochan *et al* captured an open form of *h*ERAP1 (PDBID: 3QNF, 3.0 Å) by co-crystallizing the protein in the presence of substrate peptide, and also argued that ERAP1 is inactive in the open form (5). Similarly to TIFF3, Tyr⁴³⁸ (the corresponding residue to Tyr³⁵¹ in TIFF3) shows a movement of 6 Å in the open form of ERAP1 and points away from the active site. Chen *et al* claims that their structure of a porcine APN (*p*APN) mutant (Phe⁸²Asn, Leu¹⁰⁷Phe) was crystallized in a closed form but inspection of the structure shows that it is more similar to the semi-open form of TIFF3 (144). They concluded that although their closed form of *p*APN is not similar to that of ERAP1, it still has the potential to open like ERAP1, and consequently modelled an open form of *h*PAN based on the ERAP1 open structure (144). To date, there are no reported structures or evidence to support an open form of any parasite and / or bacteria M1 aminopeptidase.

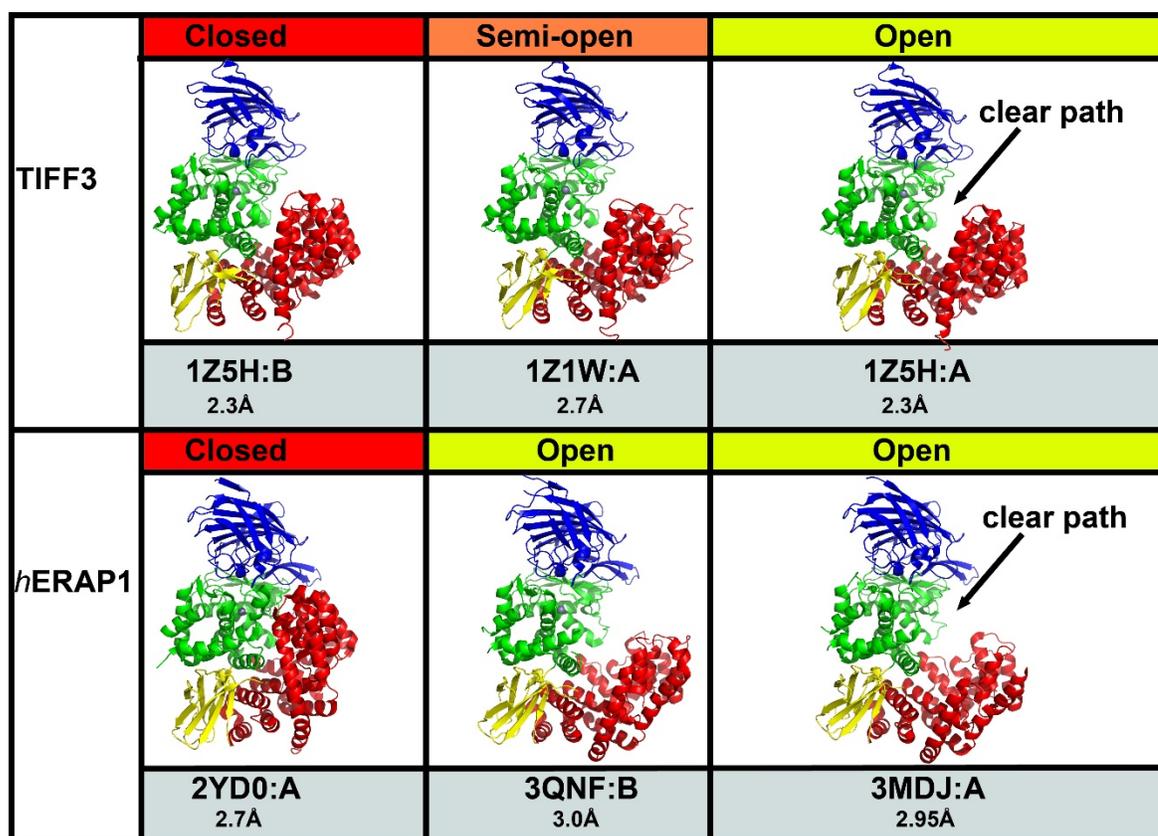


Figure 1.6. The structures of the closed and open forms of representative M1 aminopeptidases. The structures are colored by domain with domain I in blue, II in green, III in yellow and IV in red. The PDB ID, chain and resolution for each M1 aminopeptidase structure are indicated under the image.

1.3.5 PfA-M1

In *P. falciparum*, a single M1 aminopeptidase was identified nearly 20 years ago (150). As annotated by PlasmoDB (151), the PfA-M1 gene (MAL13P1.56) is located on chromosome 13 of *P. falciparum*. The gene is 3,257 bp in length and encodes a 1,085 amino acid protein of approximately 126 kDa (151). It shares ~ 70 % sequence similarity with M1 aminopeptidase orthologues from other *Plasmodium* species including *P. berghei*, *P. chabaudi chabaudi*, *P. vivax*, and *P. yoelii* (4). The N-terminus of the protein (~ 194 amino acids) is not conserved and has three asparagine-rich low-complexity regions (LCRs) and a putative transmembrane domain (4, 150). In *P. falciparum*, the low-complexity regions are removed by proteolytic processing by an as yet unidentified protease, to produce the active aminopeptidase that is often referred to in the literature as p96 (for 96 kDa) (150, 152). The

mature enzyme is thought to be membrane bound (4, 150). The Florent group also proposed the existence of a 68 kDa (p68) version of the protein (150, 152), however, this form of the protein has never been isolated away from the p96 version and is presumed to be an artifact (153). The recombinant form of the protein produced in McGowan *et al.*, removed both the low complexity regions and the transmembrane domain from the N-terminus and purified only the ecto-domain of *PfA-M1* (4).

During the erythrocytic life cycle, transcription of the *PfA-M1* gene reaches its maximum in schizonts (152), which is when digestion of Hb occurs. *PfA-M1* functions in the final stages of Hb digestion, to presumably provide free amino acids for parasite growth and replication (96). The P1 substrate specificity of the enzyme is broad (100); it can rapidly hydrolyze substrates containing a neutral P1 residue such as Leu, Ala, Phe, Tyr, and Ser as well as basic P1 residues such as Asn, Arg, and Lys (100). It is unable to process Glu or Asp residues in the P1 position (100).

The intra-cellular location of *PfA-M1* has been of interest for many years. Florent and colleagues used immuno cytochemical studies to localize *PfA-M1* and reported that the enzyme was localized to the parasite cytosol and accumulated around the digestive vacuole, presumably tethered *via* its transmembrane domain (152, 154). McGowan *et al.*, (4) studied the cellular location of the *PfA-M1* aminopeptidase by using live microscopy of parasites transfected with a GFP-*PfA-M1* conjugate and observed GFP in the cytosolic compartment only, consistent with the earlier studies of Florent and colleagues (152, 154). In contrast, the Klemba group has repeatedly reported that *PfA-M1* is localized to the inside of the digestive vacuole (155-157) as well as the parasite nucleus (157). Currently, the location of the active form of *PfA-M1* remains unresolved with an argument that the N-terminal transmembrane anchor may localise the enzyme to either side of the digestive vacuole membrane. Both positions would be amenable to Hb peptide degradation.

The structure of ectodomain of *PfA-M1* (Fig. 1.7A) is more closely related to the bacterial M1 aminopeptidases (Fig. 1.7B) than mammalian homologs (4). *PfA-M1* possesses the characteristic bacterial aminopeptidase N fold (158-160), and like the other M1 aminopeptidases, *PfA-M1* consists of four domains (Fig 1.7A), with 26 α -helices and 7

β -sheets (4). The second domain is the catalytic domain, which possess a zinc-binding motif $H^{496}EYFHX_{17}KE^{519}$ and a conserved substrate recognition motif $Gly^{490}AMEN$ ((158-160). The only zinc ion is coordinated by $His^{496}:N\epsilon 2$, $His^{500}:N\epsilon 2$, and one of the carboxyl $O\epsilon$ atoms on Glu^{519} (Fig 1.7C). A water molecule is visible in the X-ray structure between the zinc ion and carboxyl $O\epsilon$ of both Glu^{497} and Glu^{463} (Fig 1.7C). This water molecule is hypothesized to act as the nucleophilic agent in the hydrolytic activity of *PfA*-M1 (4). As mentioned above, the structure of *PfA*-M1 is in a closed state, and no enzyme in an open state has ever been observed in the parasite and bacteria kingdom.

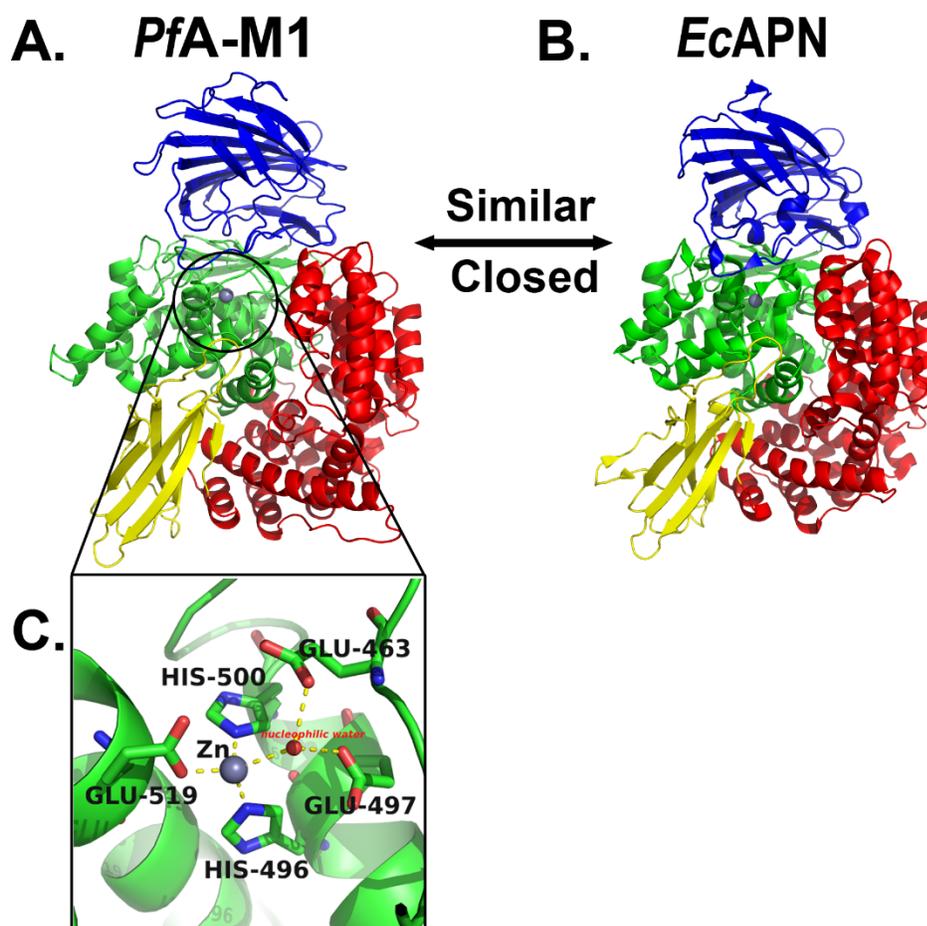


Figure 1.7. The structure of *PfA*-M1 and *E. coli* APN (*EcAPN*). (A) Overall structure of *PfA*-M1. (B) Overall structure of *EcAPN*. (C) The zinc-binding site of *PfA*-M1. The zinc ion is shown as a grey sphere and the nucleophilic water molecule as a red sphere. The carbon atoms of the residues coordinating the zinc ion are shown in stick. Domain colors are same a previous.

1.4. The M17 family of Leucine Aminopeptidases.

The M17 leucine aminopeptidases (LAPs) are a different family of metallo-aminopeptidases that belong to the MF clan (115). LAPs are ubiquitous in nature and members can be found in the kingdoms of animals, plants, protozoa and bacteria (161). The M17 aminopeptidases are cytosolic exopeptidases that form biologically functional hexamers, which can generally release any N-terminal amino acid from dipeptides and polypeptides, although often prefer leucine, and rarely process proline in P1' position (162). They have extensive physiological roles in the metabolism of hormones and neurotransmission, cell maturation, and turnover of proteins, including utilization of exogenous proteins as nutrient substances and elimination of nonfunctional proteins (79).

1.4.1 The role of the M17 aminopeptidases

In animals, the M17 aminopeptidases appear to have key roles in the development of eye lens cataracts as well as in the processing and determination of immune-dominant antigenic peptides (163-167). The human adipocyte-derived leucine aminopeptidase (A-LAP) is involved in AngII metabolism, cell migration and antigen presentation (168). A-LAP activity has been shown to be elevated in some cancers and is proposed to have a role in cancer progression (169). In plants, M17 aminopeptidase family members appear to play a role in salvaging nitrogen and carbon resources from the cotyledons of germinating seedlings (170). In bacteria, the M17 aminopeptidases appear to have a role in virulence with inhibition of their activity showing antibacterial effects (171-173). The *E. coli* leucine aminopeptidase (PepA) also has a role within the Xer site-specific recombination and can control the transcription of the *carAB* operon (174). The DNA-binding activity of the bacterial LAPs appears to be in addition (and / or unrelated) to their proteolytic activity and often occurs via the formation of a larger oligomeric quaternary structure (170). In the protozoa kingdom, the best-characterized leucine aminopeptidase is PfA-M17 from *P. falciparum* that will be discussed in more detail below in Section 1.4.4.

1.4.2 The structure of the M17 aminopeptidases

The M17 / LAP monomeric subunit is composed of an N-terminal regulatory domain and a C-terminal catalytic domain (Fig 1.8A) (81, 172, 175-177). The N-terminal domain generally contains a central six-stranded β -sheet and three α -helices, and is linked to a C-terminal catalytic domain via a long helix (Fig 1.8A, yellow helix). The catalytic domain is generally composed of nine α -helices and with a two-stranded β -sheet capping one end and a central eight-stranded β -sheet flanked on both sides. The complete active site of the M17 monomers is contained within the C-terminal domain (81, 170, 176, 178).

The M17 aminopeptidases have two divalent cations in their active site (Fig 1.8B) (171-173). A conserved aspartate residue coordinates both metal ions while conserved lysine and glutamate residues contribute to the coordination of metal ion 1, while a different (but still conserved) aspartate and glutamate residue coordinate metal ion 2 (Fig 1.8C). A carbonate ion is also always found in the active site and is coordinated by a network of hydrogen bonds with active site residues (81, 176, 179, 180). The carbonate ion is common to all M17 crystal structures and is proposed to act as a general base, accepting the proton and activating a metal-bridging water in the cleavage reaction (179).

The M17 hexamer is a dimer of trimers (181). The trimers have ~ 120 degrees between each monomer (Fig 1.8C) and two trimers interact each other and form the large hexameric quaternary structure. This assembly produces a large inner cavity where the zinc ions and active sites are protected from bulk solvent (Fig 1.8C) (170, 176, 178).

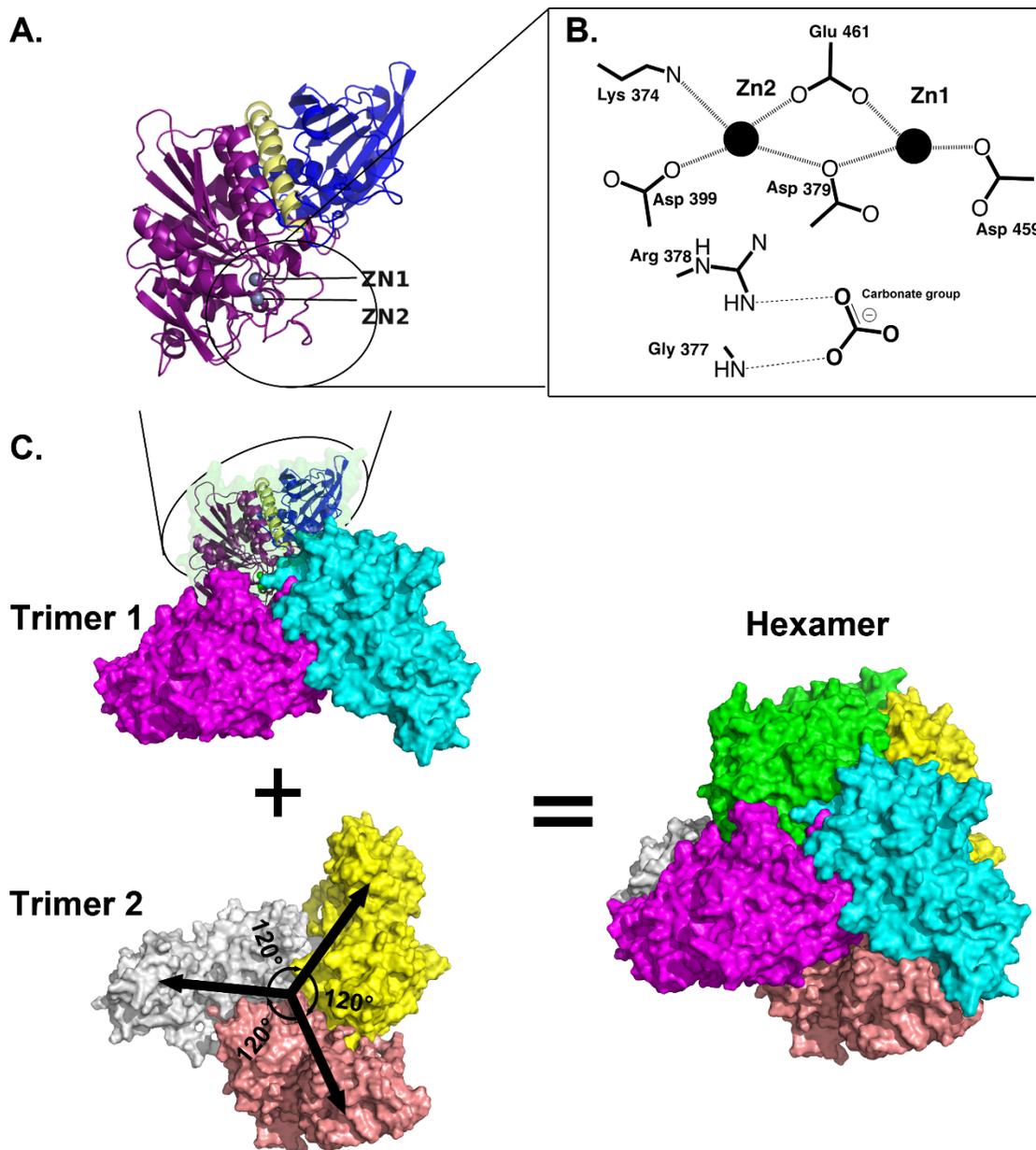


Fig 1.8. The structure of the M17 aminopeptidases. (A). The monomer of *PfA-M17* shown in ribbon where the N-terminal domain is shown in blue, the linker helix between N-terminal and C-terminal is in yellow and the catalytic domain is in magenta. (B) The molecular architecture of the active site. (C) The hexameric assembly is a dimer of trimers where trimer 1 (top) is shown in space-filled cyan, green and magenta and trimer 2 (bottom) in grey, yellow and wheat.

1.4.3 The catalytic mechanism of the M17 aminopeptidases.

The binuclear metal center of the M17 family is pivotal to its enzymatic function (176, 179, 180). In the proposed catalytic mechanism, a water molecule bridges the two metal ions and acts as a nucleophilic agent. It attacks the carbonyl oxygen of the scissile bond of the P1 position of the peptide substrate, which is initially coordinated by one metal and the nucleophilic water (Fig 1.9A). A single proton is transferred from the water molecule to the nearby carbonate ion, reducing the nucleophilic water to a hydroxide ion (179) and producing a bicarbonate ion from the original carbonate (179). The metal-bound hydroxide and the activated scissile carbonyl carbon of the peptide substrate, form a *gem*-diolate intermediate complex that is stabilized by coordination of both oxygen atoms to the double zinc ions (Fig 1.9B). Finally, the two zinc ions complete the deprotonation of the nucleophile via the coordination of a lysine residue and release the cleaved peptide (179). A second water molecule enters the active site and bridges the two metal ions again, whilst a carbonate replaces the bicarbonate inside the acidic catalytic domain (Fig 1.9).

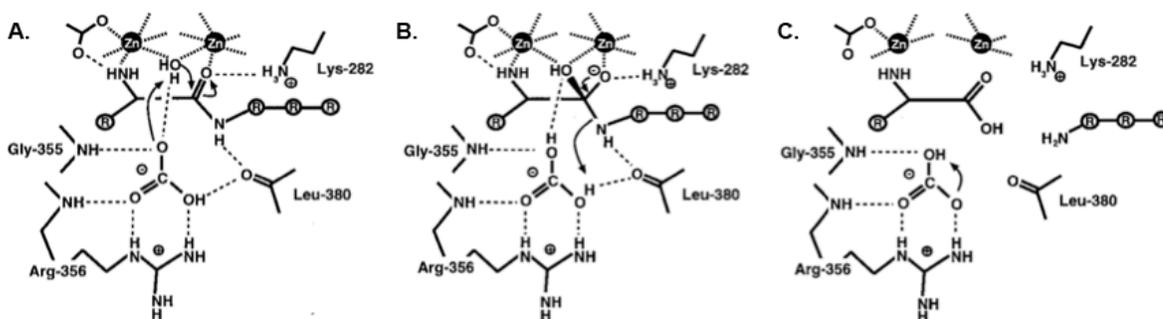


Fig 1.9. Scheme for the catalytic mechanism of M17 aminopeptidases. (A) Bicarbonate ion next to Arg356 (numbering according to PepA) acts as a general base and accepts a proton from the metal-bridging water nucleophile. (B) The proton is shuttled to the leaving peptide amino group to facilitate breakdown of the tetrahedral *gem*-diol intermediate. (C) The product amino acid is released. The figure is reproduced from (179).

1.4.4 PfA-M17

The only M17 aminopeptidase in *P. falciparum*, PfA-M17, was identified nearly a decade ago (98). The gene (*Pf14_0439*) annotated by PlasmoDB (www.plasmodb.org,

(151)) is located on chromosome 14 of *P. falciparum* with a predicted molecular weight of 67.8 kDa (182). The enzyme functions as a homo-hexamer in the parasites (98). *PfA-M17* is located in the parasite cytosol and its function is not fully understood (2, 81, 98). One of the hypothesized roles of *PfA-M17* is in the final stages of Hb digestion (81, 98). A possible secondary or alternative function is hypothesized to occur early in the development of the parasite, as specific inhibition of *PfA-M17* is lethal to parasites early in the intraerythrocytic life cycle (96).

The substrate specificity of *PfA-M17* is severely restricted to N-terminal exposed leucine amino acid (98). Substrates with a P1 Ala residue are reported to be cleaved at a rate ~ 850 fold slower than Leu, and substrates which contain Arg, Val, Pro, Gly, Ile, Glu, or Asp in the P1 position cannot be hydrolysed at all (98). The preference for leucine is postulated to relate to the need to provide isoleucine to the parasite (183). *Plasmodium* parasites use Ile in their own proteins; however, it is the only amino acid not available from Hb digestion (98, 183). Thus, the leucine preference by *PfA-M17* may be related to the demand for free leucine within the parasite cytosol to exchange for isoleucine (81).

A 2.0 Å X-ray crystal structure of *PfA-M17* showed the same hexameric arrangement of monomers as the physiological assembly described for other LAP family members (81, 170, 176, 178). In the recombinant form of *PfA-M17*, each monomeric subunit consists of residues 85-605. The N-terminal domain (residues 85-276) consists of six-stranded β -sheets and three α -helices. A central helix (residue 277-300) bridges the N-terminal and the catalytic C-terminal domain (residues 301-605). An eight-stranded β -sheet and nine α -helices form the catalytic domain, where the active site and its two zinc ions are located. Within the hexamer assembly, the six active sites are located within the interior of the quaternary structure. It is reported that the active site of *PfA-M17* contains one exchangeable site and a tight-binding site (93). The removal of metal in the exchangeable site slows catalytic activity, whilst removal of both metal ions leads to an irreversibly inactive enzyme (93).

1.5. Molecular dynamics of proteins

X-ray crystal structures present proteins as static models and fail to provide a dynamic model with the internal motions that may play an essential role in their function. Molecular dynamic (MD) simulations are an important tool for providing insights and models of the physical basis of the structure and function of biological macromolecules, which may not be accessible from classical laboratory experiments (184). MD is a computational biology tool that uses molecular mechanics (MM) to mimic the physical motions of atoms in the protein molecule present in the actual environment (185). MM are widely used in molecular structure refinement, MD simulations, Monte Carlo (MC) simulations, and docking simulations (185). The molecular mechanics force fields (FF) are the foundation of the biomolecular MD simulations, which provide information about the potential energy of a system of particles. Parameters of FF can be obtained from experimental and quantum mechanical studies of capped amino acids for proteins and deoxyribose nucleoside triphosphates (dNTPs) for DNA. These parameters can then be transferred to the larger biomolecules, such as protein and DNA. There are two kinds of interaction terms in the force field functions: the bonded and non-bonded interaction terms. Bonded interactions include harmonic oscillator energy of bond lengths, bond angles, and improper dihedrals (fixed terms), torsional dihedral angles (mobile terms, sometimes including improper dihedrals), The nonbonded terms include Van der Waals interactions and electrostatic interactions (186). Currently, there are several well-developed FFs to describe protein structure and motion, including GROMOS (187), AMBER (188), CHARMM (189), OPLS (190, 191). These FFs are refined and improved every year as research continues and computing power improves.

Using these simple models and FFs, Newton's equations of motion can be solved by a MD simulation, therefore allowing structural fluctuations to be observed with the evolution of time (184, 192-194). Dynamic simulation methods are widely used to obtain information on conformations of proteins and other biological macromolecules over time (184, 192-194) as well as kinetic and thermodynamic information. Simulations can therefore provide fine detail concerning the motions of individual particles as a function of time (185).

1.5.1 MD simulation of metalloproteins

The available and well-defined FFs (188, 191, 195) of bio-systems do not currently include parameters for certain chemical groups or elements, for example transition heavy metals. Over time, three methods of heavy metal FF development have been proposed and include a “nonbonded” (196, 197) and “semi-bonded” (198-200) model and more recently a “bonded” model (201-203). In each of these methods, the transition metal ion(s) has to be either tethered in the binding pocket by artificial bonds (the bonded model) (201-203) or constrained in the pocket by external forces (the nonbonded and semi-nonbonded models) (196, 197, 199, 200, 204). This means that all three models assume that the coordination of heavy metal(s) is unchanged during the simulation time. This is necessary because the implemented FF parameters (charge distribution and force constants) of the transition metals has to change when the coordination of transition metals varies (201). However, it is impossible to change a FF at each time step during a classical MD simulation (205). Therefore, classical MD with metal FFs cannot simulate the situation where a metal ion changes its coordination (206). This is a significant limitation of the current approach that must be considered during analysis of the MD simulations. This may be addressed by hybrid methods such as quantum mechanics/molecular mechanics (QM/MM) simulations, however, these approaches are incredibly expensive with regard to computational power (205, 207).

A MD study of *PfA*-M1 used a semi-bonded model as the heavy metal FF for a classical MD simulation (208). In this study, a single 75 ns trajectory was performed a *PfA*-M1 system. Recent advances in MD of metalloproteins means that this FF approach is now out of date, and the single trajectory might not sample enough possible conformations during the simulation. Moreover, for a considerably large system like *PfA*-M1 (~ 890 residues), it is unlikely that 75 ns would be able to sample any large domain motions. To date, there have been no MD studies to investigate any M17 family member.

In order to investigate the protein dynamics of *PfA-M1* and *PfA-M17*, this study will use the currently most advanced metal force field method to perform a classical MD simulation both proteins.

1.6 Aims and Objectives

The motivation to understand how protein structure and dynamics can control or regulate protein function has increasingly demanded the use of a combined computational and experimental approach. The two neutral aminopeptidases from *P. falciparum*, *PfA-M1* & - *PfA-M17*, have been extensively characterized by biochemistry and X-ray crystallography, however, this has provided a static snapshot of the structure of the enzymes in both their unbound and inhibitor-bound forms. It is still unclear if protein dynamics has a role in the enzymatic activity and biological function(s) of the two enzymes. Understanding the relationship between motion, structure and the function will provide new knowledge for these protease superfamilies that will be of benefit to dissect their biological roles and in the design of potent and selective inhibitors.

In order to explore the above themes, this thesis will focus on the following three main areas:

- a) To explore the dynamics of the M1 and M17 aminopeptidase superfamilies through MD simulation by producing appropriate FF parameters and completing all-atom MD simulations (Chapters 2 and 6).
- b) Investigate the dynamics of *PfA-M1* and how this relates to structure and function, specifically its aminopeptidase activity, substrate guidance and inhibitor binding (Chapter 3, 4 and 7).
- c) Investigate the dynamics of *PfA-M17* to answer the fundamental question of why this enzyme forms a hexamer and how the hexamer is related to enzymatic function (Chapter 5).

In summary, this thesis aims to advance our understanding of the M1 and M17 aminopeptidase superfamily by exploring their dynamics and how this relates to function. These data will be of interest to protease biologists as well as drug discovery programs

through the mapping of dynamics of the whole protein as well as the specific pharmacophore. In addition, the computational tools developed in this study will be of broad interest to the field.

CHAPTER 2 - METHODS

2.1. Computational methods

2.1.1 Preparation of input models

Initial structure geometries for simulations in this study were obtained from the high-resolution crystal structures, which can be found in Table 2.1. Missing atoms and residues were rebuilt using the default settings from Modeller v9.11 (209) and Chimera (210) (Table 2.1). Residues were protonated according to their states at pH 7.0 using the PDB2PQR server (211). The residue names of the metal coordination atoms were manually changed according to our generated FFs (212).

Artificial metal-bonds were generated for *PfA*-M1 (ZN1085-H496_{NT}, ZN1085-H500_{NB}, ZN1085-D519_{OB}) and the equivalent conserved residues (HXXHX₍₁₈₎E) for other M1 family members. *PfA*-M17 had artificial bonds between ZN1-D379_{OA}, ZN1-D379_{OA}, ZN2-K374_{NW}, ZN1-D459_{OC}, ZN1-E461_{OD}, ZN2-D399_{OB}, and ZN2-E461_{OD}. TIP3P water model was used and no artificial bonds were introduced for water molecules (213). All the atoms and residues involved in the artificial metal-bonds were then assigned new amber internal atom types and new residue names were generated in order to differentiate the residues from the classical AMBER FF.

The residues with artificial bonds to the metal ions were capped with acetyl (ACE) and N-methylamine (NME) by the metal center parameter builder (MCPB). Single point calculations of the capped structures were performed using B3LYP function and 6-31G* Gaussian basis set and provided optimized metal binding geometry (214). The charge distribution of the artificial bonds (including the side-chains of the metal coordinated residues and the metals) was obtained using charge model B (ChgModB) of RESP (203). For protein backbone atoms involved in artificial bonds to metal, the atomic charges were taken from the AMBER94 FF (215), as required by MCPB. The charges of the two generated FFs (metals and side-chains) are provided in Amber prepri format while force constants of the metal-binding area are in Amber 'frcmod' format. These files are available at doi.org/10.1080/07391102.2017.1364669 (212)

Table 2.1. Structures simulated in this study.

Protein	Abbreviation	PDB ID / resolⁿ / chain Missing residues rebuilt by Modeller	Ref	Simulⁿ architecture	Simⁿ time (ns)
<i>P. falciparum</i> M1	<i>PfA</i> -M1 closed	3EBG, 2.10 Å, Chain A. No missing residues.	(4)	VLSCI <i>avoca</i> Blue Gene x86, 2.15 G Hz, 512 cores	500
	<i>PfA</i> -M1-bestatin closed	3EBH, 1.85 Å, Chain A. No missing residues.	(4)	Desktop 1*GTX1080, 1.71GHZ	500
<i>P. falciparum</i> M17	<i>PfA</i> -M17 hexamer	3KQZ, 2.39 Å, Chains A-F. Chain A: K257-M261; Chain B: T255-E262; Chain C: No missing residues; Chain D: T255- V259; chain E: T255-M261; T255-M261.	(81)	NCI Rajin hp cluster x86, 2.8G Hz, 512 cores	400
	<i>PfA</i> -M17 monomer	3KQZ, 2.39 Å, Chain A. K257-M261.	(81)	VLSCI <i>avoca</i> Blue Gene x86, 2.15 G Hz, 512 cores	200
<i>E. coli</i> APN	<i>Ec</i> APN	2DQ6, 1.5 Å, Chain A No missing residues.	(158)	VLSCI <i>avoca</i> Blue Gene x86, 2.15 G Hz, 512 cores	200
<i>Porcine</i> APN	<i>p</i> APN	4FKE, 1.85 Å, Chain A No missing residues.	(144)	VLSCI <i>avoca</i> Blue Gene x86, 2.15 G Hz, 512 cores	200
<i>Human</i> ERAP1	<i>h</i> ERAP1 open	2YD0, 2.7 Å, Chain A C486-G514; S553-P557.	(5)	VLSCI <i>avoca</i> Blue Gene x86, 2.15 G Hz, 512 cores	200
	<i>h</i> ERAP1 closed	3MDJ, 2.95 Å, Chain A H417-F433; C486-R515; G552-G555; F864- G867; G893-R906	(216)	Desktop 1*GTX1080, 1.71GHZ	200
<i>Human</i> APN	<i>h</i> APN	4FYQ, 1.9 Å, Chain A Y891-S899.	(1)	VLSCI <i>avoca</i> Blue Gene x86, 2.15 G Hz, 512 cores	200
<i>Human</i> APA	<i>h</i> APA	4KX7, 2.15 Å, Chain A S608-P611.	(147)	VLSCI <i>avoca</i> Blue Gene x86, 2.15 G Hz, 512 cores	200
<i>Thermoplasma acidophilum</i> TIFF3	TIFF3 open	1Z5H, 2.3 Å, Chain A No missing residues.	(149)	VLSCI <i>avoca</i> Blue Gene x86, 2.15 G Hz, 512 cores	200
	TIFF3 closed	1Z5H, 2.3 Å, Chain B No missing residues.	(149)	Desktop 1*GTX1080, 1.71GHZ	200

Calculating the correlation between QM and MM frequencies

The frequency modes of the FF areas that were capped with ACE and NME for each system based on B3LYP/ 6-31G* QM calculation were obtained and visualized as QM frequency modes. The MM frequencies were obtained from normal mode (NM) calculations on the metal centers alone, from a globally minimized structure. The correlation between the data-sets (QM vs. MM) was calculated via linear regression in GraphPad Prism 7.

2.1.2. Bestatin parameters

The structure of peptidomimetic inhibitor bestatin was optimized by HF/6-31G* to collect RESP charges. The dihedral N2-C1-C2-C3 bonds were frozen during the Gaussian calculation to prevent the possible self-protonation of bestatin, where the ammonia group can transfer a proton to the carbonyl group. The bond constant parameters for bestatin were defined by Parm10 (217).

2.1.3. Input set up for MD simulations

Our M1 Zinc AMBER FF (ZAFF) and M17-ZAFF were used to define the metal centers of the proteins, whilst the remaining protein elements were defined by AMBER FF12SB (218-221). Each system was solvated in a water cubic box consisting of TIP3P water molecules (213) with Na⁺ ions added to neutralize any charge. For M1 systems, 13 sodium ions were added. For the M17 systems, 36 sodium ions were added to the hexamer and 6 to the monomer. The minimum distance from the surface of each complex to the faces of the water box was set to 12 Å. The M1 systems consisted of approximately 130,000 atoms with the periodic box of 117 × 125 × 120 Å³. The hexameric PfA-M17 system consisted of approximately 294,000 atoms with a periodic box of 167 × 166 × 121 Å³.

2.1.4. MD simulation protocol

All atom MD simulations were performed using NAMD 2.9 MD package (222) on the computer architecture outlined in Table 2.1. Equilibration was performed in three stages: first potential steric clashes in the initial configuration were relieved with 50000 steps of

energy minimization. Initial velocities for each system were assigned randomly from a Maxwell–Boltzmann distribution. Each system was then heated to 300 K over 0.1 ns, with the protein harmonic constraints of 10 kcal/mol, under the canonical ensemble (NVT) conditions. Following this, each system was simulated for another 500 ps under the isothermal-isobaric ensemble (NPT) conditions with applied constraints gradually reduced from 10 to 0 kcal/mol. For each system, I performed three replicates each with different initial velocities. The production stage was then run at constant temperature (300 K) and pressure (1 atm) by NPT simulations, using Langevin damping coefficient of 0.5 fs^{-1} . The integration time step of the simulations was set to 2 fs, and the nonbonded cut-off length was set to 10 Å. The thermostat and barostat control was used by Berendsen pressure compressibility at $4.57\text{E-}5 \text{ bar}^{-1}$, and Berendsen pressure relaxation time at 100 fs. For each simulated system, periodic boundary conditions (PBC) were used together with the Particle-Mesh Ewald (PME) method for electrostatic interactions (223). Conformations of the systems were saved every 10 ps for subsequent analysis. Equilibration data from prior to the production phase are shown in Table 2.2.

Table 2.2. The equilibration stage and production phase for each simulation.

Protein	Abbreviation	Production phase (3 replicate)	
		Equilibration (ns)	Data collection (ns)
<i>P. falciparum</i> M1	<i>PfA</i> -M1 closed	0 - 50	50 - 500
	<i>PfA</i> -M1-bestatin closed	0 - 50	50 - 600
<i>P. falciparum</i> M17	<i>PfA</i> -M17 hexamer	0 - 50	50 - 400
	<i>PfA</i> -M17 monomer	0 - 100	100 - 200
<i>E. coli</i> APN	<i>EcAPN</i>	0 - 25	25 - 200
Porcine APN	<i>pAPN</i>	0 - 25	25 - 200
Human ERAP1	<i>hERAP1</i> open	0 - 25	25 - 200
	<i>hERAP1</i> closed	0 - 25	25 - 200
Human APN	<i>hAPN</i>	0 - 25	25 - 200
Human APA	<i>hAPA</i>	0 - 25	25 - 200
<i>Thermoplasma acidophilum</i> TIFF3	TIFF3 open	0 - 25	25 - 200
	TIFF3 closed	0 - 25	25 - 200

2.1.5. RMSD calculations

To check structural stability during the simulations, root-mean-square deviation (RMSD) conformational analysis was performed using the gromos method (224) implemented in GROMACS (gmx rms) (225, 226). The reference structure for each system was the starting crystal structure (Table 2.1) and the backbone atoms ($C\alpha$, N, C, and O) were used to measure the deviations from the input model. All RMSD from the systems were calculated from production phase simulations. All the MD plots were generated by Xmgrace (227). The color codes for simulations are: black (run 1), red (run 2), and green (run 3).

2.1.6. RMSD clustering analysis

RMSD clustering analysis was carried out all proteins simulated. To perform this analysis, each snapshot from the trajectory is superimposed using all of the $C\alpha$ atoms from the enzyme to remove overall translation effect. A least-squares-fitting procedure is performed to remove overall rotations. A value of 1.4 Å was chosen as the optimal RMSD cutoff after evaluation of the dependence of the cluster populations against the total number of clusters in the range 1.0-1.5 Å. In the GROMOS-clustering algorithm, the conformation with the highest number of neighbors, identified within the chosen RMSD cut-off, is chosen as the center of the first cluster. All the neighbors of this conformation are then removed from the ensemble of conformations (224).

2.1.7. RMSF calculations

The root-mean-square fluctuations (RMSF) were calculated from the production phase (Table 2.2) of the MD trajectories using the GROMACS 'gmx rms' command. The fluctuations were examined on a residue basis and plotted along the residue index of each simulation system.

2.1.8. Radial Distribution Function calculations

The radial distribution function (RDF) or pair correlation function $g(r)$ was used to calculate how the density of water molecules varies as a function of distance from the zinc

ion(s). For all systems the Gromacs command 'gmx rdf' was used to calculate the RDF and zinc ions were selected as the reference particle and waters as the selected paired particle type.

2.1.9. Principle Component Analysis (PCA)

PCA was based on the calculation of diagonalization of the atom displacement matrix C , whose elements are defined as

$$C_{ij} = \langle [(X_i - \langle X_i \rangle_t)(X_j - \langle X_j \rangle_t)] \rangle_t, \quad (i, j = 1, 2, 3, \dots, 3N) \quad (1)$$

where the angle brackets $\langle \rangle_t$ indicate the time-average, and i and j run from $1 \dots 3N$, i.e. over all three dimensions of the N atoms considered. Where X_i and X_j are Cartesian coordinates of the $C\alpha$ atoms, N is the total number of the $C\alpha$ atoms examined and X_j represents the time, averaged over all the configurations captured in the simulation.

$3N$ eigenvectors \mathbf{v}^i are ranked by eigenvalues λ_i , thus λ_1 represents the largest eigenvalue. The eigenvectors and eigenvalues of C yield the modes of collective motion and their amplitudes, respectively. The principle components (PCs) of the data $\mathbf{X} = (X_1, X_2 \dots, X_{3N})^T$ can be used for subsequent essential dynamics analysis (EDA). For EDA, the trajectories were projected on to the studied V_i of the PC_i to represent the atomic dynamics found along the PC_i .

$$V_i = \mathbf{v}^{(i)} \mathbf{X} \quad (2)$$

PCA was performed by using `g_covar` command of GROMACS 4.07 (195, 228). The modes generated by the first two (and largest) eigenvalues of each system were analyzed using the "NMWiz" plugin of VMD and ProDy software (229). To avoid disturbance by highly flexible residues, PCA calculations were performed using the snapshots from the production phase of the MD simulation.

2.1.10. Calculating free energy landscapes

The energy landscapes were calculated by the following equation

$$\Delta G(r) = -K_b T [\ln P(r) - \ln P_{max}] \quad (3)$$

Where $\Delta G(r)$ is the free energy of the molecule,

K_b is the equilibrium constant,

K_b is the probability distribution of the molecular system along the theoretical multidimensional coordinates (r) of the molecule,

P_{max} denotes the maximum probability, which is normalized by subtracting energy of P_{max} to ensure that $\Delta G = 0$ for the minimum free energy.

To reduce the dimensionality of the multidimensional coordinates (r), we used a 2D projection of the PCA. The free energy surface on the first 2 dimensions is given by (230)

$$\Delta G(v^1) \propto -K_b T \ln \int dv^2 P(v^1, v^2) \quad (4)$$

Where v_1 and v_2 are the first two eigenvectors. The 2D projection of $V(v^1, v^2)$ was finally converted to an energy landscape by SigmaPlot.

2.1.11. Molecular modelling of PfA-M1-Hb6

Snapshot models of PfA-M1 were obtained from the run 2 of the unbound PfA-M1 MD simulation at $t=58$ ns of run2, and $t=384$ ns of run3 and were aligned to the pAPN template structure PDB ID 4NAQ using Pymol 1.8 (231). The structure alignments were both centered on the catalytic domain of the proteins. The peptide substrate in 4NAQ was manually truncated to six residues and the sequence mutated to that of Hb6 (D₂P₃E₄N₅F₆) using Pymol 1.8 (231). The N-terminal valine residue (V₁) was modeled manually with reference to other peptide-bound M1 aminopeptidases (144, 216, 232). The Leap program of AmberTools15 (221) was used to add N- and C-terminal groups of the modelled substrate. The Amber14SB FF (221) was used to define the system, with the exception of the metal center where we used the M1-ZAFF generated in this study. A TIP3P solvent box was used to fully solvate the two-modeled complexes.

Three steps were used to optimize the models produced. First, the models were minimized by minimizing the water molecules by restraining the protein for 50000 steps, the water molecules were relaxed for 20000 steps using a NTP ensemble after the 50000 cycles

of system minimization on the whole system. Second, the model was heated from 0 to 300 K within 50 ps by a NVT ensemble with the restraint of 10 kcal/mol. Third, the whole system was relaxed by NPT simulation at 300 K for 500 ps with restraints gradually released from 10 kcal/mol to 0 kcal/mol on the protein heavy atoms. A 5 ns NPT simulation was used to equilibrate the system with no restraints. The last frame of each simulation was extracted as the final models of *PfA-M1c-Hb6* and *PfA-M1o-Hb6*.

2.1.12. Steered MD (SMD) simulations

SMD simulations were performed to apply a directional force or ‘pull’ to the Hb6 peptide and observe the dissociation of the peptide. Two directional forces were applied to the peptide bound in both the *PfA-M1c-Hb6* and *PfA-M1o-Hb6* models (Table 2.3). All SMD simulations were performed using the CUDA accelerate version of the NAMD 2.11 MD package (222). The configuration files contained the direction vector, spring constant and velocity (Table 2.3). SMD simulations were performed for 3 ns.

Table 2.3. Details of SMD simulations performed in this study.

SMD	Starting model	Direction/vector	Spring constant (kcal / mol / Å ²)	Velocity (Å / ps)	Channel length (Å)
1	<i>PfA-M1c-Hb6</i>	C-terminal pulling force Atom A: <i>PfA-M1</i> :Y ₅₈₀ @CB Atom B: Hb6: F ₆ @Cα Direction \overline{AB} : (1.886, -4.276, -0.284)	10	0.007	59
2	<i>PfA-M1c-Hb6</i>	N-terminal pulling force Atom A: <i>PfA-M1</i> : L ₄₂₂ @CD1 Atom B: Hb6: F ₁ @CB Direction \overline{AB} : (9.144, 15.758, 5.807)	10	0.007	49
3	<i>PfA-M1o-Hb6</i>	C-terminal pulling force Atom A: <i>PfA-M1</i> : Y ₅₈₀ @CB Atom B: Hb6: F ₆ @Cα Direction \overline{AB} : (-22.762, 11.851, -6.591)	10	0.007	60
4	<i>PfA-M1o-Hb6</i>	N-terminal pulling force Atom A: <i>PfA-M1</i> : A ₃₂₀ @Cα Atom B: Hb6: F ₁ @CB Direction \overline{AB} : (-12.603, -3.814, -5.462)	10	0.007	44

2.1.13. Calculating the external force and work of SMD

To calculate the external force and work of the SMD simulations, an external work analysis was performed. The external forces along the trajectories were extracted along the SMD simulation time to the SMD log files. The displacements along the direction of external force were also generated and recorded in the SMD log files. The configuration file originally from the NAMD SMD packages, was manually edited. To control the velocity and direction of the force, I edited the SMD configuration file performed these simulations. There are only four inputs to this script. Input 1 (script lines 21-23) is the coordination of the first atom (A), which indicates the atom I applied the external force to. The second input (script lines 25-27) is the coordination of a second atom (B). The two atoms (AB) determine the direction of the external force (\overline{AB}). Input 3 (script lines 30) is the external force (in kcal/mol/Å² for constant force or in Å / timestep for the pulling velocity) that describes the quantity of the external force. From there, I scripted lines 46-80 to project the force information onto X, Y and Z with the direction defined by input 1 and 2.

2.1.14. MD analysis

Dynamics cross-correlation matrices (DCCM) (233) were produced from the simulations of PfA-M1 (Chapter 3) by using Bio3D v2.4.0 package (<http://thegrantlab.org/bio3d/>) from R studio (233, 234). Dyndom server (235) and Hingefind tcl script (236) were used to predict the hinge residues for the M1 homologs. To run the Dyndom server, the two extreme coordinates detected from each PC of M1 simulations were used as the two different conformations, and uploaded to Dyndom server. The two rigid domains were then recognised and depicted in blue (steady region) and red (moving region) regions. The Hingefind tcl script was then used in the TCL console of software VMD1.9.3 (237). The two conformations for Dyndom server were loaded into VMD and superposed automatically by the Hingefind tcl script, and the coordinates of the hinge helices were then generated.

All the simulation trajectories were analyzed using the GROMACS 5.14 simulation package (226). Graphs and plots were produced with Xmgrace (227) and GraphPad Prism7.

Molecular graphics were prepared with PyMOL 1.8.23 (231). For the images of the M1 aminopeptidase protein structure, a standard color code was generally used where domain I is in blue, domain II green, domain III yellow, and domain IV red. The protein assembly interface of *PfA-M17* were explored by PDBePISA (238). PROMALS3D server was used to perform the structural sequence alignment of the selected M1 proteins (239).

2.2. EXPERIMENTAL METHODS

The methods described in Sections 2.2.1, 2.2.2 and 2.2.3 were performed by other members of the McGowan lab and are briefly included for clarity. I either performed or assisted with experiments outlined in Sections 2.2.4 and 2.2.5.

2.2.1. Site-directed mutagenesis

Amino acid substitutions were performed using a PCR-based method. Nucleotide changes to the *PfA-M1* gene within the expression plasmid, pSM139, were introduced using primers complementary to the *PfA-M1* sequence, except at the position of the desired mutation (lower case letter in Table 2.4). The presence of mutants was confirmed via DNA Sanger sequencing.

Table 2.4. Oligonucleotides used for *PfA-M1* mutants

<i>PfA-M1</i> mutagenesis	Sequence (5' – 3')
A453P For	CTTGAACCAAgGACAAAGGCTG
A453P Rev	GTCGGAAACAGgAACCAACTTG
E463Q For	GTTGGTGCTATGcAAAACAAGGG
E463Q Rev	CCCTTGTTTTgCATAGCACCAAC
E497Q For	GGTCACcAGTACTTCCACCA
E497Q Rev	TGGTGGAAGTACTgGTGACC

2.2.2. Protein production

PfA-M1 and mutants were expressed in *Escherichia coli* BL21 and purified by the Monash University Protein Production Unit via a two-step purification (HiTRAP Ni-NTA-agarose column followed by Superdex-200 size-exclusion chromatography). Purity of the purified protein samples was analyzed by 12 % SDS-PAGE. Protein concentration was determined via a NanoDrop, measuring absorbance at 280 nm.

2.2.3. Enzyme kinetics

Aminopeptidase activity was determined by measuring the release of the fluorogenic-leaving group, 7-(4-methyl) coumarylamide (NHMeC), from the fluorogenic peptide substrates H-Leu-NHMeC (10 μ M). Assays were carried out in 96-well microtiter plates (200 μ L total volume) in 50 mM Tris-Cl, pH 8.0 at 37 °C and activity was monitored until steady-state was achieved. Fluorescence was measured using a FluoroStar Optima plate reader (BMG Labtech), with excitation and emission wavelengths of 355 nm and 460 nm, respectively. The Michaelis constant, K_M , was calculated from the initial rates over a range of substrate concentrations (0.2–500 μ M) with a fixed enzyme concentration (20 nM for the WT, 125 nM for A453P, 25 μ M for E463Q and E497Q mutants). The enzyme turnover rate, k_{cat} , was calculated from Michaelis-Menten with non-linear regression protocols by using GraphPad Prism 7.

2.2.4. Aminopeptidase activity against Hb6 substrate

The peptide H-LDPENF-OH (Hb6) was purchased from Mimotopes and solubilised in MQ water to a concentration of 4.1 mM. Purified *PfA-M1* protein in 50 mM Hepes, pH 8.0, 0.3 M NaCl, 0.25 M imidazole, 5% (v/v) glycerol was concentrated to 38.7 μ M. 100-fold molar excess of Hb6 peptide (4.2 mM) was added to *PfA-M1* (42 μ M) and the reaction allowed to proceed for 2 hrs at 37°C. The reactions were stopped by heat treatment at 100°C for 10 mins. Resulting peptide products were analyzed by Liquid Chromatography-Mass Spectrometry (LC-MS). LC-MS was performed by the Monash Proteomics Platform who mixed each sample with a 1:1 ratio of Matrix, 10 mg/mL α -cyano-4-hydroxycinnamic acid in 50 % acetonitrile 0.1% TFA and spotted onto the MALDI target plate. Samples were

subsequently analyzed in reflector mode with a mass range of 100 Da to 1000 Da using plate model calibration against 4700 mix peptide standards by using the Instrument of 4700 Proteomics Analyzer MALDI TOF/TOF (AB Sciex). Data processing were processed with 4000 Series Explorer version 3.0 and calibrated against 4700 mix peptide standards.

2.2.5. Protein Crystallography

PfA-M1 and mutants were crystallized by hanging drop using previously established reservoir conditions that ranged from 12 – 22 % PEG8K, 0.2 M MgCl₂, 0.1m Tris pH 7 - 8.5 with 5 % v/v glycerol buffer. Co-crystallography experiments involved the addition of the peptide Hb6 to the drop at a molar ratio of 20:1 peptide: protein. Crystals grew within two days and were dehydrated against reservoir buffer with 15 % v/v glycerol for 16 hours prior to data collection. Crystal soaks were performed with the peptide Hb6 at a molar ratio of 5:1 and soaks were performed for various times (10 mins – 2 hrs) prior to data collection (after dehydration step). Diffraction data was collected under a continuous liquid nitrogen stream at the MX1 beamline at the Australian Synchrotron.

2.2.6. Structure refinement

Diffraction images were processed by using iMOSFLM (240) and from the CCP4 6.5 suite (241). The FreeR set was obtained from previous structure solutions and 5 % of each dataset was flagged for the calculation of an R_{free} (242) with no low resolution cut off applied to the data. Structure solution used the molecular replacement program PHASER and the crystal structure of *PfA*-M1 (PDB ID 3EBG) was set as the search model. Phenix 1.9.1 interface was used to perform refinement and COOT 0.82 was used for inspection and manual (243). Waters were added to the model when the R_{free} was < 30 and hydrogen-bonding geometry contacts of 2.5–3.5 Å were used to retain solvent molecules in the model. The quality of the final model was assessed using MolProbity (244).

CHAPTER THREE
INVESTIGATING THE PROTEIN DYNAMICS OF *PfA-M1*

3.1. INTRODUCTION

The MA clan, family M1 aminopeptidases are a large superfamily (245). X-ray crystal structures from higher order species (including insects, animals and humans) indicate that the proteases are able to undergo a large conformational change from a 'closed' to 'open' form (5, 144, 149, 216, 246, 247). Given the fact that the active site of the enzyme is buried, this motion has important implications for the proteolytic mechanism as well as substrate entry and product egress. It is generally accepted that the open form of M1 is in an inactive state that allows substrates direct access to the catalytic domain through the gap produced as the C-terminal domain IV moves away from the body of the molecule. The closed form of M1 is the active form and has an ordered active site and specificity pockets, ready to position the substrate for hydrolysis (144, 149, 246).

For *PfA-M1*, the crystal structure clearly showed a closed conformation of the aminopeptidase and no 'open' conformation has ever been observed (4, 248). A previous classical all-atom MD study on *PfA-M1* failed to identify any large domain movements in the protein (208). This study, however, performed a single replicate on a relatively short time scale (simulation time = 75 ns) (208). Given the size of the protein system (890 residues), the time-scale sampled may preclude interpretations on the basis of slow collective domain dynamics, which are often accessed at higher timescales. Therefore, to date, it is assumed that *PfA-M1* cannot undergo the conformational change that other members of the M1 aminopeptidase family can perform. This then prompts the question, how else might *PfA-M1* differ from the superfamily and what are the implications of this variation for its biological function and possible future drug development?

In this chapter, I present the findings of an MD study where I performed three independent all-atom MD simulations on *PfA-M1*. As *PfA-M1* contains an essential zinc atom, I used my own hybrid bonded/nonbonded model for the generation of the metal parameters (212). The results show that the hypothesis that *PfA-M1* does not undergo a large conformational change was incorrect as one of our trajectories clearly showed a large 'opening' motion. The implications of this discovery are discussed.

3.2. RESULTS

3.2.1 MD simulations

In this chapter, I performed three independent MD simulations on *PfA*-M1 (time = 500 ns each) to investigate the protein dynamics that may contribute to enzyme function. I assessed the overall stability of the *PfA*-M1 MD simulations by analysing the root mean square deviation (RMSD) and root mean square fluctuations (RMSF) of the C α atoms of *PfA*-M1. The inspection of the RMSD shows that *PfA*-M1 reached equilibrium after 50 ns (Fig. 3.1A). A single stable state (from 100 – 500 ns) was observed on the first and second runs, during which minor rearrangements were observed (1.8 ± 0.2 Å and 2.3 ± 0.3 Å, respectively). The third *PfA*-M1 replicate displayed two equilibrium states, the first from 50 to 240 ns and a second from 300 to 500 ns (Fig. 3.1A). The large RMSD variation (~ 4.1 Å) between the conformations from each state suggests the occurrence of a large conformational change in this trajectory. The RMSF of the stable phase (50 - 500 ns) of each trajectory show that domains I and IV were markedly more flexible in run 3 (green line, Fig 3.1B) than both runs 1 and 2. Taken together, the two equilibrium states of run 3 may be a result of large rigid body changes of position by domains I and IV.

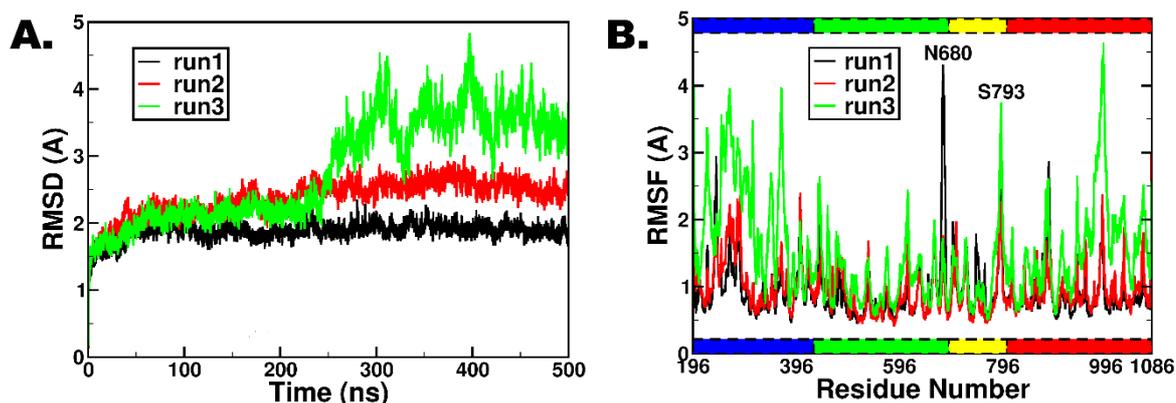


Figure 3.1. The RMSD and RMSF of the C α atoms of PfA-M1 in MD simulations. (A) RMSD and **(B)** RMSF from three MD simulations (run 1 = black, run 2 = red and run 3 = green). The color bars on the top and bottom of panel B indicate the domain boundaries where domain I is blue, II is green, III yellow and IV red.

3.2.2. Nucleophilic water behavior throughout simulation.

For the MD simulations, I chose to use my hybrid bonded/non-bonded model to generate the metal FF parameters for M1 aminopeptidases for two major reasons. First, that a bonded model of protein-metal interactions is the current gold standard for parameterizing metals in MD (201, 249, 250) and secondly, to leave the waters in the simulation as non-bonded interactions. The solvent water molecules in our simulation were all treated equally as non-bonded TIP3P waters. The reason I wanted to do this was to accurately depict a functional active site of the protease (where there is a nucleophilic water present) as well as allow the parameters to be used later for modelling inhibitors within the active site (where I want to be able to investigate their possible dissociation). Therefore, for the PfA-M1 simulations, I needed to ensure that the parameters would accurately reproduce the metal-water coordination accurately and realistically represent a functional metalloprotease. To assess the water behavior within the active site during the MD simulations, I performed a radial distribution function analysis (RDF) for all the water molecules that were within 14 Å proximity of the catalytic zinc ion (Fig 3.2). The RDF plot confirmed that throughout all of my simulations and for the duration of the time sampled, one water molecule was located within the first shell of zinc ion (2.0 ± 0.08 Å, Fig 3.2). This

water is the catalytic water (Fig 3.2) and bridges between E497 and the zinc. The water is exchanged ~ 30 times in each replicate simulation. It is worth noting that the distribution of water 1 has a RDF water number of 1.25 (Fig 3.2). This could be explained by the process involved in the water site substitution at this position ($2.0 \pm 0.08 \text{ \AA}$ to zinc ion) in some of the snapshots found from the trajectory where two waters may be each located within the first shell were captured from some snapshots, making the RDF water number greater than 1. A second water molecule was also present in the second shell from the zinc ion ($2.5 \pm 0.09 \text{ \AA}$) throughout the entire simulation and was coordinated by E463 and E497 (Fig 3.2).

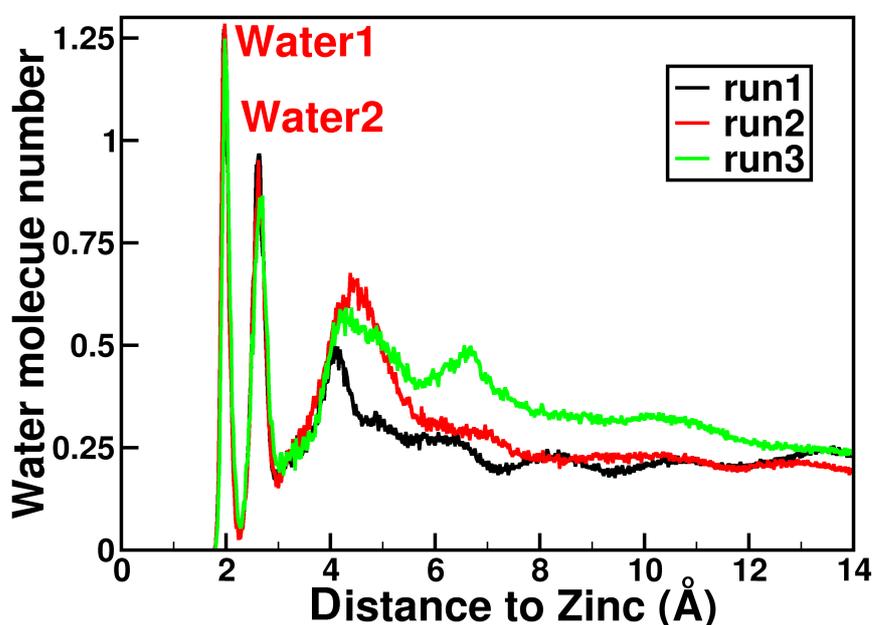


Fig 3.2. RDF of water molecules in proximity to zinc ion.

3.2.3. Dynamic cross-correlation matrix of *PfA-M1* simulations.

I was interested to know what types of coarse-grain dynamics the domains of *PfA-M1* underwent in the simulations, so I generated a residue-by-residue dynamic cross-correlation matrix (DCCM) for each trajectory. The DCCM shows the relationship between the movements of residue pairs and summarises which direction the residue pair is moving relative to each other (Fig 3.3A). The DCCM does not provide exact movements but can define a detailed elastic network for each trajectory to get direct visualisation of the

movement relationships (Fig 3.3B) (251). From this analysis, run 1 shows no major correlated motion (Fig 3.3B), whilst run 2 exhibits a weak dynamic cross-correlation in domain I. In contrast, run 3 showed strong correlation of domain I (residues 196 to 391) and the C-terminal region of domain IV (residues 800 to 1084). In run 2, a weak inverse correlation between domain IV residues 750-990 and domain III was observed. The cross correlations observed indicate that slow frequency motions of domains I and IV may be occurring in runs 2 and 3.

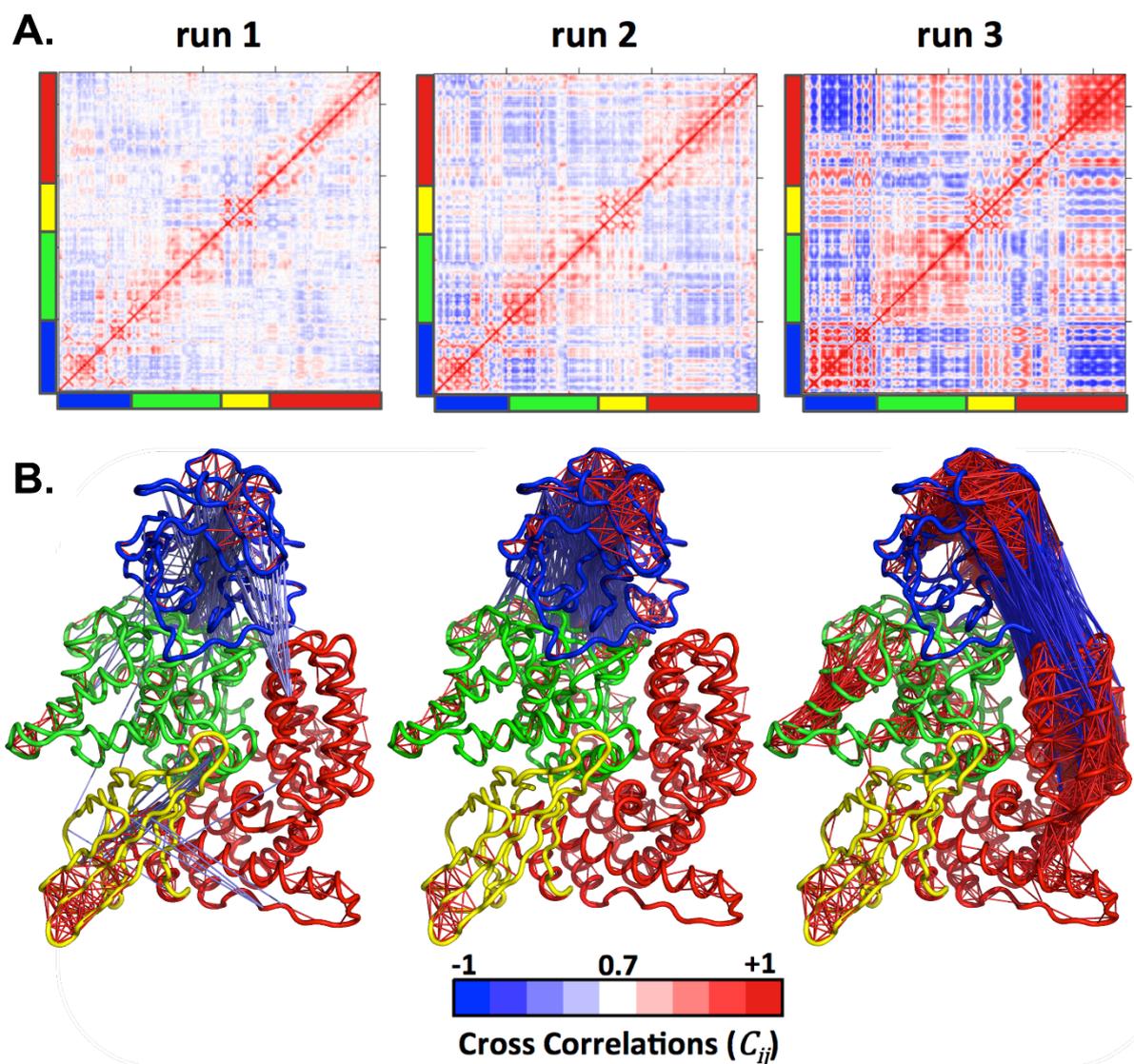


Figure 3.3: Mapping the correlated movements of *PfA-M1*. (A) Residue-residue DCCM was generated from each MD trajectory, as indicated above in each heatmap. *PfA-M1* domains are mapped onto the heatmap by color code on each axis. The DCCM of *PfA-M1* simulations show residues that move in the same direction (red), inverse or opposite direction (blue) and no correlated movement (white). The color scale is indicated at the bottom of panel B. (B) Visualization of the cross correlations plotted onto the *PfA-M1* structure with the same color code as in A (red = same direction, blue = opposite direction).

3.2.4. Principle component analysis of *PfA-M1* simulations.

To identify the essential dynamics in the simulations, we performed a Principal Component Analysis (PCA) by deconvolution of the DCCM. To focus on the major movements observed throughout the simulation, we only analyzed the top two principle components (PCs) from this analysis and projected the trajectories (runs 1, 2 & 3) onto their PCs, respectively, to make a porcupine plot (Fig 3.4). For run 1, PC1 and PC2 together accounted for 26 % of the total covariance, run 2 was 36 % and run 3, 69 % (Fig 3.4). As apparent from the DCCM and the low combined percentage of PC1 and PC2 of run 1 and run2 can be seen to show little movement throughout the trajectory (Fig 3.4A). The PC1 of run1 indicates movement in domain III, where the loop encompassing residues 680-690 moves away from the body of the molecule (Fig 3.4A). PC2 of run 1 report less movement than PC1 and described a rigid body movement of the entire of domain I where the domain moved up and down relative to the body of the molecule (Fig 3.4A). In both runs 2 and 3, PC1 accounts for a higher proportion of the overall covariance (Fig 3.4B & C). For run 2, the movement described by PC1 shows that domains I and IV move toward each other to a “more” closed structure, relative to the starting position (as defined by the X-ray crystal structure) (Fig 3.4B). The PCA of run 2 is interesting as it indicated that the apparent closed form of *PfA-M1* in crystal structure can move to an even more compact “closed” conformation. In PC2 of run 2, we saw a similar movement to that of PC2 in run 1, with a rigid body, hinging movement of domain I. In run 3, the same movement along the PC1 vector significantly exaggerates this hinging motion, allowing the protein to sample an ‘open’ conformation where domains I and IV move away from each other (Fig 3.4C). The PC2 in run 3 describes a twisting motion between domains I and IV (Fig 3.4C). The PCA of run 3 is particularly fascinating as it shows an open form of *PfA-M1*, indicating that the hypothesis that *PfA-M1* does not open is wrong.

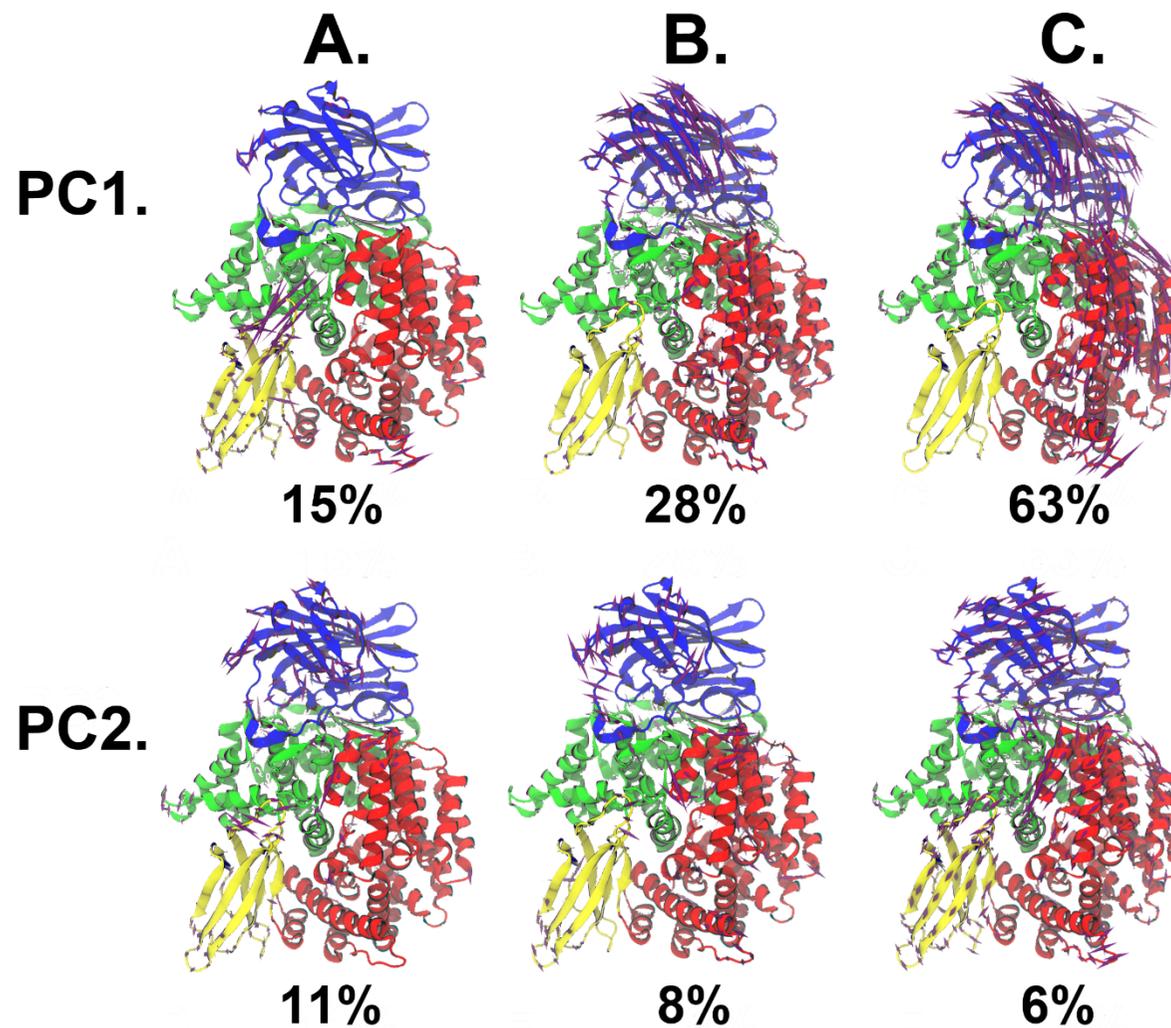


Figure 3.4. A porcupine plot of PC1 and PC2 from run 1 (A), run 2 (B) and run 3 (C). Motion > 1.5 Å is shown as unidirectional purple spikes where the length of each spike is proportional to the range of motion.

3.2.5 Mapping the movement and free energy landscape of run 3.

To understand how *PfA-M1* can move from a closed state, similar to the crystal structure, to the open state that was sampled in run 3, I projected the run 3 trajectory onto the PC1 and PC2 vector defined subspace (Fig. 3.5A). Then by coloring each snapshot of the trajectory dots by its own time step, I was able to visualize the pattern of movement observed during run 3 (Fig 3.5A). The first motion observed is along PC2 with a 15° clockwise rotation of domain I and an equivalent and simultaneous counter-clockwise shift in domain IV (step a to b, Fig 3.5). This movement occurs in the first 50 ns, before the system reached its first equilibrium stage, and may represent the release of crystal contacts from the starting position. Next, *PfA-M1* transitions from a closed state to an open state (steps c – e, Fig 3.5) by rotating domain I up by 33 degrees and domain IV down by 26 degrees. The opening motion is along the PC1 vector. Finally, the protein reverse steps a and b to twist domains I and IV by 15 degrees in opposite directions, whilst retaining an overall open conformation (step f, Fig 3.5).

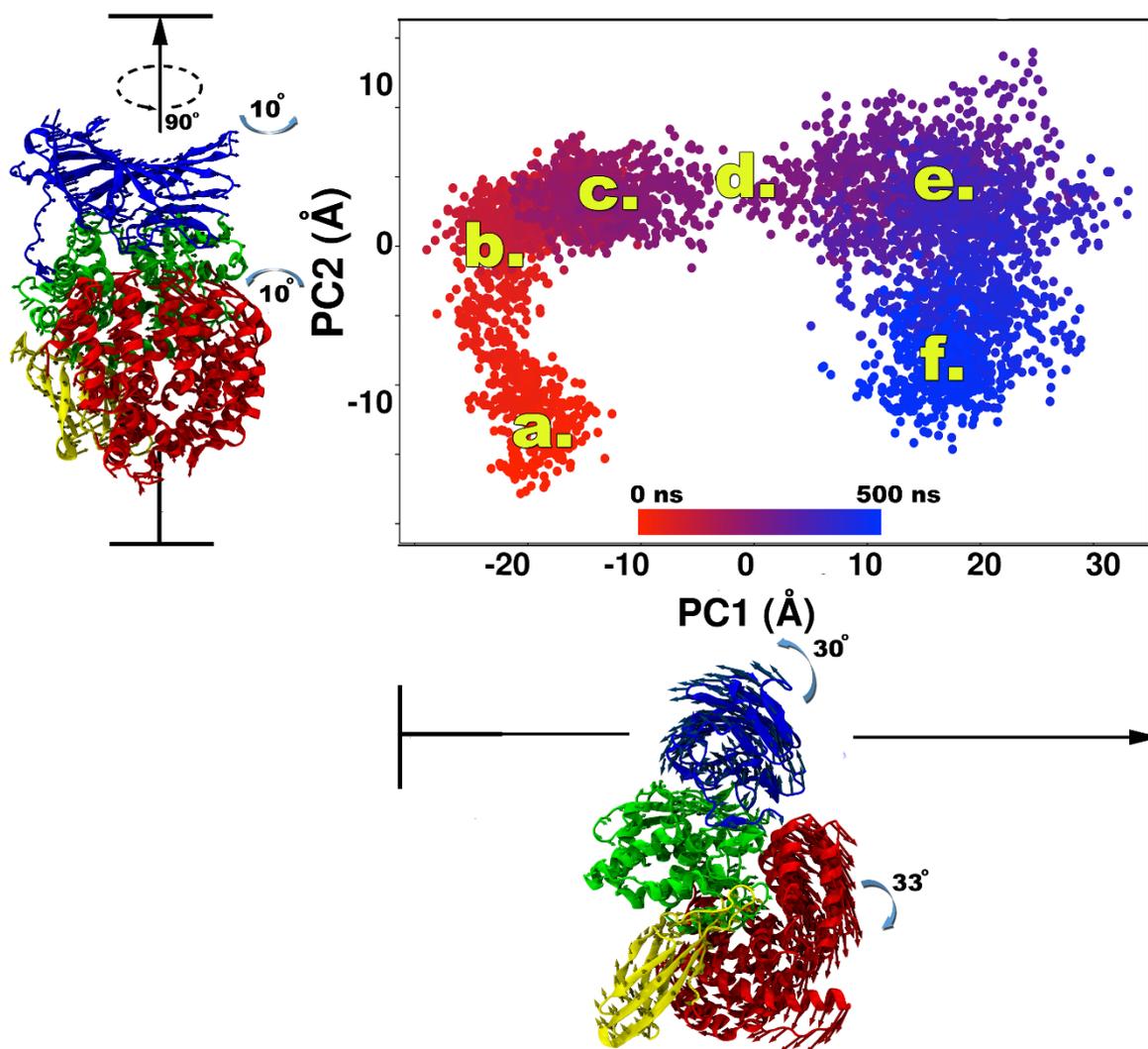


Figure 3.5. The 2D projection of run 3 with time stamps. Dot color is defined by simulation time from 0 (red) to 500 ns (dark blue). Selected time stamps in ns are indicated where a = 0, b = 50, c = 170, d = 236, e = 290 and f = 450. A porcupine plot of each PC is shown next to the relative axis to assist in visualization of the movements occurring. The porcupine arrows on the cartoon structures show the direction of movement and are colored by domain. Motion >1.0 Å is shown as directional arrows on C α atoms where the length of each arrow is scaled to the movement observed.

To understand any potential barriers to the opening and closing of *PfA-M1*, I produced a probability-dependent free energy landscape based on the PC1 and PC2 defined subspace (252) (Fig. 3.6) This plot describes a similar pattern of movement to that shown in Figure 3.5 but now assigns free energy of the pathway. The darker colored

snapshots indicate the conformations of low energies and provide a low energy pathway of the most possible movements (Fig.3.6A). Combining the time-lapse data from Figure 3.5 and 3.6A, I was then able to map the free energy versus the time of the simulation (Fig 3.6B). The results show that there is a large energy barrier to overcome to convert *PfA*-M1 from the closed to open form of *PfA*-M1 ($\Delta G = 4.53$ kT, Fig 3.6B). Firstly, the closed starting conformation (a) rapidly shifts to a more stable conformation (b) which has the lowest free energy of the system ($\Delta G = -9.63$ kT). From there, the protein transitions from a closed conformation to the open form (states b-e). In the conformational transition, the angle between domain I and IV increases by 66° and expose the catalytic domain to bulk solvent (state e, Fig 3.6B). The final protein transition occurs in the last 210 ns (e-f) of run 3 and has a final free energy of $\Delta G = -7.43$ kT (state f, Fig 3.6). During this time, the protein twists domain I and IV counter-clockwise by 15° and stays in the open form. The low free energy observed for the open conformations of *PfA*-M1 ($\Delta G = -7.43 - -8.1$ kT) provide some support that this form may exist in reality.

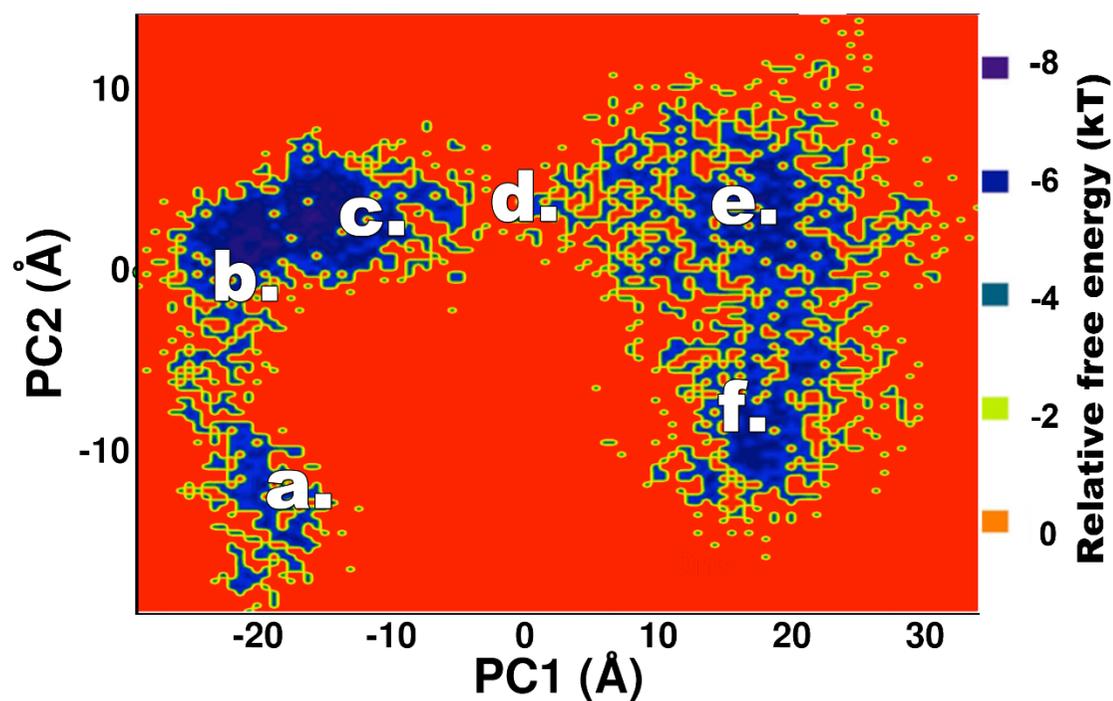
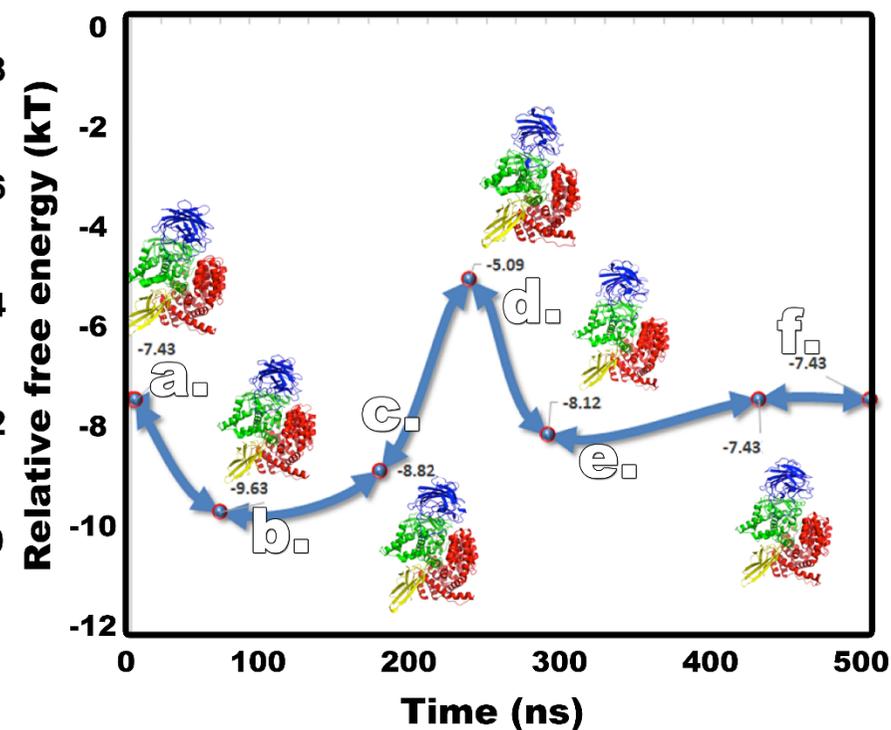
A. Free Energy Land Landscape (kT) of Run3**B. Free Energy Along Run3**

Figure 3.6. Probability dependent free energy plots of run 3. The energies were generated by converting probabilities of snapshots in the subspace defined by PC1 and PC2 (see method 2.1.9). **(A)** The free energy landscape contoured heatmap projected onto PC1 and PC2. Free energy unit is shown in kT and the color of the dots indicates energy from red (0 kT) to dark blue (-10 kT). **(B)** The free energy mapped against simulation time with states a to f shown as per Fig 3.5B.

3.2.6. Prediction of residues involved in the hinging motion of *PfA-M1*.

MD simulations identified that, compared to the crystal structure, *PfA-M1* can move from a more 'closed' conformation to a completely 'open' conformation. Members of the M1 aminopeptidase family protein have already reported both open and closed states (5, 144, 149, 216, 246, 247), which provide insight into their protease function. The next section describes the use of computational methods that would allow me to experimentally validate my computational prediction that *PfA-M1* can open and identify residues or regions within *PfA-M1* that may contribute to the dynamics of the protein.

The identification of potential "hinge" residues in *PfA-M1* would provide targets for mutagenesis to sabotage the hinge motion. If mutagenesis was able to alter the enzymatic activity or structure of *PfA-M1*, this would provide support for the fact that motion is intrinsic to *PfA-M1* function. To try and identify such residues I used the dyndom server (253) to detect the residues that may be involved in a hinge within *PfA-M1*. To achieve this, I uploaded the coordinates of the most closed state (run 2 at 343.4 ns) and the open state (run 3 at 348.4ns) that were produced by the simulations. The dyndom server performs an automated best fit alignment of the two structures and detects a fixed domain (blue), a moving domain (red), and a series of flexible residues that may bend in a hinge motion. The results from the server reported a rotation of 43.1° between the two states of *PfA-M1* and identified a series of residues that may be involved in the movement (Fig 3.7). Seven separate regions within *PfA-M1* were identified to potentially contribute to the open-to-close motion of the protein, however, 5 of these regions overlapped with the GAMEN motif (Fig 3.7A) potentially identifying this region as important for dynamics.

In order to narrow down the list of potential hinge residues, Hingefind (236) were also used to draw a hinge axis between closed and open state of *PfA-M1* sampled from the simulations (Fig 3.7C). The Hingefind algorithm performed its own best fit alignment of the two structures. The results showed the hinge axis only penetrates the M462 and A453 residues (Fig. 3.7C) These two residues were also within regions identified by the DynDom

server. Therefore, from this work, I wanted to investigate whether M462 and A453 are required for the movement of *PfA*-M1.

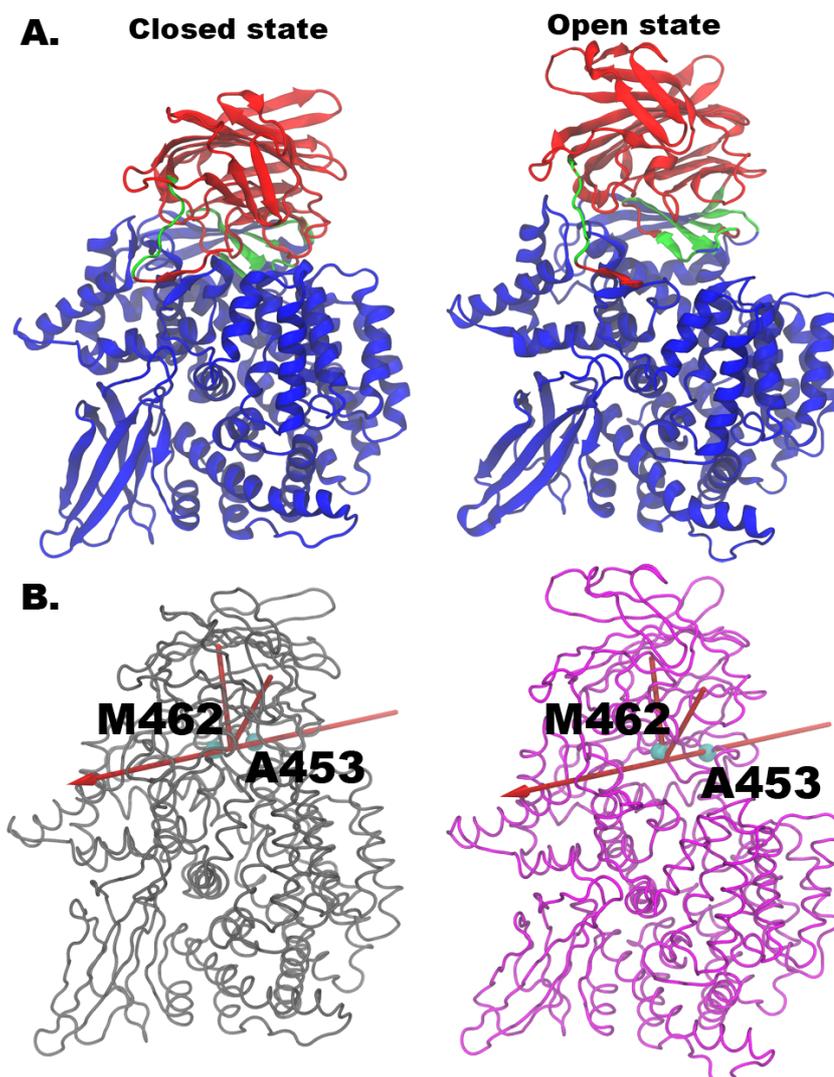


Figure 3.7. Dyndom and Hingefind prediction of the potential hinge residues in *PfA*-M1. (A) The closed and open states of *PfA*-M1 where hinge regions were detected by dyndom server. The fixed domain is shown in blue, moving domain is in red, and bending residues are in green. **(B)** Hingefind predicts the hinge axis (red arrow) between the closed (grey) and open (magenta) states of *PfA*-M1. The scale of the rotation is indicated by the two red cylinders. The Ca atoms of M462 and A453 were shown as cyan spheres. The plots were drawn with VMD using the Hingefind.tcl plugin (236).

3.2.7. Testing if residues M462 and A453 contribute to the dynamics of *PfA*-M1.

To test if M462 and A453 are involved in the dynamics of *PfA-M1*, I designed point mutations that would alter the nature and size of the residues. For M462, alteration to serine (M462S) or lysine (M462K) was selected so to shorten (serine) or charge (lysine) the side chain. The idea was that a serine would allow more room for the localized neighboring residues to move, as well possibly adding hydrogen bonds to the backbone of Q317_{OE1} or N464_{OD1} and thus limiting the movement of residue 462 (Fig 3.8). For M462K, providing a longer, positively-charged side chain may alter the electrostatics of the region as well as possibly introduce hydrogen bonds to Q317_{NE3} and C318_N, thus also potentially limiting movement (Fig 3.8). In position 453, I designed a proline substitution (A453P) to limit the flexibility of the main chain and reduce the capacity of the protein to lever domain I (Fig 3.8).

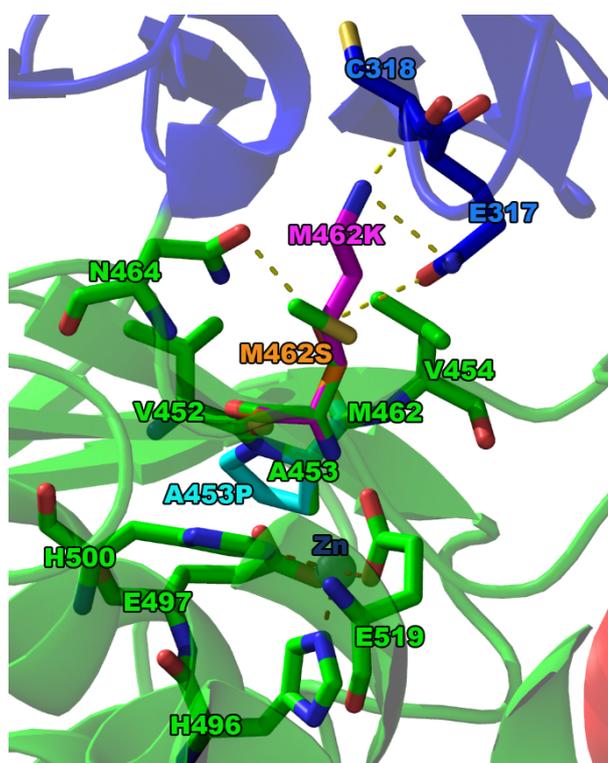


Figure 3.8. Design of potential hinge mutants in *PfA-M1*. The carbon atoms of the protein are colored by domain. M462S is shown in yellow sticks and labelled in yellow (M462S) and M462K is shown in magenta sticks and labelled in magenta (M462K). A453P is shown in cyan sticks and labelled in cyan (A453P). Hydrogen bonds are shown as yellow dashes, while the metal bonds are red dashes. The zinc ion was shown in silver sphere.

Site-directed mutagenesis was used to alter the *PfA-M1* gene. Ms Komagal Kannan Sivaraman (McGowan laboratory) generated these mutants, produced the recombinant protein and assayed the mutant proteins for their aminopeptidase activity to provide data on the affect the mutations had on the function of the protease. The enzymatic activity of each mutant is summarized in Table 3.1. The activity of the mutants was assessed in

comparison to that of wild-type enzyme. We found that the K_M values of the mutants were slightly improved compared to the wild-type enzyme ($\sim 1.5 - 2.0$ fold, Table 3.1), indicating that the mutants bind to substrate slightly better. However, the overall enzyme efficiency of mutants was generally reduced compared to the wild-type (Table 3.1, k_{cat}/K_M). This is a result of the reduced k_{cat} values obtained, indicating that the mutants do not turn-over substrate as rapidly as the wild-type protein. M462S was less efficient than either A453P or wild type with an overall efficiency of $1404 \text{ M}^{-1}\text{s}^{-1}$. The enzyme activity of M462K is only $9.81 \text{ M}^{-1}\text{s}^{-1}$, suggesting M462K hardly functions as a protease and implies that the sidechain change from hydrophobic to basic has impaired the enzyme turnover (k_{cat} only $0.083 \pm 0.001 \text{ s}^{-1}$). I found that all the mutants showed impaired enzyme activity indicating that the mutation had affected the catalytic mechanism, however, as the mutants were able to process substrates, it is likely that the essential elements of the active site were available.

Table 3.1. Aminopeptidase activity of wild-type *PfA-M1* and mutants.

<i>PfA-M1</i> protein	K_M ($\text{M} \times 10^{-6}$)	k_{cat} (s^{-1})	k_{cat}/K_M ($\text{M}^{-1}\text{s}^{-1}$)
Wild-type	120.1 ± 0.9	0.92 ± 0.003	7652
A453P	54.2 ± 1.8	0.11 ± 0.001	2080
M462S	59.4 ± 2.5	0.083 ± 0.001	1404
M462K	81.6 ± 1.7	$0.80 \times 10^{-4} \pm 0.006 \times 10^{-4}$	10

The enzyme activity of each mutant showed a deviation from the wild-type protein. To confirm whether this loss of function was a result of a structural change, I attempted to crystallize the mutant *PfA-M1* proteins. Ms Kannan Sivaraman purified the recombinant protein and I set up the hanging drop crystal trays with the assistance of Dr. Nyssa Drinkwater (McGowan laboratory). With the assistance of Drs Drinkwater and McGowan, we were able to successfully solve the structures of A453P and M462K to 2.00 \AA and M462S to 2.4 \AA (Table 3.2). The three structures were virtually identical to that of wild-type

(overall RMSD of 0.21 ± 0.03 Å) confirming that the loss of function observed was not a result of a loss in structural integrity. Excellent density was observed at each of the respective mutations (Fig 3.9). The change from A453 to P453 resulted in the loss of a single hydrogen bond with F413 (Fig 3.9A), and the other parts of the protein maintained exactly the same to the wild-type crystal structure. For M462K, I observed that the K462_{NZ} atom of the side-chain formed separate hydrogen bonds with the backbone of Y383, E463 and E319 (Fig 3.9B). The introduction of serine at position 462 changes the hydrophobicity of the local environment and two water molecules around S462 were present (Fig 3.9C). The S462 formed a hydrogen bond with I469_O. A water-mediated bond between M462_S and S316_N was also observed (Fig 3.9). The changed hydrophobicity may influence the dynamics of the potential hinge residue S462. The change of hydrophobicity and increased pocket volume might be the reason for decrease in enzymatic activity.

Table 3.2. Data collection and refinement statistics.

	<i>PfA-M1 (M462S)</i>	<i>PfA-M1 (A453P)</i>	<i>PfA-M1 (M462K)</i>
Data collection			
Wavelength (Å)	0.95370	0.95370	0.95370
Resolution (Å)	30.4 - 2.4 (2.486 - 2.4)	36.0 - 2.0 (2.072 - 2.0)	32.8 - 2.0 (2.071 - 2.0)
Space group	P 21 21 21	P 21 21 21	P 21 21 21
Unit cell	75.01 108.93 117.63 90 90 90	75.64 108.74 117.83 90 90 90	76.34 108.95 118.19 90 90 90
Total reflections	430454	882619	466546
Unique reflections	38340 (3758)	64205 (6111)	66804 (6538)
Multiplicity	6.6 (5.8)	13.7 (11.2)	7.0 (6.3)
Completeness (%)	99.92 (99.89)	96.90 (93.87)	99.34 (98.18)
Rpim	0.113 (0.786)	0.073 (0.467)	0.06 (0.351)
Mean I/sigma(I)	10.28 (2.93)	7.91 (1.55)	10.35 (2.34)
Wilson B-factor	19.58	22.23	19.55
CC1/2	0.984 (0.32)	0.993 (0.534)	0.995 (0.701)
Refinement			
R-work	0.1700 (0.2125)	0.1715 (0.2667)	0.1860 (0.2547)
R-free	0.2187 (0.2633)	0.2181 (0.3249)	0.2196 (0.2966)
Number of atoms			
non-hydrogen	7854	8014	7935
macromolecules	7279	7216	7331
ligands	21	3	19
water	554	795	585
Protein residues	889	889	889
RMS(bonds)	0.008	0.007	0.005
RMS(angles)	1.27	1.01	1.01
Ramachandran			
favored (%)	97	98	98
outliers (%)	0.56	0.11	0
Clashscore	4.36	2.73	2.75
Average B-factor	23.7	26.10	23.9
macromolecules	23.5	24.20	23.5
ligands	31.3	44.50	32.2
solvent	26.8	43.40	28.8

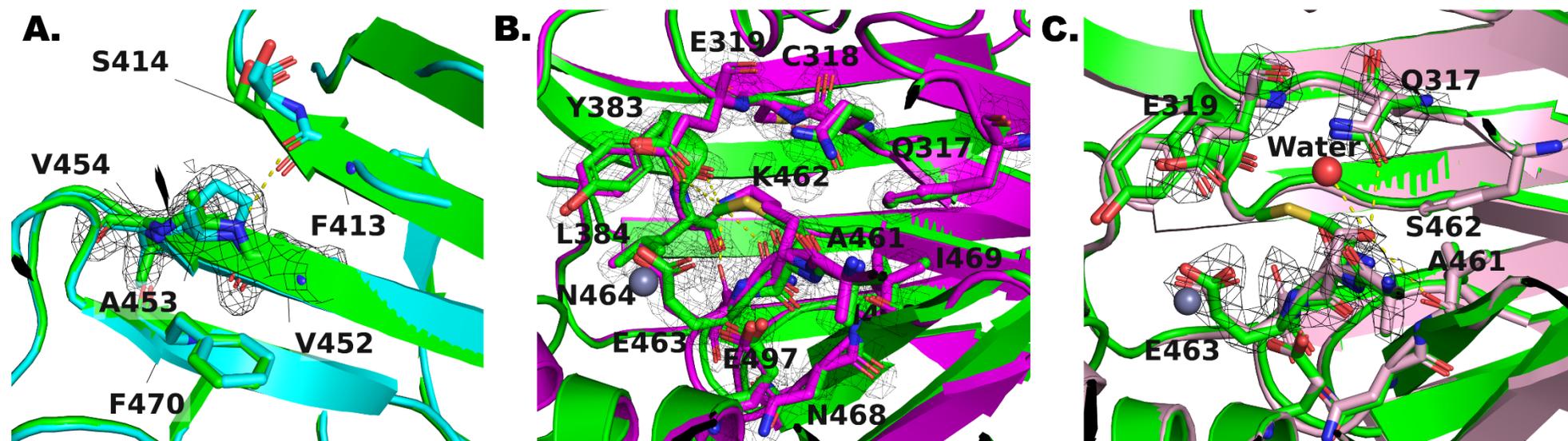


Figure 3.9. Comparison of the crystal structures and electron density shown for the mutated positions. The comparison between the wild type *PfA-M1* crystal structure (green) and (A) A453P (cyan), (B) M462K (magenta) and (C) M462S (pink). The carbon atoms of wild type structure are colored in green.

3.3. DISCUSSION

Within the M1 aminopeptidase superfamily, some members have been shown to possess two states, an open state and a closed state (5, 144, 149, 246). Former studies have concluded that the open form is the inactive form and it allows substrates to enter the active site through the gap between domains IV and I (144, 149, 160, 216). The same studies state that the closed state refers to the active form of the enzyme, allowing correct positioning of the scissile peptide bond within the active site (144, 149, 216). *PfA-M1* has only ever been observed in the closed form (4, 248). The closed form of *PfA-M1* has a competent active site with structured S1 / S1' substrate pockets, ready to cut the peptide bond, which is consistent with the literature on the structure of the M1 aminopeptidases. My MD simulations provide evidence that *PfA-M1* may also be able to adopt an open conformation. The open form of *PfA-M1* shows a large space between domains I and IV, and this structure is similar to the open form of the M1 homolog, ERAP1 (PDB ID 3MDJ) (216). The predicted free energy landscape from my MD study illustrated a free energy barrier (4.53 kT) between the open and closed form of *PfA-M1*. This energy barrier may be the reason that the open form of *PfA-M1* has never been observed experimentally in crystallographic studies (4, 248). The closed to open transition was only observed in one of the three replicates. Large domain motions are expected at μs - ms simulation time scale (184, 254) and this may be the reason I only observed the opening event in one trajectory. Increasing the length and number of trajectories that start from random velocities may increase the chance of sampling different dynamics of the protein (184, 254). Other enhanced sampling techniques of MD simulations including Hamiltonian or temperature replica exchange (255, 256), metadynamics (257), accelerated molecular dynamics (258) or molecular dynamics flexible fitting (259) might also be helpful to increase the probability of sampling the closed to open transformation.

Based on the literature, I hypothesized that if *PfA-M1* did "open", then the open state would be needed for its enzymatic function. To test the hypothesis, hinge mutants were produced to try to sabotage the hinge axis. The results showed that the functions of these

mutants (A453P, M462S, and M462K) were affected to different extents. The structures of the mutant proteins were determined and showed no loss of structural integrity and that each were maintained in the closed form of *PfA-M1*, which suggests that the loss of enzymatic activity is due to changes in the catalytic mechanism or dynamics. My hypothesis that the rigid main chain of proline introduced to position 453 would reduce flexibility and limit the movement of the protein. The k_{cat} of A453P ($0.11 \pm 0.001 \text{ s}^{-1}$) is 8-fold less than that of the wild type ($k_{\text{cat}} = 0.92 \pm 0.003 \text{ s}^{-1}$) and with no direct known involvement in the proteolytic mechanism, this may be due to dynamics. The alteration to M462 was expected to alter enzymatic activity as M462 is part of the essential GAMEN motif that positions the substrate for hydrolysis. M462S changed the hydrophobicity of the S1 pocket, increasing the hydrophilicity of the area. My hypothesis was that this may influence the dynamics of a protein hinge in the region. This may be the reason that k_{cat} of M462S was reduced to $0.083 \pm 0.001 \text{ s}^{-1}$, 11-fold less than the wild-type. Similarly, M462K altered the environment of the S1 pocket, and produced a hydrogen network with the surrounding residues (Y383, E319, and E463). The increase in interactions in the region may limit the movement of the protein.

Unfortunately, although the mutants produced in this study did show damaged aminopeptidase activity they provided no direct evidence for any conformational change. Further, former mutagenesis studies has shown that mutation of V459, E572 and M1034, all located in the S1 pocket, can affect the enzyme function to different extent (248, 260). My mutants changed the environment of the S1 pocket and also changed one of the key residues of the GAMEN. Thus the experimental data cannot prove the hypothesis that *PfA-M1* can open.

However, arising from other investigations in the McGowan laboratory, other mutants of *PfA-M1* have been produced. One of these mutants, E319P, was shown to be catalytically inactive and was crystallized in an open form (Fig 3.10). The position of this mutant, E319, is highly conserved and has been shown in previous studies of M1 aminopeptidase homologs to be involved in a key salt bridge with K572. The hypothesis for this mutant was to disrupt this salt bridge and investigate the effect. Comparison of the open conformation from simulations and the crystal structure of the open conformation of E319

showed that domain I is in the same position between the two models (Fig 3.10), although domain I from the simulation data is rotated slightly more than the crystal structure. However, different to the most open conformation from the simulations (run 3 at 348.4 ns), there is also a large space between domains II and IV of the crystal structure that is more similar to the open form ERAP1 (5, 216). This suggests the predicted open state of *PfA-M1* from my simulation data was reasonable but may not have had sufficient time to sample the more open conformation.



Figure 3.10. The comparison between the open form obtained via simulation and the open *PfA-M1*(E319P) X-ray crystal structure. The proteins are shown in cartoon. The open X-ray structure is in green and open structure from simulation in magenta. The simulation structure was selected as the extreme open conformation of PC1 in run 3 and was observed at 348.4 ns. The two structures were aligned using residues in domain II using Pymol.

This study has produced a new model of *PfA-M1* function that contradicts previous studies (4). In combination with other data from the McGowan laboratory, this work shows

that *PfA-M1* can indeed adopt a similar open conformation to that of other M1 aminopeptidase homologs (5, 149, 216). I propose a new *PfA-M1* mechanism where *PfA-M1* starts at the closed crystal structure position (Fig 3.11, a). To open and accept substrate, domain I and IV must twist in a clockwise direction. The protein can open via a rigid body movement, levering the N-terminal domain I away from the rest of the molecule to reach an open state (Fig 3.11, c). This results in an opening of *PfA-M1* allowing access to bulk solvent and presumably, to any available substrate, to the active site of the enzyme (Fig 3.11, substrate). As I performed classical MD simulations, and Newton's equations of motion are time reversible, which indicates that the movements happened in simulations should be reversible (261). I can therefore model that *PfA-M1* can close through an anticlockwise twist (Fig 3.11, d and f). This returns the molecule to a catalytically competent state for the hydrolysis (Fig 3.11, e). The protein can then release the hydrolysis products via an open conformation through the clockwise twist of domain I and IV (Fig 3.11, f and g). My hypothesis is that in the absence of substrate or a reduced concentration of substrate, *PfA-M1* remains closed which is the state that has the lowest free energy. The physiological trigger for the activation or increased activation of *PfA-M1*, which triggers it to open, remains to be proven.

3.4 CONCLUSIONS

The characterization of *PfA-M1*'s mode of action as outlined in this work could be used as a basis for future rational drug design to produce antimalarials. It provides new horizon to *PfA-M1* as the enzyme also process an open form. The protein need to twist and then achieve an open form, and such motion might involve in enzymatic function. Therefore, a future potential inhibitor might need to target the twisting motion of the protein, which might effectively prevent the protein form opening for substrates.

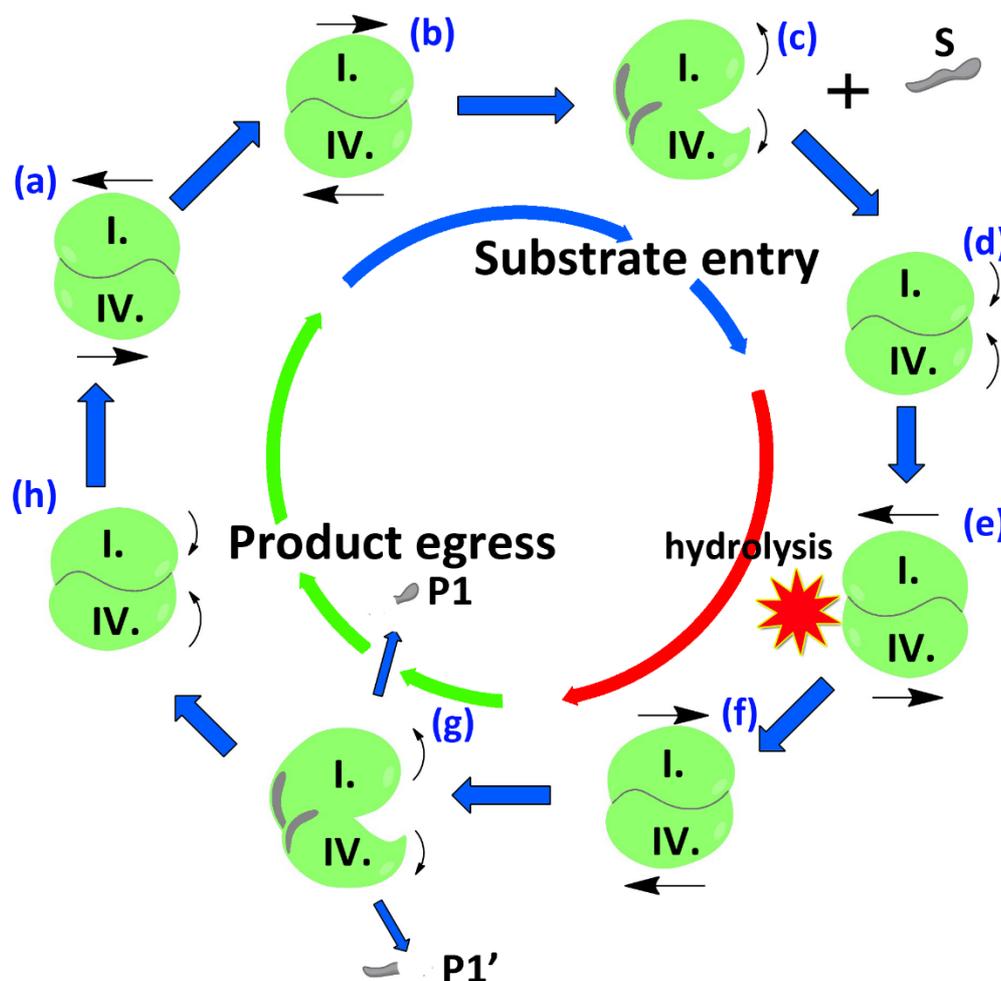


Figure 3.11. Proposed mechanism of action of *PfA-M1*. The protein is shown diagrammatically in green and domains I and IV are labelled. The direction of motion is indicated by arrows. The potential substrate is shown in grey. (a) The starting or resting conformation of *PfA-M1* is completely closed. (b) *PfA-M1* undergoes a clockwise twist to reach the transition state between open and closed conformation. (c) *PfA-M1* moves to an open conformation and is ready for substrates to enter in the interior of the protein and the catalytic domain. (d) *PfA-M1* moves to a transition state when holding the substrate inside the protein. (e) *PfA-M1* twists to a closed conformation to position the P1 / P1' residues for hydrolysis of the peptide bond. (f) *PfA-M1* then twists domains I and IV in a counter-clockwise direction. (g) *PfA-M1* then opens again to release the products of proteolysis. At this point, the protein may accept another substrate and return to state d, or may move to a transition state h and return to a closed form, a.

CHAPTER FOUR

DETERMINING THE STRUCTURAL FEATURES OF *PfA-M1* BOUND TO A PEPTIDE SUBSTRATE

4.1 INTRODUCTION

Determining the structure of an enzyme bound to its natural substrate can provide profound knowledge of the proteolytic mechanism as well as provide structural activity relationships (SAR) of the intermediate state that can aid the design of novel inhibitors. To date, there have been three structures of M1 aminopeptidases in complex with peptide substrates reported (139, 144, 262). To successfully capture the substrate bound enzyme complex for leukotriene A4 hydrolase (LTA4H) and human APN (*hAPN*), catalytic mutants of these mutants were produced and co-crystallized with peptide substrate (139, 144). The inactive catalytic mutants were produced by mutating the catalytic Glu residues in each enzyme (LTA4H(E296Q) or *hAPN*(E384Q)), allowing the enzyme-substrate structure to be produced and not processed. For the *Anopheles gambiae* M1 aminopeptidase, the determination of the first structure identified two bound substrates (PDBID: 4WZ9) that must have co-purified with the enzyme (232). For *hAPN*, the substrate angiotensin IV was captured inside the closed form of *hAPN* after the protein was treated with EDTA to chelate the zinc to effectively disable its enzymatic activity (1) .

The natural substrate for *PfA-M1* is Hb peptides (96). Previous work showed that the selective inhibition of *PfA-M1* in parasite trophozoites resulted in an accumulation of Hb peptides (96). The peptides that were found at the highest concentration were two six-mers, V⁹³DPVNF⁹⁸ from the Hb α -chain and V⁹⁸DPENF¹⁰³ from the Hb β -chain (263). The finding implies that *PfA-M1* can process longer peptides than the di- or tri-peptides that has been published previously (152, 154, 208) .

However, to date, no studies have experimentally investigated the digestion of longer peptides by *PfA-M1*. Therefore, there is no structural data to map the potential substrate pockets of the enzyme (S2, S3, S4 etc). This represents a significant gap in our knowledge of the enzyme mechanism and description of potential druggable sites within the protein. To address this, I undertook a study to investigate the activity of *PfA-M1* on a six-mer peptide that we called Hb6 (LDPENF) and attempted to model and validate a transition-state complex of *PfA-M1* bound to Hb6. The aim of this work was to identify new substrate

pockets within *PfA-M1* as well as identify how the substrate might enter the active site and the possible routes that the proteolytic products would exit from.

4.2. RESULTS

4.2.1. Hb6 is a substrate of *PfA-M1*.

A previous study had identified two six-mer peptides, V⁹³DPVNF⁹⁸ from the Hb α -chain and V⁹⁸DPENF¹⁰³ from the Hb β -chain as natural substrates of *PfA-M1* (96). To improve the affinity of the peptide for *PfA-M1*, the P1 residue was altered to leucine, for which *PfA-M1* has been shown to have a strong preference (100). Therefore, the peptide used in this study had the sequence L₁D₂P₃E₄N₅F₆, where L₁ is the P1 residue, whilst D₂ is the P1', P₃ the P2' etc. In this study, this peptide will be referred to as Hb6.

To validate that Hb6 could act as a substrate for recombinant *PfA-M1*, I used the Monash Proteomics Platform to perform liquid chromatography-mass spectrometry (LC-MS) to detect if the product of hydrolysis was present after digestion. For the LC-MS experiments, I prepared two samples with either Hb6 alone or Hb6 with *PfA-M1* with a 100-fold molar excess of peptide. The LC-MS results showed that Hb6 alone was detected at the correct molecular weight of 734.33 (Fig 4.1A) whilst the addition of *PfA-M1* resulted in the reduction of the full-length peptide and the appearance of a smaller peptide that corresponded to the molecular weight of the product, DPENF (621 Da, Fig 4.1B). These results indicated that the P1 residue of Hb6 can be hydrolysed by *PfA-M1*, however, the proteolytic activity is stopped at the P2 Asp residue. This experiment also confirmed that Hb6 is a substrate of recombinant *PfA-M1*.

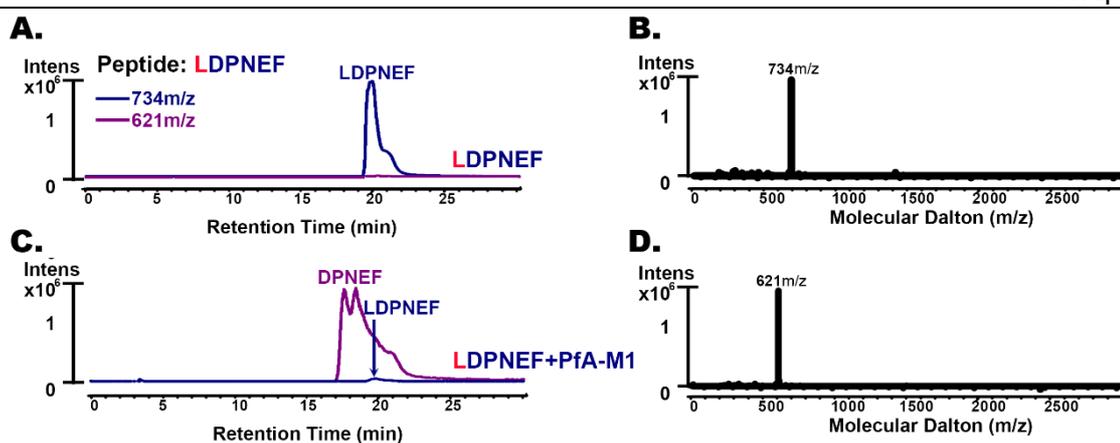


Figure 4.1. LC-MS experiments to investigate if Hb6 is a substrate for *PfA-M1*. (A) Extracted-ion chromatogram (EIC) of sample 1 (Hb6 alone). The blue trace is EIC chromatogram of 734 m/z (mass-to-charge ratio, purple trace is that of 621 m/z. (B) The total observed spectrum signals of sample 1 with a molecular Dalton range of 0 to 3000. (C) Extracted-ion chromatogram (EIC) of sample 2 (Hb6 and *PfA-M1* mixture). (D) The total observed spectrum signals of sample 2 with a mass scan range of 0 to 3000 Da.

4.2.2. Attempts to produce a co-crystal structure of *PfA-M1* bound to Hb6

To attempt to solve the structure of a *PfA-M1*-Hb6 complex, it was necessary to produce an inactive form of *PfA-M1* to prevent *PfA-M1* from digesting Hb6 during the crystallization process. *PfA-M1* requires one to two nucleophilic water molecules in the active site for hydrolysis of peptide bonds (4). Based on previous work that elucidated the catalytic mechanism of *PfA-M1* (208) as well as others (1, 144), I decided to introduce point mutations to alter the residues that coordinate the nucleophilic waters. Removing the ability of the enzyme to coordinate water in its active site should produce a catalytic (inactive) mutant. As it is not certain which Glu residue (E497 or E463) contributed the most to the coordination of the catalytic water, I designed two single point mutants, *PfA-M1*(E463Q) and *PfA-M1*(E497Q) (Fig 4.2). The choice of a glutamine side chain was based on the assumption that it would coordinate water poorly but retain a similar shape and position to the original residue.

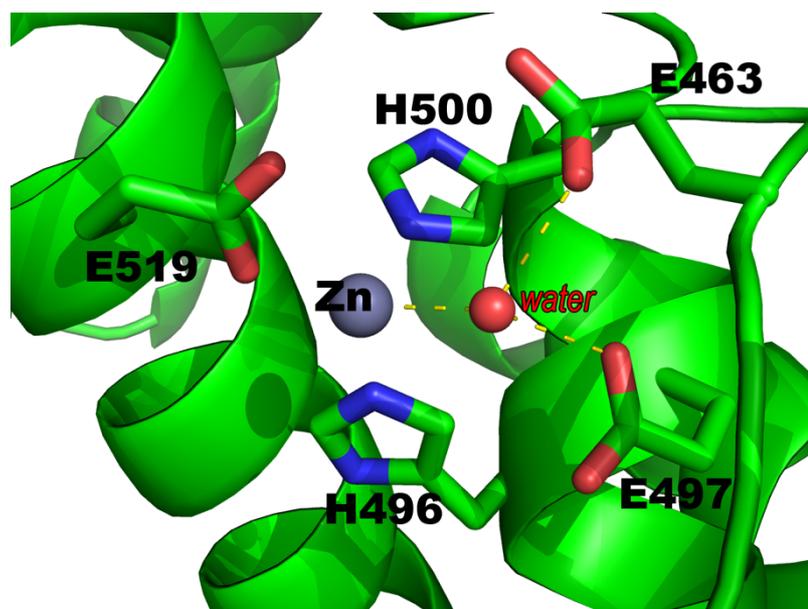


Figure 4.2. The nucleophilic water that is coordinated by E463 and E497 in *PfA-M1*. The cartoon of the protein structure is shown in green. The residues involved in metal coordination and nucleophilic water bridge were in sticks. Zinc ion was shown by sphere.

To make the mutant *PfA-M1* proteins, we used site-directed mutagenesis to alter the gene encoding the enzyme. Ms Komagal Kannan Sivaraman (McGowan laboratory) generated these mutants and produced the recombinant proteins. Ms Kannan Sivaraman also assayed the mutant proteins for their aminopeptidase activity to provide data on the effect the mutations had on the function of the protease. The activity of the mutants was assessed in comparison to that of wild-type enzyme and showed that the mutant proteins, as expected, had no detectable aminopeptidase activity (tested up to 25 μM of protein).

To ensure that the mutations had not introduced major structural changes in the active site of the enzyme, I crystallized the two mutant proteins in their unbound form. Ms Kannan Sivaraman purified the recombinant proteins E463Q and E497Q and I set up the hanging drop crystal trays with the assistance of Dr. Nyssa Drinkwater (McGowan laboratory). With the assistance of Drs Drinkwater and McGowan, we were able to successfully solve the structures of E463Q to 1.9 Å and E497Q to 2.0 Å (Table 4.1). The two structures were virtually identical to that of wild-type (overall RMSD of 0.4 ± 0.05 Å) confirming that the loss of function observed was not a result of a loss in structural integrity.

The crystal structures showed that the coordination of the water molecules in both mutants was altered (Fig 4.3). In *PfA-M1(E463Q)* mutant, the position of the side-chain is altered drastically and now points towards domain I (Fig 4.3A). In *PfA-M1(E463Q)*, the position of the side-chain is altered drastically and now points towards domain I (Fig 4.3A). The dihedral N-CA-CB-CG angle of Q463 switched to 92.9 degrees, resulting in NE2 of Q463 facing away from the metal center. This change in orientation is likely due to repulsion between the positively charged NE2 atom of Q463 and the also positively charged metal center (total charge of the metal center is +1).” This results in the loss of its interaction with the nucleophilic water (Fig 4.3A). Moreover, the GAMEN motif is closer to the zinc ion / metal centre (by $2.2 \pm 0.1 \text{ \AA}$) and M462 is actually quite close to the zinc ion (Fig 4.3A). The loss of the interaction with the water or the movement of the GAMEN motif might be the reason for the loss of activity observed in the mutant. The *PfA-M1 (E497Q)* showed a similar geometry to the wild-type with regard to the position of the mutated side-chain, however, the interaction with the nucleophilic water is similarly lost (Fig 4.3B). A water bridges to the zinc ion only through E463 and the position of the nucleophilic water has changed, moving further from the mutated Q497 (Fig 4.3B). The position of the GAMEN motif is unchanged with this mutant, suggesting that E497 may have a more direct role in activating the proteolytic attack than E463.

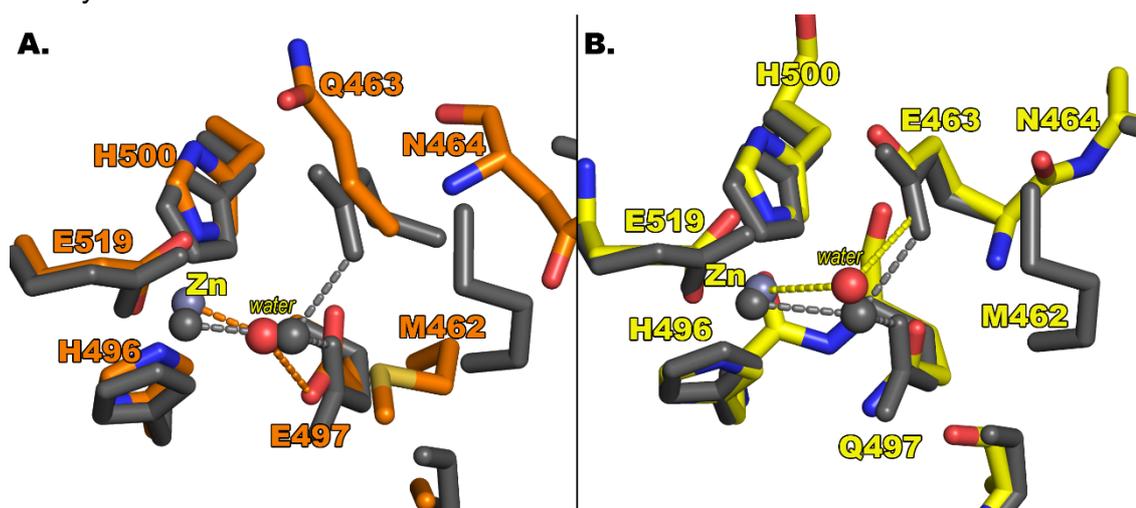


Figure 4.3. Comparison of the structures of (A) E497Q and (B) E463Q with the wild-type enzyme. Protein residues are shown in stick with carbon atoms of wild-type (grey), E463Q orange and E497Q yellow. Waters are shown in red (mutant) and grey (wild-type).

Table 3.1. Data collection and refinement statistics.

	<i>PfA-M1(E497Q)</i>	<i>PfA-M1(E463Q)</i>
Data Collection		
Wavelength (Å)	0.95370	0.95370
Resolution range (Å)	33.8 - 2.0 (2.071 - 2.0)	38.4 - 1.9 (1.968 - 1.9)
Space group	P 21 21 21	P 21 21 21
Unit cell	74.53 109.27 117.32 90 90 90	76.75 108.47 115.78 90 90 90
Total observations	958292	341512
Unique reflections	65405 (6472)	65552 (5704)
Multiplicity	14.6 (14.7)	5.2 (4.1)
Completeness (%)	100.00 (100.00)	85.39 (75.17)
Rpim	0.046 (0.412)	0.044 (0.479)
Mean I/sigma(I)	14.41 (2.03)	12.48 (1.64)
Wilson B-factor	26.43	18.55
CC1/2	0.998 (0.636)	0.996 (0.569)
Refinement		
R-work	0.1920 (0.2883)	0.1589 (0.2540)
R-free	0.2206 (0.3315)	0.2043 (0.2780)
Number of atoms		
non-hydrogen	7686	8322
macromolecules	7218	7316
ligands	21	27
water	447	979
Protein residues	889	889
RMS(bonds)	0.005	0.008
RMS(angles)	1.10	1.20
Ramachandran		
favored (%)	97	98
outliers (%)	0	0
Clashscore	4.22	3.10
Average B-factor	36.30	20.50
macromolecules	36.10	19.20
ligands	44.70	29.20
solvent	37.90	30.20

Attempts to co-crystallize the structure of catalytic mutants in complex with Hb6 were unsuccessful. I tried both co-crystallography approaches as well as ligand soaks to attempt to produce the structure. We varied the molar ratios of Hb6 to *PfA-M1* protein (both E463Q & E497Q). We also tried changing the soaking times (from 2 minutes to overnight) and pre-treatment with EDTA to try and further prevent the enzyme from functioning. However, all data collected from co-crystallography and crystal soaks showed that there was no electron density for Hb6 in the structures solved and that the overall structure of mutants was

unchanged. There were no observed changes to the arrangement of the neither asymmetric unit nor crystal packing.

4.2.3. Modelling the *PfA-M1-Hb6* complex

Since I was unsuccessful in obtaining the crystal structure of *PfA-M1-Hb6*, I chose to produce a model of the complex. My previous work (Chapter 3) had shown that *PfA-M1* could exist in both an open and closed form. Therefore, I wanted to investigate the potential binding positions of Hb6 within an open and a closed form of the protein. The coordinates of both the open (*PfA-M1o*) and closed (*PfA-M1c*) were available from my simulations. Despite the fact that the open and closed forms of *PfA-M1* have undergone a major conformational change, my current data shows that the catalytic domain II is largely unchanged. Therefore, to guide the binding pose of Hb6 to *PfA-M1*, I used the structure from the porcine Aminopeptidase N (*pAPN*) that had been co-crystallized with a poly-alanine seven-mer (PDB ID 4NAQ (144)). To model the complexes, I aligned domain II (residues 398 to 645) from the *PfA-M1* structures with the template and maintained the position of the backbone atoms of the substrate (A1 to A6 and the C-terminus of the peptide was trimmed). The substrate was then trimmed from the C-terminus to six amino acids in length and mutated to the sequence of Hb6. The two models were then minimized by Amber package (described in Chapter 2) to release any close contacts or clashes between the Hb6 and *PfA-M1*. The main clashes in the modelled structures were located between the Hb6 D₃ side chain and E497_{OE} in *PfA-M1*. The minimization step optimized the model and increased the distance between the two groups to a more appropriate distance. The optimized models obtained, *PfA-M1o-Hb6* and *PfA-M1c-Hb6*, are shown in Figure 4.4. The overall structure of the models showed little change when compared to the original coordinates (*PfA-M1c-Hb6* RMSD = 0.5 Å and *PfA-M1o-Hb6* RMSD = 0.4 Å).

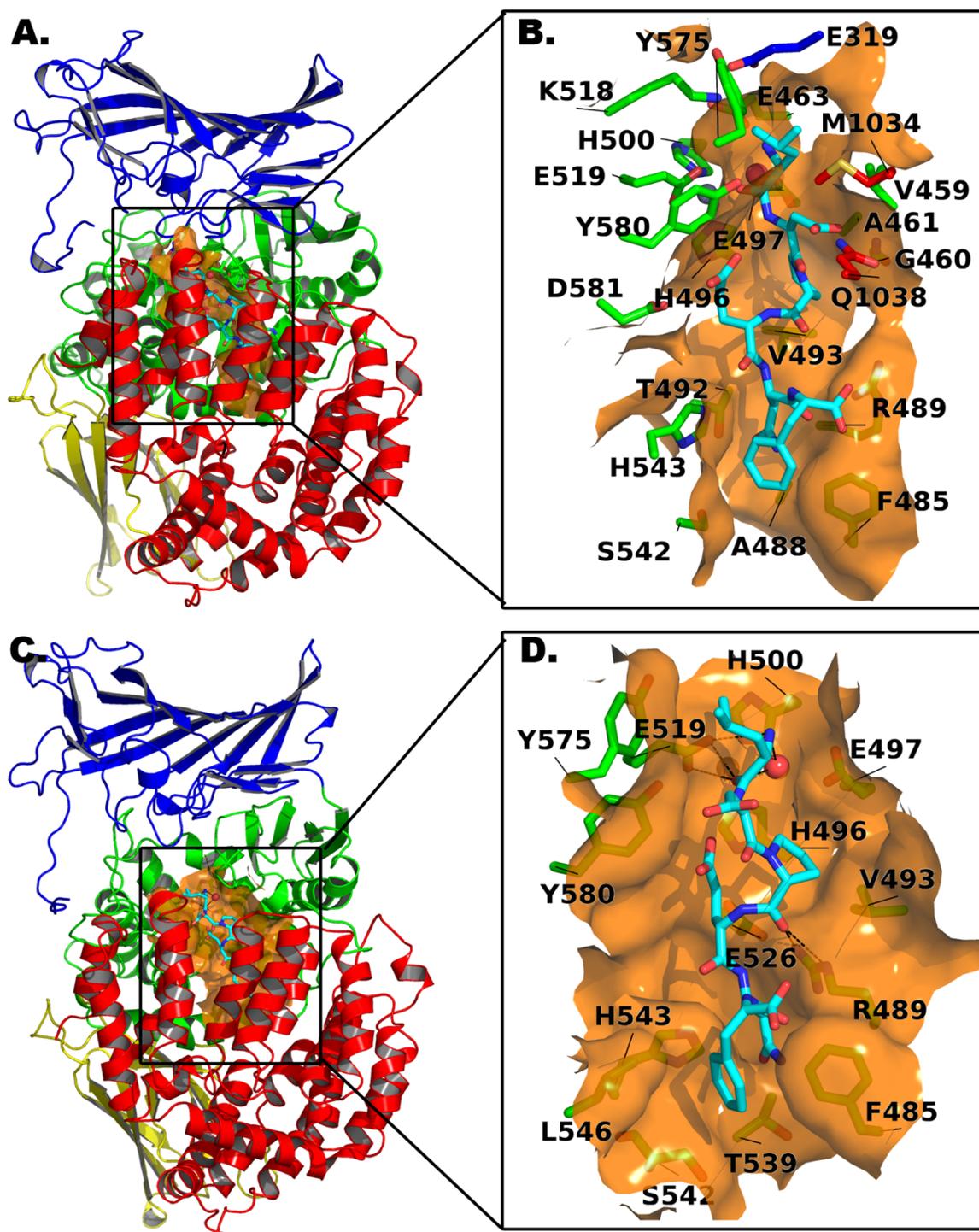


Figure 4.4. The models of Hb6 bound to *PfA-M1*. (A) *PfA-M1c*-Hb6 and (B) zoom of the active site. (C) *PfA-M1o*-Hb6 and (D) zoom of the active site. *PfA-M1* is shown in standard domain coloring, except for residues within 4.5 Å to Hb6 which are shown in orange surface. The carbon atoms of Hb6 are shown in cyan, water and metals are in spheres.

The modelled Hb6 sits in the catalytic domain alongside the GAMEN motif in both models and is in a similar position to the template substrate from *pAPN* (Fig 4.4). In *PfA-M1c*-Hb6 model, the P1 residue L₁ is located in the S1 pocket and forms a backbone hydrogen bond with GAMEN residue E463 as well as zinc-coordinating residues E519 and H496 (Fig 4.4B). The carbonyl oxygen also coordinates directly to the zinc ion (Fig 4.4B). The P2 residues, D₂, forms a hydrogen bond with GAMEN residue G460_N and the P3 and P4 residues have hydrophobic interactions with V493, A461 and G460 (Fig 4.4B). The side chain of P4 residue (E₄) shows no polar contacts to *PfA-M1* whilst N₅ forms a hydrogen bond with T492_{OG} (Fig 4.4B). The final residue of Hb6, F₆, formed three hydrogen bonds with R489 through its carbonyl oxygen (Fig 4.4B).

In the *PfA-M1o*-Hb6 model, weaker hydrophobic interactions between GAMEN motif were observed but the major difference from the open model was that the side chains of the Hb6 did not form any hydrogen bonds with *PfA-M1* (Fig 4.4D). The backbone oxygen atom of P₃ did form two hydrogen bonds with R489.

4.2.4. Steered MD study suggests N-terminal access point is preferred by Hb6.

I was then curious about how Hb6 might access the *PfA-M1* active site. *PfA-M1* has two potential access points to the active site, a C-terminal pore and an N-terminal access point formed by the interactions of domains I, II and IV (4, 208). In chapter 3, I investigated the possibility of an open form of *PfA-M1*, and believed that substrate might get access to the active site from an open form of *PfA-M1*. However, whether the substrate enters from the gap between domain I and IV, or from the C-terminal pore in the open form is still unclear. To investigate the potential path that Hb6 might travel, I used steered MD (SMD) simulations to predict whether Hb6 prefers the C-terminal or N-terminal path. In SMD simulations, an external force is added to 'pull' atoms in the direction of the force, in this case, to pull the substrate towards either the N-terminal entrance or the C-terminal pore (Fig 4.5).

I performed four separate SMD simulations using a constant pulling force. These simulations applied an external force of velocity of 0.001 Å / ps on the C α of Hb6_{L1} from the N-terminal channel for both *PfA-M1c*-Hb6 (SMD1) and *PfA-M1o*-Hb6 (SMD2), and on the

$C\alpha$ of $Hb6_{F6}$ towards C-terminal channel from *PfA*-M1c- $Hb6$ (SMD3) and *PfA*-M1o- $Hb6$ (SMD4) (See Table 2.2 for more detail). The atom that defined the direction of force is shown in Fig 4.5. Each channel in the two models are approximately linear and parallel to the force direction used (Fig 4.5A & B). The only channel that result in potential clash with side chains is the C-terminal channel of the closed model (SMD3, residues D971 and D830). Given the same force velocity, the time taken for $Hb6$ to exit the interior of *PfA*-M1 was faster for the shorter N-terminal pathway than the C-terminal (Fig 4.5). The external work was calculated from the trajectories and log files of the simulations (see Section 2.1.12) and used to generate a plot of the external work vs the displacement or SMD extension of $Hb6$ (Fig 4.5E). This plot describes the displacement of $Hb6$ between each snapshot compared to the position of $Hb6$ in the starting conformation.

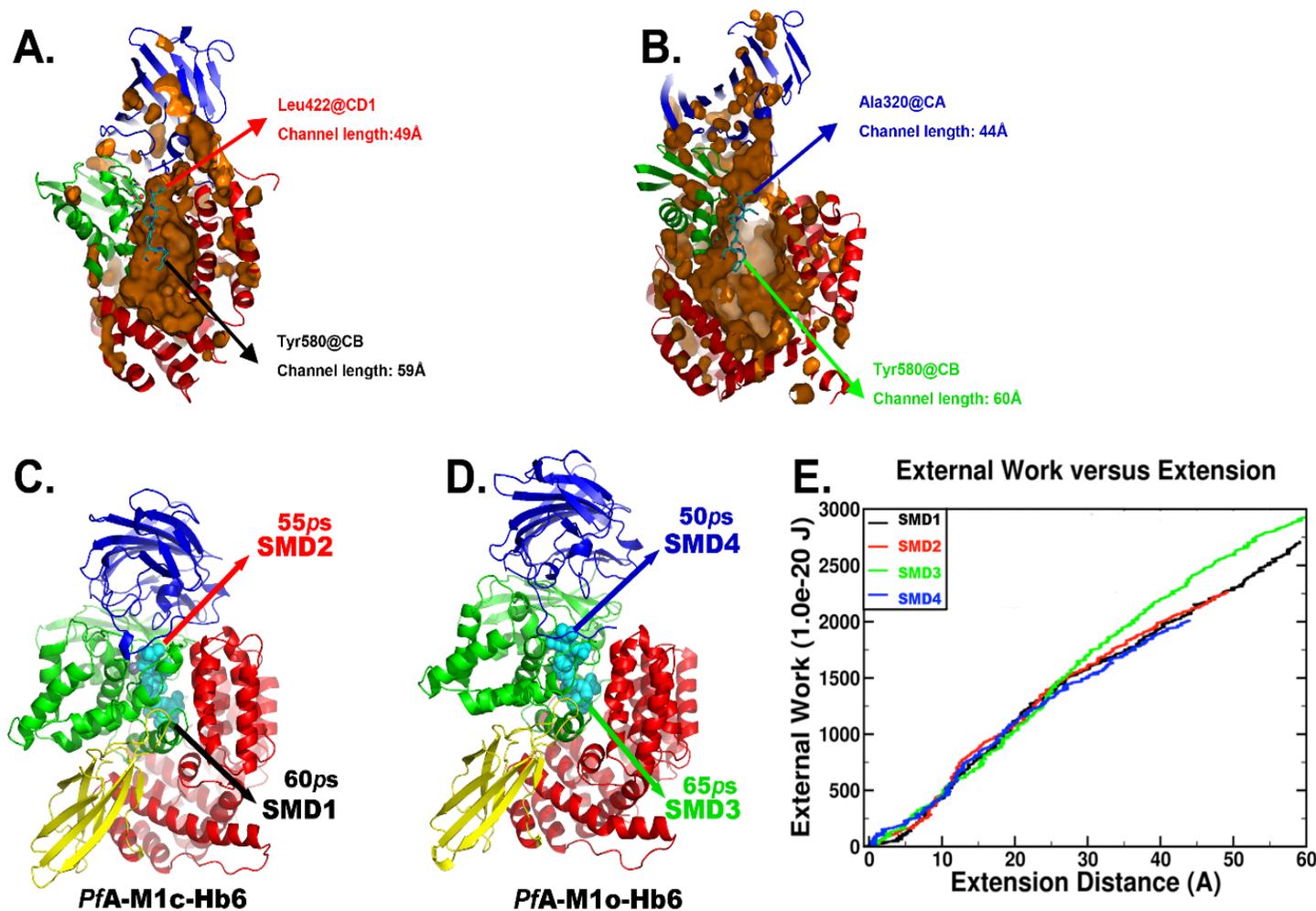


Figure 4.5. SMD of *PfA-M1* Hb6 models based for two different pathways. The force direction, channel length and constrained atom for both (A) *PfA-M1c-Hb6* and (B) *PfA-M1o-Hb6* models are shown and *PfA-M1* is in standard coloring. In brown, the protein pocket surface is indicated. In panels (C) *PfA-M1c-Hb6* and (D) *PfA-M1o-Hb6*, Hb6 is shown in spheres with the carbon atoms shown in cyan. Arrows indicate the direction of the force applied to both the N-terminal path (SMD 2 & 4) and the C-terminal path (SMD 1 & 3). (E) External work curves from the four SMD simulations (colored as indicated). The extension distance was taken from the log file of each SMD, which described the extension distance of every saved snapshot, and external work was calculated by Newtons law of $W = FS$. Extension distance unit is in Å, and the calculated external work is in 1.0 E-20J.

The external work vs extension shows that the external force (the slopes of the four SMDs) are very similar to each other at the beginning of the simulation (Fig 4.5E). However, between an extension of 22 - 40 Å, the slope of SMD3 rises more sharply than the other runs, indicating that a higher external force is required during SMD4 to keep Hb6 moving at the same speed as the other SMDs. Interestingly, this happened from 22 Å, where the F6 of Hb6 approaches the bottle neck (D971 and D830) of the C-terminal channel in the *PfA-M1o-Hb6* model. Analyzing the bottleneck radius between the two C-terminal channels of the models suggested that the pore in the open form is narrower compared to that from the closed form (1.94 Å in the open and 2.45 Å in the closed). SMD 2 and 4 required a smaller external work (2.1 E-17 J for SMD2, and 1.7 E-17 J for SMD4), compared to the external work on Hb6 towards the C-terminal channel of both models (2.9 E-17 J for SMD1 and 2.7 E-17 J for SMD3). SMD4 has the least external work. This suggested that the N-terminal channel of an open *PfA-M1* is the most preferred way for Hb6 to migrate through.

4.3 DISCUSSION

In this chapter, the peptide known as Hb6 was shown to be a substrate for recombinant *PfA-M1*. The protease was able to cleave the P1 residue but unable to keep processing through the P1' Asp (D2). This result supports the previous study by Harbut *et al.* that identified a six residue peptide as the likely substrate for *PfA-M1* in a chemical genetics study in *P. falciparum* (96). The inability to cleave the P1' Asp residue is also consistent with earlier studies showing that *PfA-M1* does not readily cleave an Asp residue (100). Whether *PfA-M1* shows processive activity, i.e. the ability to cleave the P1 residue and continue to process the same peptide substrate, requires further work with residues that can be cleaved (for example, change the D₂ to a L₂ in Hb6).

My attempts to produce a crystal structure of catalytic mutants of *PfA-M1* bound to Hb6 were unsuccessful. There are many reasons why crystallization was unsuccessful and may include lack of affinity of Hb6 to *PfA-M1* as well as sub-optimal crystal packing. The McGowan laboratory has produced numerous co-crystal structures of *PfA-M1* bound to

potent small molecule inhibitors. Moderate inhibitors (micromolar range) have rarely been successfully resolved. Currently, the affinity of Hb6 for *PfA*-M1 has not been determined. Previous attempts to use ITC to determine binding of a known peptidomimetic to *PfA*-M1 have been unsuccessful with no enthalpic change observed despite nanomolar inhibition. Therefore, future work may want to try and determine the affinity of the peptide for the proteins. The other reason that may influence our success in producing crystals of the complex is the crystal packing. The conditions used in this study were similar to those used to produce standard *PfA*-M1 crystals. This may limit the space available for the substrate to enter the active site (if soaked into a pre-formed crystal) or may also prevent a conformational change that is required for binding. The catalytic mutants produced in this study, E463Q and E497Q, were both found to be in the closed forms. However, in the structure of seven alanine peptide bound to *pAPN*, the protein was crystallized in its open form, where there is a clear gap between domain II and domain IV to allow the substrate to access the active site (144).

I also attempted to produce a complex structure by extraction of the essential zinc ion. This method was successfully used to produce the structure of Ang IV bound to *hAPN* (1). However, for *PfA*-M1, our EDTA treatment was obviously insufficient as electron density for the zinc ion was observed in all datasets collected. Future work to produce a structure of *PfA*-M1 complexed to substrate would need to focus on improving the affinity of the substrate for the protease, screening alternate conditions to allow for a possible conformational change or find a more effective method of chelating the protein.

I then modelled the complex structure of *PfA*-M1c-Hb6 and *PfA*-M1o-Hb6 from the template *pAPN*-7Ala. The model firstly describes how a natural substrate might bind inside *PfA*-M1. The detailed interaction analysis indicated that, apart from S1 and S1' pocket region, residues A468, T492, V493, R489, S542, and H543 are involved in the substrate binding, and help position P3 to F6 of Hb6 at C-terminal channel. Among these residues, T492 and V493, provide hydrophobic interactions to P3. These residues are conserved in *pAPN* (corresponding residues = T379 & V380) was also found to provide hydrophobic interactions to the backbone of P3 of the 7Ala substrate. R489 coordinates the backbone

oxygen of E4 in Hb6. The R489 position is conserved in *PfA-M1*, *pAPN*, *pAPA* and *hAPN*. The corresponding residue (R434) in *pAPN* coordinates A6 of 7Ala. Similarly, R442 in *hAPN* coordinates the backbone nitrogen of F6 in Ang IV (1). This location of P3 to F5 is also consistent to the proposed long substrate binding mechanism inside *hERAP1* (216). It has been proposed that *hERAP1* uses its C-terminal channel for long substrates and a separate substrate pathway (between domains II and domain IV) for the short substrates (216). Interestingly, the corresponding residue of *PfA-M1*-R489 in *hERAP1* is an alanine (A407), which is located at the branch point between the two substrate pathways. It is possible that the substitution of the Arg for Ala at position 407 is a reason why *hERAP1* has two substrate pathways but other M1 aminopeptidases do not.

The SMD study suggested that N-terminal channel is a more likely pathway for the substrate to travel in *PfA-M1*, especially when the protein is open. More hydrogen bonds were found in *PfA-M1c*-Hb6 model and more external work was required to pull the substrate out of the *PfA-M1c*-Hb6. This suggests that the closed conformation is better for maintaining the substrate in the correct position for catalysis, supporting that the closed conformation is an active form. The limitations of SMD simulations are the value and velocity of the external force applied. If it were possible to sample numerous SMD at varying external force, the potential mean force (PMF) could be calculated and ideally the simulations performed can be viewed as spontaneous process (264-267). The PMF is basically the free energy profile along the simulated biological event (268). The information in a PMF plot might explain better which channel the protein prefers from a computational biology aspect (268). The result of SMD is also biased based on the direction of the force applied and the starting models created for this purpose.

4.4 CONCLUSIONS

The results of the experimental work in this chapter prove that *PfA-M1* can cleave a peptide of up to six amino acids long. The computational work in this chapter provides meaningful data on the nature of peptide binding, in particular of peptides longer than 2 amino acids. The simulation of the potential molecular pathways that the peptide substrates

may travel may identify as yet unknown druggable sites in *PfA*-M1. Finally, the work in this chapter asks the question as to whether *PfA*-M1 may have two potential substrate routes, like *hERAP1*, and therefore, may process longer peptides than tested here.

CHAPTER FIVE

INVESTIGATING THE DYNAMICS OF *PfA*-M17

5.1 INTRODUCTION

Previous crystallographic studies of the unbound and inhibitor-bound enzyme confirmed that *PfA-M17* forms a hexameric quaternary structure (81). The six active sites of the *PfA-M17* hexamer are arranged in a disc-like fashion, orientated inwards to form a central catalytic cavity. Flexible loops, located at the external surface of the hexamer, are proposed to guard each of the six entrances and are proposed to govern access to the catalytic cavern (81). The dynamics and motion of the protein complex in relation to its enzymatic function remain unclear. In particular, questions arise as to how peptide substrates can enter the catalytic interior and why does it need to be a hexamer? To unveil the mysteries of *PfA-M17*, I used all-atom MD simulations to study the movements of *PfA-M17*.

5.2 RESULTS

5.2.1 Nomenclature used to describe the structure of *PfA-M17* in this study.

In this chapter, I have undertaken an extensive study of the dynamics of *PfA-M17* using various MD simulation experiments. To assist in the explanation and analysis of this data, I have mapped the secondary structure topology and defined the secondary structural elements of the recombinant monomeric structure by PDBsum server (269) (Fig 5.1). The recombinant *PfA-M17* protein has a truncated N-terminus and residues 1-84 are not included in the recombinant protein (81). The recombinant monomeric *PfA-M17* contains 28 loops (L1-L28), 20 β -strands (β 1- β 20), and 15 helices (α 1- α 15). Residues 1-84 are truncated in the recombinant protein and not included in this topology. The N-terminal domain (amino acids 85-271) is composed of 6 α -helices, 10 loops and 7 β -strands. The C-terminal domain (amino acids 323-601) consists of 7 α -helices, 18 loops and 13 β -strands and is linked to the N-terminal domain via a long α -helix, however, this helix is bent in the middle and is detected as two helices according to the PDBsum server (α 7-8) (81).

To clarify the description of the arrangement of the hexameric assembly of *PfA-M17*, I have also defined the monomers of the hexamer by a simplified format that is based on a premise that *PfA-M17* and the M17 superfamily is composed of a dimer of trimers (81, 170, 176, 178). Therefore, I refer to the trimeric subunit as having an ABC composition and the second trimer as $A_2B_2C_2$. In this nomenclature, the hexamer is described as an $(ABC)_2$ composition (Figure 5.2). I found that the use of chain IDs (A – F for the hexamer) as per the crystal structure was confusing when investigating the subunit assembly and behavior (as described in this chapter).

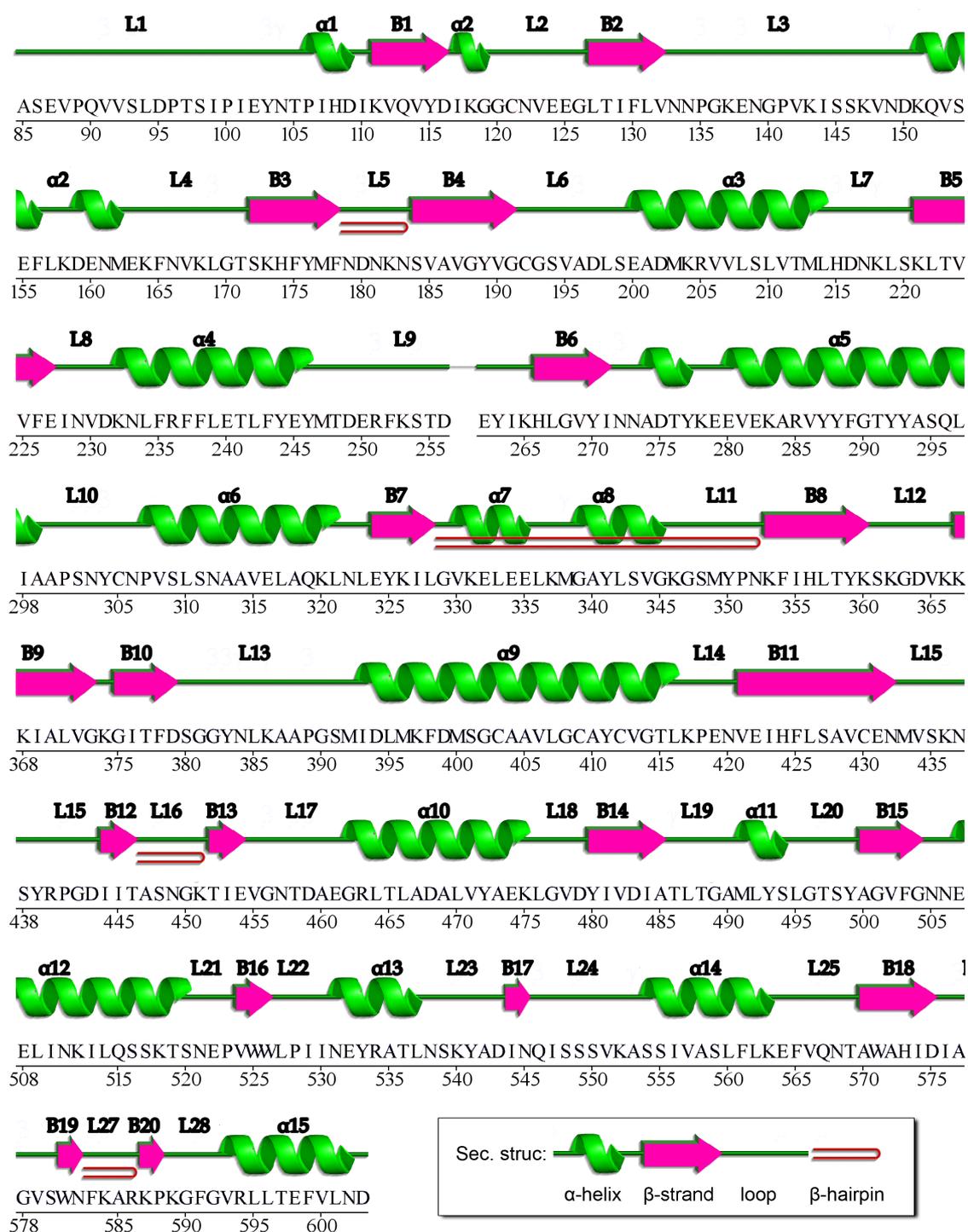


Figure 5.1. The topology index of the secondary structure used in the *PfA-M17 MD analysis*. The residue range is from 86 to 601. The plot is generated by PDBsum server. Structural elements are labeled and indicated with α -helices in green and β -strands in pink.

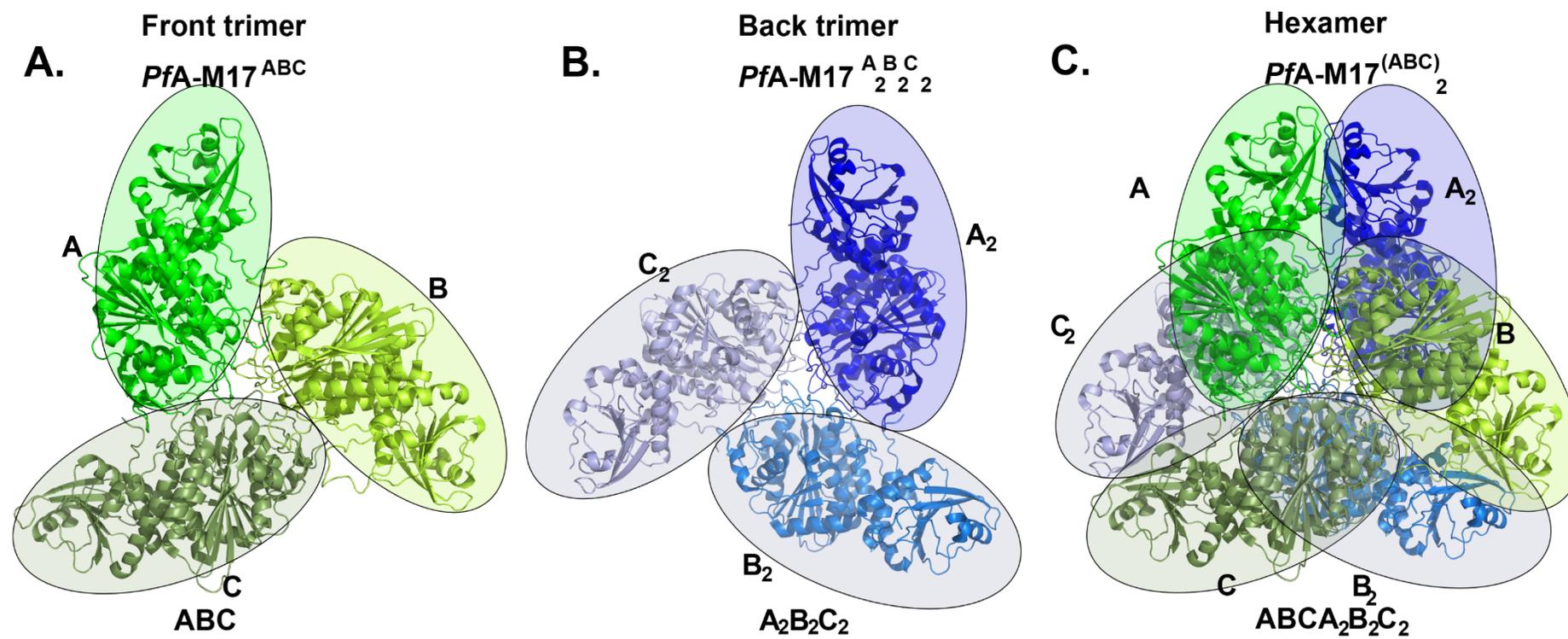


Figure 5.2 The molecular assembly of *PfA-M17*. (A) The front trimer (ABC), (B) the back trimer ($A_2B_2C_2$) and (C) the hexamer $(ABC)_2$.

5.2.2. MD simulations of the *PfA-M17* monomer and hexamer.

To complete the MD simulations of *PfA-M17*, it was necessary to generate an appropriate zinc FF for the two cations present in the active site of the M17 aminopeptidases. This work is described in (212). This FF uses a hybrid bonded/non-bonded model and was used to complete all-atom MD for both the monomeric A-subunit of *PfA-M17* ($n=3$, 200 ns) system as well as the biologically active hexamer, *PfA-M17*^(ABC)₂ ($n=3$, 400 ns). Each system was simulated from three independent replicates using different velocity seeds at the beginning of each simulation.

The overall stability of each system was analyzed by calculating the root mean square deviation (RMSD) of the C α atoms during the simulation. Analysis of the overall plasticity of the *PfA-M17* hexamer showed that it did not undergo large movements in the time-frame sampled (average RMSD of hexamer = 2.9 Å, Fig 5.3A). This was not the case with simulations of the *PfA-M17* monomer, where removal of the protein-protein interactions of the hexamer appeared to allow structural changes in comparison to the crystal structure (average RMSD of monomer = 4.2 Å, Fig 5.3B).

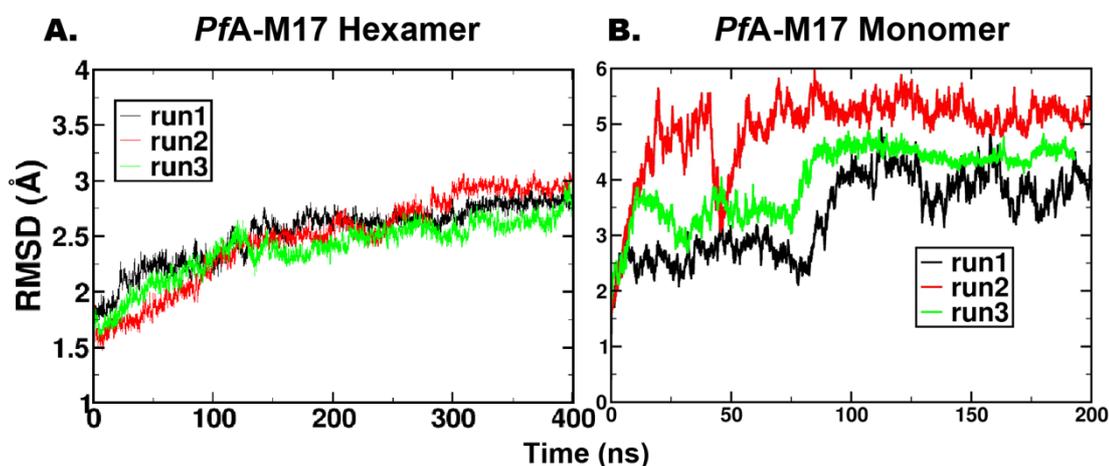


Figure 5.3. The RMSD of C α atoms during the (A) hexamer and (B) monomer simulations. The comparison of RMSD traces of C α atoms relative to the crystal structure as a function of simulation time. The time interval is 10 ps per sampled frame. Run 1 is shown in black, run 2 in red and run 3 in green.

The mobility of individual residues throughout the simulations was also assessed via calculation of a root mean square fluctuation (RMSF) of the C α atoms (Fig 5.4). The RMSF plots showed that all three triplicate runs were consistent in their movement and were generally cross-consistent between a monomer alone and the monomers within the hexamer (Fig 5.4). However, in the hexamer simulations, there are some differences between the trimers, but consistencies within each trimer, suggesting that the trimers might have a different scale of fluctuation. The RMSF results also correlate with the RMSD findings that the hexamer is more stable than the monomer. Smaller fluctuations were observed in each chain of the hexamer than were observed in the monomer. The largest variation in all systems was the N-terminus of the protein (L1, Fig 5.4) but this was disregarded due to the truncation of the N-terminus. Two other loop regions, L9 and L13, also showed large fluctuations in nearly all simulations (Fig 5.4). Typically, residues 257-261 were very flexible (average fluctuation of 5.4 Å in monomer simulations, 4.3 Å in hexamer simulations). These residues were known to be flexible as they could not be placed into electron density in the original crystal structure (81). Residues 141-145, that are located at a protein-protein interface in L3 loop, also experienced fluctuations in both the hexamer and monomer simulations (Fig 5.4)

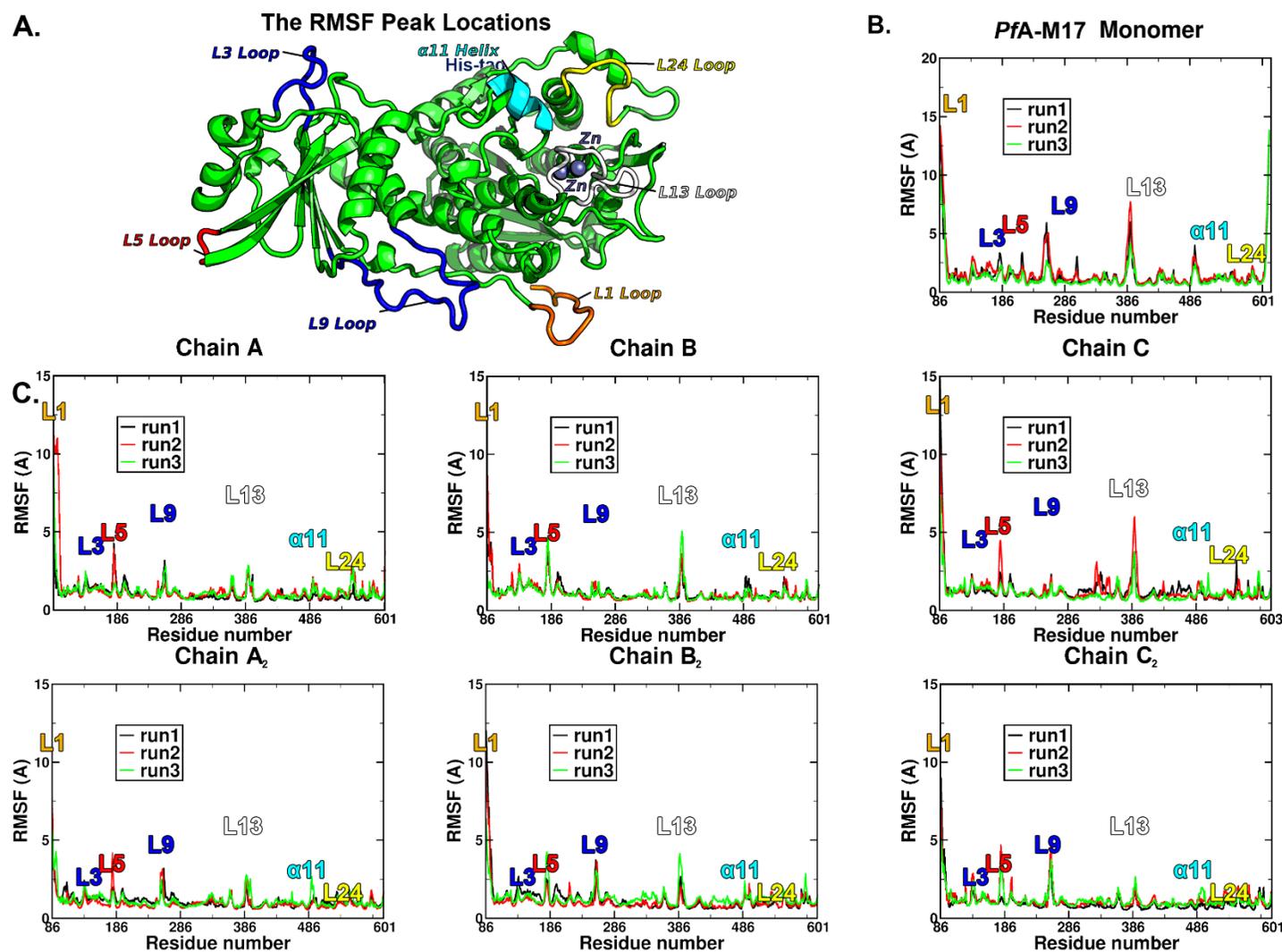


Figure 5.4. RMSF analysis of both monomer and hexamer simulations. (A) Cartoon model of the *PfA-M17* monomer (green) showing RMSF peak locations in different colors (labelled). (B) RMSF of $C\alpha$ atoms in monomer simulations. (C) RMSF of $C\alpha$ atoms in hexamer simulations shown by monomer ID (A, B, C, A₂, B₂, C₂). Run 1 is shown in black, run 2 in red and run 3 in green.

5.2.3. The hexamer protects the active site environment.

The active site of each *PfA*-M17 monomer has one or two cations, one carbonate ion and generally one catalytic water molecule (Fig 5.5), (81). The active form of the enzyme is presumed to have two cations within the active site (93) and all my *PfA*-M17 simulations had two cations in the active site (212). My choice to use my own FF for the MD meant that the metal ions in each simulation were effectively tethered to their coordination residues and as such, could not move throughout the simulation (212). Therefore, the only events I could observe were the overall stability of the active site residues, and the behavior of the water molecules and carbonate ion throughout the simulations.

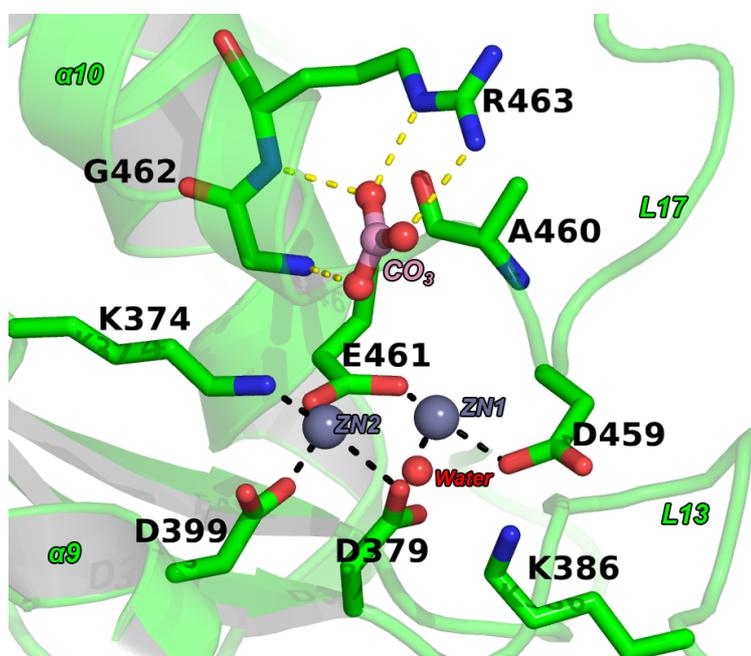


Figure 5.5. The architecture of the *PfA*-M17 active site.

The zinc ions and catalytic water are shown in grey and red spheres respectively. The nearby residues are shown in stick with the protein carbon atoms in green and the carbonate ion in pink. The metal bonds are shown as black dashes and hydrogen bonds as yellow dashes.

The residues that coordinate the metal ions show very little turbulence in both monomer and hexamer simulations (Fig 5.4). This is likely due to my choice to use a bonded model for the first shell protein atoms that meant that these residues were effectively tethered to their position. The substrate specificity pocket, or S1 pocket, that coordinates the P1 residues of the substrate is defined by residues M392, M396, F398, G489, and A577 (81). I found that the S1 pocket residues M392, Met396, P398 in L13 and T486, G489 and L492 in α 11 show more fluctuation in the monomer simulations (average RMSF 4.5 ± 0.5 Å) than in hexamer simulations (average RMSF 1.2 ± 0.4 Å). This indicates that the

hexamer may have a role in the maintenance of the shape and size of the S1 pocket for *PfA-M17*, thus possibly influencing substrate specificity.

In the M17 aminopeptidase superfamily, the carbonate ion found in the active site acts as a general base to accept protons from the nucleophilic water (178). For *PfA-M17*, the carbonate ion is not explicitly added to the recombinant enzyme, however, scavenges the ion from the environment (to date, it is always present in the active site of all structures solved (81, 107, 113, 114)). In the *PfA-M17* crystal structures, it is coordinated within the active site via water-mediated interactions to R378 and G377 (Fig 5.5). In all of the MD simulations (both monomer and hexamer), the carbonate ion was readily lost from the active site (within ~ 50 ns in all systems, the average distance between the CO₃ and the ZN2 is 23.5 Å at the end of the simulations). The loss of the CO₃ ion fits with the reported catalytic mechanism of M17, where it is proposed to act as a proton-acceptor and once this event has occurred, it would leave the active site to allow a new CO₃ ion in to continue to proteolytic process (179, 180, 270). In my simulations, we do not observe the 'replacement' of the CO₃ ion, however, I am unable to simulate the proton transfer event and also did not add excess CO₃ ions to the simulation protocol. Therefore, it is not surprising that this event was not observed. The addition of excess CO₃ ions to the simulation protocol would be useful if future studies were to investigate the precise atomic details of the catalytic reaction.

The behavior of the water molecules within the MD simulation was not restricted due to the decision to leave them non-bonded (212). Interestingly, I found that the water molecules in and around the active site behaved differently between the hexamer and monomer simulations (Fig 5.6). In the monomer simulations, zinc 1 (Zn1, coordinated by D379, E461 and D459, Fig 5.5) was coordinated by a water molecule. My MD showed that this water was maintained in only one trajectory (run 2) and had only ~ 80 % occupancy (Fig 5.6). This might be because there is more fluctuations within the structure of the active site in the monomer simulations than in the hexamer simulations. In contrast, all three trajectories of the hexamer had an exchangeable water site that was stably occupied and coordinated to the metal throughout each hexamer simulation in each of the six metal binding sites (100 % occupancy, Fig 5.6). I propose that the water molecule in this position

can act as the nucleophilic water for *PfA-M17*, and that the hexameric assembly appears necessary to stabilize the nucleophilic water site. This conclusion is supported by early work done by Chen *et al.*, who proposed that a bridging water molecule becomes terminally coordinated to only one of the zinc ions before the nucleophilic attack (271). The zinc 2 ion (Zn2, coordinated by K374, D379 and D399, Fig 5.5) was not coordinated by any water molecule throughout any of the simulations performed.

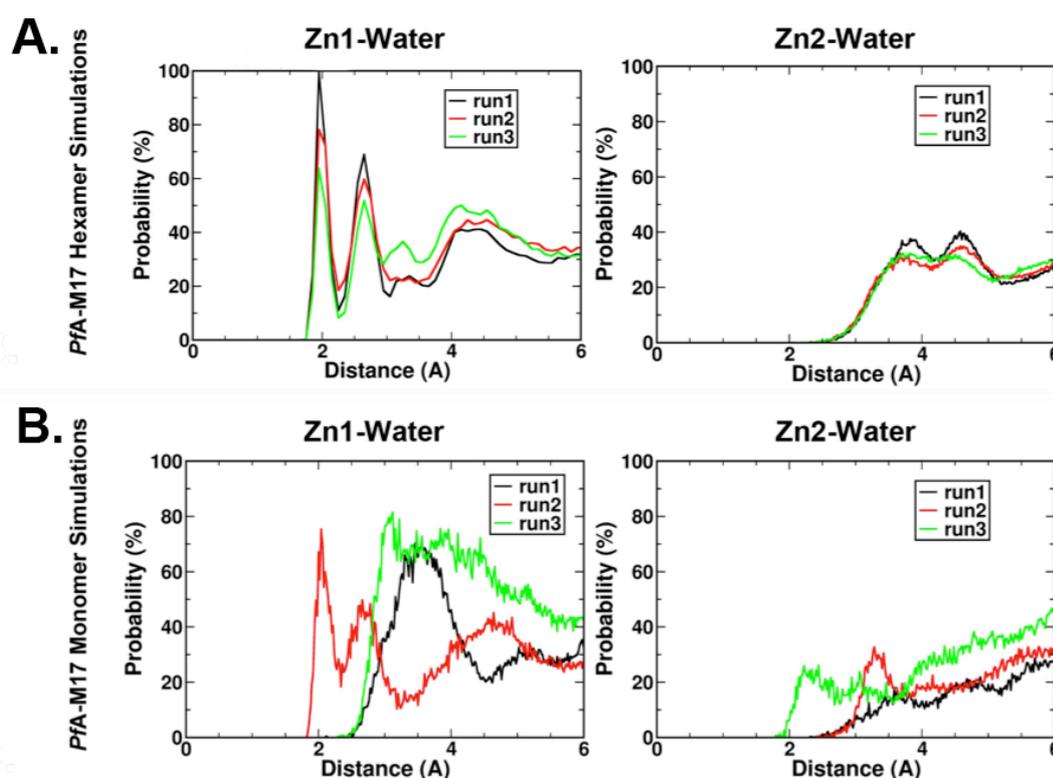


Figure 5.6. The behavior of water molecules in the active site of *PfA-M17*. Radial distribution function (RDF) plots of the water molecules around the zinc ions for hexamer (A) and monomer (B) simulations. The x-axis shows the distance between the water molecules and zinc atoms and the y-axis shows the probability of water molecules being found at that position.

5.2.4 Principle component analysis of the MD simulations

I was interested to see if there were any large conformational changes that were occurring within both our monomer and/or hexamer MD simulations. I therefore performed two separate principle component analyzes (PCA) on the simulation trajectories. The first

PCA focused on the dynamics of the hexamer (PCA_{hex}) and was calculated from the atomic displacement matrix of the hexamer backbone and zinc atoms, using the production phase of hexamer simulations (120,000 frames in total) as the sampling pool. The backbone and zinc atoms of the crystal structure of the *PfA*-M17 hexamer was used as the reference structure. The second PCA focused on the A-subunit type movements in both hexamer and monomer simulations (PCA_{mon}). To do this, I extracted the coordinates pertaining to the backbone and zinc atoms from each monomer within the hexamer simulation trajectories to create six A-type trajectories of backbone and zinc atoms. These six trajectories were then combined with the trajectories from the monomer simulations. The total combined A-type trajectory was then taken as the conformational sampling pool, and the PCA_{mon} of the backbone and zinc ion atoms was performed with the backbone and zinc ion of chain A from the crystal *PfA*-M17 as the reference structure

PCA defines a set of eigenvectors derived from the matrix of pairwise correlated motion of atoms. The eigenvectors, or principle components (PCs), are ranked according to the amplitude of the protein motions they describe. In general, the first eigenvectors that account for 50% or more of protein fluctuations are thought to describe the most significant motion within protein simulation (272-274). The contributions of the PCs from both of my PCA are shown in Table 5.1. I can see that the PC1 of both the hexamer and monomer PCA account for the majority of the overall covariance of each conformational sampling pool (Table 5.1). The percentage of PC2 values drops significantly and PC3 takes less than 7% in each PCA. Therefore, to analyze the global conformational transitions that were observed in my MD simulations, I chose to use only PC1 and PC2 in further analyzes.

Table 5.1. The percentage of the motion accounted for by the three PC values from the PCA_{hex} and PCA_{mon}.

	PCA _{hex}	PCA _{mon}
PC1	62.00%	50.78%
PC2	10.00%	12.59%
PC3	6.71%	6.52%

5.2.5. The essential dynamics of the *PfA-M17* monomer.

I was then curious about protein dynamics of the monomer alone, in the absence of the protein-protein interactions of the hexamer. In my PCA_{mon} (Fig 5.7), I analyzed the data from the monomer only simulation as well as the individual A like monomers within the hexamer. When all the A-type trajectories from monomer and hexamer simulations were plotted along PC1_{mon} and PC2_{mon} (Fig 5.7A), it can clearly be seen that the monomers that are not subjected to protein-protein interactions of the hexamer show very little motion in common with that of monomers from the hexamer (Fig 5.7A). The movements that projected onto PC1_{mon} from the monomer simulations (PC1_{mon(mon)}) showed an expansion of up to ~ 6 Å in size (Fig 5.7B), however, unsurprisingly the expansion for monomers within the hexamer was less than that of the monomers alone (Fig 5.7 A&C). L13 and L24 form part of the S1' specificity pocket for *PfA-M17*, and their dramatic movement in monomer alone simulations provide further support to the fact that the hexamer is involved in stabilizing the size and shape of the catalytic machinery. The PC2_{mon} identified significant different movements between monomer simulations (PC2_{mon(mon)}) and hexamer simulations (PC2_{mon(hex)}) within the C-terminal catalytic domain (Fig 5.7A, D, E). In the isolated monomers, L9 were observed to move towards α3 by 4.6 Å, while in the A type units of hexamer, L9 were observed to move apart from α3 by 3 Å. L13 and L24 were observed to move towards each other in the isolated monomer PC2_{mon(mon)}, which squeezed the S1' specificity pocket for *PfA-M17*. However, a subtler scale of opposite movements within L13 and L24 were observed in PC2_{mon(hex)}.

In summary, the isolated monomer shows extreme expansion both in the N and C-terminal domains, including the metal binding centers. However, I saw that the active site residues and metal binding center show little movement at all in the A-like units of hexamers. Different movements were observed between the two in the L9, L13, and L24. There is no significant movement observed for L24 nor the metal center, supporting the hypothesis that the hexamer acts to stabilize the active site and metal center for proteolysis.

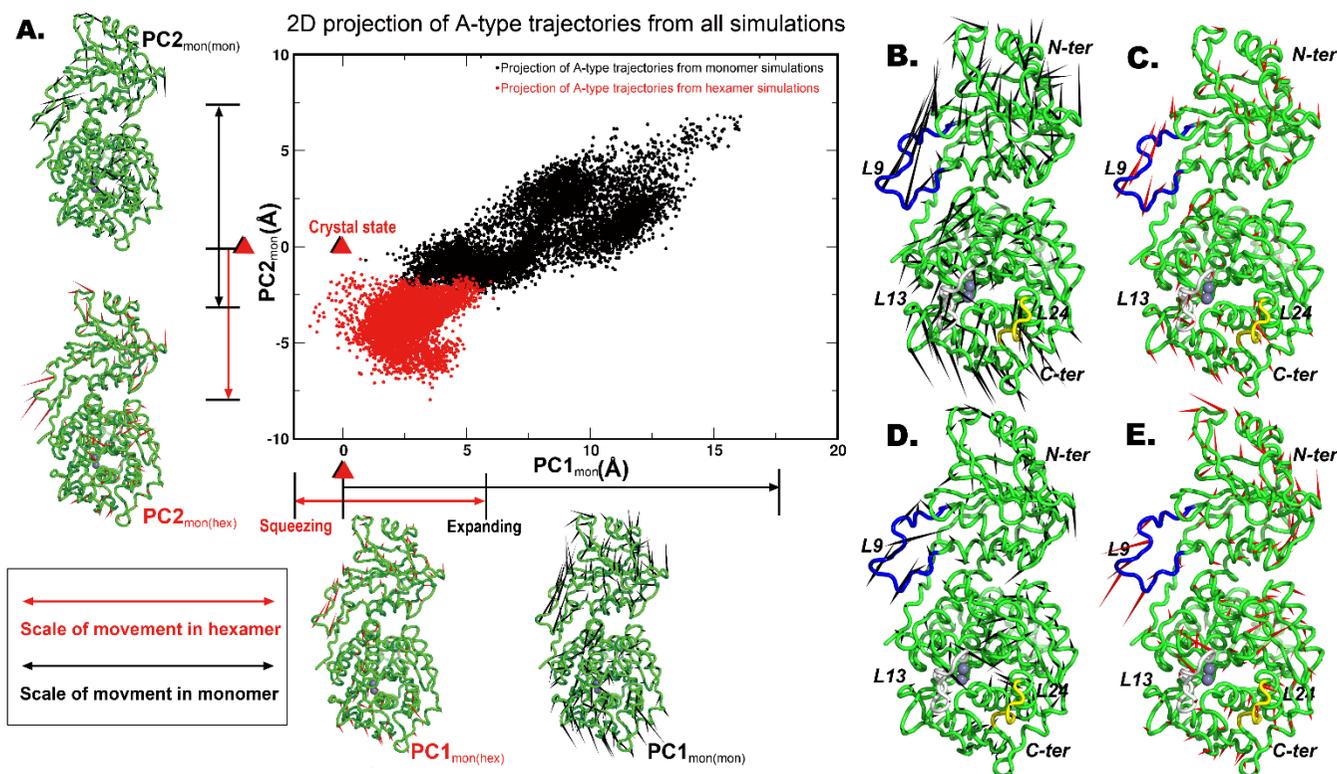


Figure 5.7. The PCA_{mon} on the A-like trajectories from monomer and hexamer simulations. (A)

The 2D projection of PC1_{mon} and PC2_{mon} using the A-type trajectories from all simulations.

The snapshots of the A-type trajectories from monomer and hexamer simulations are shown as projected dots by PC1_{mon} and PC2_{mon} (center) and the projected movements are shown by

porcupine plots surrounding. The porcupine arrows in the protein are proportional to the scale of the obtained PC movements. The protein is from chain A of the crystal structure. The schematic graphs of the single unit *PfA*-M17 are shown along the X- axis and Y- axis. The PC1_{mon(mon)} and PC2_{mon(mon)} are the movements that projected using the monomer simulations, while the PC1_{mon(hex)} and PC2_{mon(hex)} movements are projected using the A-like units from hexamer simulations. The detail movements are shown in panel (B) PC1_{mon(mon)} (C) PC1_{mon(hex)}, (D) PC2_{mon(mon)}, and (E) PC2_{mon(hex)}.

5.2.6 The essential dynamics of the *PfA*-M17 hexamer.

The essential dynamics of the *PfA*-M17 hexamer were projected onto the top two PCs from PCA_{hex}. The PC1_{hex} projected dynamics shows that the hexamer expands by an average of ~ 6 Å on each side (Fig 5.8). This is a known behavior in the simulation of crystal structures and reflects the correction in atom volume due to the release of crystal contacts (275). Analysis of PC2_{hex} showed a movement in L5^A that was connected to a rigid body movement of the entire N-terminal domain from the equivalent trimer subunit (Fig 5.8B). To describe this in more detail, the L5^A moves 4.4 Å from its original position (as per the crystal structure) toward the body of the hexamer. This appears to allow or facilitate a concomitant domain shift of the entire N-terminal domain by an equivalent amount (~ 5.2 Å) in the same direction (Fig 5.8E). I then focused on the movements of the neighbor units (A and A₂, B and B₂, C and C₂) by superposing the PC2_{hex} movements (the neighbor units B and B₂, C and C₂ were superposed on A and A₂). What was also interesting about this observation was that this movement was unidirectional and was not perpetuated throughout the entire hexamer. I did not observe the N-terminal domain rigid body movement in either A, B or C subunits, only in A₂ and B₂ (no movements were observed in either C or C₂, despite similar fluctuations of L5). Investigating why this might be showed me that the hexamer assembly is not perfectly symmetrical and the length of molecule from A to B is slightly smaller than the other two sides (Fig 5.8A). Such slight changes in length may result in variation in the motion observed.

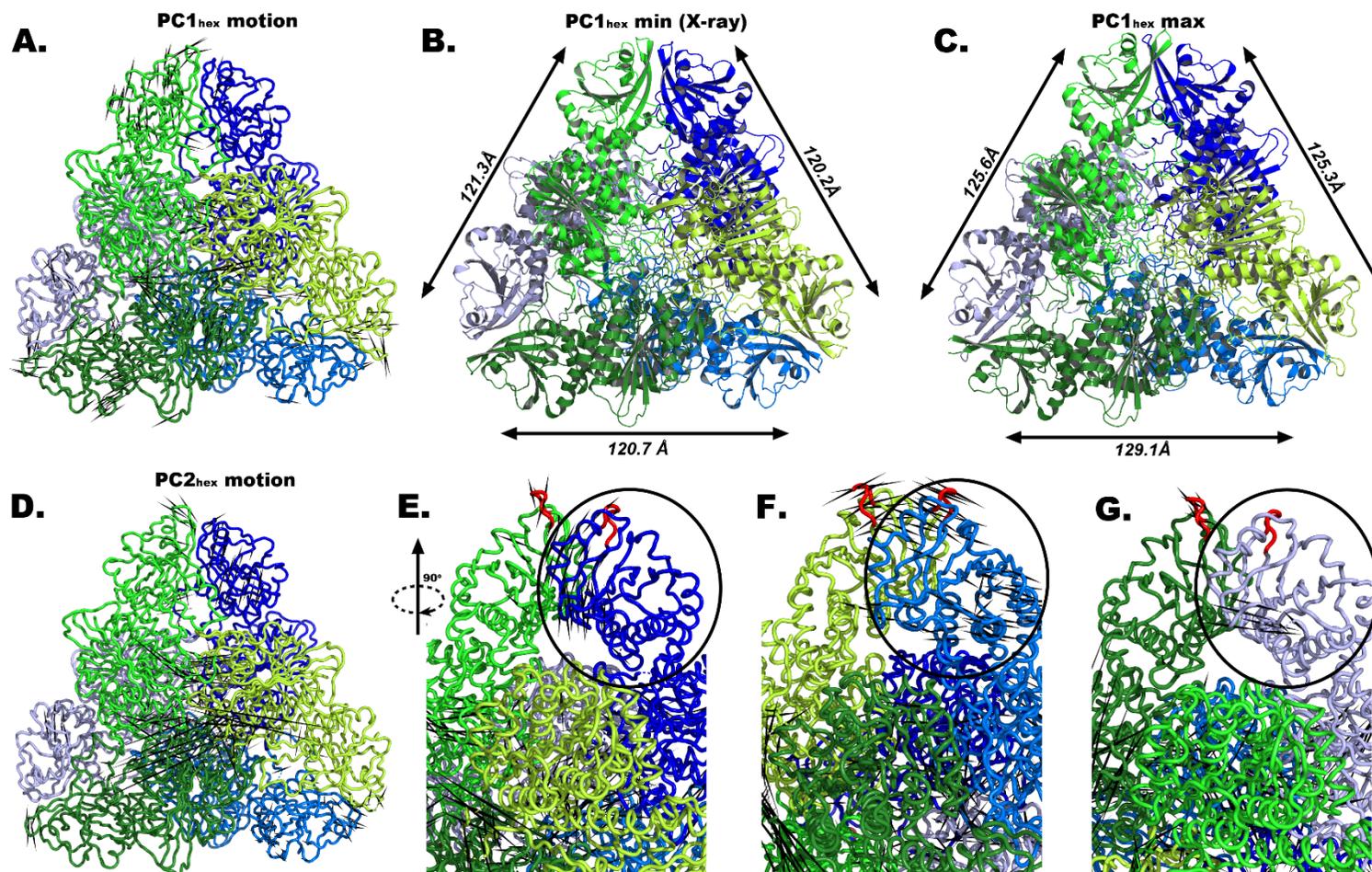


Figure 5.8. The essential dynamics of the *PfA-M17* hexamer. (A) PC1_{hex} shows the expansion of the three sides, which results in an increase in the overall size of *PfA-M17*. (B) The size of minimum structure (X-ray structure) in PC1_{hex}. (C) The size of maximum structure in PC1_{hex}. (D) Porcupine plots of the movements along PC2_{hex} showing synergistic movements between (E) A and A₂, (F) B and B₂ and (G) C and C₂. The black porcupine arrows on C α atoms represent the direction of movement and their length is proportional to the scale of detected motion. To reduce the complexity and focus on the significant movements, only movements that are greater than 1.5 Å were shown by the arrows. In (E), (F) and (G), L5 is shown in red cartoon and the N-terminal domain is circled in black.

In order to link the two major movements found in PC1 and PC2, I calculated a free energy landscape (FEL) derived from projecting the hexamer trajectories onto the PC1 and PC2 defined subspace (276-278). The FEL was then calculated based on the probability of each sampled conformation in the 2D projection. The result shows that from the conformation captured in the crystal structure, *PfA-M17* relaxes through a transition state before sampling two significant conformations (Fig 5.9). Comparing the Gibbs free energy of the states that are close to the starting structure ($G = 1739.96$ kcal/mol) and the transition state ($G = 1622.85$ kcal/mol) shows a ΔG of 117.11 kcal/mol and explains the relaxation of the structure.

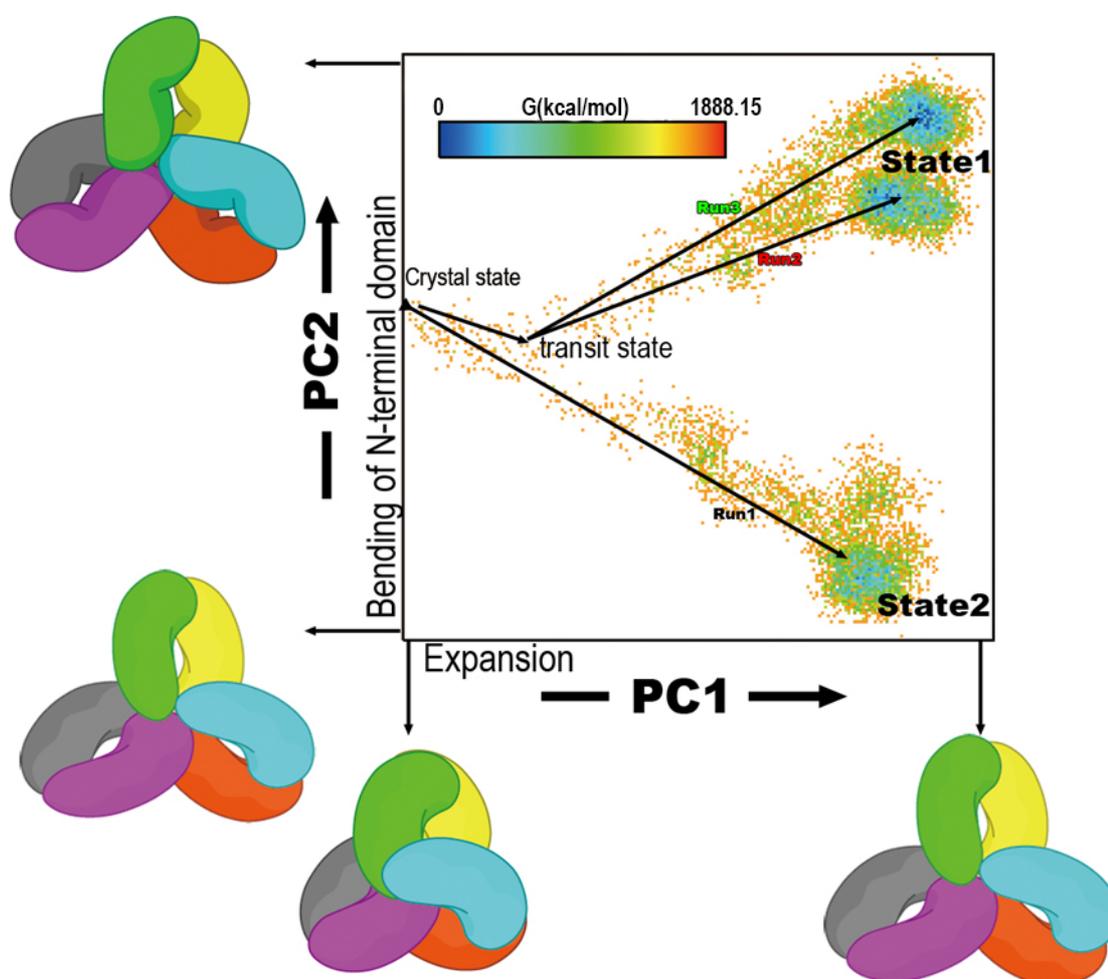


Figure 5.9. FEL of hexameric *PfA-M17*. The Gibbs free energy projection onto the space defined by PC1 and PC2. The schematic images of the *PfA-M17* conformational states are shown alongside the x- and y- axis. The color for Gibbs free energy ranges from blue (0 kcal/mol to red (1888.15 kcal/mol, the relative highest energy in the FEL.)

5.2.7. Hexamer movement is achieved by independent trimer movements.

To ascertain what was the driving force behind the movements of the hexamer further, I dissected the assembly of the *PfA*-M17 hexamer, and analyzed the dynamics of the sub-assembly of *PfA*-M17 in the hexamer trajectories. PDBePISA (238) was used to analyze the details of the structural stability of the possible sub-assemblies of the *PfA*-M17 hexamer (Table 5.2). The results showed that the most stable assembly was the hexamer with the internal energy $\Delta G_{\text{int}} = -110.2$ kcal/mol and a dissociation energy of $\Delta G_{\text{diss}} = 92.2$ kcal/mol. In addition to the hexamer, two other assemblies were identified; a trimer ($\text{ABC}/\text{A}_2\text{B}_2\text{C}_2$) or a dimer ($\text{AB}_2/\text{BC}_2/\text{CA}_2$). The trimers ($\Delta G_{\text{int}} = -24.6 \pm 0.2$ kcal/mol) appeared to be slightly more stable than the predicted dimers ($\Delta G_{\text{int}} = -22.8 \pm 0.5$ kcal/mol), however, the values were very similar (Table 5.2). The detected dissociation energy of the trimers drops drastically compared to that of a hexamer, but is still higher than that of the dissociation of the dimer (Table 5.2). This means the dimer assembly is likely to be the first to dissociate within the possible M17 oligomers.

Table 5.2. Analysis of potential *PfA*-M17 assemblies identified from PDBe-PISA.

Assembly types	Composition	ΔG_{int}^1 (kcal/mol)	ΔG_{diss}^2 (kcal/mol)
Hexamer	(ABC) ₂	-110.2	92.2
Trimer	ABC/A ₂ B ₂ C ₂	-24.6 ± 0.15	21.9 ± 3.5
Dimer	AB ₂ / BC ₂ / CA ₂	-22.8 ± 0.5	10.3 ± 0.9

¹ ΔG_{int} is the internal free energy of the given assembly type.

² ΔG_{diss} is the calculated average dissociation energy.

I then trimmed the hexamer trajectories into separate trimer and dimer trajectories and analyzed the two different assemblies to investigate their dynamics and identify the drivers of hexameric movement. I performed PCA on the trimer (PCA_{tri}) and dimer (PCA_{dim}) trajectories using ABC and AB₂ of crystal structure as references, respectively. For the trimers, the top 3 PCs take 51.3 % of the overall covariance in trimer trajectories (Table 5.3), however, the top 3 PCs for the dimer only account for 42.12 % of the dimer covariance

(Table 5.3), indicating that the significant movements in dimer cannot be reasonably separated by only the top three PCs.

Table 5.3. The percentage of the top three PC values from the PCA_{tri} and PCA_{dim}.

	PCA _{tri}	PCA _{dim}
PC1	29.57%	19.91%
PC2	14.42%	11.34%
PC3	8.32%	10.87%

The trimer and dimer trajectories were then projected onto their top two PCs, respectively (Fig 5.10). The distribution of trajectory snapshots on PC2 represents the conformational similarity and difference. That is to say when the distribution of snapshots is highly crowded in certain areas of the PC defined subspace, the snapshots are in close proximity to each other, whereas a high desecration of snapshots distribution in the subspace means the conformations of the snapshots are different. Interestingly, the result of trimer 2D projection showed that the conformations of two trimers were separated clearly by the PC_{tri1} vector (Fig 5.10A). This indicates that the trimers are moving independently from one another within the hexameric arrangement. However, the 2D projection of dimer trajectories (Figure 5.10B) suggests that even though PC1_{dimer} only takes 19.9% of overall covariance, and dimer PCA cannot separate the unique motions among the three dimer trajectories. Moreover, there are no independent or dependent movements that can be dissected as the projections of the three dimers do not clearly overlap nor are completely separate from each other.

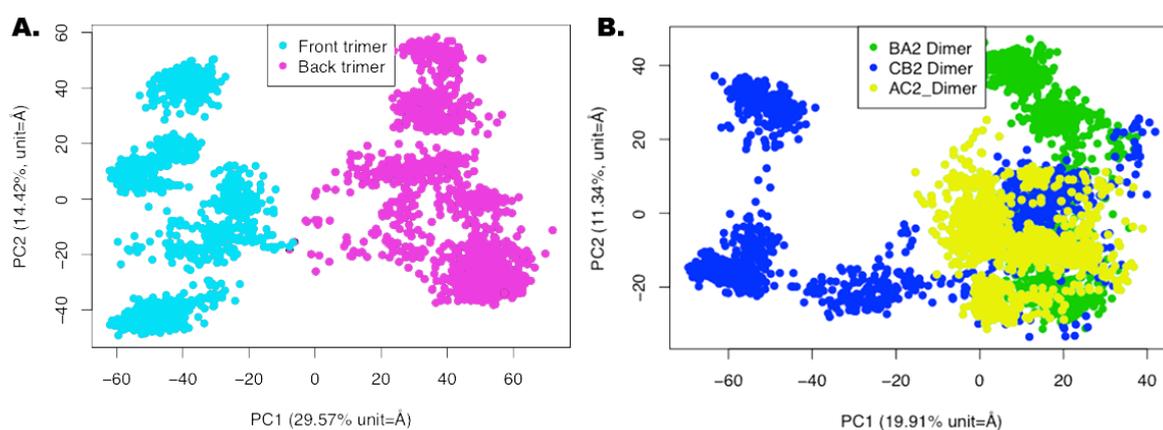


Figure 5.10. The 2D projection of PC1 and PC2 for the (A) PCA_{trimer} and (B) PCA_{dimer} .

Conformations from the front and back trimer trajectories were projected as cyan and magenta dots in panel A, respectively. Conformations from the BA₂, CB, and AC dimer trajectories were projected as green, blue, and yellow dots in panel B, respectively.

5.2.8. *PfA-M17* has an entrance to the active site cavity that is occluded by cooperative movements of L9 and L3.

My work indicates that the dynamics of the trimers are the driving force behind hexamer movements, I was then curious about the correlated dynamics between the front trimer (ABC) and back trimer (A₂B₂C₂) that might account for the relevant biological events such as substrate guidance and processing. Therefore, I firstly looked for such trimer communication from the projected movements along PC1 from the PCA_{hex} . Interestingly, I found that there is a small alternate but coordinated motion of L9 with L3 of all the six units (Fig 5.11). It was postulated that movement of L9 (residues 247-265) might have a role in guiding substrates entering in *PfA-M17* (81). L3 (residues 137-145) is located at a trimer protein-protein interface between A and A₂ and has previously also been identified as flexible in crystal structures (81). The movement indicated that when L9 and L3 of the front trimer are moving close to each other (L9^A and L3^{A₂}, L9^B and L3^{B₂}, L9^C and L3^{C₂}), the L9 and L3 from the back trimer are moving away from each other (L3^A and L9^{A₂}, L3^B and L9^{B₂}, L3^C and L9^{C₂}) (Fig 5.11). In order to monitor the movements in the two loops along PC1 sampled from all hexamer simulations, the distance between the C α atoms of N258 and N139 were measured from the two most extreme structures in PC1 (Table 5.4). It can be

seen that the distance between these residues on the ABC trimer decreases by $2.8 \pm 2.5 \text{ \AA}$ from max to min conformation. At the same time, the distances of the back entrances increase by $7.1 \pm 3.7 \text{ \AA}$ (Table 5.4). However, the loop movement is not uniform between the subunits as I noticed that the distance between Chain A N258:C α and Chain A₂ decreased (Table 5.4). This door-like motion between trimers displays a coordination between the dimer of trimers and supports the hypothesis that this region is involved in regulation of the enzymatic function by acting to regulate the entrance of peptide substrate to the inner catalytic sub-compartment (81).

Table 5.4. The distances between C α of N258 and N139 from the two extreme structures of PC1_{hex}. Min and max represent the two extreme structures of the PCA_{hex}.

Chains		A and A ₂	B and B ₂	C and C ₂	Average (Å)
Residues		A:N258 – A ₂ :N139	B:N258 – B ₂ :N139	C:N258 – C ₂ :N139	
Front entrance	max	10	6	12	9.33 ± 2.5
	min	6.6	6.5	6.6	6.57 ± 0.05
	Δ Dist	3.4	-0.5	5.4	2.77 ± 2.5
Residues		A:N139 – A ₂ :N258	B:N139 – B ₂ :N258	C:N139 – C ₂ :N258	
Back entrance	max	7.9	7.9	6.3	7.37 ± 0.8
	min	11.8	13	18.5	14.43 ± 2.9
	Δ Dist	-3.9	-5.1	-12.2	-7.07 ± 3.7
Chains		A and A ₂	B and B ₂	C and C ₂	
Residues		A:N258 – A ₂ :N139	B:N258 – B ₂ :N139	C:N258 – C ₂ :N139	

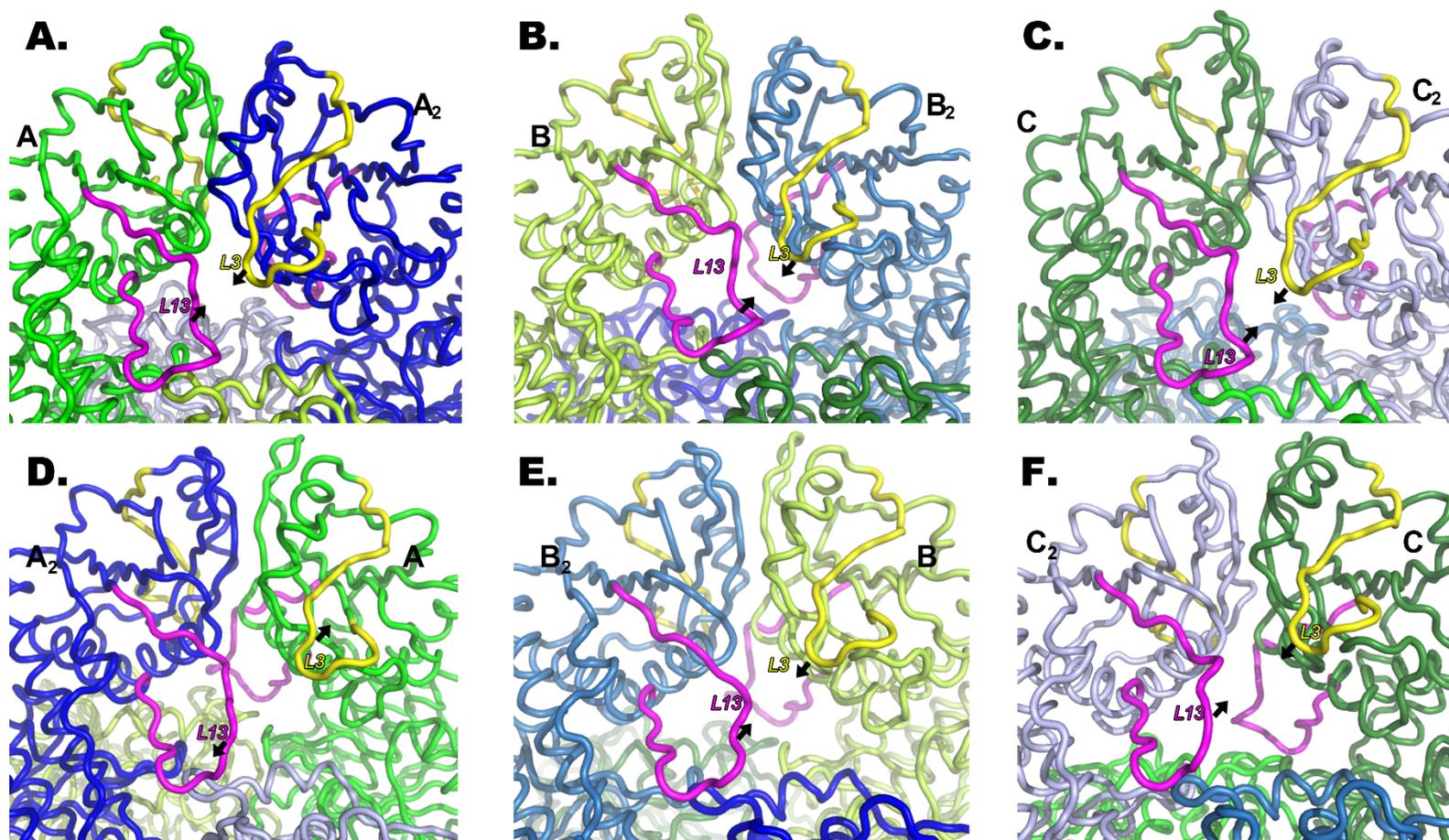


Figure 5.11. Representative movements of L9 and L3 along PC_{hex1} . The L13 and L3 loops are shown in magenta and yellow, respectively. The black arrows indicate the loop movements. The movements between (A) A^{L9} and A_2^{L3} , (B) B^{L9} and B_2^{L3} , (C) C^{L9} and C_2^{L3} , (D) A_2^{L9} and A^{L3} , (E) B_2^{L9} and B^{L3} and (F) C_2^{L9} and C^{L3} .

5.2.9. Cooperativity in the conserved metal binding sites of the hexamer.

PC2 of PCA_{hex} also identified communication between the six catalytic domains in the projected movement. Fascinatingly, I observed opposing movements of L13 from each subunit within its respective trimer, where L13^A and L13^B move together to the same direction and appear to push L13^C toward the catalytic zinc ions in the C active site (Fig 5.11B). A similar phenomenon was observed in the back trimer with the concerted motion of L13 of A₂ and B₂ resulting in the movement of L13 C₂ to occlude the zinc ions of the chain A₂ active site (Fig 5.11). At the same time, I observed that α 11 and L24 move towards the N-terminal domain of their respective A-type unit in the hexamer dynamics (Fig 5.11). These synergistic movements indicate an important role for L13 in the catalytic mechanism and / or regulation of activity within the M17 superfamily.

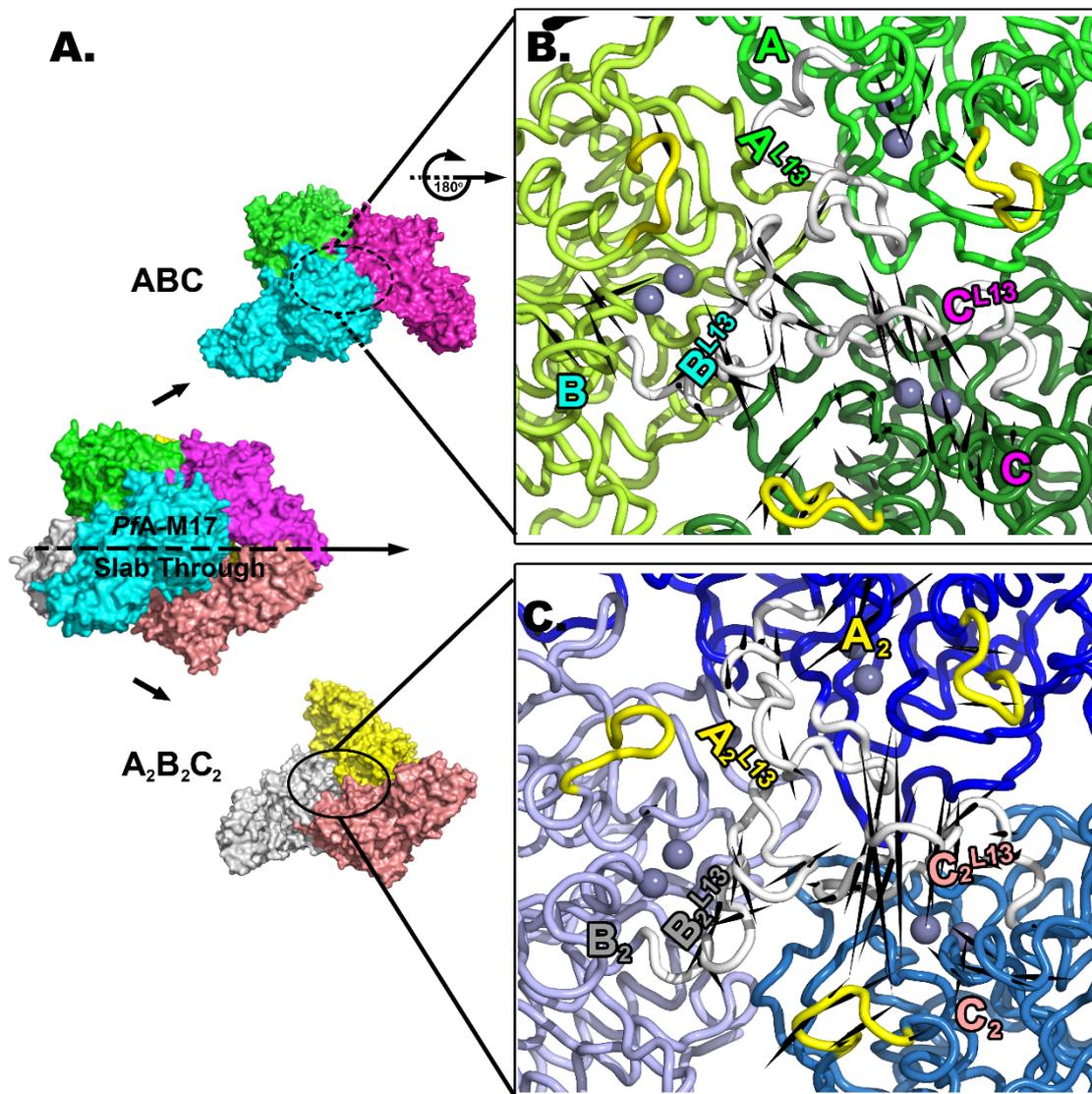


Figure 5.11. *PfA-M17* may have cooperative dynamics within the active sites. The catalytic movements captured in the front (ABC) and the back (A₂B₂C₂) trimer along PC2. (A) The slab through view of the hexamer *PfA-M17*. (B) The catalytic movements in the front trimer. (C). The catalytic movements in the back trimer. The porcupine arrows are black, and the L13 in each unit are in white tubes, and the L24 are in yellow tubes. The zinc ions are shown as spheres.

5.3. DISCUSSION

5.3.1. Why does *PfA-M17* form a hexamer?

Many M17 proteins are reported to be functional hexamers (81, 170, 176, 178). A major question for *PfA-M17* was why does the malaria parasite invest so much energy to

produce the complexity and size of the hexameric protease? Another question is how the protein dynamics contribute to the function of *PfA-M17*? Therefore, I performed MD simulations of both monomer and hexamer *PfA-M17* to find the answers to these questions. The recombinant *PfA-M17* is truncated due to the presence of low complexity regions. I am unable to consider the implications of this truncation in my study. I did observe that L1 in N-terminal domain was extremely flexible in both hexamer and monomer simulations. This is likely due to the truncation of the N-terminus and the removal of protein / peptide that it might interact with. Simulation of the hexamer showed that it supported the proper shape of the substrate binding pockets, as the loops that involved in forming the substrate-binding pocket in monomer simulations moved more than in the hexamer simulations. The lack of stability of the monomer simulation also indicated that protein-protein interactions are important for enzyme stability.

I also found that the hexamer can help to protect the active site environment, as the important nucleophilic water is lost in the monomer simulation, but retained in the hexamer simulations. In the other M17 proteins, the nucleophilic water was also found in between the zinc ions of each A-type units (167, 172, 279-281), our results indicate that the assembly of a hexamer is important to maintain the position of the catalytic water molecule for M17 aminopeptidase family.

5.3.2. Do trimers regulate the activity of *PfA-M17*?

My investigation into the different assemblies within the hexamer indicated that the trimers may be capable of communication. The putative substrate entrance regions showed opposing motion between the trimers, suggesting that the trimers drive the movement of the substrate guiding loops in a hexamer *PfA-M17*. The L9 loop was found to be very important for this motion and it appears that flexibility or dynamics in L9 is conserved among the M17 aminopeptidase family. Among the M17 aminopeptidase crystal structures available, all of the structures have high B-factors or missing residues in L9 (81, 167, 172, 280, 282, 283), except for the M17 from *Trypanosoma brucei* (284). In this M17, the L9 residues (170-177) form an extensive hydrogen network from main chains atoms and likely

cause a more rigid loop structure. The sequence of the L3 region is not conserved shared by the family members (81, 167, 172, 280, 282, 283), however, high B-factors suggest that the L3 dynamics also exists in the bovine lens LAP (167) and the tomato LAP (282). However, in *Pseudomonas putida* (280), *Staphylococcus aureus* (283) and *Helicobacter pylori* (172), L3 is much shorter or even missing altogether. Thus, I suspect that substrate guidance / entry is controlled by the coordinated motion of the L3 and L9 loops, and that in higher organisms such as parasites, mammalian and plants, communication between the trimers helps regulate substrate movement. However, in lower organisms, such as bacteria, there may exist a different regulation of substrate entry and exit that does not rely on the dynamics of L3.

5.3.3. A model for the catalytic mechanism of the M17 aminopeptidase superfamily.

The L13 loop (residues 388-391) appears to have a dynamic and fundamental role for the M17 proteolytic mechanism. In the original crystal structure, the L13 loops had high B-factor and this supports my results from PC2_{hex} that L13 undergoes a radical movement. The position of L13 is different among the available crystal structures of different M17 aminopeptidases and alignment of the structures shows the different positions (81, 167, 172, 280, 282, 283). In the *Pseudomonas putida* LAP, the position and stability of the loop has been shown to be affected by the environmental pH (279, 280). Sequence alignment of the loop region shows sequence diversity (Fig 5.13,(172). This could mean that dynamics may be driving evolution or vice versa and providing a mechanism for different catalytic mechanisms, regulation or even substrate specificity. My data shows that K386 is highly mobile during the L13 movement of PC2_{hex}. The equivalent lysine is highly conserved and has been captured interacting with the P1 substrates in several studies (167, 172, 279, 280). It is known to be important for enzymatic activity. Therefore, from my current study, I hypothesize that the role of K386 in PfA-M17 is to hold the substrate at an appropriate position for hydrolysis as well as potentially guide the product away from the active site after hydrolysis.

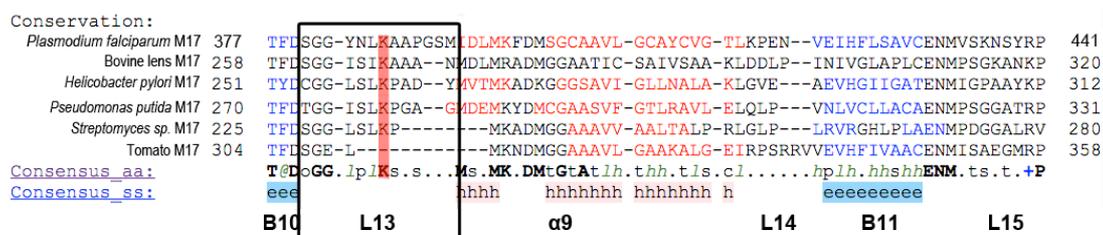


Figure 5.13. A structure-based sequence alignment of L13 regions of M17 aminopeptidases by PROMALS3D server. Consensus_aa is the conserved amino acids among the given eight M17 proteins. In bold, the single letter amino acid code means the amino acid is conserved at that position. Consensus_ss is the consensus secondary structure. Shown at the bottom is the topology index defined in this study. The structural information used here are *PfA*-M17 (PDB ID 3QKZ) (81), bovine lens M17 (PDB ID 2EWB) (167); *H. pylori* M17 (PDB ID 4ZLA) (172); *Caenorhabditis elegans* M17(PDB ID 2HB6), *P. putida* M17(PDB ID 3H8G) (280), *Streptomyces* sp. M17 (PDB ID 5LHJ) (283), and tomato M17 (PDB ID 5D8N) (282).

5.4 CONCLUSIONS

Summarising the biologically relevant protein dynamics from this study, my results suggest that substrate entrance and product exit from the hexamer of *PfA*-M17 requires communication between the trimers. This in turn, provides some rationale for the need for a hexameric assembly for proteolysis. The hexamer further acts to stabilize the catalytically competent structure of *PfA*-M17 and maintain a catalytic water at all times in the active site. It is possible that assembly and / or dissociation of the hexamer may be a mechanism by which the parasite controls proteolysis by *PfA*-M17 throughout different life-cycle stages. From a biophysical perspective, future studies would need to employ hybrid QM/MM simulations to consider the actual reaction dynamics and capture any possible movement of the zinc ions. However, the size of this particular system currently makes such a study prohibitively expensive in terms of computational power. Further, longer time scales for all-atom MD as performed here may also benefit this work but have the same cost associated with them. The current bottleneck of the approach is the development of affordable supercomputers and improved simulation methods.

CHAPTER SIX

PROTEIN DYNAMICS OF THE CLAN MA, FAMILY M1

AMINOPEPTIDASES.

6.1. INTRODUCTION

The clan MA, family M1 aminopeptidases are recognized as attractive drug targets for various disease states including infectious diseases and cancer (4, 81, 87, 88, 285, 286). To date, the structures of 13 different M1 aminopeptidase enzymes have been reported from humans, animals, insects, parasites, bacteria and archae (1, 4-6, 144-148). Four have been captured in both their 'open' and 'closed' forms (144, 149, 216, 247). Studies on these enzymes have postulated that the open form is catalytically inactive and exists to guide the peptide substrate to the active site (144, 147). The closed form is considered the catalytically active form where the hydrolysis of the P1-P1' peptide bond occurs (144, 147). However, the molecular mechanism that underlies the transition between the two states is still unclear. In my previous study (chapter 3), I showed the *PfA*-M1 appears to undergo a similar transition from closed to open. Prior to my study, it was assumed that the bacterial and parasite M1 aminopeptidases may only occupy a closed form. To try and understand how the protein dynamics may vary in different species of M1 aminopeptidase, I sought to characterize the protein dynamics of selected representatives of the family. Understanding the protein dynamics of different M1 aminopeptidases may prove beneficial to future drug design due to differential dynamics producing variation in druggable pockets between the family members.

6.2. RESULTS

6.2.1. Selection of representative M1 aminopeptidase structure for analysis.

All-atom MD simulations of large proteins remains computationally expensive and this meant I was unable to simulate every unique M1 aminopeptidase that was present in the protein databank (PDB). Therefore, to study the diversity in protein dynamics throughout the whole M1 aminopeptidase family, I selected seven representative ectodomain proteins for which there was structural information. Whilst all M1 aminopeptidases to date have been found to have an N-terminal membrane bound anchor, the challenge of structurally characterizing this region of the protein has precluded any structural information to be gathered on the region. Therefore, my analysis will only consider the ectodomain of the M1

aminopeptidase family. The selected seven structures had the following three features, (1) coverage of the various species within kingdoms of life, (2) each must contain the four complete domains that is characteristic of the aminopeptidase N fold and (3) had PDB coordinates that were available to the public prior to the commencement of this study (January 2015). It is worth noting that I excluded any member of the leukotriene A4 hydrolase family due to the fact that many members of that sub-family lack domain III (and therefore did not fit the criteria for inclusion).

Given those caveats, I chose nine types of M1 aminopeptidases to analyze that came from four different mammalian homologs, one bacterial and one archaea enzyme (Table 6.1). The selection includes two open and seven closed forms of M1 aminopeptidase. The two open forms (TIFF3 and *hERAP1*) also have their closed forms available for comparison. As described before, M1 aminopeptidases exist as both monomers and dimers (Fig 4A and 4B), and the dimers are generally membrane bound and can be found surface-exposed or within the cytoplasm (145, 146). In this work, I focus on the dynamics of the monomer, as not all of the selections exist in a dimer assembly and monomers are found to be biologically functional. The dimer interactions and modelled dimer dynamics based on monomer dynamics will be addressed later in this chapter.

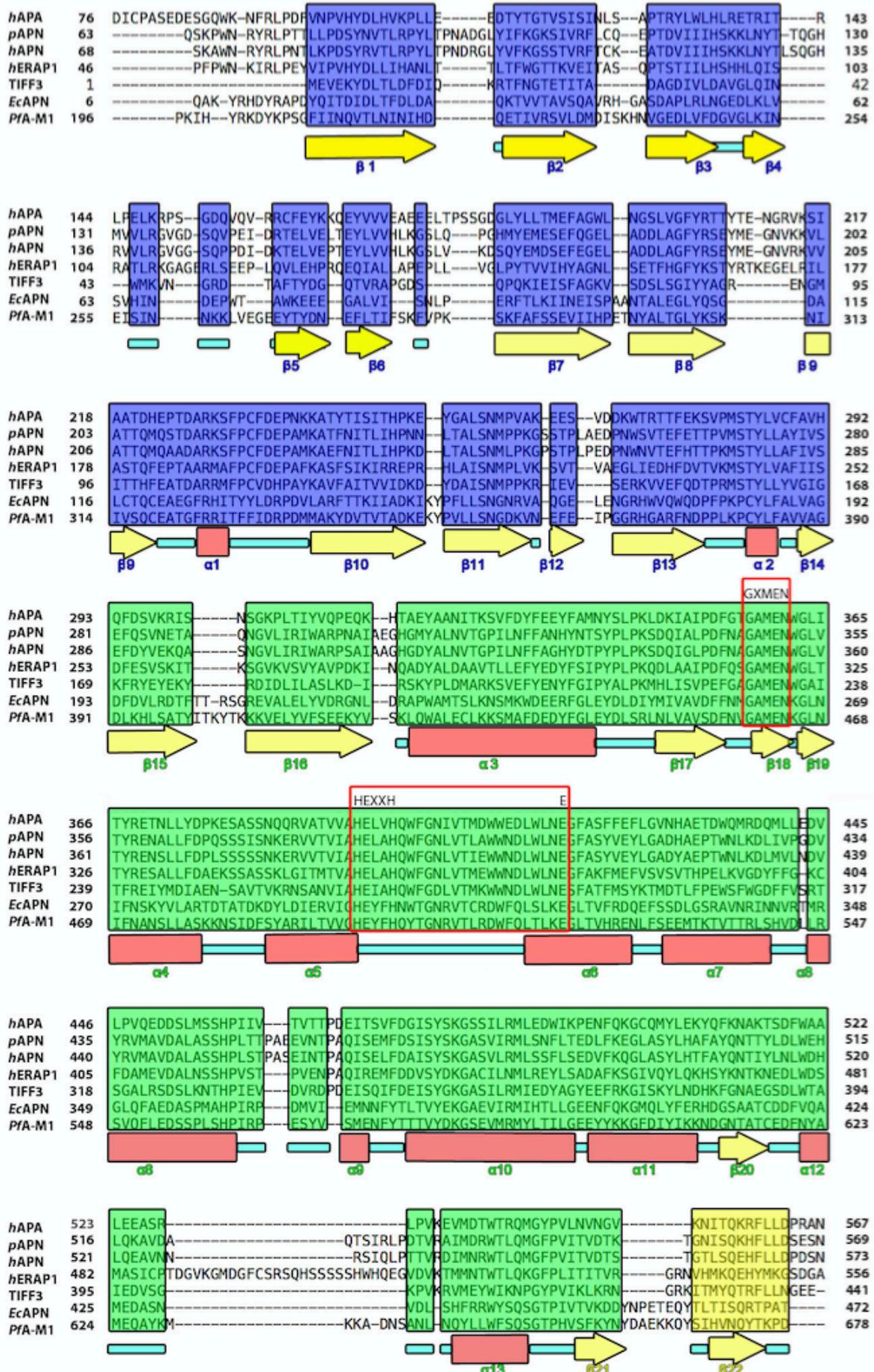
To compare the variation in secondary structure, I performed a structure-based sequence alignment of the selected M1 representatives (Fig. 6.1) by putting the PDB structures in PROMALS3D server (239). I chose to use a structure-based sequence alignment as a sequence only alignment (E.g. BLAST) would not take into consideration the known structural features of the proteins. I did attempt a BLAST alignment of the sequences, however, there was significant mismatch in domains III and IV, which disagreed with the known domains within crystal structures. The alignment showed that the core conserved elements included 14 β -sheets and 2 α -helices in domain I; 6 β -sheets and 10 α -helices in domain II; 5 β -sheets in domain III and 17 α -helices in domain IV. To allow me to discuss conserved secondary structural elements across the representatives, I have renamed the conserved secondary structures for use in later analysis (Fig 6.1). Finally, for domain IV of the M1 aminopeptidases, I have dissected the domain into the two modules

that were first reported in TIFF3 (149). In the alignment shown in Figure 6.1, the first module is α 14 - 19, and the second module α 20 - 30.

Table 6.1. Selected M1 aminopeptidase structures analyzed in this study.

Protein	Species	Abbreviation used in this study	PDB ID: chain	Resolution (Å)	Open or closed?	Residues rebuilt by Modeller	Reference
Tricorn Interacting Factor F3	<i>Thermoplasma acidophilum</i>	TIFF3	1Z5H:A	2.3	open	-	(149)
Tricorn Interacting Factor F3	<i>T. acidophilum</i>	TIFF3	1Z5H:B	2.3	closed	-	(149)
Aminopeptidase N	<i>E. coli</i>	<i>EcAPN</i>	2DQ6:A	1.5	closed	-	(158)
<i>PfA-M1</i>	<i>P. falciparum</i>	<i>PfA-M1</i>	3EBG:A	2.1	closed	-	(4)
Aminopeptidase A	<i>Homo sapien</i>	<i>hAPA</i>	4KX7:A	2.15	closed	608-611	(147)
Aminopeptidase N	<i>H. sapien</i>	<i>hAPN</i>	4FYQ:A	1.9	closed	891-898	(1)
Endoplasmic Reticulum Aminopeptidase 1*	<i>H. sapien</i>	<i>hERAP1</i>	2YD0:A	2.7	open	486-513, 553-557	(5)
Endoplasmic Reticulum Aminopeptidase 1*	<i>H. sapien</i>	<i>hERAP1</i>	3MDJ:A	2.95	closed	417-433, 486-514, 552-555, 864-867, 893-906	(216)
Aminopeptidase N	<i>Sus scrofa</i>	<i>pAPN</i>	4FKE:A	1.85	closed	-	(144)

* Structure has bestatin bound.



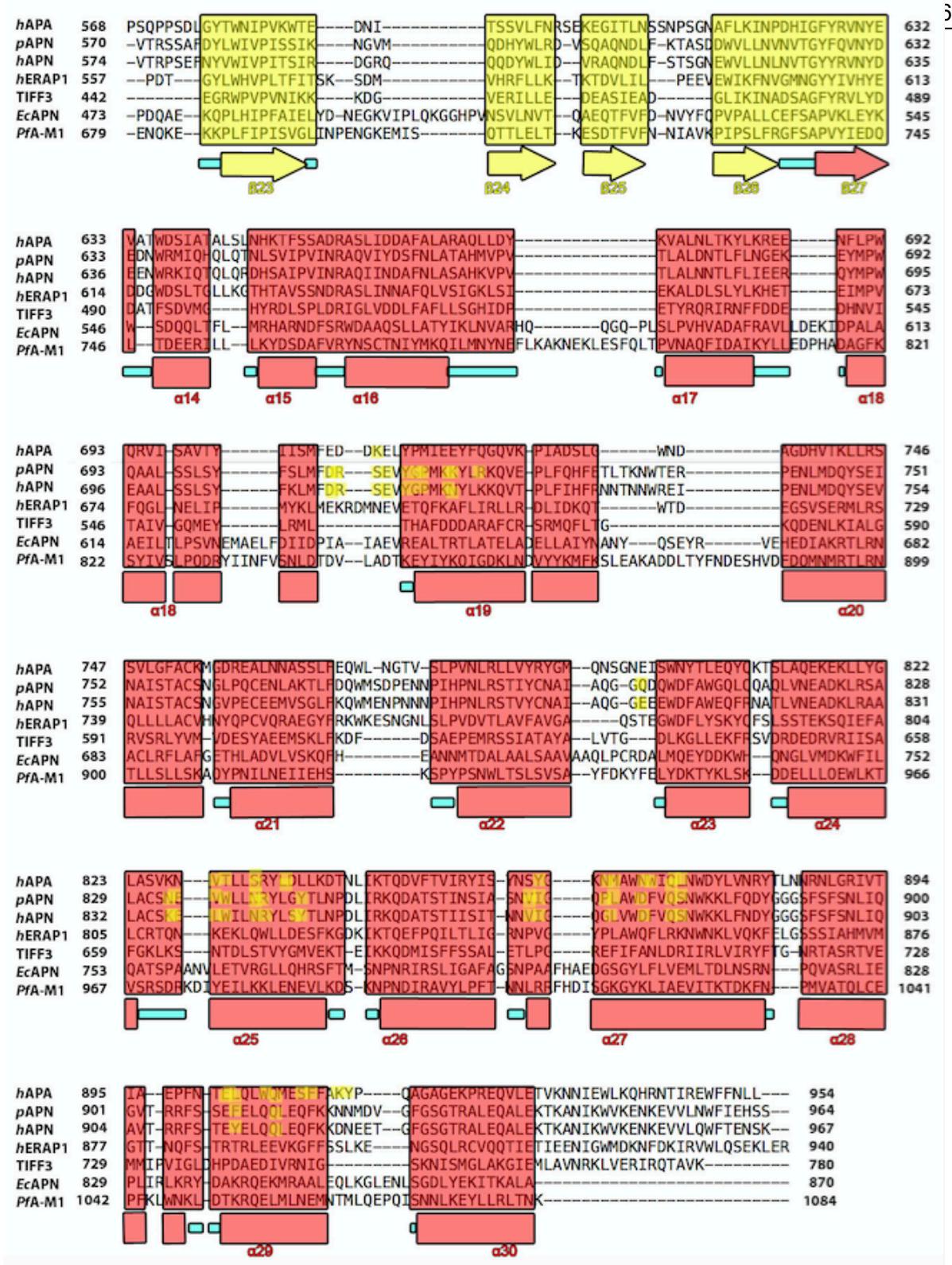


Figure 6.1. Structure-based sequence alignment of selected M1 aminopeptidases.

Protein names are shown on the left with the respective numbers of each protein. Conserved secondary structure elements are boxed and nomenclature indicated at the bottom of each box. Color codes apply as per standard M1 coloring (domain I blue, domain II green, domain III yellow and domain IV red). Residues highlighted in yellow are involved in dimer interfaces. GAMEN and zinc-binding motif are in red box and indicated.

6.2.2. MD simulations of selected M1 aminopeptidase structures.

In this chapter, I have investigated the dynamics of M1 aminopeptidases by performing MD simulations on the selected structures (Table 6.1). I performed 200 ns of MD simulations, each with three replicates, on the closed form of *EcAPN*, *hAPN*, *hERAP1*, *hAPA*, *pAPN*, and *TIFF3* systems and the open forms of *hERAP1* and *TIFF3*. The results of the *PfA*-M1 simulations presented in chapter 3 are also included for comparison purposes.

To begin the analysis of M1 aminopeptidase simulations, the stability of the simulations was checked by RMSD analysis. The RMSD values of the protein C α atoms were calculated and compared to the C α atoms of each starting snapshot (Fig 6.3). The results indicate that the RMSD of the C α atoms in the systems increases gradually in the first 25 ns and then remains stable to the end of the simulations. Comparing of the RMSD average values among the production phase of the six systems, an immediate observation is that four M1 systems (3.5 ± 0.4 Å, for closed *hERAP1*, 3.4 ± 0.5 Å for the open *hEPRA1*, 3.9 ± 0.3 Å for *pAPN*, 2.8 ± 1.4 Å for open *TIFF3*) have higher RMSD values than that of the other four M1s (*EcAPN* 2.1 ± 0.3 Å, *hAPA* 2.0 ± 0.2 Å, closed *TIFF3* 2.4 ± 0.5 Å and *hAPN* 2.3 ± 0.2 Å). This indicated that *EcAPN*, *hAPA*, closed *TIFF3* and *hAPN* have less global conformational changes during the simulations. Run 2 of closed form *hERAP1* has a larger RMSD (average 3.4 Å ± 0.3 in production phase) than that of the other two replicates of *hERAP1*, suggesting that the conformations in run 2 are different from the other two runs. The three runs in *pAPN* quickly increased to 3.8 Å in the first 19 ns and stayed steady at 3.9 ± 0.3 Å throughout the remaining simulation, indicating that a large conformational change apparently takes place rapidly before equilibration is reached. The average RMSD of the open *TIFF3* triplicates all differed with run 1 showing a stable RMSD of 2.0 Å but runs 2 and 3 show a slow increase in RMSD at different times (run2 ~ 29 ns and run3 ~ 70 ns) indicating a large movement within the protein at these times.

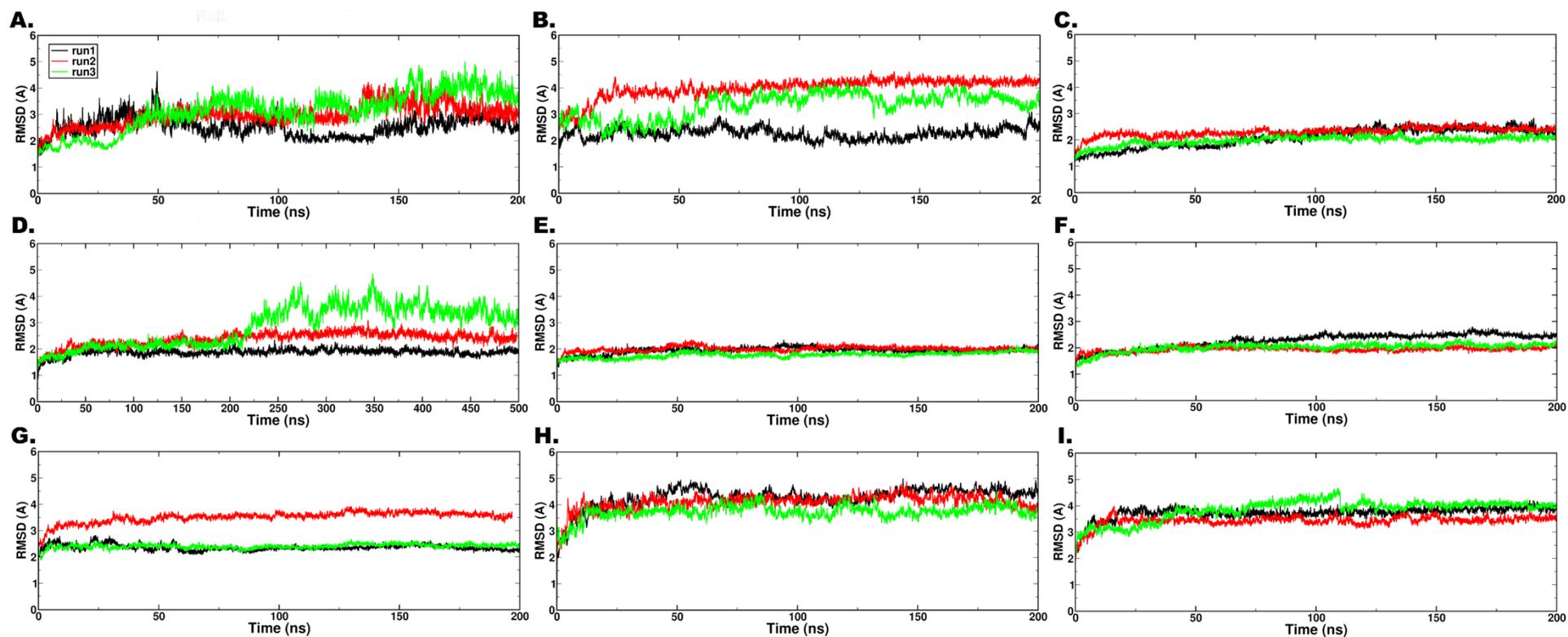


Figure 6.3. The RMSD of each M1 aminopeptidase MD simulation. (A) closed TIFF3, (B) open TIFF3, (C) *EcAPN*, (D) *PfA-M1*, 500 ns from chapter 3, (E) *hAPA* (F) *hAPN* (G) closed *hERAP1* (H) open *hERAP1* and (I) *pAPN*. RMSD were calculated based on C α atoms compared to the starting crystal structures of each system. Replicates indicated by color where run 1 = black, run 2 = red and run 3 = green.

The detailed analyzes of root-mean-square fluctuation (RMSF) of the backbone atoms versus the protein residue numbers from the equilibrium phase of the production MD simulation (25-200 ns) are depicted in Figure 6.4. The immediate finding is that the catalytic domain II and C-terminal domain IV are highly stable for all enzyme representatives, however, with the exception of ERAP1 and TIFF3, the residues of the third domain of M1 aminopeptidases are subject to fluctuation. In both open and closed *hERAP1*, the residues 486–513 were missing in the PDB coordinates and were re-built with ModellerV9.13 prior to MD simulation. The RMSF results for *hERAP1* support that this region is highly disordered with high RMSF values for both the open and closed forms of the protein (Fig 6.4). In both closed and open TIFF3 simulations, a relatively higher overall RMSF of domains I and IV than the other parts of the protein provides hints that these two domains undergo a large conformational change during the simulation.

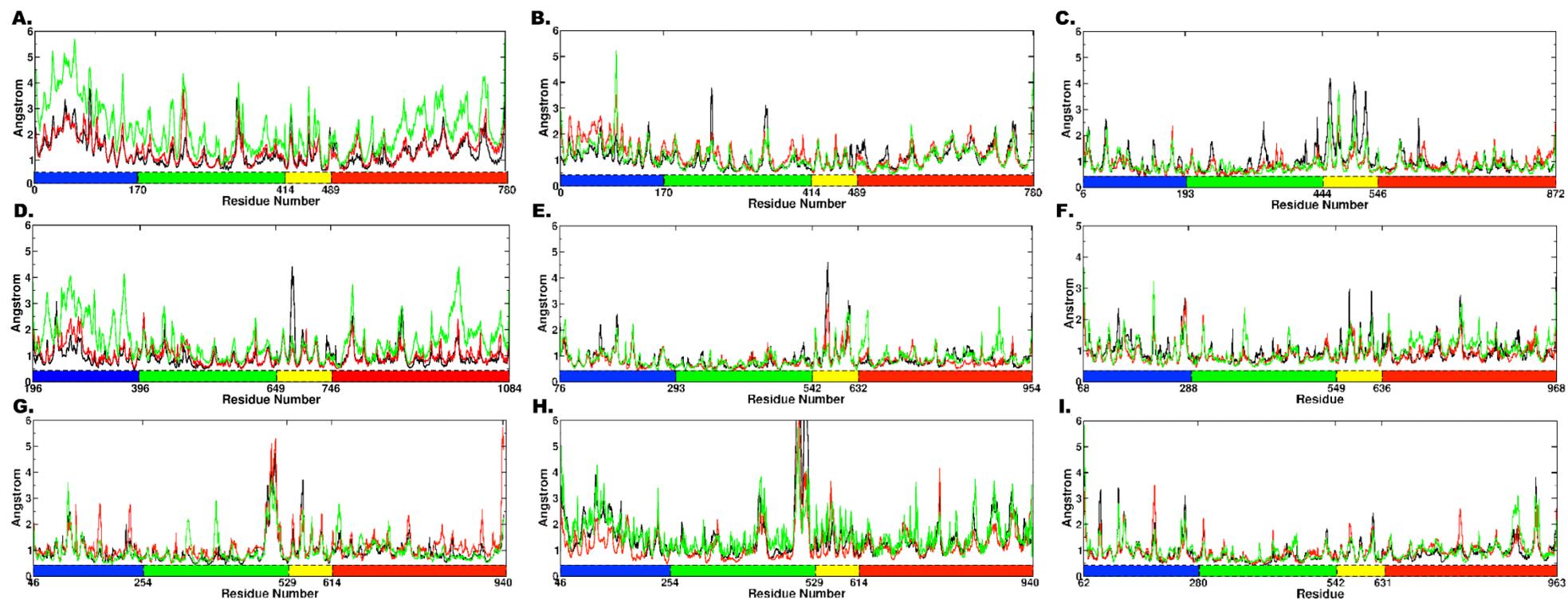


Figure 6.4. RMSF of each MD simulation. (A) closed TIFF3, (B) open TIFF3, (C) *Ec*APN, (D) *PfA*-M1, 500 ns from chapter 3, (E) *h*APA, (F) *h*APN, (G) closed *h*ERAP1, (H) open *h*ERAP1 and (I) *p*APN. RMSF were calculated from the production stage. Replicates indicated by color where run 1 = black, run 2 = red and run 3 = green.

6.2.3. PCA of the simulated M1 systems.

Separate PCA was performed on combined trajectories of each type of M1 system. The atomic covariance matrix of the atomic displacement was generated and its eigenvectors were extracted and ranked by the magnitude of the $3N$ eigenvalues. The identification of relevant motion is usually described by the first few PCs that have large contributions to the overall variance in highly complex trajectories (273, 287). To identify how many PCs I needed to analyze, I first compared the fractional contribution of the first three PCs extracted from each trajectory (Table 6.2). Most of these reduced subspaces account for at least 58 % of overall PCs, indicating that PC1-3 can be used to identify the characteristic motion of the whole system being analyzed.

Table 6.2. Fractional contribution of each PC to overall variance

System	Contribution to overall variance		
	PC1	PC2	PC3
TIFF3 (closed)	0.56	0.18	0.06
TIFF3 (open)	0.57	0.14	0.04
<i>Ec</i> APN	0.28	0.26	0.10
<i>PfA</i> -M1 (500 ns, chapter 3)	0.48	0.21	0.05
<i>h</i> APA	0.33	0.18	0.07
<i>h</i> APN	0.30	0.28	0.07
<i>h</i> ERAP1 (closed)	0.43	0.19	0.07
<i>h</i> ERAP1 (open)	0.78	0.11	0.04
<i>p</i> APN	0.38	0.19	0.08

6.2.3.1 TIFF3

The TIFF3 simulations were the most mobile of the systems I studied. I performed separate PCA on the two different simulations that had started from an open or closed state of the protein (Table 6.2). The closed PCA was dominated by $PC1_{TIFF3 (closed)}$ that provided 56 % of the overall variance and showed an overall closing motion, so that the final conformation was more closed than the starting position. Interestingly, in both $PC1_{TIFF3 (closed)}$ and $PC2_{TIFF3 (closed)}$, domain I and regions of domain II (β -strands 15 to 19 and α -helices 3 to 5) move as a rigid body. In $PC1_{TIFF3 (closed)}$, this is by 7.5 Å away from the body of the molecule and shows the GAMEN motif (domain II) moving by 6.7 Å (in the same direction) (Fig 6.5A). In $PC2_{TIFF3 (closed)}$, these elements swing up and down by 11° (Fig 6.5B).

Meanwhile, domain III, together with α -helices 14 to 16 of domain IV, rotate down by 3.5° in $PC1_{\text{TlFF3 (closed)}}$ (Fig 6.5A), but $PC2_{\text{TlFF3 (closed)}}$ shows a swing of up to 7.5° (Fig 6.5B). In $PC1$, α -helices 21-28, the majority of domain IV module II (α 20-30), move 7.4 \AA toward the catalytic domain II (Fig 6.5A) but then in $PC2$, these α -helices rotate inside and outside of the catalytic pocket by 8.6° . Overall, $PC1_{\text{TlFF3 (closed)}}$ shows a closing of the structure whilst $PC2_{\text{TlFF3 (closed)}}$ maintains a distance between the junctions of domains I, II and IV. $PC3_{\text{TlFF3 (closed)}}$ showed a similar movement pattern to that of $PC2_{\text{TlFF3 (closed)}}$ but domain III remained still (Fig 6.5C). Combined, the movements result in a reduction in the size of the S1 pocket, making TlFF3 more compact (closed) than its starting position. The Hingeplot and Dyndom servers reported two potential hinges in the $PC1$ movement (Fig 6.5A). The first hinge is unique amongst our selected M1 aminopeptidases and appears to account for how domain I and domain II β -strands 15 to 19 and α -helices 3 to 5 move as a rigid body in both $PC1$ and $PC2$ (Fig 6.5A). The second hinge was located within domain IV and appears to account for the inside and outside of the catalytic pocket rotation in α -helices 21 to 28 of domain IV (Fig 6.5A).

The PCA of TlFF3 simulations starting in an open conformation also showed that the molecule 'closed' during the simulation. Along the $PC1_{\text{TlFF3 (open)}}$ (57%, Fig 6.5E) movement, β 1-6 of domain I tilts up by 3.3 \AA , but the rest of domain I, together with parts of domain II (β -strands 15 to 19 and α -helices 3 to 5) rotates down by 4.3° and is similar to that found in the closed TlFF3 PCA. The domain III swings outward by 3.7° . At the same time, α 20-29 and the upper part of α 30 in domain IV rotates up toward the active site by 27° to produce the closed conformational state of TlFF3. This conformation is similar to the final "more closed" form of TlFF3 that was identified at the end of the closed TlFF3 simulations. The RMSD between the closed form from closed simulations and open simulations was only 1.2 \AA versus 2.3 \AA between the closed X-ray structure of TlFF3 and closed state found in open form of TlFF3 simulations.

Hingeplot and Dyndom servers reported a hinge axis during the closing motion that is buried deep in domain IV (Fig 6.5F). The hinge is similar to the hinge in $PC1_{p\text{APN}}$ (see section 6.2.3.7) and the second hinge in $PC1_{h\text{ERAP1 (open)}}$ (See section 6.2.3.6) but is different

to the hinge identified in PC1_{TIFF3(closed)} (Fig 6.5B). Among the residues identified to bend is F516 (Fig 6.5F). F516 is conserved among *hAPA* (F663), *pAPN* (F663), *hAPN* (F666), and *hERAP1* (F644) and therefore may contribute to the opening and closing of module II (α 20-30) of domain IV within the M1 aminopeptidases. In the PC2_{TIFF3 (open)} movement, domain I together with part of domain II (β 15-17 and α 3-4) twisted anticlockwise by 23° (Fig 6.5G). The loop V334-D347 (corresponding to α 6 of *EcAPN*, *hAPN*, and *pAPN*) of domain II rotates down by 8.3°. The domain III swings down by 6.5° and the upper part of α 9-15 in domain IV moves in toward the plane of the figure (Fig 6.5G). Taken together, this means that all of domain I and part of domain II (β 1-4 and α 1-3) twist against domain IV (Fig 6.5G). In the PC3_{TIFF3 (open)} movement, the same rigid body (domain I together with β 1-4 and α 1-3 of domain II) rotates down by 24° (Fig 6.5H). The loop V334-D347 of domain II moves up by 8.2 Å towards domain I (Fig 6.5H). At the same time, domain III swings down by 6.3 degree and the entire domain IV move up by 4.7 Å (Fig 6.5H).

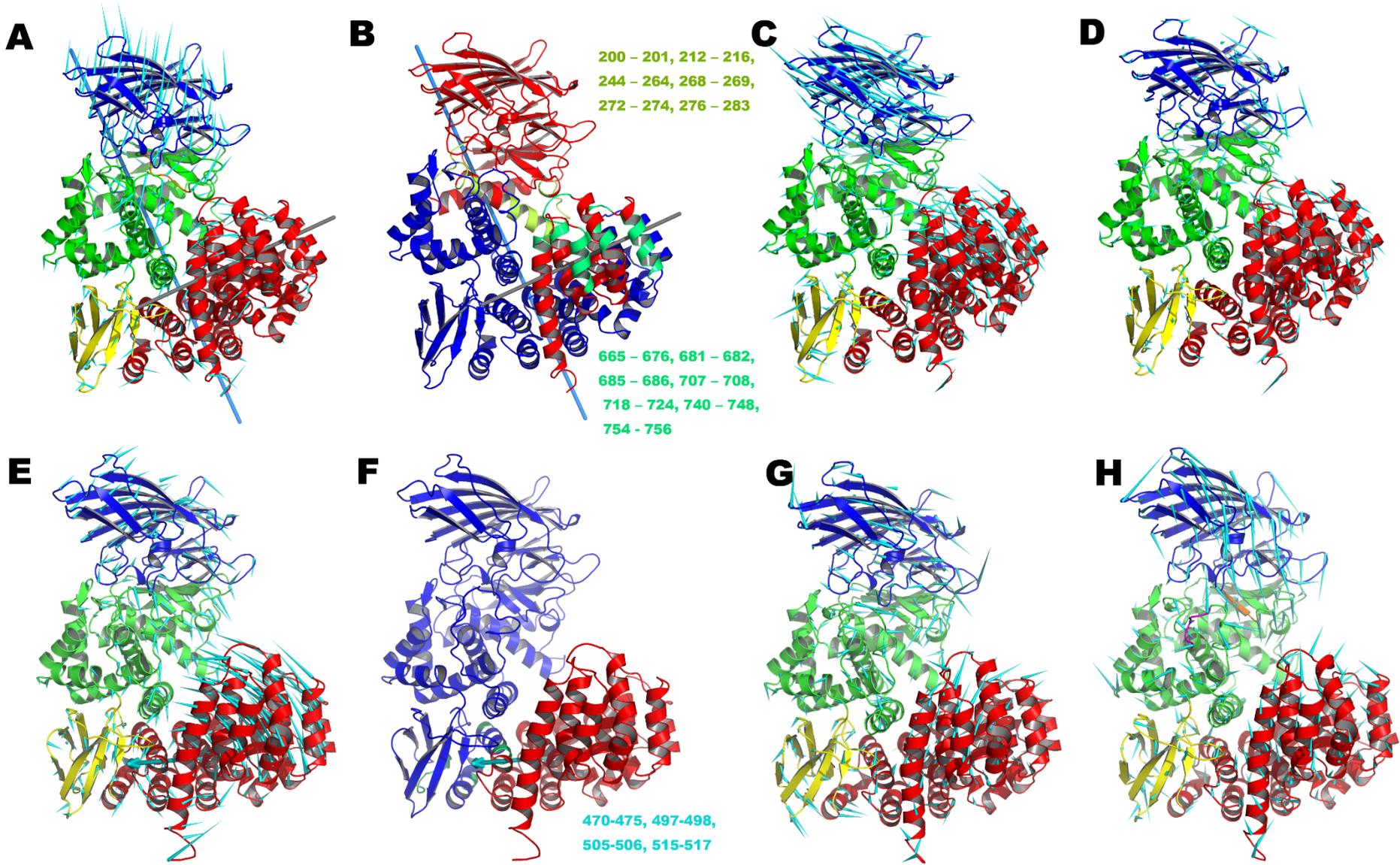


Figure 6.5. The porcupine plot showing the projected movements of TIFF3 as described by PCA for both open and closed starting positions. TIFF3 is shown in cartoon and colored by domain (I blue, II green, III yellow, and IV red) unless otherwise indicated. The porcupine arrows are in cyan and scaled proportionally to the magnitude of C α atom movement. For clarity, only movements larger than 1.25 Å are displayed by arrows. On the top panel is the PCA of the closed TIFF3 conformation where (A) is PC1, (B) is the potential hinges and bending residues identified, (C) PC2 and (D) PC3. In image A, the first hinge is shown with blue stick, the second hinge is shown with grey stick. The GAMEN motif is shown in orange, α 14-16 in magenta, α 21-28 in wheat color. In panel B, the hinges detected in PC1_{TIFF(closed)} by the Dyndom server are shown. The blue cartoon shows the fixed region, and red is the mobile region. The light green shows the potential hinge residues for the first hinge, and dark green shows the second. The potential hinge residues that are in same colors were shown nearby each hinge region. On the bottom panel is the PCA of the open TIFF3 conformation where (E) is PC1, (F) is the potential hinges and bending residues identified, (G) PC2 and (H) PC3. In E, (PC1_{TIFF(open)}) α 20-29 are shown in wheat. In F, the hinges detected in PC1_{TIFF(open)} are shown. The hinge is shown with cyan stick. The blue cartoon is the fixed region, and red is the mobile region, and hinge residues were in cyan color.

6.2.3.2 *Ec*APN

Overall, *Ec*APN was largely stable, maintained a closed conformation and had little to no movements within the active site. PC1 and PC2 were both significant contributors to the overall movement identified, however, neither resulted in large domain motions. The projected movement on PC1 indicated that domain I twists horizontally and the α -helix 11-12 (residues 396 - 431) move 4.2 Å up toward domain I (Fig 6.6A). This movement is accompanied by a 3.3° swing in domain III, a 5 Å loop movement (residues 629-634, between α 18 and 19) and a 2.6 Å movement of α -helices 29-30 (residues 836-870) in domain IV away from the active site (Fig 6.6A). PC2 showed a 4.1 Å rotation of domain I toward the body of the molecule and a similar swing in the position of domain III to that identified in PC1 (Fig 6.6B). PC3 only accounted for 10 % of the overall motion but identified a more significant movement of domain III than in either PC1 or PC2 (Fig 6.6C). Analysis of the movement using the Dyndom server, identified residues 333-338, 347-348, 444-445, 469-480, 539-540 and 543-546 as being important to the hinging motion and two residues (T445 and V540) that might act as hinge axis (as predicted by Hingeplot, Fig 6.6D).

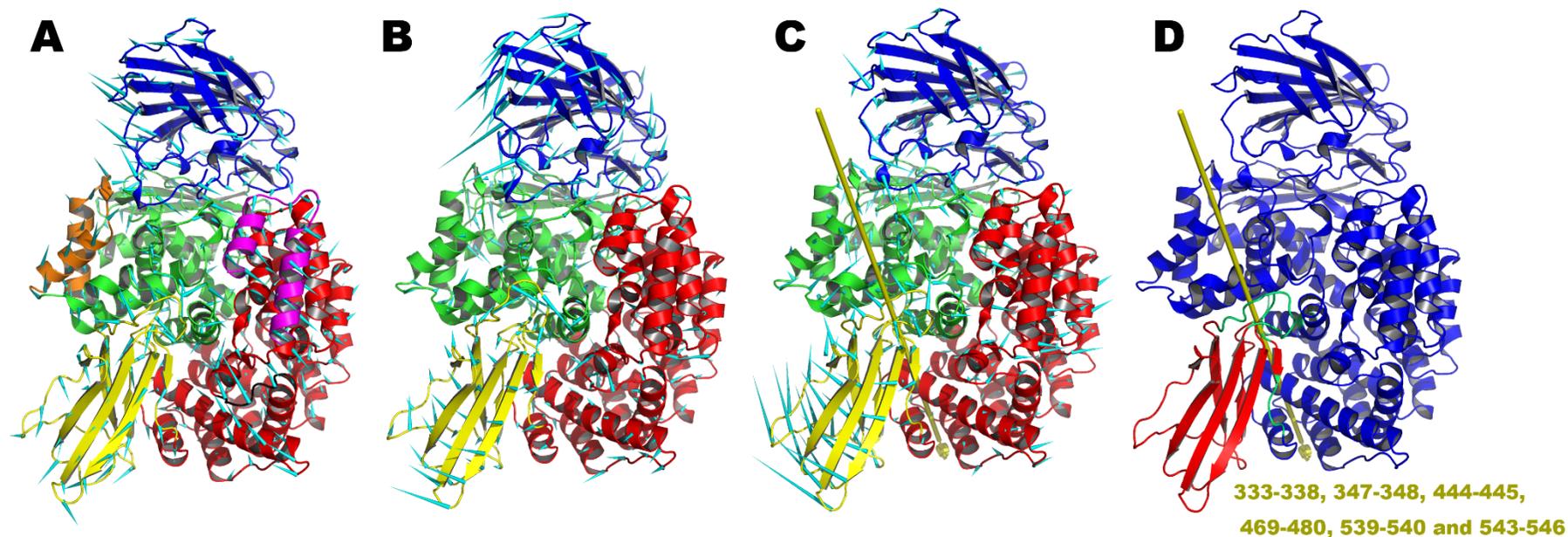


Figure 6.6. The porcupine plot showing the projected movements of *EcAPN* as described by PC1 (A), PC2 (B) and PC3 (C). *EcAPN* is shown in cartoon and colored by domain (I blue, II green, III yellow, and IV red) unless indicated otherwise. The porcupine arrows are in cyan and scaled proportionally to the magnitude of C α atoms movement. For clarity, only movements larger than 1.25 Å are displayed by arrows. In panel A, residues 396-431 are shown in orange, loop residues 629-634 (between α 18 and α 19) in domain IV are in black. and helices α -29 to 30 in magenta. (D) A putative hinge was identified in the PC3 projection movement. The hinge axis is shown as a yellow cylinder. The blue cartoon is the fixed region and red is the mobile region. The hinge residues are indicated in yellow.

6.2.3.3 *PfA-M1*

All analysis of *PfA-M1* is outlined in Chapter 3. The *PfA-M1* data has been included in parts of this chapter for comparison only.

6.2.3.4 *hAPA*

No large domain motions were captured in *hAPA* simulations and *hAPA* stayed in its closed conformation. The major motion observed for *hAPA* was a large movement in the long loop within domain III (residues 559-580). PC1 accounted for 33% of the overall movement of *hAPA* and consisted of a 2.1 Å downward movement in domain I with a concomitant 15° outward rotation of $\alpha 9$ and the loop the links $\alpha 9$ to $\alpha 8$ (residues 445-474) in domain II (Fig 6.7A). At the same time, the long loop of domain III swings up by 32° whilst the rest of domain III rotates down by 7.0° (Fig 6.7A). In domain IV, $\alpha 22$ to $\alpha 29$ and the upper part of $\alpha 30$ move away from the body of the molecule by 3.0 Å (Fig 6.7A). Together with the movements of $\alpha 8$ -9, the S1 substrate pocket gains volume due to PC1 in these simulations. In PC2, domain I slightly twists (3.7°) in a horizontal direction and the domain III long loop moves randomly (Fig 6.7B). PC3 also sees domain I rotate down toward the body of the molecule by 4.0° (Fig 6.C). In summary, PC1-3 indicated that the protein did not experience large domain movements but experienced an overall expansion or inflation, mainly due to the movement of $\alpha 8$ -9 (corresponding to S1 pocket helix in *PfA-M1*) and the movement of $\alpha 22$ -30 (corresponding to the module II of domain IV). This motion is likely related to the release of crystal constraints / crystal packing.

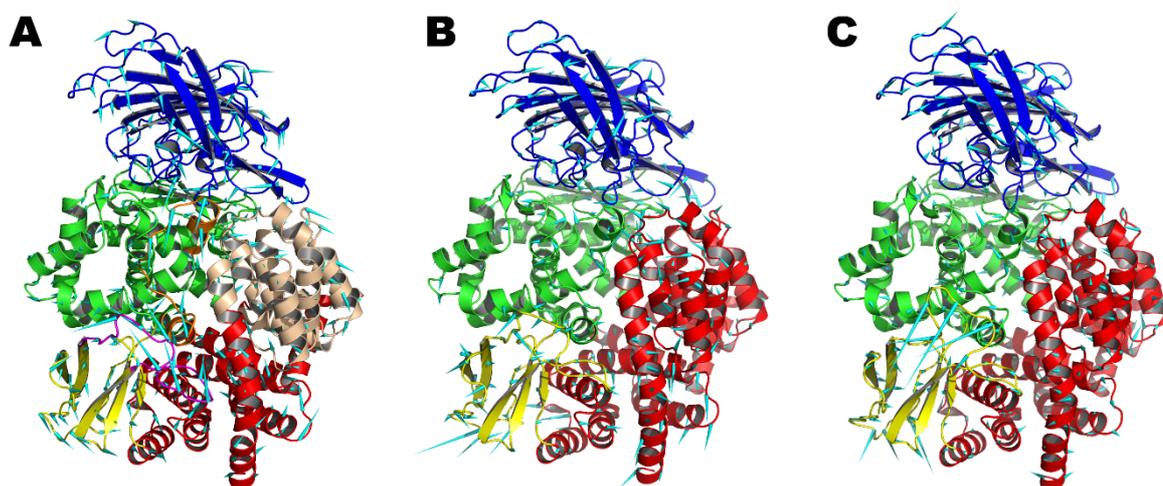


Figure 6.7. The porcupine plot showing the projected movements of *hAPA* as described by PC1 (A), PC2 (B) and PC3 (C). *hAPA* is shown in cartoon and colored by domain (I blue, II green, III yellow, and IV red) unless indicated otherwise. The porcupine arrows are in cyan and scaled proportionally to the magnitude of C α atoms movement. For clarity, only movements larger than 1.25 Å are displayed by arrows. In panel A, α 9 and the loop between α 8-9 are shown in orange, and α 22-29 and the upper part of α 30 are in wheat.

6.2.3.5 *hAPN*

No large domain motions were captured in *hAPN* simulations and *hAPN* stayed in its closed conformation. PC1 and PC2 provide similar contributions to the overall movement (30 and 28 % respectively). PC1 sees the entire domain I move up by 1.5 Å, and α -helices 25 to 29 of domain IV toward the catalytic centre by 3.4 Å (Fig 6.8A). In PC2, domain I rotates down by 5.6° and α -helices 25 to 27 within domain IV move up by 2.2 Å, but α -helices 28 and 29 shift down by 3.3 Å (Fig 6.8B). This split in the movement of domain IV (the entire domain is not experiencing a rigid body motion) allows a small expansion of the S1' pocket of *hAPN* (Fig 6.8B). PC3 shows rotation in the N-terminal region of domain I by 9.1° that allows β -strand 11 and 12 to shift to the right by 3.3 Å (Fig 6.8C). The α -helices 28 and 29 from domain IV also shift in the same direction as domain I but by 4.2 Å (Fig 6.8C). Interestingly in PC3, a α 9 in domain II moved, resulting in the residue F472 (corresponding to the Y575 in *PfA-M1* on S1 pocket helix) moving closer to the zinc center residues. In total, the PCA identified three major patterns of movement in the *hAPN* simulations. First, that the domain I rotates up and down as a rigid body. Second, that domain II remains rigid in each of the top three PCs and three, that the solvent exposed long loop of domain III (residues 565 to V586) moves in apparently random directions.

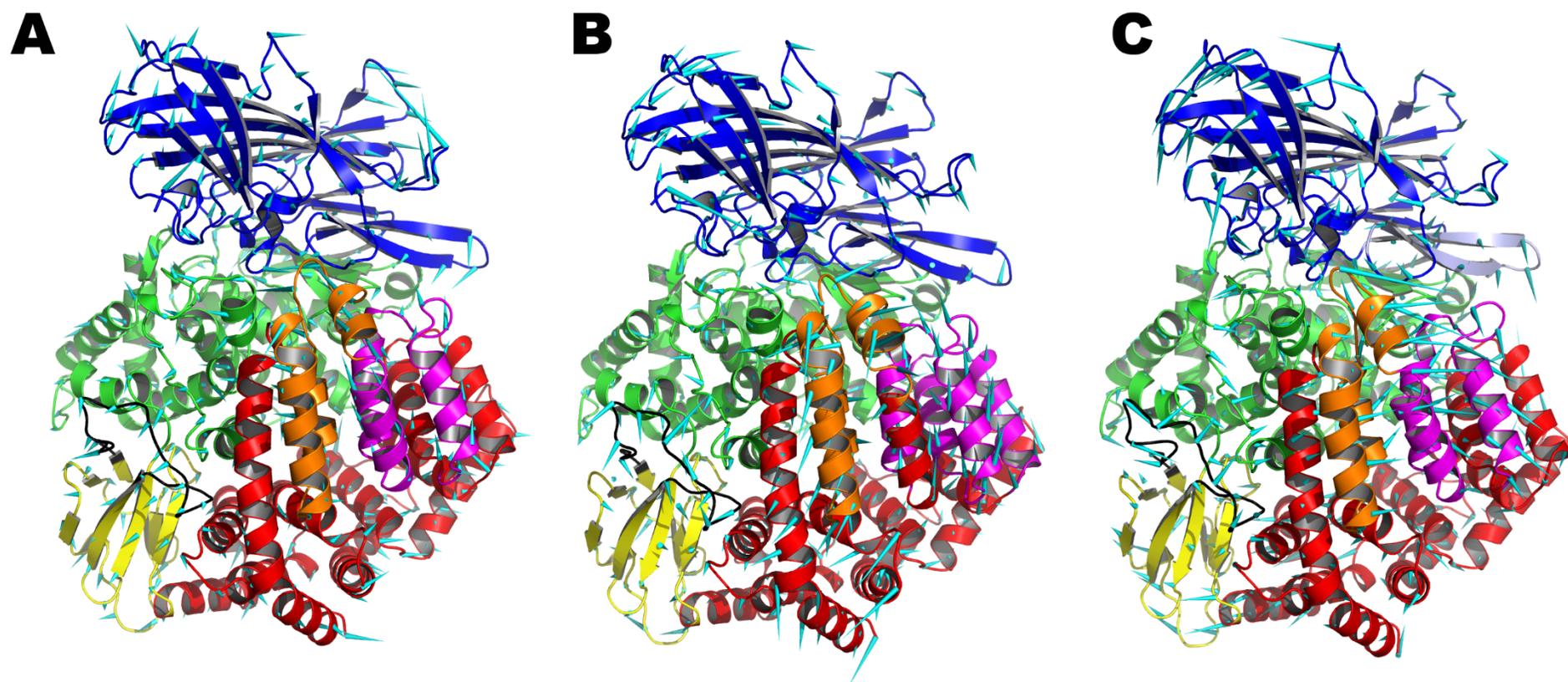


Figure 6.8. The porcupine plot showing the projected movements of *hAPN* as described by PC1 (A), PC2 (B) and PC3 (C). *hAPN* is shown in cartoon and colored by domain (I blue, II green, III yellow, and IV red) unless indicated otherwise. The porcupine arrows are in cyan and scaled proportionally to the magnitude of C α atoms movement. For clarity, only movements larger than 1.25 Å are displayed by arrows. In all panels, α 25-27 of domain IV is in magenta, α 28-29 in orange and the long loop (565-586) of domain III in black. In panel C, β 11-12 is in light blue.

6.2.3.6 *h*ERAP1

Similar to TIFF3 (See section 6.2.3.1), I performed two separate PCA on the combined closed and open forms of the simulations of *h*ERAP1. The outcomes of these two PCAs were also similar to that of TIFF3 in that both simulations were observed overall to undergo a ‘closing’ motion. Compared to the other M1 aminopeptidases analyzed in this study, *h*ERAP1 has a large loop insertion in domain II (residues 478–517, see Fig 6.1). The two PCAs showed that this loop possesses apparently random movements in the top three PCs of both PCAs. The flexibility observed in this loop during the MD simulations is consistent with the experimental crystallography data, where residues 485-514 could not be resolved in the electron density (5, 216).

The PCA on *h*ERAP1 showed that none of the closed simulations reached an open conformation, but each of the open simulations did reach a closed conformation in the last 50 ns. Specifically, in PC1_{*h*ERAP1 (closed)} (43%, Fig 6.9A), domain I rotates down by 3.2°, the $\alpha 6$ of domain II together with the upper part of $\alpha 23$ -26 of domain IV shift to left by 3.1 Å, which constrains the S1’ pocket (Fig 6.9A). The GAMEN motif of domain II also moves down by 3.5 Å. The whole protein moves to a more compact form than the closed X-ray structure of *h*ERAP1 (Fig 6.9A). In PC2_{*h*ERAP1 (closed)} (Fig 6.9B) and PC3_{*h*ERAP1 (closed)} (Fig 6.9C), only random motions of the flexible loop were sampled, and the rest of the protein remained relatively still.

In the open *h*ERAP1 simulations, along PC1_{*h*ERAP1 (open)}, domain I rotates down by 5°, domain II and domain III barely move except that the long loop (478-518) of domain II moves randomly (Fig 6.9D). At the same time, $\alpha 21$ -30 significantly rotate toward the metal centre by 34°. The upper part of $\alpha 21$ -30 shifts 9.2 Å during this rotation and effectively closes *h*ERAP1. This conformation is more similar to the closed crystal state found in the closed *h*ERAP1 simulations than the closed crystal structure. The RMSD between the closed state sampled in both open and closed simulations is only 1.5 Å versus 2.3 Å between the closed state sampled in open *h*ERAP1 simulation and closed crystal structure of *h*ERAP1. A group of flexible bending residues that might be involved in the closing motion observed in PC1_{*h*ERAP1 (open)} were detected by the Dyndom server and a putative hinge axis

generated by Hingeplot (Fig 6.9D). The first group of residues account for the rotation of domain I (180-187), and include E183 that corresponds to E319 of *PfA-M1* (Fig 6.9E). Mutation of E319 to proline was the mutation that resulted in capturing the open state of *PfA-M1* (see Chapter 3). The hinge residues of a second putative hinge are also listed in Fig 6.9D. These residues allow for the rotation between α 14-19 and α 21-30 of domain IV.

In the $PC2_{hERAP1}$ (open) movement, domain II rotates up by 2.4° and the domain II loop incorporating 419-434, (corresponding to α 9 or the S1 helix of domain II in the other M1s) moves to left by 8.9 Å. This movement is accompanied by the upper part of α 13-15 shifting to the left by 5.5 Å (Fig 6.9F). In the $PC3_{hERAP1}$ (open) movement, domain I rotates up and down by 4° and the rest of the protein remained stable except for random turbulence of the flexible loop in domain II (Fig 6.9G). In summary, the major movement from PC1-3 indicate the protein can move to a closed form that is more compact than the crystal form *hERAP1*. This movement may be mediated by the two separate hinge regions.

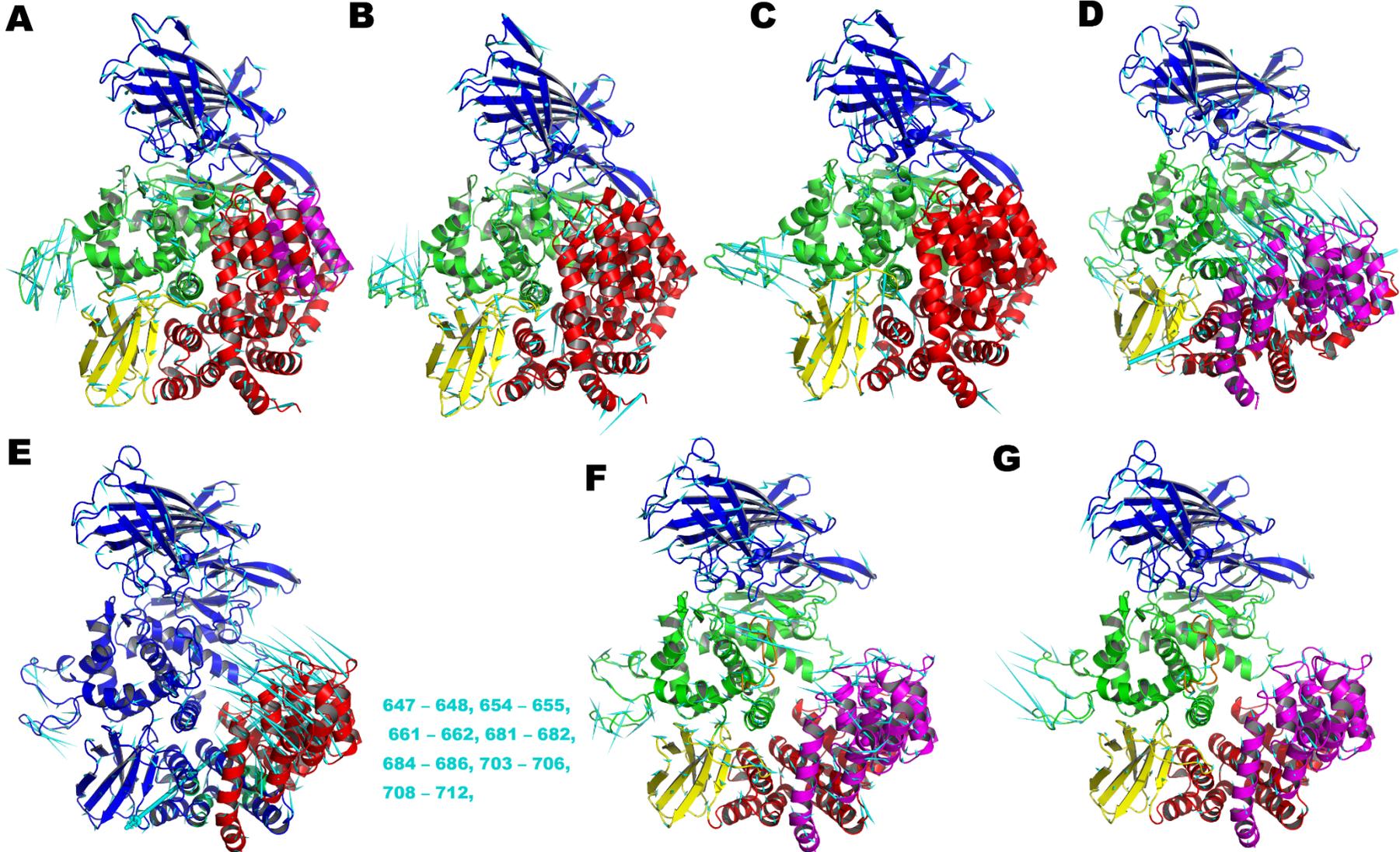


Figure 6.9. The porcupine plot showing the projected movements of *h*ERAP1 as described by PCA for both open and closed starting positions. *h*ERAP1 is shown in cartoon and colored by domain (I blue, II green, III yellow, and IV red) unless otherwise indicated. The porcupine arrows are in cyan and scaled proportionally to the magnitude of C α atom movement. For clarity, only movements larger than 1.25 Å are displayed by arrows. On the top panel is the PCA of the closed *h*ERAP1 conformation where (A) is PC1, (B) PC2, (C) PC3 and (D) is the potential hinges and bending residues identified in PC1_{*h*ERAP1 (closed)}}. In panel A, α 23-26 of domain IV are in magenta, α 8-9 in orange, the GAMEN motif in white and loop residues 478–517 in black. In panel D, F and G, α 21-30 of domain IV are in magenta. The hinge axis is shown in cyan stick, the blue cartoon shows the fixed region, and red is the mobile region. The Dyndom detected hinge residues indicated in cyan text. On the bottom panel is the PCA of the open *h*ERAP1 conformation where (E) is PC1, (F) PC2 and (G) PC3.

6.2.3.7 *p*APN

The overall finding of the PCA for *p*APN is that the simulations captured a more compact *p*APN molecule than the starting crystal structure. Specifically, in PC1, the entire domain I undergoes a rigid body shift and rotates upward (away from the body of the molecule) by 4°, whilst $\alpha 9$ of domain II also moves up by 14° (Fig 6.10A). At the same time, $\alpha 21$ to 29 and the upper part of $\alpha 30$ in domain IV undergo a large movement toward the catalytic centre by 12.5 Å, closing the molecule further and producing a more compact closed state for *p*APN (Fig 6.10A). Using both DynDom and Hingeplot, I identified a hinge axis in domain IV and key hinge bending residues that might be required for the motion (Fig 6.10B). The overlapping residues from these two methods (F662, I944, K945, and G800) might act as the hinge axis (Fig 6.10B). Interestingly, this closed form *p*APN is very similar to the recent closed *p*APN dimer crystal structure (RMSD = 1.2 Å). The major difference is the upper part of $\alpha 27$ -30 in the closed simulation seems to constrain the size of the S1' pocket more than what was observed in the crystal structure (PDB ID 5LDS, (247)). In PC2, domain I rotates down by 20° and $\alpha 25$ to 27 of domain IV move up by 4.4 Å (Fig 6.10C). In PC3, domain I undergoes a 2.3 Å rigid body movement but the loop incorporating L88-Y96 shifts down dramatically (12.6 Å) toward domain IV (Fig 6.10D). In summary, domain I in *p*APN can also rotate as a rigid body, and the closed form from PC1 supports the recent crystallographic data on the closed form of *p*APN. In my simulations, the closing motion is achieved by domain I rotating downwards while the second module of domain IV ($\alpha 21$ to 29) swings up toward the active site. The hinge axis identified for *p*APN is similar to those detected in the open conformation simulations of both TIFF3 and *h*ERAP1.

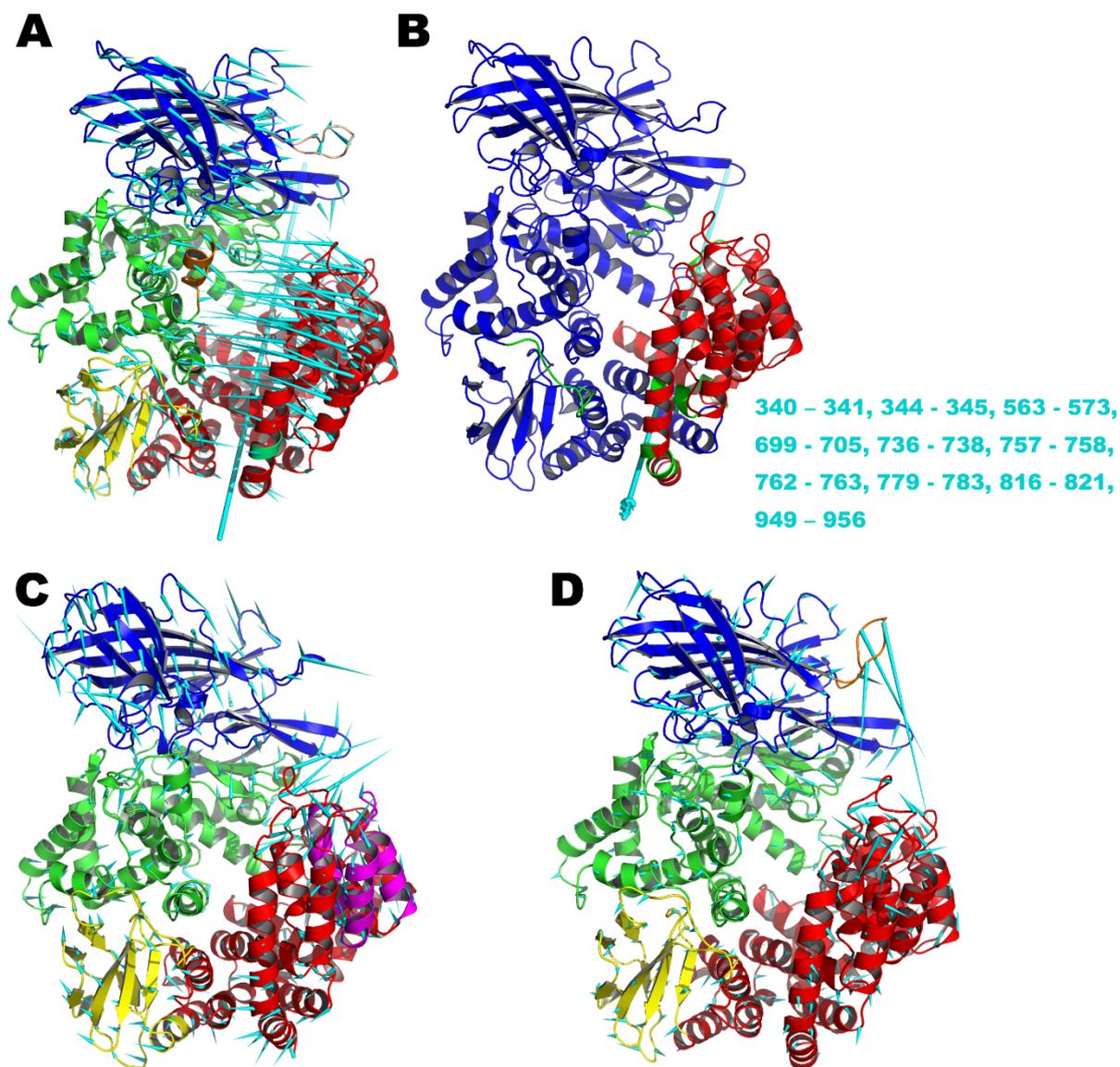


Figure 6.10. The porcupine plot showing the projected movements of pAPN. pAPN is shown in cartoon and colored by domain (I blue, II green, III yellow, and IV red) unless otherwise indicated. The porcupine arrows are in cyan and scaled proportionally to the magnitude of C α atoms movement. For clarity, only movements larger than 1.25 Å are displayed by arrows. (A) PC1 where the S1 helix α 9 is in orange. (B) The A putative hinge identified in the PC1 movement by Dyndom. The hinge axis is shown in cyan stick, the blue cartoon shows the fixed region, and red is the mobile region. The hinge residue numbers are indicated in cyan text. (C) PC2 where α 25 to 27 are shown in magenta. (D) PC3. where L88-Y96 are shown in orange.

6.2.4 The mammalian M1 aminopeptidases dimers might possess the same opening movement as were found in monomers.

The selected M1 structures for *hAPA*, *pAPN*, *hAPA* were crystallized as homodimers and have been reported to be functional as membrane-anchored homodimers on the cell surface (1, 144, 147). It has been postulated that the dimerization facilitates the conformational changes without change to the dimer interface ((147). I was curious about the dynamics of the dimers but time did not permit the MD simulation of dimer structures. Therefore, I decided to investigate how the monomeric dynamics may translate to a dimer. To do this, I first looked into the conservation of the protein-protein interfaces between the homodimers (Fig 6.11A). I found that the protein-protein interactions are located at domain IV and mediated by the two modules of domain IV (Fig 6.11). For the protein-protein interface of *pAPN* and *hAPN*, module I of domain IV is mostly basic and module II is acidic (Fig 6.11C). The two regions have opposing charges, which result in attractive forces that facilitate the interactions of the two monomers, as the dimer is linked by domain IV of the monomers with twisted orientations (Fig 6.11). Two glutamine residues involved in the dimer interface were found to be conserved in the M1s from higher organisms studied here (Fig 6.1 & Fig 6.11). Two conserved salt bridges were also identified in the dimer interface of *pAPN* and *hAPN* dimers (Fig 6.11) and hydrogen bonds (10 in *hAPN*, 6 in *hAPA*, 4 in *pAPN*) and water bridges (15 for *pAPN*, 8 in *hAPA*, 8 for *hAPN*) were pivotal to the interactions (Fig 6.11). The solvent accessible surface areas (SASA) of the protein-protein interface were mapped and calculated for the three dimers and showed a similar value (1750 – 2050 Å², Fig 6.11).

In the simulations of the monomeric form of these proteins, I found that the low RMSF of the conserved Glu residues within the dimer interface (*hAPA* Q872 = 0.38 ± 0.01 Å and Q907 = 0.42 ± 0.02 Å; *hAPN* Q881 = 1.10 ± 0.02 Å and Q917 = 0.9 ± 0.03 Å; *pAPN* Q878 = 0.82 ± 0.03 Å and Q914 = 0.43 ± 0.01 Å) suggests that the movement sampled from my simulations may not interfere the dimerization of M1 aminopeptidases.

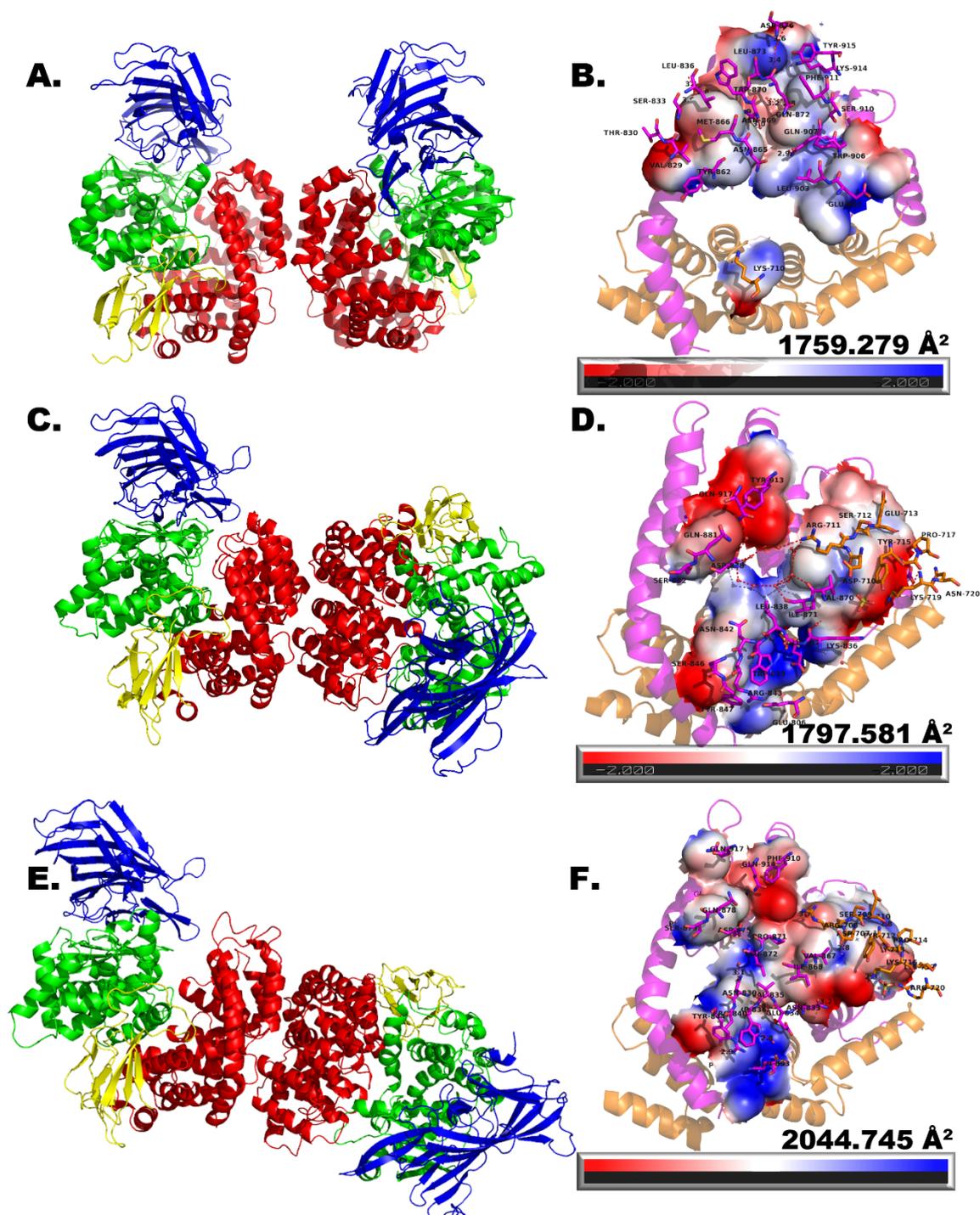


Figure 6.11. The dimerization of *hAPA*, *hAPN* and *pAPN*. The overall structure of the homodimers in the crystal structures are shown in (A) *hAPA*, (C) *hAPN*, and (E) *pAPN*. The surface of the dimer interactions are shown in (B) *hAPA*, (D) *hAPN*, and (F) *pAPN*. The electrostatic potential map shows residues involved in dimerization. In B, D and F, the modules of domain IV were shown by orange (module I) and magenta (module II). The range of the electrostatic potential map is -2 to 2 kT. The surface area is indicated beside the electrostatic potential legend.

6.3. DISCUSSION

In this chapter, I performed all-atom MD simulations of a selection of six different M1 aminopeptidases starting either from their open or closed forms, and analyzed their major movements. I used PCA to investigate if the proteases underwent a major conformational change and found that for all six systems simulated, no closed starting conformation was found to sample an open form. Further, the two systems that started open, each closed during the simulation time. However, despite not sampling an opening movement, common motions were identified within the M1 aminopeptidase family and suggest that given the right circumstances, all M1 aminopeptidases can sample both an open and closed form.

In this study, I observed that there are three dynamic relationships or 'rules' that appear to be important to allow the opening and closing motion of the M1 aminopeptidase family. One, that domains I and IV have transient interactions with each other. In the closed form M1 simulations (*hAPA*, *hAPN*, *hAPN*, *pAPN*), domain I was observed to interact with domain IV in the crystal structure, but in the beginning of each simulation, these interactions were quickly lost and domain I was then observed to move as a rigid body throughout the rest of the simulations. Second, that domain IV can pivot on a flexible hinge that appears to be located within α 14-16, and that this hinge is what allows the enzyme to 'open'. This is consistent with the crystallographic studies of TIF3 and *hERAP1* that also showed a clear gap between domain IV and domain I in the open form (149, 216). My hinge residue analysis suggested that higher organisms (*hERAP1*, *hAPN*, *pAPN*) have the similar hinge axis for the movement and conserved hinge residue were put forward, however, TIF3 has a different hinge axis in α 14-16, which might due to the loosely packed helical assembly of domain IV. Finally, three, that the module II of domain IV can move towards metal center and allows the C-terminal helix of the whole protein to make contact or interact with α 9 of the catalytic domain II, thereby sealing the S1' pocket. Interestingly, a phenylalanine residue that was identified as a key bending residue to produce this movement is highly conserved across the mammalian and archael M1 aminopeptidases (*hAPA*, F663, *pAPN*, F663, *hAPN* F666, *hERAP1*, F644 and TIF3, F516).

6.3.1 Six movements can summarize the dynamics that were commonly observed in the M1 aminopeptidase representatives.

My PCA of MD simulations of seven different M1 aminopeptidases identified a total of six possible movements within the superfamily (Fig 6.12). In the TIFF3 simulations performed in this study, I observed all six movements from the one protease (Fig 6.12). The first is a hinged vertical swing of domain I, identified as one of the most important motions required for the conformational change (as mentioned above). The common hinge for this movement appears to be located between $\beta 9$ and $\alpha 1$ in domain I. Within this hinge axis, TIFF3, *PfA-M1*, *hAPA*, *hERAP1* and *EcAPN* all have a glutamic acid at the same position that is the predicted hinge residue. In *pAPN* and *hAPN*, this position has a glutamine residue. A salt bridge between this conserved glutamic acid and a lysine residue in domain II is known to be critical to the stability and activity of *EcAPN* (288). It has previously been postulated that this conserved Glu residue (E121 in *EcAPN*) interacts with the amino-terminus of the substrate. My results here would also imply that disruption of this salt bridge would disrupt a stabilizing feature of the closed ensemble. This is supported by results for *PfA-M1* also, where mutation of the same conserved residue (E319 in *PfA-M1*) resulted in an inactive, open conformation of the protein. The conservation of these elements within the whole family implies that dynamics in this area are integral to the superfamily and that a similar vertical swing of domain I, as observed in TIFF3, would be common to all the M1 aminopeptidases.

The second commonly observed motion was a vertical rigid body shift of domain I, which is controlled by the conserved helices $\alpha 1$ and $\alpha 2$ in domain I. Sequence conservation of $\alpha 1$ and $\alpha 2$ is very high (Fig 6.1) and may explain why the movement is common amongst all the candidates. The third common motion is a hinge-like swing of domain III that is accompanied by (four) a shift in the first module of domain IV (domain III drags at first module of domain IV), and five, a hinge-like swing of the entire domain IV toward domain II in either direction (to close further or open). Simulation of the closed form of *PfA-M1* also suggests that domain IV moves towards domain II (Chapter 3). Finally, a horizontal rigid body shift of module II of domain IV toward domain II. Taking these results together, and

assuming that these movements can go in both directions, I can postulate the dynamics of opening and closing of the M1 aminopeptidases. Thomas Schaffee (La Trobe University, Australia), a collaborator of the McGowan laboratory, produced a tree that represents evolutionary relationships between the M1 aminopeptidases used in this study. The structural analysis that produced this tree will be reported elsewhere, however, I have used the tree to show the differences in the dynamics between the M1 aminopeptidases from different species (Fig 6.12).

The dependencies or relationships between these movements also asks the question of the evolution of the dynamics and the relationships between the enzymes. The M1 aminopeptidase superfamily is large and diverse and how might the dynamics have evolved over time? When I analyze the movements observed in each species and their structural relationships (Fig 6.12), I can see that the hinged vertical swing of domain I is connected to the movement of domain IV. For TIFF3, when domain I moves up, domain IV swings down, but in *hERAP1*, *pAPN*, *hAPN*, and *hAPA* when domain I moves, only module II of domain IV moves away from $\alpha 5$ of domain II. The largest swing motion of domain III was observed in the lower organisms (TIFF3, *EcAPN* and *PfA-M1*) and a rotation of $\alpha 9-15$ of domain IV was observed in *hERAP1*, *pAPN*, and TIFF3. Therefore, it might be due to the request of performing enzymatic tasks that involve conformational changes. Moreover, the common dynamics in higher and lower organism are not always the same, and this might be result from the evolutionary preference.

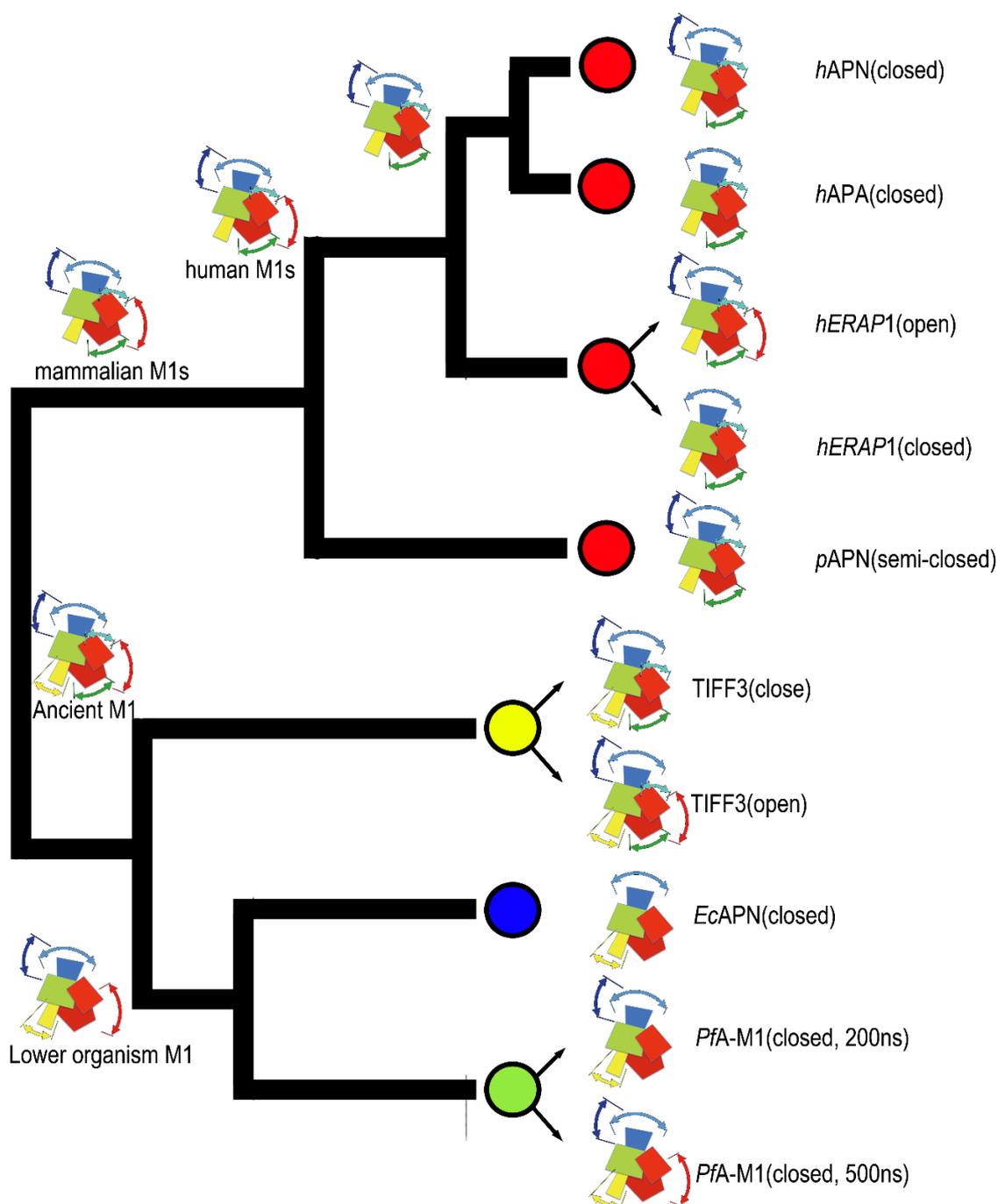


Figure 6.12. Hierarchical clustering of structural similarity RMSD displayed as unrooted dendrogram (as produced by Thomas Schaffee, La Trobe University). The M1 aminopeptidases are shown schematically with standard coloring. Domain IV is divided into module I and II (as shown by two rectangles). The red circles indicate mammalian enzymes, yellow is archae, blue bacteria and green parasite. The six commonly observed motions are shown by arrows (color coded to domain).

6.3.2. The unique movements among the M1 aminopeptidase candidates.

Whilst there were common movements identified within all the M1 aminopeptidase studies here, there were also elements within individual enzymes that moved entirely uniquely to that of the other family members. The movement of β 8-9 in domain I showed major differences in its movements between the representatives. In mammalian enzymes (*hAPA*, *hERAP1*, *hAPN* and *pAPN*), β 8-9 is longer than the other candidates and maintains a relatively stationary position throughout simulations. β 8-9 in mammalian M1s is located directly above the upper part of the module II in domain IV, and helps shape the groove between domain I and domain IV. In contrast to the mammalian enzymes, *EcAPN*, *PfA-M1* and *TIFF3*, have a shorter β 8-9 that does not form contacts with domain IV and is more flexible. In *EcAPN* and *PfA-M1*, there is ~ 9 Å between β 8-9 and domain IV. The differences in the interactions of β 8-9 and domain IV between the mammalian and lower species provides an explanation as to why domain I in the mammalian enzymes show less vertical rotation dynamics compared to what was found in *EcAPN*(PC1_{*EcAPN*}), and *PfA-M1*(PC2_{*PfA-M1*}). Therefore, the evolutionary trend that favours a longer β 8-9 to pack against domain IV may slow the dynamics of domain I, promoting a closed state.

In domain II, the GAMEN motif and zinc metal center were found to be very rigid in *EcAPN*, *hAPA*, *pAPN*, and *hAPN*. However, in both open and closed form of *hERAP1* and *TIFF3*, and the MD trajectory that showed an open *PfA-M1* (chapter 3), I witnessed a shift in the position of the GAMEN motif that appeared to follow domain I dynamics. The stability of the GAMEN motif appears connected to the opening and closing motion. Once the GAMEN starts to move, the whole enzyme will undergo a conformational change, and following this means that when the GAMEN is stable, the enzyme is closed and catalytically competent. Given the function of the GAMEN motif is to correctly position the substrate for hydrolysis, this finding supports previous literature that that closed form of the enzymes are the active forms but further implies that dynamics may be intrinsically linked to catalysis.

The most varied motions among all the systems studied came from domain IV. It appears that the packing and size of the helices in domain IV can influence the dynamics of the protein. Domain IV of TIFFF3 has the least number of residues and helices are loosely packed in comparison to the other candidates. The average length of the helices is also shorter than other candidates. Both open and closed form of TIFFF3 were sampled from the simulations, however, the open form of TIFFF3 was not retained during the simulation and the protein closed in each trajectory within 200 ns. TIFFF3 is from *Thermoplasma acidophilum*, an archaeobacteria that lives in a high temperature, acidic environment. Proteins from this organism therefore have to function in this environment and the high temperature conditions may give rise to more fluctuations in protein conformations than proteins that exist in more temperate environments. From a thermodynamics aspect, the loosely packed secondary structure of TIFFF3 may assist the enzyme in such an environment and may be an evolved strategy for its function. In more 'normal' conditions (as sampled in my MD simulation), this may translate to a protein that is more flexible or dynamic than other representatives.

The conical domain IV also appears to have variable packing of its helices that may account for the different movements observed. In the original analysis of TIFFF3, it was proposed that domain IV has two modules, module I (α 14-19) and module II (α 20-30) (149). My analysis of the M1 aminopeptidases in this study would concur that most of the domain IV dynamics are module dependent. Simulation and X-ray crystal structure data suggested that the higher organisms (*p*APN, *h*APN and *h*ERAP1) have unique rotation between these two modules, and one conserved phenylalanine residue is detected as a mutual hinge residue for the rotation. For *Ec*APN and *Pf*A-M1, I did not identify movement between the modules nor is the phenylalanine residue conserved. This might be because in *Ec*APN and *Pf*A-M1, module I is larger and has more contacts with domain II. The major difference of module I between *Ec*APN and *Pf*A-M1 and the mammalian enzymes is the position and length of α 18 and α 19 that appear to sterically block rotation between modules I and II. In the mammalian organisms, the α 18 helix is an average of 5 amino acids longer than in *Ec*APN and *Pf*A-M1 and maintains a stable distance of 12 Å to domain II. α 19 is very short

and easily bridges to module II. Rotation of these two modules is observed in simulation. In *EcAPN* and *PfA-M1*, $\alpha 18$ moves toward $\alpha 10$ in domain II, reducing the size of the S1' pocket. These enzymes also have a very long $\alpha 19$ in order to bridge module I to module II of domain IV. This long $\alpha 19$ would appear to preclude any rotation between the modules. The long $\alpha 19$ in *EcAPN* and *PfA-M1* also extends beyond the globular body of the molecule, which may be the reason why dimers are not observed in these species. The dimeric interface of the mammalian enzymes are at the exact point that $\alpha 19$ protrudes in *EcAPN* and *PfA-M1*. A previous study on *PfA-M1* also suggested that the protruding tail of $\alpha 19$, may have a role in substrate guidance (208), however, my results in chapters 3 and 4 would suggest that substrates are more likely to access the active site via an opening of the protein.

In mammalian M1 aminopeptidases, $\alpha 30$ of module II in domain IV is ~ 20 residues longer than that of the *EcAPN* and *PfA-M1*. This long helix lends itself to more mobility and can move towards $\alpha 9$ of domain II as part of the conformational change. However, the equivalent helix in *EcAPN* and *PfA-M1* has a tight interaction with $\alpha 9$ in domain II mediated by hydrogen bonds. This may also explain why it is harder for the less complex organisms to open domain IV in the conformational change.

Domain III of *EcAPN* has the longest β -sheets of all the candidates and was observed to have the largest swinging motion of domain III. A similar motion was observed in *PfA-M1* and this was the second largest among the candidates. I suggest that the dynamics of domain III might be required to assist in the opening of domain IV in the lower organisms, as module I of domain IV swings together with domain III. The domain III loops between $\beta 22$ and 23 for both *EcAPN* and *PfA-M1* move in concert with domain II, while the corresponding but longer loop in higher organisms (*hERAP1*, *hAPA*, *hAPN*, and *pAPN*) moves randomly through the simulations. This concerted movement makes me suspect that the movement may assist the movement of the domain IV.

6.4. CONCLUSIONS

I found that the domain I of the M1 aminopeptidases can rotate and twist as a rigid body and that the rotation of domain I may be associated with the rotation of domain IV to aid the protease to open or close. My findings suggest that all M1 aminopeptidase enzymes can occupy an open or closed conformation, however, the open form of M1 aminopeptidase is the unstable form, while closed form of M1 is the stable, catalytically competent form.

CHAPTER SEVEN

MAPPING THE PATHWAY AND DYNAMICS OF BESTATIN INHIBITION OF THE *Plasmodium falciparum* M1 AMINOPEPTIDASE, Pfa-M1.

Submitted and under review, Plos Computational Biology

Mapping the pathway and dynamics of bestatin inhibition of the *Plasmodium falciparum* M1 aminopeptidase, PfA-M1.

Wei Yang¹, Blake T. Riley², Benjamin T. Porebski^{2,3}, Itamar Kass^{2,4}, Ashley M. Buckle²,
Sheena McGowan^{1*}

¹ Infection and Immunity Program, Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, VIC 3800, Australia

² Infection and Immunity Program, Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, VIC 3800, Australia

³ Current Address: Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge, CB2 0QH, United Kingdom

⁴ Victorian Life Sciences Computation Centre, Monash University, Clayton 3800, Victoria, Australia

*To whom correspondence should be addressed: sheena.mcgowan@monash.edu; +613 9902 93009

KEYWORDS: Bestatin; M1 aminopeptidase; molecular mechanisms; molecular dynamics; Markov models.

Short title: Dynamics of bestatin bound to PfA-M1

Abbreviations: HMM: hidden Markov state model, MD: molecular dynamics, MSM: Markov state model, PfA-M1: *Plasmodium falciparum* aminopeptidase M1, SAR: structure-activity relationships.

Abstract

The M1 metallo-aminopeptidase from *Plasmodium falciparum*, PfA-M1, is an attractive drug target for the design of new antimalarials. Bestatin, a broad spectrum metallo-protease inhibitor, is a moderate inhibitor of PfA-M1, and has been used to provide structure activity relationships to inform drug design. The crystal structure of PfA-M1 with bestatin bound within its active site has been determined, however, dynamics of the inhibitor and the association or dissociation pathway have yet to be characterised. Here I present an all-atom molecular dynamics study where we have generated a hidden Markov state model from 2.3 μ s of molecular dynamics simulation. Our hidden Markov state model identifies five macrostates that clearly show the events involved in bestatin dissociation from the PfA-M1 active site. The results show for the first time that bestatin can escape from the active site of PfA-M1, primarily due to weak interactions within the substrate pockets. Our approach identifies individual dynamics and control of substrate / inhibitor movement that could be exploited to produce selective inhibitors that can differentiate between similar members of the M1 aminopeptidase superfamily.

Author Summary

The development of inhibitors/drug candidates typically involves understanding the precise molecular details of their interactions with their targets. This requires a full 3-dimensional image of the candidate bound to the target to be obtained, most commonly using X-ray crystallographic analyses. This approach, whilst it reveals all the atomic interactions between the two molecules, is a static 'snapshot' of the final state and provides no information on how the events before, or in indeed after association. What is essentially needed is a 'movie' describing how the inhibitor/drug engages and disengages with its target. Using molecular simulation techniques, we have mapped, in 3-dimensions, the route by which the inhibitor bestatin can dissociate from the new antimalarial drug target, PfA-M1. This analysis provides important knowledge that may be exploited in the design of new inhibitors/drugs that target PfA-M1 and the large M1 aminopeptidase superfamily.

INTRODUCTION

Bestatin, a natural product isolated from *Streptomyces olivoreticuli*, is a Phe-Leu dipeptide analogue that contains an α -hydroxy- β -amino acid [1]. Bestatin is a broad-spectrum inhibitor of metallo-aminopeptidases and acts via coordination of the zinc ions within the active site of these enzymes [1]. Bestatin has been used to identify and map protease substrate-binding pockets, providing structure activity relationships (SAR) of various M1 family aminopeptidases, including the new malarial drug target, *PfA-M1* [2-6].

PfA-M1, the M1 aminopeptidase from *Plasmodium falciparum*, is involved in the digestion of hemoglobin in blood stage malaria parasites and is an attractive drug target for the development of novel antimalarials [2, 7, 8]. *PfA-M1* belongs to the clan MA family of M1 aminopeptidases that are characterized by two conserved motifs within their catalytic domain. The first motif, HEXHX₍₁₈₎E, is responsible for the coordination of an essential zinc ion and the second, GXMEN, is a substrate guiding motif [9]. *PfA-M1* consists of a transmembrane anchor and a large ectodomain with a canonical 4-domain fold (Fig 1A). The X-ray crystal structure of the ectodomain identified that the active site was buried within the core of the enzyme, with putative access points formed either between domain interactions or through the completely helical C-terminal domain IV that forms a pore with direct access to the active site [2].

The crystal structure of *PfA-M1* bound to bestatin is consistent with a competitive mechanism of inhibition, with the oxygen atoms of the α -hydroxy- β -peptide bond coordinated to the essential zinc ion (Fig 1B) [2]. When bound to bestatin, the coordination geometry of the *PfA-M1* zinc ion is altered from tetrahedral to a pentahedral geometry that is thought to lock the enzyme in an inactive transition state [2]. The S1 and S1' substrate specificity pockets of the enzyme have also been mapped using bestatin (Fig 1B) [2]. The Phe moiety of bestatin occupies the S1 pocket that is predominantly formed by a flexible loop incorporating residues 572-575 as well as a conserved methionine residue (M1034) from domain IV (Fig 1B) [2, 10]. The S1' pocket houses the bestatin Leu moiety and is a large cavity that is open to the C-terminal domain (Fig 1B) [2]. The G₄₆₀AMEN₄₆₄ motif forms

hydrogen bonds with the amide nitrogen of the α -hydroxy- β peptide bond as well as the bestatin carboxylic acid group (Fig 1B).

Work from our laboratory has focussed on the design of novel inhibitors of *PfA-M1* [11, 12]. A previous study modelled a peptide substrate into the active site of *PfA-M1* using bestatin as a guide [13]. A classical molecular dynamics (MD) simulation of this substrate-bound model indicated that the peptide substrate maintained both its conformation and pentahedral metal coordination during a short time-scale (75 ns) [13]. To investigate potential ligand dynamics and diffusion in and around the active site at more relevant biological timescales, we chose to use bestatin as a model system with which to study the binding dynamics between *PfA-M1* and small molecule inhibitors.

Due to the presence of a zinc ion in the active site of *PfA-M1*, we recently developed a hybrid bonded/non-bonded MD simulation model to investigate the behavior of small molecule inhibitors that are bound to the active site of M1 aminopeptidases [14]. In our current study, we use these parameters to perform a series of MD simulations of *PfA-M1* bound to bestatin, producing a total simulation time of 2.3 μ s. This sample size provided us with a large ensemble of conformations (~230,000 snapshots) that we used to perform time-structure independent component analysis (tICA). For our purposes, tICA is a superior form of data analysis in comparison to principle component analysis (PCA) as it allows exploration of the independent movements that occur between protein and ligand over time [15] and is thus well suited to studying the relatively subtle, slow active-site dynamics occurring during enzyme-ligand interactions [16]. Further, it permits generation of a Markov state model (MSM) to identify conformational states of protein and ligand, and transitions between these states [17, 18]. Our analysis identified five key, metastable intermediates in the *PfA-M1*-bestatin binding pathway as well as the atomic interactions that influence the stability of the small molecule in the active site pocket. Furthermore, the transition probability predicted that when these interactions are lost, bestatin can leave the active site of *PfA-M1*. The results from this study provide a route to potent, selective inhibitors of *PfA-M1* by comprehensive structure-activity relationships generated from both interactions and dynamics between the enzyme and inhibitor.

METHODS

System preparation for MD. The starting model for MD of PfA-M1-bestatin was PDB ID 3EBH [2]. Missing atoms and residues were modelled using Modeller 9v11 [19], and charged amino acids were modelled in their protonated states as obtained on the PDB2PQR web server at physiological pH 8.5 [20]. The metal centre was defined as described previously [14] and the protein defined by Amber FF14SB. The O2 and O3 atoms of bestatin were left as non-bonded atoms (no metal-bonds were linked to the zinc ion). Bestatin underwent geometry optimisation using the HF/6-31G* basis set using Gaussian 09 D.01 [21], with atomic charges determined by restrained electrostatic potential fitting (RESP) [22]. During geometry optimization the dihedral N2-C1-C2-C3 bond was frozen to prevent the possibility of self-protonation. The bond constant parameters for bestatin were defined by Parm10, the same set of parameters that used in Amber FF14SB from Amber Tools package [23]. The protein-ligand systems were solvated by TIP3P waters in a cubic water box with a minimum of 12 Å between the edge of the box and the protein-ligand complex. Na⁺ ions were added to neutralize any charge.

MD simulation protocol. All atom MD simulations were performed by using NAMD 2.9 [24] on an IBM Blue Gene/Q cluster using our own Amber force field parameters (212). Equilibration was performed in three stages. First, potential steric clashes in the initial configuration were relieved with 50000 steps of energy minimization. Initial velocities for each system were then assigned randomly according to a Maxwell–Boltzmann distribution at 100 K. Each system was then heated to 300 K over 0.1 ns, under the isothermal-isometric ensemble (NVT) conditions, with the protein atoms (excluding hydrogens) harmonically restrained (with a force constant of 10 kcal mol⁻¹ Å⁻²). Following this, each system was simulated for 100 ps under the isothermal-isobaric ensemble (NPT) with applied harmonic restraints that were reduced from 10 to 2 kcal mol⁻¹ Å⁻². The above equilibration process was performed three times from the same starting structure in order to initiate three production simulations with different initial velocities. For production simulations, the time step was set to 2 fs and the SHAKE algorithm was used to constrain all bonds involving

hydrogen atoms [25]. All simulations were run at constant temperature (300 K) and pressure (1 atm), using a Langevin damping coefficient of 0.5 fs^{-1} [26], and a Berendsen thermostat relaxation time of $\tau_P = 0.1 \text{ ps}$ [27]. The Particle-Mesh Ewald (PME) method was used to set the periodic boundary conditions (PBC) that were used for long-range electrostatic interactions [28] and a real space cut-off of 10 \AA was used. Conformations were sampled every 10 ps for subsequent analysis. In the first round of MD simulations, three complete simulations for PfA-M1-bestatin were performed for 600 ns. Five representative conformations, extracted by the K-means algorithm [29], which are the central conformations of the five most populated clusters from the first round of simulations, were used to launch a second round of MD simulations and two replicates of each representative structure was simulated for 50 ns (see MSM methods for more detail). All frames with time interval of 10 ps were saved to disk.

MD Analysis. Simulation trajectories were analyzed using the GROMACS 5.14 simulation package [30]. For principle component analysis (PCA), $3N \times 3N$ atom covariance matrices of the protein displacement in simulations were generated based on backbone atoms (N, C α , C, O) of the PfA-M1-bestatin crystal structure. Principle Components (PCs), that taken together accounted for more than 50% of the overall covariance, were chosen for essential dynamics analysis. The GROMACS 5.14 simulation package [30] was used to project the trajectory onto the top PCs. Graphs and plots were produced with Xmgrace [31] and GraphPad Prism7. Molecular graphics were prepared with PyMOL 1.8.23 [32] and VMD1.9.3 [33].

Time-structure based independent component analysis. Pyemma [34] was used to perform time-structure independent component analysis (tICA) of the combined MD trajectories. The tICA in this study was constructed based on the methods described previously using a lag time of 2 ns [16] [35], which was optimized for studying ligand-binding states. The RMSD of bestatin and the metal centre residues (H496, H500 E519, and zinc) from the reference starting conformation (from the crystal structure) was then used as the 'feature' parameter for tICA. The covariance matrix from the entire trajectory data was calculated by eq. (1) [36].

$$C_{ij}(\tau) = \langle r_i(t)r_j(t+\tau) \rangle_t \quad (1)$$

Where C_{ij} is the covariance matrix, τ is the lag time (2ns), and r_i is the input RMSD feature, which are defined as the RMSD of the selected atoms from the reference structure at time t . The tIC eigenvectors, as well as the tIC eigenvalues were obtained by singular value decomposition of the covariance matrix. The top two tICs (tIC1 and tIC2) were dominant (59.9%) to the overall covariance, thus were used for the analysing essential dynamics of bestatin binding. The MSM was constructed upon the subspaces defined by the top two tICs.

Construction of MSM. The K-means algorithm [29] within the software program Pyemma [34] was used to divide the conformational space into 1265 microstates based on the top two tICs from the total MD simulations (2.3 μ s). Time was then coarse-grained into discrete intervals τ . A MSM that can predict longer timescale dynamics was subsequently built via the first-order master equation eq. (2).

$$p(n\tau) = T^n \tau p(0) \quad (2)$$

where $p(n\tau)$ is a vector describing the state population and $T^n \tau$ is the transition probability matrix of the lag time τ .

Validation and evaluation of MSM. The generated MSMs were then interrogated by implied timescale plots using a selection of random lag times (< 20 ns) to ensure that the model was Markovian. The timescales (τ_k) were obtained from the transition probability matrix T_{ij} and defined by eq. (3).

$$\tau_k = -\frac{\tau}{\ln \mu_k(\tau)} \quad (3)$$

where (μ_k) is the eigenvalue of the transition matrix with the given lag time τ .

The MSM was then evaluated by the Chapman-Kolmogorov test to ensure the model to be Markovian [37]. Bayesian MSMs were used for the display of statistical errors in the test [38]. The MSM refinement followed the protocol outlined previously by Jiang *et al.*, [39]. Unfortunately, validation of our MSM using the Chapman Kolmogorov test showed that the predicted model was not Markovian in nature, as none of the implied timescales

reached a plateau and also showed poor discretization at any of the provided lag times (lag times tested from 0.1 – 20 ns, Supp. Fig 1A). The Chapman-Kolmogorov evaluation test [37] showed that the MSM from 1.8 μ s of MD, with lag time of 5 ns, resulted in a prediction that was too fast (Supp. Fig 1B). Therefore, our initial triplicates of 600 ns were not sufficient for the MSM so we chose to perform additional MD sampling. We then divided the conformations obtained from the initial triplicate simulations into 5 clusters using the K-means algorithm [29] and choose the middle conformation of each cluster as a new starting point for a second round of MD simulations. Each new starting conformation was used (with reassigned initial velocities) to perform two independent replicates of 50 ns of MD simulation, with two replicates that had different initial velocities being completed for each cluster. We then combined our total MD simulation (first round = 600 ns in triplicate, a second round = 50 ns in duplicate for five different clusters). The combined MD trajectories consisted of 2.3 μ s or 230,000 frames with a timestep of 10 ps. Clustering the conformations [29] identified 1265 microstates. The model was tested again via the Chapman-Kolmogorov test and found to be Markovian (Supp Fig 1C & D) so we pooled the total 1265 microstates into 5 metastable macrostates using the Robust Perron Cluster Analysis (PCCA+) algorithm [40]. These five macrostates were then used to build a coarse-grained hidden Markov state model (HMM) using a lag time of 5 ns [41] which described transition probabilities between the five macrostates.

RESULTS

Construction and validation of a Markov state model (MSM)

To investigate the binding and conformational dynamics of bestatin bound to PfA-M1, we performed all-atom MD simulations of the PfA-M1–bestatin complex for a total of 2.3 μ s. To investigate the behavior of the complex in more detail and access functionally-realistic timescales, we performed time-structure based independent component analysis (tICA) on this dataset using the RMSD between bestatin and the PfA-M1 metal centre (H496, H500, E519 and zinc ion) as the tICA “feature” parameter. The top two tICs explain 59.9%

of the movement observed between bestatin and the metal centre of PfA-M1. To generate a free energy landscape of bestatin binding to PfA-M1, we projected our MD trajectories onto the tIC 1 and 2 vectors and produced a 2D projection of the conformational space that was sampled (Fig 2A). The free energy landscape identifies two major populations within the trajectories that are separated by a large energy barrier (Fig 2A). Further, the X-ray crystal structure was not the lowest energy conformational state observed in our study (Fig 2A).

To understand the details of the binding interactions and dynamics of the bestatin molecule within the free energy landscape, we built a MSM with 1265 microstates on the top two tICs. Evaluation of the implied timescale showed convergence at a lag time of 5 ns and the produced model was Markovian as the plot was shown to plateau after 5 ns (Supp Fig 1C). Validation, as detailed in the Methods, showed that the new MSM predicts slightly faster kinetics than the MD simulations (Supp Fig 1D). The same trend was observed in other recent studies using MSMs [39]. To group the 1265 microstates into meaningful macrostates for further analysis, we constructed a coarse-grained MSM by pooling the total 1265 microstates into 5 metastable macrostates in a hidden Markov state model (HMM), as described in Methods.

Pentahedral coordination of the zinc ion is not stably maintained

In the crystal structure, bestatin is coordinated to the zinc ion via two bonds, one from the inhibitors' hydroxyl group ($\text{Bes}_{\text{O}_2}\text{-Zn}$) and a second from the carbonyl group ($\text{Bes}_{\text{O}_3}\text{-Zn}$) [2] (Fig 1B). Whilst the starting conformation of our MD simulation was taken from this X-ray crystal structure, our MSM analysis showed that this binding pose is not stable. In state 1 ($\Delta G = 4.32$ kT, Fig 2B) we observed two binding conformations within the population, that we have termed substate 1a and substate 1b (Supp. Fig 2A and 2B). In both substates, the zinc ion remains pentahedrally coordinated via bestatin, however, the position of the oxygen atoms that coordinate the zinc ion changes. Substate 1a is the equivalent of the X-ray crystal structure, with a $\text{Bes}_{\text{O}_2}\text{-Zn}$ bond of 2.2 ± 0.06 Å and a longer $\text{Bes}_{\text{O}_3}\text{-Zn}$ bond of 2.5 ± 0.2 Å. In substate 1b, this configuration switches so that Bes_{O_3} moves closer to the

zinc ion ($2.05 \pm 0.1 \text{ \AA}$) and Bes_{O_2} moves away, lengthening the metallo-bond to $2.5 \pm 0.2 \text{ \AA}$. In the substate 1b conformation, the Phe moiety of bestatin loses its π - π interactions with the phenol group of Y575 and can be seen to rotate within the S1 pocket (Supp. Fig 2B). This results in the movement of E572, a residue known to be flexible [10], as well as M1034, causing a change in the size and shape of the S1 pocket. The bestatin Leu moiety maintains a similar position in both substates 1a and 1b. The hydrogen bonding network of substate 1b is also altered compared to 1a, with Bes_{O_2} changing its H-bond coordination to E497_{OE} , E463_{OE} and E463_{N} whilst Bes_{O_3} lost its hydrogen bond with Y580_{OH} but has interactions with H500_{N_2} , and E497_{OE} (Supp. Fig 2B, Supp. Table 1). The carboxylic acid of bestatin ($\text{Bes}_{\text{O}_1/\text{O}_4}$) forms only one hydrogen bond with A461_{N} (Supp. Fig 2B, Supp. Table 1).

In addition to the dynamics of the zinc-ligand interactions, our results also show that there is a rapid transition between states 1 and 2, with state 2 representing the lowest energy conformation for bestatin (Fig 2B). State 2 has a very low free energy ($\Delta G = 0.51 \text{ kT}$) compared to state 1 ($\Delta G = 4.32 \text{ kT}$) and a forward transition rate of $9.1 \mu\text{s}^{-1}$ (reverse transition is $6.0 \mu\text{s}^{-1}$). State 2 therefore represents the most stable binding pose and the largest population observed in our study. Within state 2, 30 % of the conformations sampled had lost their Bes_{O_3} -Zn metal bond (distances $> 3.6 \pm 0.5 \text{ \AA}$), retaining only a tetrahedral zinc coordination via the Bes_{O_2} -Zn bond ($2.1 \pm 0.1 \text{ \AA}$). The loss of the Bes_{O_3} -Zn metal bond appears to allow the carboxylic group ($\text{Bes}_{\text{O}_1/\text{O}_4}$) to move deeper into the S1' pocket via the formation of two hydrogen bonds with G460_{N} (Table S1). This movement of the carboxyl group also drags the bestatin Leu moiety down into the large S1' pocket (Supp. Fig 2C). We also observed that the bestatin Phe group rotates similarly to that observed in state 1 (Supp. Fig 2C). There was an overall loss of four hydrogen bonds, with new bonds from Bes_{N_2} moving closer to E497 and forming a new hydrogen bond with E497_{OE} . The hydrogen bond network of Bes_{O_2} is maintained as described in both substates of state 1 and Bes_{O_3} still forms a hydrogen bond with Y580_{OH} (Supp. Table 1).

Bestatin must overcome a large energy barrier to lose zinc coordination.

State 3 in our model represents the loss of the final metallo-ligand bond between bestatin and the zinc ion, initiating the diffusion of bestatin away from the active site. When the bond between $\text{Bes}_{\text{O}2}\text{-Zn}$ is lost, the side-chain of E497 moves and fulfills the tetrahedral coordination of the zinc ion whilst also maintaining a hydrogen bond with $\text{Bes}_{\text{O}2}$. The transitions between states 2 and 3 are very rare events (Fig 2B). Both the forward and backward transition rates calculated in our study were only $0.15 \mu\text{s}^{-1}$, however, once obtained, state 3 is quite stable ($\Delta G = 1.8 \text{ kT}$). The loss of the zinc coordination sees the entire bestatin molecule move away from the active site pulling the Phe moiety out of the S1 pocket (Supp Fig 2D). In this state, compared to the crystal structure, S1 pocket residues E572 and M1034 have moved $2.3 \pm 0.2 \text{ \AA}$ from their original position, enlarging the size of the pocket. The movement out of the S1 pocket appears related to a new hydrogen bond that is formed between $\text{Bes}_{\text{N}2}$ and GAMEN residue A461_N. The original hydrogen bonds between the GAMEN motif and the carboxylic ($\text{Bes}_{\text{O}1/\text{O}4}$) oxygens are lost, and G460 and A461 move 4.1 \AA away from the active site. The carboxyl group $\text{Bes}_{\text{O}1/\text{O}4}$ moves further down the large S1' pocket and forms new hydrophobic interactions with V459 (Supp Fig 2D). The Leu group also swings towards the Phe moiety and has hydrophobic interactions with the backbone atoms of T577 (Supp. Fig 2D).

Bestatin has two potential exit points from the active site

Once the zinc coordination is completely lost, bestatin appears to be able to move away from the active site in two different directions (Fig 2B). States 4 and 5 represent diffusion away from the active site and are the two smallest populations observed in the HMM (Fig 2B). This is likely because the backward transition rates (returning to state 3) for both state 4 ($5.78 \mu\text{s}^{-1}$) and 5 ($2.2 \mu\text{s}^{-1}$) are significantly higher than their forward rates (1.1 and $0.35 \mu\text{s}^{-1}$ respectively, Fig 2B), indicating that state 3 is the favored position. In both states 4 and 5, the geometry of the zinc ion reverts to what is observed in the unbound X-ray crystal structure of *PfA*-M1 (PDB ID 3EBG) where a water molecule coordinates the zinc ion. We observe that a water molecule moves into the active site in states 4 and 5, displacing the direct interaction of the side-chain of E497 with the zinc in state 3, and forming

a water-mediated bridge between E497 and the zinc ion (Supp. Fig 2E). This returns the active site to a proteolysis ready position, with the nucleophilic water in position.

In state 4, bestatin moves further into the S1 pocket, moving toward an apparent exit to solvent that is formed by the junction of domains I, II and IV (Fig 3). In our study, bestatin is captured at a maximum distance of 4.5 ± 1.7 Å from the zinc ion and the Phe group forms hydrophobic interactions with Q317, E319, M1034, Q1038 and V459 (Supp. Fig 2E). The $\text{Bes}_{\text{O}2}$ maintains a hydrogen bond with E519, but $\text{Bes}_{\text{O}3}$ only retains a hydrogen bond with Y580_{OH} (Supp. Table 1). $\text{Bes}_{\text{N}2}$ only forms one hydrogen bond with E319_{OE}. The carboxyl group $\text{Bes}_{\text{O}1/\text{O}4}$ continues its way down to the deep S1' pocket, losing all hydrogen bonds and hydrophobic interactions to the pocket. The Leu group swings down to accompany the carboxylic group but has hydrophobic interactions with T576 and T577.

State 5 is the least populated state found in our study (Fig 2B). At the maximum displacement from the zinc ion as measured in this study, bestatin has moved 6.4 ± 0.3 Å from the zinc ion and is migrating toward the S1' pocket / channel. Extrapolation of this directional vector means that the molecule would leave the enzyme via the C-terminal domain IV pore as previously hypothesized [2]. Both the Phe and Leu moieties are in the S1' pocket and the whole molecule is observed to migrate away from the metal center. The Phe group has hydrophobic interactions with V493, G460 and A461 and the Leu moiety also has hydrophobic interactions with S1' pocket residues 575-577 and Y580 (Supp. Fig 2F). The $\text{Bes}_{\text{O}2}$ and $\text{Bes}_{\text{O}3}$ form no hydrogen bonds in state 5, whilst $\text{Bes}_{\text{N}2}$ reforms a hydrogen bond with E497_{OE}. The carboxylic oxygens ($\text{Bes}_{\text{O}1/\text{O}4}$) move towards Q1038, however, measurements indicate that they are still too far to form a hydrogen bond. Finally, in contrast to all other states, bestatin does not form a hydrogen bond to Y580.

Overall dynamics of PfA-M1 throughout simulations

We investigated the overall stability and dynamics of the PfA-M1 protein throughout the MD simulation. Root mean square deviation (RMSD) of the backbone atoms in simulations stabilized at 2.1 ± 0.1 Å indicating that the protein underwent no large rigid body motion (Supp Fig 3A). Any potential global conformational transitions were investigated by

principle component analysis (PCA) with a distance matrix defined by the protein backbone atoms. Our PCA identified that principle component 1 (PC1) and 2 (PC2) contributed 45 % to the overall variance whilst PCs 3-5 accounted for 11.3 %. Inspection of PC1 and 2 shows a small concerted motion of all domains that likely results from the release of crystal contacts (Supp Fig 3B). Investigation at the residue level identified that the volume of the S1 pocket is increased due to outward movement of S1 helix (residues 570-575) and M1034 (Supp Fig 3C). The GAMEN motif residues also moved marginally with E463 and N464 moving 0.5 Å towards the S1 pocket, while G460 and A461 move up by 0.7 Å (Supp Fig 3C). The overall result of PC1 and PC2 dynamics produces an S1 pocket that is altered compared to the crystal structure and indicates that the slightly confined S1 pocket observed in crystal structures may be due to packing constraints.

A model for the binding and dissociation of bestatin

The five states identified in the HMM offer a chance to provide a reasonable model for the association and dissociation of bestatin to *PfA-M1*. Whilst our data only models diffusion away from the active site, the transition rates in combination with previous studies indicate a probable series of events.

The most stable interaction observed throughout our entire study was a hydrogen bond between Y580 and bestatin. Y580 is a key residue in the catalytic mechanism of the M1 aminopeptidases [2, 5, 42], where it binds to the substrate P1 carbonyl oxygen in the tetrahedral transition state, reducing the activation energy of the hydrolysis reaction [13, 42]. If one considers then, that the interaction between Y580 and bestatin is one of the primary drivers of position and stability (as supported by our current MD data), then state 5 may represent the approach of the inhibitor to the active site (pre-binding). In all other states in our study, Y580 shares at least one hydrogen bond with bestatin and could be considered associated, if not bound, to the active site.

The transition from state 5 to state 3 is highly favoured, indicating that this is the likely direction of movement of bestatin (Fig 2B). In state 3, we observe the first interaction with the GAMEN motif, a recognized substrate guiding region (Supp. Table 1) [2, 5, 42].

From state 3, there are two main options for the inhibitor, neither of which occurs rapidly (Fig 2B). It can move into a more favourable energetic position and form more interactions with the GAMEN motif and other catalytic residues, including the zinc ion (state 2) or it can move past the active site toward the N-terminal exit (state 4). The transition rates for either of these directions are low (compared to a return to state 3) and represent very different outcomes. A move to state 2, that has an equally high energy barrier to any direction of movement, represents true binding of the inhibitor to the active site. The equal rates of transition between directions is consistent to the biological activity observed in experimental inhibition assays, where a linear inhibitory kinetics are observed [10]. A move to state 4 may represent the dissociation of the inhibitor from the enzyme, and ultimately an egress route.

State 1 can be accessed with equal efficiency from state 2 and likely exists due to the crystallography data. Analysis of the interconverting binding poses within state 1 (1a and 1b) shows that the bestatin pose is not stable and our PCA shows us that the size of the S1 pocket increases in the absence of crystal contacts. The hydroxyl ketone oxygens move between a first and second shell coordination of the zinc ion, resulting in the two distinct populations that we have called substate 1a and 1b. Transition from state 1 to 2, our lowest energy state, shows that 70 % of the population retain this dynamic pentahedral coordination, however, the remaining 30 % revert to a tetrahedral. The same tetrahedral geometry is also observed in state 3, where E497_{OE} fulfills the zinc coordination, suggesting that the loss of zinc coordination by bestatin may be due to competition by E497_{OE}.

Previous studies by Jones *et al.*, showed that pentahedral coordination of the zinc ion was unchanged throughout their MD simulation of a modelled substrate-bound PfA-M1 [13]. However, their MD parameters placed dummy atoms around the zinc ion (a non-bonded model) and charge distributions were determined empirically by test simulations [13]. This raises the concern of unrealistic charge distributions amongst first shell atoms [43] and may account for the difference in the stability they observed versus the dynamic coordination observed in our current study. Our hybrid force field parameters are appropriate for the tetrahedral metal geometry of the M1 aminopeptidase active site and allow non-polarised waters near the zinc ion to move, describing a more realistic protease

active site. In validation studies of our force field parameters, two waters were identified in the second shell of Harding's bond cut-off (2.1 Å, occupancy of 100% and 2.6 Å, occupancy of 90%) [14]. This finding suggests that our force field parameters can simulate both tetrahedral and pentahedral geometry of zinc binding area but may show a slight preference for retention of tetrahedral geometry. In this study, our starting model had pentahedral geometry between bestatin and PfA-M1. One limitation of our study is that the non-polarizable force-field used to model coordination bonds might decrease the accuracy of modelling the real electric distribution around the metal ion. However, this remains a challenge to all of the parameterization methods of metallo-proteins in classical MD simulations [44-49]. The polarization of the inhibitor changes depending on its vicinity to the metal ion, when coordinated it is polarized but if it moves away from the metal ion, it will become non-polarized. Given that we are unable to change the polarization state of the inhibitor during a classical MD simulation, we chose to model bestatin in a non-polarized state so that when the inhibitor moves away from the metal centre, the non-polarized inhibitor is appropriate within the area defined by a classical force field.

DISCUSSION

Excluding viruses, M1 aminopeptidases are distributed throughout all phyla, and have been implicated in a wide range of functions including cell maintenance, growth and development, and defense [50]. The catalytic domain and active sites of the M1 aminopeptidases share the highest % sequence identity within the superfamily and includes conservation of key residues involved in the proteolytic mechanism [50]. Understanding the dynamics of the active site and S1 / S1' substrate pockets may provide new avenues to selectivity and potency of inhibitors, not only for the production of antimalarials, but also for inhibitors of the human cancer target, Aminopeptidase N (APN) [51].

With the exception of APN, the bestatin bound M1 aminopeptidases show a conserved binding mode wherein the backbone hydroxy ketone of bestatin coordinates the active site zinc ion [2, 5, 42, 52-57]. In APN, it is the terminal carboxylic acid that coordinates the zinc, which results in a completely different binding pose [6]. Comparison of the

canonical vs APN binding modes indicates that the APN substrate-binding pocket is wider than those in homologous enzymes, which allows or even promotes a different bestatin binding pose. Interestingly in our study, relaxation of the crystal contacts also produced an increase in size of the S1 substrate pocket that immediately resulted in fluctuations in the position of the Phe ring. Rotation of the Phe ring when bestatin is bound is also not unique to *PfA-M1*, as other crystal structures have identified different positions of the Phe ring [2, 5, 42, 52-57]. Positioning of the Phe ring relies on π - π stacking in *PfA-M1* as well as in porcine Aminopeptidase N and the human Aminopeptidase A [4, 58]. Improving the strength of the interactions with S1 pocket residues would stabilize the inhibitor directly via stronger bonds as well as by limiting the flexibility of the pocket.

Dynamics within the S1 pocket of *PfA-M1* has been observed previously but were assumed to be result of an induced fit of bulky inhibitor molecules [10]. However, our analysis in this study shows that the substrate specificity pockets may be inherently flexible and not simply 'pushed' by a bulky inhibitor. E572 and M1034 have been identified as flexible cap residues in multiple studies [10, 13, 59, 60]. In a recent study, Rosati *et al* confirmed that E572 and M1034 promote broad specificity and enhance catalysis but that replacement of these residues with those of the *E. coli* aminopeptidase N did not promote catalytic efficiencies that are observable for the *E. coli* enzyme. Their conclusion was that there are factors beyond sequence that drive catalytic efficiency in the M1 aminopeptidases [60]. Our study shows that at least one other factor to consider is protein dynamics. Inducing rigidity into the substrate pockets, via engineering stronger interactions between ligand and protein, may aid potency of inhibitors.

Random accelerated MD of APN bound to bestatin suggested that there were three potential pathways for ligand dissociation [61]. The major channel was the large C-terminal domain IV pore [61]. The other two less populated routes were located at the junction of domains I, II and IV and from the catalytic domain [61]. Our current study suggests a model for bestatin binding and dissociation that involves only two different intramolecular routes. The route that the ligand takes appears dependent on the dynamics of the S1 pocket. In all observations made in this study, the S1 pocket is made larger by the movement of E572

and M1034 away from the zinc ion. Once these residues are 2.7 Å from their starting position, the GAMEN motif is also able to move away from the zinc and destabilizes the position of the bestatin carboxylic acid. The concurrent movement of the bestatin molecule allows E497_{OE} to compete with the Bes_{O2} for coordination to the zinc, and if successful, can break metal coordination with the ligand. From this point, E572 and M1034 act as gates. If both continue to move and open, the S1' pocket is no longer contained and the ligand can migrate through the S1 and exit the molecule via the smaller channel formed by the junction of domains I, II and IV. However, if the position of E572 and M1034 stabilizes, then diffusion of the molecule occurs via the internal channel formed by the helices of domain IV.

Some members of the M1 aminopeptidase family have been shown to undergo a radical conformational change that 'opens' the protease, exposing the active site to solvent [5, 55, 62]. Analysis of 'open' conformation of the M1 aminopeptidases shows that the S1 pocket is poorly organised [62, 63]. An open conformation of *PfA*-M1, nor any bacterial homologs characterised to date, has never been observed experimentally, despite resolution of numerous X-ray crystal structures [2, 10-12, 59, 64-66]. Despite these differences, it appears that dynamics in the structure, shape and size of the S1 pocket of these enzymes is still key to their specificity and ultimately, their inhibition.

CONCLUSIONS

The results of this study raise significant findings with regard to how to model inhibitor behavior for the M1 aminopeptidases. The method of choice for structure-activity relationships is often X-ray crystallography, which provides powerful atomic information about key interactions between enzyme and ligand. However, as shown here, dynamics should not be overlooked. Our results show that our crystal structure of *PfA*-M1 bound to bestatin is likely artificially restrained by crystal contacts and consequently displays a smaller S1 pocket. The change in size of the S1 pocket has profound effects on the movement of the P1 inhibitor moiety, the Phe ring. Likewise, dynamics of the S1 pocket also appear to dictate inhibitor egress from the active site (and possibly entrance) and show key residues to target to improve potency by stabilizing both the inhibitor and target. Finally,

the identification of preferred routes of entrance and egress provides new avenues to promote selectivity within the enzyme family. Individual dynamics and control of substrate movement could be exploited to produce selective inhibitors that can differentiate between similar members of the M1 aminopeptidase superfamily.

Acknowledgements: This work was supported by the Australian National Health and Medical Research Council (Project grant 1063786 to SM). AMB acknowledges fellowship support from the NHMRC (1022688). WY is supported by postgraduate scholarships from Monash University. This work was supported by the Victorian Life Science Computation Initiative (VLSCI, vlsci.org.au) and National Computation Infrastructure (NCI, nci.org.au).

Author Contributions: WY and SM conceived and designed the experiments, analyzed the data, and wrote the paper. WY and IK performed the experiments. WY, BR, BP analyzed the data.

References:

1. Umezawa H, Aoyagi T, Suda H, Hamada M, Takeuchi T. Bestatin, an inhibitor of aminopeptidase B, produced by actinomycetes. *J Antibiot (Tokyo)*. 1976;29(1):97-9. PubMed PMID: 931798.
2. McGowan S, Porter CJ, Lowther J, Stack CM, Golding SJ, Skinner-Adams TS, et al. Structural basis for the inhibition of the essential *Plasmodium falciparum* M1 neutral aminopeptidase. *Proc Natl Acad Sci U S A*. 2009;106(8):2537-42. doi: 10.1073/pnas.0807398106. PubMed PMID: 19196988; PubMed Central PMCID: PMC2636733.
3. McGowan S, Oellig CA, Birru WA, Caradoc-Davies TT, Stack CM, Lowther J, et al. Structure of the *Plasmodium falciparum* M17 aminopeptidase and significance for the design of drugs targeting the neutral exopeptidases. *Proc Natl Acad Sci U S A*. 2010;107(6):2449-54. doi: 10.1073/pnas.0911813107. PubMed PMID: 20133789; PubMed Central PMCID: PMC2809755.

4. Chen L, Lin YL, Peng G, Li F. Structural basis for multifunctional roles of mammalian aminopeptidase N. *Proc Natl Acad Sci U S A*. 2012;109(44):17966-71. doi: 10.1073/pnas.1210123109. PubMed PMID: 23071329; PubMed Central PMCID: PMC3497818.
5. Nguyen TT, Chang SC, Evnouchidou I, York IA, Zikos C, Rock KL, et al. Structural basis for antigenic peptide precursor processing by the endoplasmic reticulum aminopeptidase ERAP1. *Nat Struct Mol Biol*. 2011;18(5):604-13. doi: 10.1038/nsmb.2021. PubMed PMID: 21478864; PubMed Central PMCID: PMC3087843.
6. Wong AH, Zhou D, Rini JM. The X-ray crystal structure of human aminopeptidase N reveals a novel dimer and the basis for peptide processing. *J Biol Chem*. 2012;287(44):36804-13. doi: 10.1074/jbc.M112.398842. PubMed PMID: 22932899; PubMed Central PMCID: PMC3481283.
7. Skinner-Adams TS, Lowther J, Teuscher F, Stack CM, Grembecka J, Mucha A, et al. Identification of phosphinate dipeptide analog inhibitors directed against the *Plasmodium falciparum* M17 leucine aminopeptidase as lead antimalarial compounds. *J Med Chem*. 2007;50(24):6024-31. doi: 10.1021/jm070733v. PubMed PMID: 17960925.
8. Flipo M, Beghyn T, Leroux V, Florent I, Deprez BP, Deprez-Poulain RF. Novel selective inhibitors of the zinc plasmodial aminopeptidase *PfA-M1* as potential antimalarial agents. *J Med Chem*. 2007;50(6):1322-34. doi: 10.1021/jm061169b. PubMed PMID: 17326615.
9. Rawlings ND, Morton FR, Barrett AJ. MEROPS: the peptidase database. *Nucleic Acids Res*. 2006;34(Database issue):D270-2. doi: 10.1093/nar/gkj089. PubMed PMID: 16381862; PubMed Central PMCID: PMC1347452.
10. Velmourougane G, Harbut MB, Dalal S, McGowan S, Oellig CA, Meinhardt N, et al. Synthesis of new (-)-bestatin-based inhibitor libraries reveals a novel binding mode in the S1 pocket of the essential malaria M1 metalloaminopeptidase. *J Med Chem*. 2011;54(6):1655-66. doi: 10.1021/jm101227t. PubMed PMID: 21366301; PubMed Central PMCID: PMC3516848.

11. Mistry SN, Drinkwater N, Ruggeri C, Sivaraman KK, Loganathan S, Fletcher S, et al. Two-pronged attack: dual inhibition of Plasmodium falciparum M1 and M17 metalloaminopeptidases by a novel series of hydroxamic acid-based inhibitors. *J Med Chem*. 2014;57(21):9168-83. doi: 10.1021/jm501323a. PubMed PMID: 25299353.
12. Drinkwater N, Vinh NB, Mistry SN, Bamert RS, Ruggeri C, Holleran JP, et al. Potent dual inhibitors of Plasmodium falciparum M1 and M17 aminopeptidases through optimization of S1 pocket interactions. *Eur J Med Chem*. 2016;110:43-64. doi: 10.1016/j.ejmech.2016.01.015. PubMed PMID: 26807544.
13. Jones PM, Robinson MW, Dalton JP, George AM. The Plasmodium falciparum Malaria M1 Alanyl Aminopeptidase (*Pf*A-M1): Insights of Catalytic Mechanism and Function from MD Simulations. *PLoS One*. 2011;6(12):e28589.
14. Yang W, Riley BT, Lei X, Porebski BT, Kass I, Buckle AM, et al. Generation of AMBER force field parameters for zinc centres of M1 and M17 family aminopeptidases. *J Biomol Struct Dyn*. 2017:1-10. doi: 10.1080/07391102.2017.1364669. PubMed PMID: 28782419.
15. Naritomi Y, Fuchigami S. Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: the case of domain motions. *J Chem Phys*. 2011;134(6):065101. doi: 10.1063/1.3554380. PubMed PMID: 21322734.
16. Plattner N, Noe F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat Commun*. 2015;6:7653. doi: 10.1038/ncomms8653. PubMed PMID: 26134632; PubMed Central PMCID: PMC4506540.
17. Swope WC, Pitera JW, Suits F. Describing protein folding kinetics by molecular dynamics simulations. 1. Theory. *J Phys Chem B*. 2004;108(21):6571-81. doi: 10.1021/jp037421y. PubMed PMID: WOS:000221546200008.
18. Swope WC, Pitera JW, Suits F, Pitman M, Eleftheriou M, Fitch BG, et al. Describing protein folding kinetics by molecular dynamics simulations. 2. Example applications to alanine dipeptide and a β -hairpin peptide. *J Phys Chem B*. 2004;108(21):6582-94.

19. Eswar N, Webb B, Marti-Renom MA, Madhusudhan M, Eramian D, Shen My, et al. Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics*. 2006;5.6. 1-5.6. 30.
20. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res*. 2004;32(Web Server issue):W665-7. doi: 10.1093/nar/gkh381. PubMed PMID: 15215472; PubMed Central PMCID: PMC441519.
21. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al. *Gaussian 09*. Wallingford, CT, USA: Gaussian, Inc.; 2009.
22. Bayly CI, Cieplak P, Cornell WD, Kollman PA. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J Phys Chem*. 1993;97(40):10269-80. doi: Doi 10.1021/J100142a004. PubMed PMID: WOS:A1993MA74100004.
23. Case De, Darden T, Cheatham T, Simmerling C, Wang J, Duke R, et al. *Amber 11*. University of California, 2010.
24. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005;26(16):1781-802. doi: 10.1002/jcc.20289. PubMed PMID: 16222654; PubMed Central PMCID: PMC2486339.
25. Ryckaert J-P, Ciccotti G, Berendsen HJ. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys*. 1977;23(3):327-41.
26. Kawasaki K. Simple Derivations of Generalized Linear and Nonlinear Langevin Equations. *J Phys A*. 1973;6(9):1289-95. doi: Doi 10.1088/0305-4470/6/9/004. PubMed PMID: WOS:A1973Q738100004.
27. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. Molecular-Dynamics with Coupling to an External Bath. *J Chem Phys*. 1984;81(8):3684-90. doi: Doi 10.1063/1.448118. PubMed PMID: WOS:A1984TQ73500045.

28. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A Smooth Particle Mesh Ewald Method. *J Chem Phys.* 1995;103(19):8577-93. doi: Doi 10.1063/1.470117. PubMed PMID: WOS:A1995TE36400026.
29. Gonzalez TF. Clustering to Minimize the Maximum Intercluster Distance. *Theoretical Computer Science.* 1985;38(2-3):293-306. doi: Doi 10.1016/0304-3975(85)90224-5. PubMed PMID: WOS:A1985ASV1200008.
30. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem.* 2005;26(16):1701-18. doi: 10.1002/jcc.20291. PubMed PMID: 16211538.
31. Turner P. XMGRACE, Version 5.1. 19. Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR. 2005.
32. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. 2015.
33. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph.* 1996;14(1):33-8, 27-8. PubMed PMID: 8744570.
34. Scherer MK, Trendelkamp-Schroer B, Paul F, Perez-Hernandez G, Hoffmann M, Plattner N, et al. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J Chem Theory Comput.* 2015;11(11):5525-42. doi: 10.1021/acs.jctc.5b00743. PubMed PMID: 26574340.
35. Silva DA, Bowman GR, Sosa-Peinado A, Huang XH. A Role for Both Conformational Selection and Induced Fit in Ligand Binding by the LAO Protein. *PLoS Comp Biol.* 2011;7(5):e1002054. doi: ARTN e1002054
10.1371/journal.pcbi.1002054. PubMed PMID: WOS:000291015800028.
36. Hyvärinen A, Karhunen J, Oja E. Independent component analysis: John Wiley & Sons; 2004.
37. Prinz JH, Wu H, Sarich M, Keller B, Senne M, Held M, et al. Markov models of molecular kinetics: generation and validation. *J Chem Phys.* 2011;134(17):174105. doi: 10.1063/1.3565032. PubMed PMID: 21548671.

38. Trendelkamp-Schroer B, Wu H, Paul F, Noe F. Estimation and uncertainty of reversible Markov models. *J Chem Phys.* 2015;143(17):174101. doi: 10.1063/1.4934536. PubMed PMID: 26547152.
39. Jiang HL, Sheong FK, Zhu LZ, Gao X, Bernauer J, Huang XH. Markov State Models Reveal a Two-Step Mechanism of miRNA Loading into the Human Argonaute Protein: Selective Binding followed by Structural Re-arrangement. *PLoS Comp Biol.* 2015;11(7):e1004404. doi: ARTN e1004404
10.1371/journal.pcbi.1004404. PubMed PMID: WOS:000360620100037.
40. Deuffhard P, Weber M. Robust Perron cluster analysis in conformation dynamics. *Linear Algebra and Its Applications.* 2005;398:161-84. doi: 10.1016/j.laa.2004.10.026. PubMed PMID: WOS:000227411500010.
41. Lange OF, Grubmuller H. Full correlation analysis of conformational protein dynamics. *Proteins.* 2008;70(4):1294-312. doi: 10.1002/prot.21618. PubMed PMID: 17876828.
42. Ito K, Nakajima Y, Onohara Y, Takeo M, Nakashima K, Matsubara F, et al. Crystal structure of aminopeptidase N (proteobacteria alanyl aminopeptidase) from *Escherichia coli* and conformational change of methionine 260 involved in substrate recognition. *J Biol Chem.* 2006;281(44):33664-76. doi: 10.1074/jbc.M605203200. PubMed PMID: 16885166.
43. Hu L, Ryde U. Comparison of Methods to Obtain Force-Field Parameters for Metal Sites. *J Chem Theory Comput.* 2011;7(8):2452-63. doi: 10.1021/ct100725a. PubMed PMID: 26606619.
44. Peters MB, Yang Y, Wang B, Fusti-Molnar L, Weaver MN, Merz KM, Jr. Structural Survey of Zinc Containing Proteins and the Development of the Zinc AMBER Force Field (ZAFF). *J Chem Theory Comput.* 2010;6(9):2935-47. doi: 10.1021/ct1002626. PubMed PMID: 20856692; PubMed Central PMCID: PMC2941202.
45. Lin F, Wang R. Systematic Derivation of AMBER Force Field Parameters Applicable to Zinc-Containing Systems. *J Chem Theory Comput.* 2010;6(6):1852-70. doi: 10.1021/ct900454q. PubMed PMID: 26615845.

46. Pang YP. Successful molecular dynamics simulation of two zinc complexes bridged by a hydroxide in phosphotriesterase using the cationic dummy atom method. *Proteins*. 2001;45(3):183-9. doi: Doi 10.1002/Prot.1138. PubMed PMID: 11599021.
47. Hancock RD. Molecular Mechanics Calculations as a Tool in Coordination Chemistry. *Prog Inorg Chem*. 1989;37:187-291. doi: Doi 10.1002/9780470166383.Ch4. PubMed PMID: WOS:A1989DV50200004.
48. Aqvist J, Warshel A. Free energy relationships in metalloenzyme-catalyzed reactions. Calculations of the effects of metal ion substitutions in staphylococcal nuclease. *J Am Chem Soc*. 1990;112(8):2860-8.
49. Duarte F, Bauer P, Barrozo A, Amrein BA, Purg M, Aqvist J, et al. Force field independent metal parameters using a nonbonded dummy model. *J Phys Chem B*. 2014;118(16):4351-62. doi: 10.1021/jp501737x. PubMed PMID: 24670003; PubMed Central PMCID: PMC4180081.
50. Drinkwater N, Lee J, Yang W, Malcolm TR, McGowan S. M1 aminopeptidases as drug targets: broad applications or therapeutic niche? *FEBS J*. 2017;284(10):1473-88. doi: 10.1111/febs.14009. PubMed PMID: WOS:000401328000006.
51. Wickström M, Larsson R, Nygren P, Gullbo J. Aminopeptidase N (CD13) as a target for cancer chemotherapy. *Cancer Sci*. 2011;102(3):501-8.
52. Addlagatta A, Gay L, Matthews BW. Structure of aminopeptidase N from *Escherichia coli* suggests a compartmentalized, gated active site. *Proc Natl Acad Sci U S A*. 2006;103(36):13339-44. doi: 10.1073/pnas.0606167103. PubMed PMID: 16938892; PubMed Central PMCID: PMC1569165.
53. Chen YY, Farquhar ER, Chance MR, Palczewski K, Kiser PD. Insights into Substrate Specificity and Metal Activation of Mammalian Tetrahedral Aspartyl Aminopeptidase. *J Biol Chem*. 2012;287(16):13356-70. doi: 10.1074/jbc.M112.347518. PubMed PMID: WOS:000302903700071.
54. Helgstrand C, Hasan M, Uysal H, Haeggstrom JZ, Thunnissen MMGM. A Leukotriene A(4) Hydrolase-Related Aminopeptidase from Yeast Undergoes Induced Fit

- upon Inhibitor Binding. *J Mol Biol.* 2011;406(1):120-34. doi: 10.1016/j.jmb.2010.11.059. PubMed PMID: WOS:000287281800010.
55. Kochan G, Krojer T, Harvey D, Fischer R, Chen L, Vollmar M, et al. Crystal structures of the endoplasmic reticulum aminopeptidase-1 (ERAP1) reveal the molecular basis for N-terminal peptide trimming. *Proc Natl Acad Sci U S A.* 2011;108(19):7745-50. doi: 10.1073/pnas.1101262108. PubMed PMID: 21508329; PubMed Central PMCID: PMC3093473.
56. Stsiapanava A, Tholander F, Kumar RB, Qureshi AA, Niegowski D, Hasan M, et al. Product formation controlled by substrate dynamics in leukotriene A4 hydrolase. *Biochim Biophys Acta.* 2014;1844(2):439-46. doi: 10.1016/j.bbapap.2013.12.003. PubMed PMID: 24333438.
57. Thunnissen MM, Nordlund P, Haeggström JZ. Crystal structure of human leukotriene A4 hydrolase, a bifunctional enzyme in inflammation. *Nat Struct Mol Biol.* 2001;8(2):131-5.
58. Yang Y, Liu C, Lin YL, Li F. Structural insights into central hypertension regulation by human aminopeptidase A. *J Biol Chem.* 2013;288(35):25638-45. doi: 10.1074/jbc.M113.494955. PubMed PMID: 23888046; PubMed Central PMCID: PMC3757224.
59. Harbut MB, Velmourougane G, Dalal S, Reiss G, Whisstock JC, Onder O, et al. Bestatin-based chemical biology strategy reveals distinct roles for malaria M1- and M17-family aminopeptidases. *Proc Natl Acad Sci U S A.* 2011;108(34):E526-34. doi: 10.1073/pnas.1105601108. PubMed PMID: 21844374; PubMed Central PMCID: PMC3161592.
60. Rosati M, Dalal S, Klemba M. Two cap residues in the S1 subsite of a *Plasmodium falciparum* M1-family aminopeptidase promote broad specificity and enhance catalysis. *Mol Biochem Parasitol.* 2017;217:7-12. doi: 10.1016/j.molbiopara.2017.08.001. PubMed PMID: 28811124; PubMed Central PMCID: PMC5633526.

61. Liu Y, Tu GG, Lai XP, Kuang BH, Li SH. Exploring ligand dissociation pathways from aminopeptidase N using random acceleration molecular dynamics simulation. *J Mol Model*. 2016;22(10):236. doi: Artn 236
10.1007/S00894-016-3105-Y. PubMed PMID: WOS:000385140700004.
62. Kyrieleis OJ, Goettig P, Kiefersauer R, Huber R, Brandstetter H. Crystal structures of the tricorn interacting factor F3 from *Thermoplasma acidophilum*, a zinc aminopeptidase in three different conformations. *J Mol Biol*. 2005;349(4):787-800. doi: 10.1016/j.jmb.2005.03.070. PubMed PMID: 15893768.
63. Stamogiannos A, Maben Z, Papakyriakou A, Mpakali A, Kokkala P, Georgiadis D, et al. Critical Role of Interdomain Interactions in the Conformational Change and Catalytic Mechanism of Endoplasmic Reticulum Aminopeptidase 1. *Biochemistry*. 2017;56(10):1546-58. doi: 10.1021/acs.biochem.6b01170. PubMed PMID: 28218509.
64. Kannan Sivaraman K, Paiardini A, Sienczyk M, Ruggeri C, Oellig CA, Dalton JP, et al. Synthesis and structure-activity relationships of phosphonic arginine mimetics as inhibitors of the M1 and M17 aminopeptidases from *Plasmodium falciparum*. *J Med Chem*. 2013;56(12):5213-7. doi: 10.1021/jm4005972. PubMed PMID: 23713488.
65. Ruggeri C, Drinkwater N, Sivaraman KK, Bamert RS, McGowan S, Paiardini A. Identification and Validation of a Potent Dual Inhibitor of the *P. falciparum* M1 and M17 Aminopeptidases Using Virtual Screening. *PLoS One*. 2015;10(9):e0138957. doi: 10.1371/journal.pone.0138957. PubMed PMID: 26406322; PubMed Central PMCID: PMC4583420.
66. Dalal S, Ragheb DR, Schubot FD, Klemba M. A naturally variable residue in the S1 subsite of M1 family aminopeptidases modulates catalytic properties and promotes functional specialization. *J Biol Chem*. 2013;288(36):26004-12. doi: 10.1074/jbc.M113.465625. PubMed PMID: 23897806; PubMed Central PMCID: PMC3764805.
67. Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures.

PLoS Comput Biol. 2012;8(10):e1002708. doi: 10.1371/journal.pcbi.1002708. PubMed
PMID: 23093919; PubMed Central PMCID: PMC3475669.

Figure legends

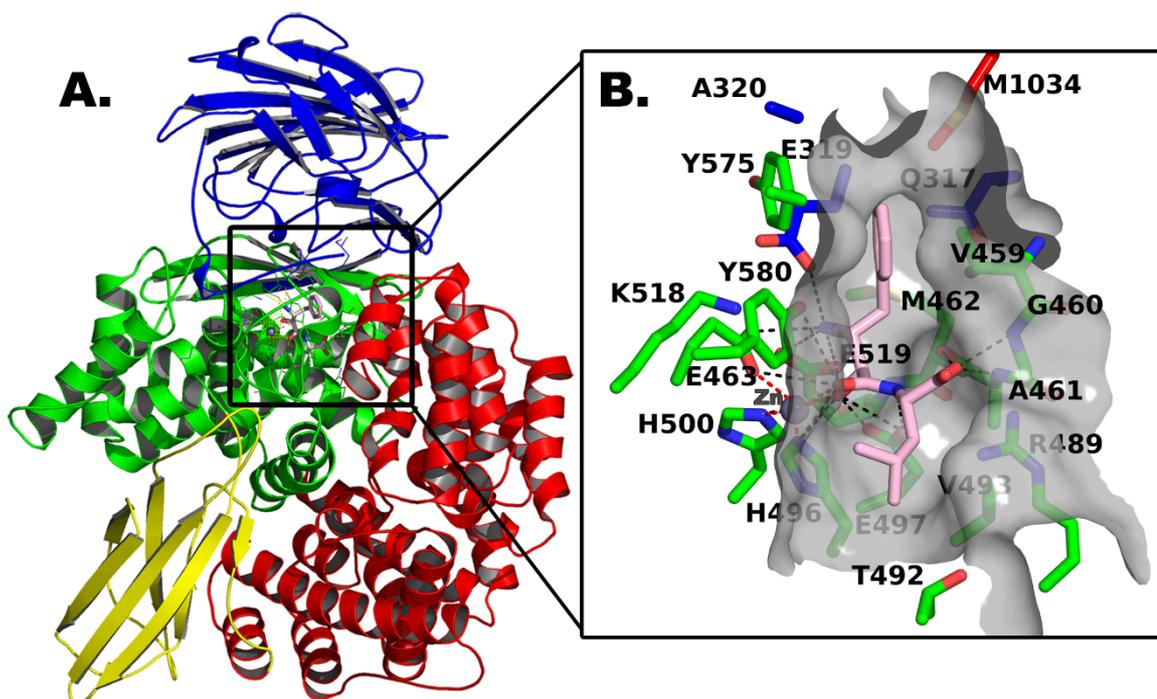


Figure 1. The crystal structure of *PfA-M1* bound to bestatin. (A) A cartoon model of the X-ray crystal structure of *PfA-M1* bound to bestatin, where the protein is colored by domains (I blue, II green, III yellow, IV red) and the zinc ion is shown as a grey sphere. **(B)** Inset zoom shows the carbon atoms of bestatin as pink sticks and *PfA-M1* carbon atoms by domain color. Hydrogen bonds are shown as black dashes and metal bonds as red dashes.

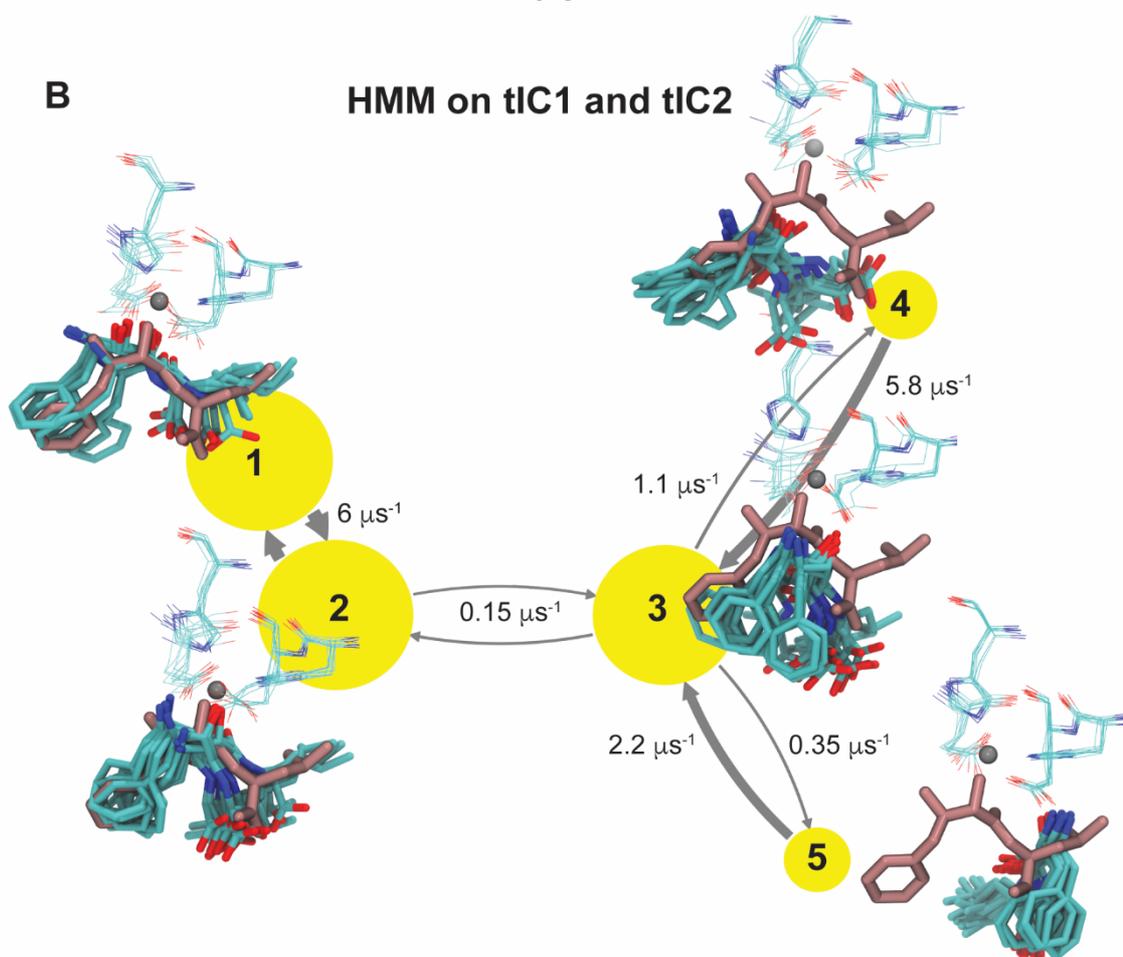
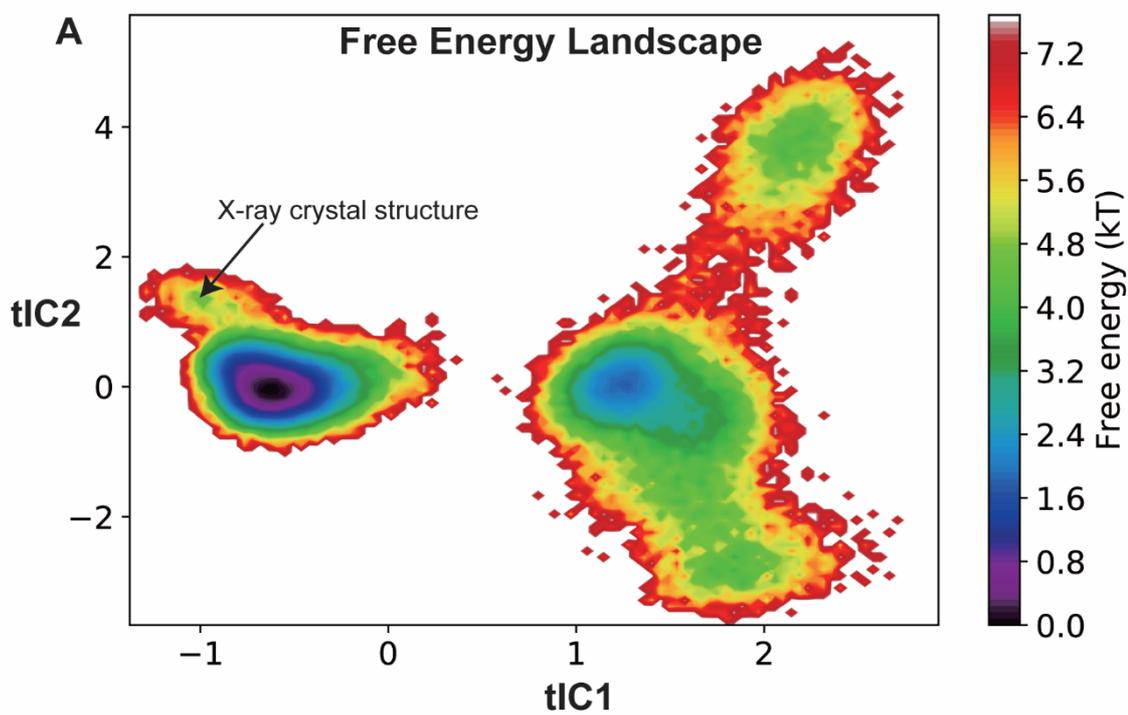


Figure 2. The free energy landscape and HMM *PfA-M1*-bestatin simulations. (A) The free energy landscape projected onto tIC1 and tIC2 as indicated on the axis. Free energy (kT) color legend is shown on the right of panel. (B) The estimated coarse-grained HMM of the bestatin diffusion process. The yellow circles represents a bestatin macrostate (numbered 1 – 5) and the diameter of the circle is scaled to be proportional to the population size. The estimated exchange macrostates were shown by the gray arrows and the thicknesses of the arrows are proportional to the exchange rates. Actual rates are indicated along arrows in text (μs). The metal center residues (H496, H500, E519) are shown in lines (representative from 10 frames automatically generated by the PCCA module of Pyemma) and the zinc ion in grey sphere. Carbon atoms of bestatin are shown in pink (X-ray crystal structure position) or cyan (representative from 10 frames from respective state). The molecular images were generated by VMD1.9.2 software (237).

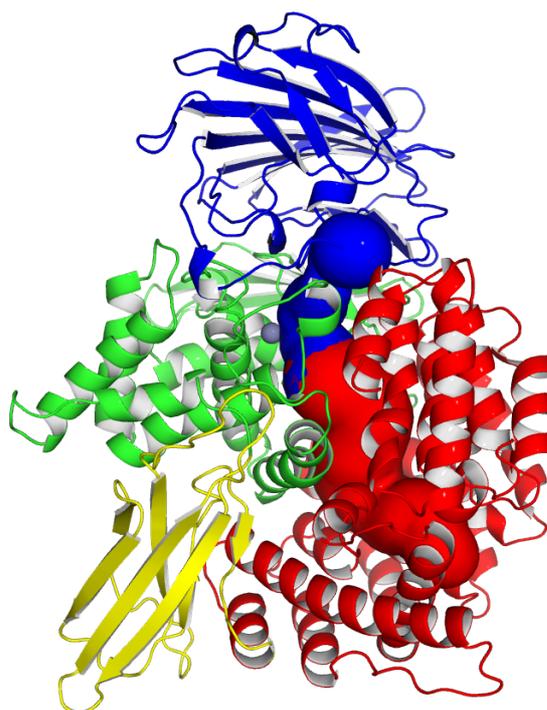
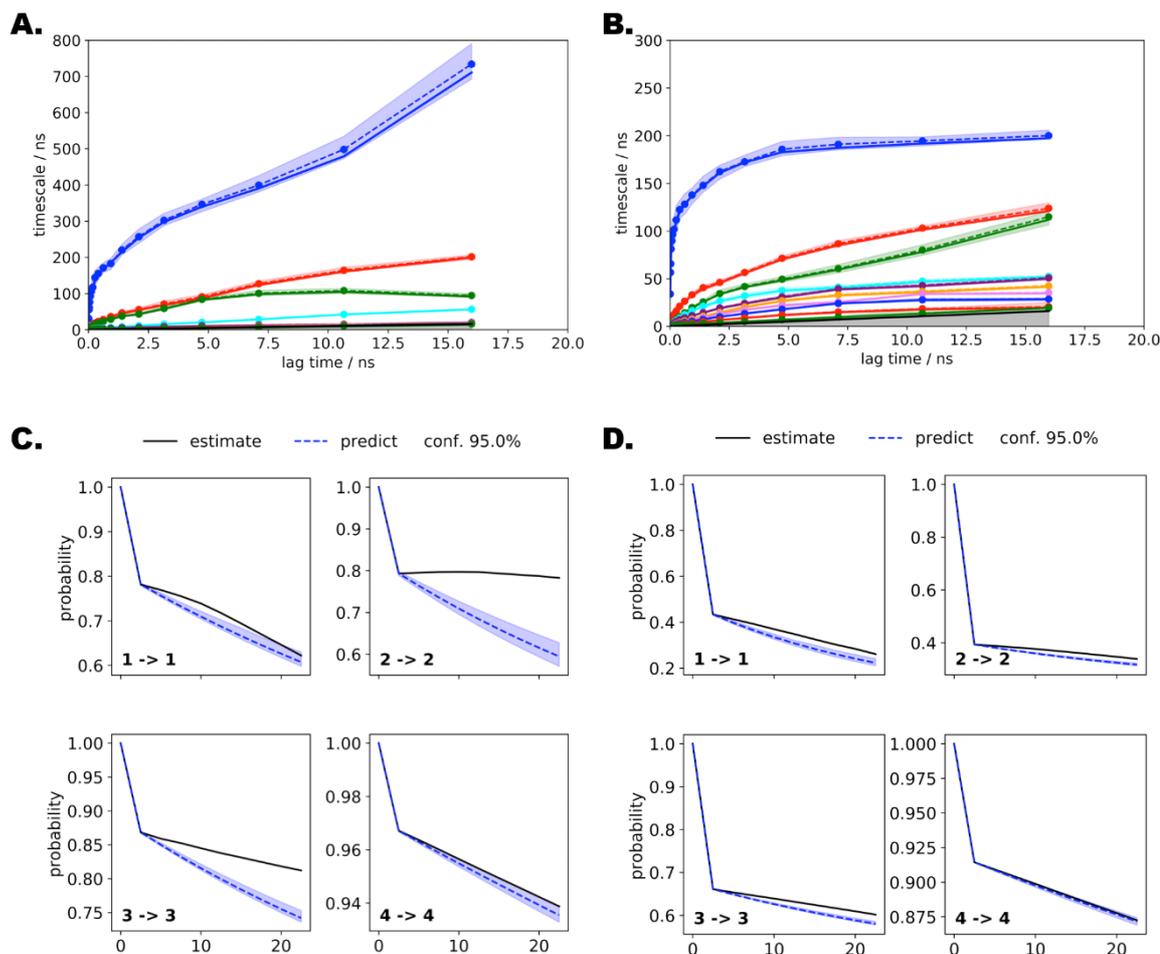
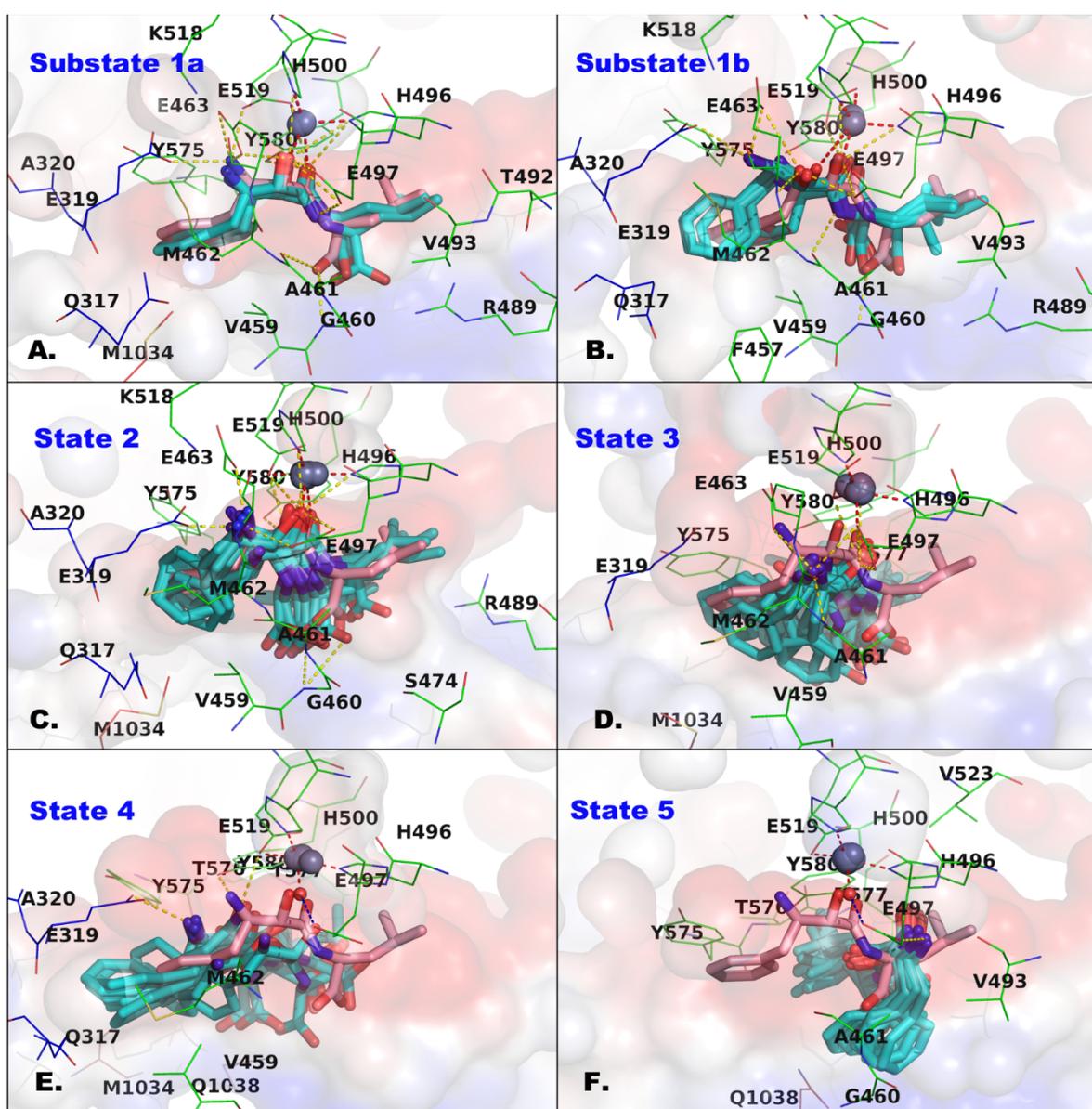


Figure 3. Cartoon model of the channels within *PfA-M1*. The protein is colored by domain. The channel formed by the junction of domains I, II and IV is shown in solid blue and the C-terminal channel in red. The image was generated using CAVER3.0.2 software (289) from 3EBH.

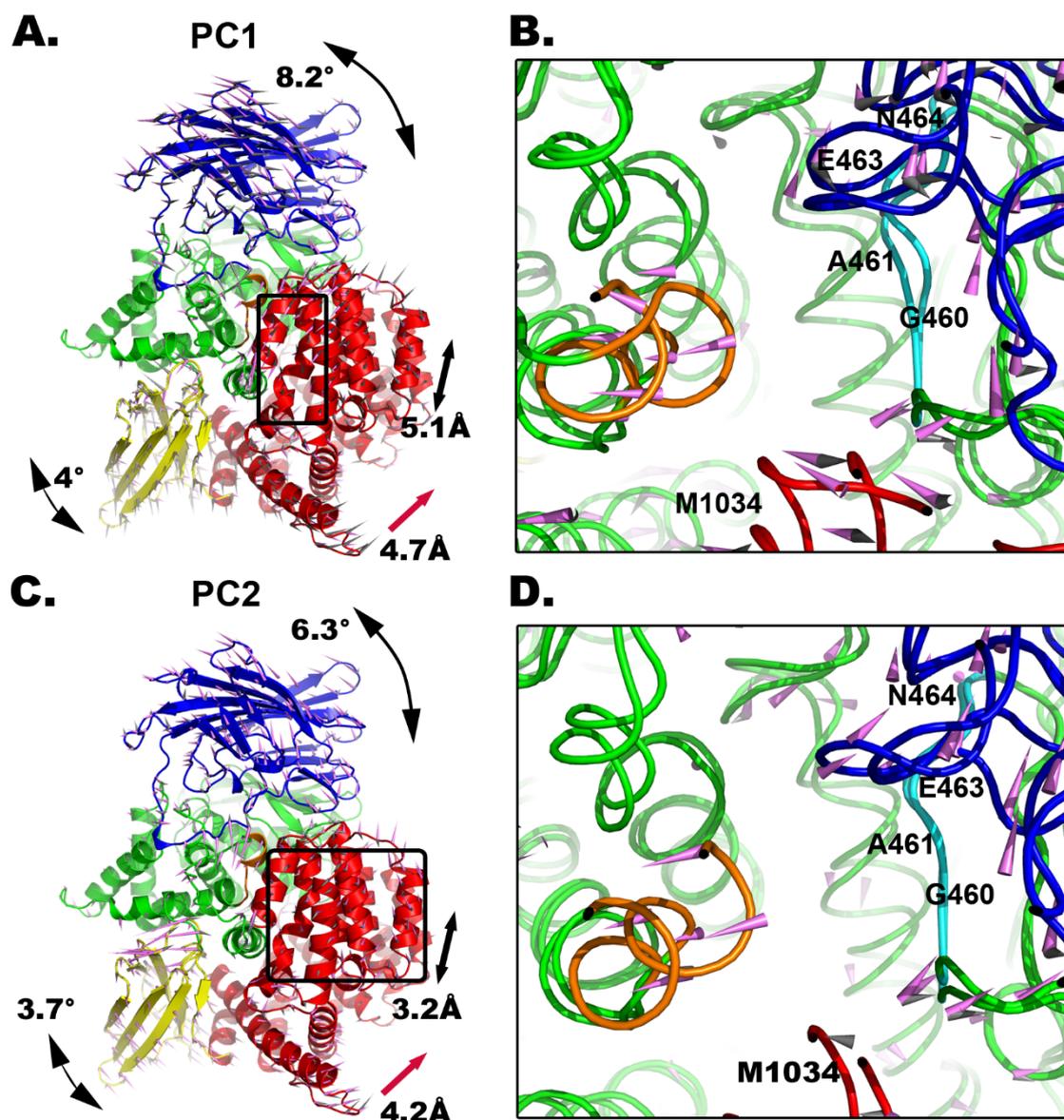
Supplementary Information



Supplementary Figure 1. Construction and validation of the MSM. Implied timescale test based on **(A)** first round MD simulations and **(B)** combined (two) rounds MD simulations. The color lines are automatically generated from the 10 representative slowest implied timescales with a range from 1 to 20 ns. The solid lines are from MD data and the dashed lines from the prediction of timescales. The shaded area is the error bars assigned by bootstrapping calculations, which represent the uncertainties in the values of prediction of MD data. The Chapman-Kolmogorov test based on **(C)** first round MD simulations and **(D)** combined (two) rounds of MD simulations. The solid black line is the prediction from MD data and dashed lines the prediction of the constructed MSM. Panels show the prediction of the given MSM model quantity for lag time $k\tau$ (1->1, 2->2, 3->3, and 4->4) as well as the comparison to an independently estimated a model at $k\tau$.



Supplementary Figure 2. The protein-ligand interactions of HMM macrostates. The interactions between bestatin (crystal structure is pink stick, HMM conformations from 10 frames are cyan sticks) and *PfA*-M1. Residues within 4.5 Å of the bestatin are shown in lines, colored by domain. The zinc ions (grey) and water molecules (red) from macrostates are shown in spheres. The internal surface of *PfA*-M1 pockets is shown by electronic charge distributions. Hydrogen contacts are in yellow dashes. Metal-bonds are in red dashes. Water bridges were in blue dashes. **(A)** Sub-state 1a. **(B)** Substate 1b. **(C)** State 2. **(D)** State 3. **(E)** State 4. **(F)** State 5.



Supporting Figure 3. The projected PC1 and PC2 movements of *PfA-M1*. Porcupine plots of PC1 (**A**) and PC2 (**C**) show the cartoon model of *PfA-M1* colored by domain. Pink arrows depict positive PC directions and gray arrows negative PC direction. Arrows are proportional to size of movement. Overall size of domain movements indicated by arrows and measurements. In the inset zoom boxes, the details of the S1 pocket expansion in both PC1 (**B**) and PC2 (**D**). In orange, the starting position of the S1 helix shows that it moves outward, making the pocket bigger. In cyan, shows the movement of the GAMEN residues.

Supplementary Table 1: Metal and H-bonding network present in HMM macrostates where grey shading indicates the bond is present. Dark grey indicates 100% occupancy, while light grey indicates 30% occupancy.

HMM state Bond Info.	1		2	3	4	5
	1a	1b				
Metal Bonds						
H496 _{NE} -ZN						
H500 _{NE} -ZN						
E519 _{OE} -ZN						
Bes _{O2} -ZN						
Bes _{O3} -ZN			30%			
E497 _{OE} -ZN						
Water _O -ZN						
Hydrogen bonds						
Bes _{N2} -E519 _{OE}						
Bes _{N2} -E463 _{OE}						
Bes _{N2} -E319 _{OE}						
Bes _{N2} -E497 _{OE}						
Bes _{O2} -H500 _{NE}						
Bes _{O2} -H496 _{NE}						
Bes _{O2} -E497 _{OE}						
Bes _{O2} -E463 _{OE}						
Bes _{O2} -E463 _N						
Bes _{O2} -E519 _{OE}						
Bes _{O2} -Y580 _{OH}						
Bes _{O3} -Y580 _{OH}						
Bes _{O3} -H496 _{NE}						
Bes _{O3} -E497 _{OE}						
Bes _{O3} -H500 _{NE}						
Bes _{O3} -E519 _{OE}						
Bes _{N1} -A461 _O						
Bes _{N1} -E497 _{OE}						
Bes _{O1/O4} -A461 _O						
Bes _{O1/O4} -A461 _N						
Bes _{O1/O4} -G460 _N						

Hydrogen bonds defined as $3.0 \text{ \AA} \leq \text{distance between donor and receptor} \leq 3.5 \text{ \AA}$.

Residues = GAMEN

CHAPTER EIGHT - DISCUSSION

Members of the M1 and M17 aminopeptidase superfamilies play crucial roles in the cellular metabolism. Inhibition of their activity can modulate chronic and infectious diseases, for example modulation of hyper tension, control of tumor growth and treatment of malaria (4, 98, 104, 125, 126, 131). The development of specific, potent inhibitors of the two types of protease offers a new route to treatment for such diseases (4, 98, 104, 125, 126, 131). The ability to produce selective and potent inhibitors for different members of a superfamily requires precise knowledge of the mechanism of action, architecture of the target and their potential dynamics, key data that can be exploited in future inhibitor design.

The *PfA*-M1 and -M17 proteins are attractive new antimalarial drug targets. The protein dynamics that relate to their enzymatic function remain unknown. This knowledge is important to broaden our understanding of the protein superfamilies and may contribute to the design of new inhibitors / drugs. With the new parameters for the MD simulation of M1 and M17 aminopeptidases designed and implemented in this research, I performed classical MD simulations on seven M1 aminopeptidase enzymes (including *PfA*-M1) and *PfA*-M17. These studies shed light on the intramolecular pathways that the substrate may take to enter the active sites of *PfA*-M1 and -M17 and address the evolution of protein dynamics within the M1 aminopeptidases. In light of the different protein dynamics identified from this thesis, the results are important new target residues for the production of potent and selective inhibitors to target *PfA*-M1.

8.1 The hybrid FF for the simulation of M1 and M17 aminopeptidases.

MD simulation of the M1 and M17 aminopeptidases by a classical MD approach meant that I had to tackle one of the toughest problems in classical MD, how to accurately parameterize the metal ions (197, 199, 201, 249, 290, 291). Most classical FFs used in MD simulations describe only covalent bonds and electrostatic interactions, and lack the ability to correctly model the dynamic polarization of coordinated metal centers (292). Three main strategies have been employed to generate models of metal-containing sites for simulations including “nonbonded” (197), “semi-bonded” (293) and “bonded” (196) models. Whereas nonbonded models use non-covalent interactions to define metal centers (294), the semi-

bonded model places dummy atoms around metals to mimic the electron orbitals (204). In contrast, the recent Zinc Amber Force Field (ZAFF), an extension to the AMBER FF for simulating zinc-containing proteins, uses a bonded model to define the metal-ligand bonds (201). This model extracts artificial metal-bond constants (angles, bond fluctuations) and partial charge distribution for the metal centre as defined by the required quantum mechanics (QM) single point and frequency calculations (201). The accuracy of the FF can be determined by comparing the normal mode (NM) frequencies of molecular mechanics (MM) to the QM frequencies (201). The ZAFF describes 12 unique zinc-containing proteins with a bonded-model, but did not include the M1 and M17 aminopeptidase metal centres.

In this study, I considered the advantages and disadvantages of the three models available and chose to use a hybrid bonded/non-bonded approach to generate unique parameters to accurately model the zinc-protein interactions, while leaving all other ligands unrestrained. The hybrid bonded/non-bonded approach (recently described in (292)) allows us to more accurately study the behavior of weakly-bound small-molecule inhibitors, as well as the exchange of the nucleophilic water molecule(s) in the active sites of both enzymes, providing the basis for a more accurate depiction of the catalytic mechanism of M1 and M17 aminopeptidases. When a M1 or M17 aminopeptidase is simulated bound to ligand, there is likely to be a small cost to the accuracy of the metal center geometry due to the non-polarized metal center. The benefits of my hybrid FF include that the metal center geometry is maintained throughout the simulation and that the FF can be used to do comparative studies of unbound and ligand-bound targets.

The hybrid FF parameters are appropriate for the tetrahedral metal geometry in M1 and zinc 1 in the M17 aminopeptidase active sites. In the evaluation studies of our FF parameters, two waters were identified in the second shell of Harding's bond cut-off both in M1 (2.1 Å, occupancy of 100% and 2.6 Å, occupancy of 90%) and zinc position 1 in M17 (2.1 Å, occupancy of 100% and 2.7 Å, occupancy of 80%) (212). This finding suggests that the hybrid FF parameters can simulate both tetrahedral and pentahedral geometry of zinc binding area but may show a very slight bias for retention of tetrahedral geometry. However, neither the bonded model nor semi-bonded model is able to simulate the two situations

within one simulation, and supports that my decision to produce a hybrid FF is the most appropriate for this biological system.

8.2. The protein dynamics of the clan MA, M1 aminopeptidase superfamily.

The protein dynamics of seven selected M1 aminopeptidases were studied in this thesis. The results show that all M1 aminopeptidases likely have the potential to sample both an open and closed conformation. In my study, I performed MD simulations of between 200 – 600 ns and in all systems and all replicates, I observed that only *PfA*-M1 went from a closed to open conformation in one trajectory. However, the simulations of the open form M1 aminopeptidases (*TIFF3*, *hERAP1*) showed that the proteins rapidly closed, reaching a closed conformation within 20 - 100 ns. This result indicates that open state of M1 is energetically unfavorable and may be sampled rarely. Future drug development should therefore still focus more on obtaining high affinity to the closed form M1 aminopeptidases or looking for inhibitors for the twisted closed form M1 in order to prevent M1 from opening again. Overall, my MD simulation of the M1 aminopeptidases shows that it is likely that simulation of their opening and closing events would require MD simulations of longer time-scales, likely microseconds to even seconds. Enhanced sampling and especially biased enhanced sampling might be helpful to access meaningful comparison to biological timescales. This could be achieved using next generation GPU accelerated supercomputer clusters in the future. Other factors that may influence the opening and closing might include physiological changes or concentration of the substrate present in the local environment.

To compare the various M1 aminopeptidase representatives I studied in this thesis, I produced a conserved M1 aminopeptidase 'core' structure (Fig 6.1). The analysis of the dynamics of the M1 aminopeptidase core structure indicated the highly conserved catalytic domain II was rigid and showed the least dynamics. Domain II has the highest sequence identity amongst the M1 aminopeptidases. The dynamics of domain I showed a common rotation and hinge motion. The scale of this motion in *TIFF3* was larger than in the other representatives, likely due to the fact that domain I is more loosely packed. The loosely packed domain I can be explained by the evolutionary trend of M1, which tends to form

larger domains I and IV with more α -helices and β -sheets for higher organisms but smaller domains in the more ancient types. The fact that TIFF3 exists in high temperature and low pH may also be related to its more loosely packed structure. The loose packing of TIFF3 might be a built-in strategy for the protein to make conformational changes more easily under such conditions.

In the bacteria and parasites, the β -sheets of domain III were larger than in the mammalian enzymes and found to be more mobile. They are 2+ more β -strands that are each 4 amino acids longer in each of the major β -sheets, and have similar flexible loop connections to domain II and IV to that of the higher organisms. This might account for the scale of domain III movement being larger in lower organisms than in the higher organisms. The dynamics of domain III might be an important aspect to help the lower organisms undergo the conformational change to open. The lower organisms appear to rely on rigid body movement of domain IV to open, which is different to the higher organisms. It is possible that the larger domain III swing helps to pivot domain IV in bacteria and parasites.

Domain IV shows complexity in its dynamics that appears related to the packing of the domain IV helices. Domain IV can be effectively divided into two modules, module I (α 14-19) and module II (α 20-30). Module I in *EcAPN* and *PfA-M1* is larger and contains the long α 19 that is easily identified in all representatives studied here. The packing of the module II helices 20-30 in the bacterial and parasite enzymes was also found to be tighter than the higher organisms and the archae. For example, in *EcAPN* and *PfA-M1*, α 18 is close to domain II, while the corresponding helix in the higher organisms is 5 amino acids longer in length and maintains a distance of 12 Å to domain II. Module I of the bacterial and parasite enzymes moves in concert with the swing movement of domain III. Whilst the opening of *PfA-M1* relies on the movement of the entire of domain IV, the higher organisms rely more on the movement between the two modules in domain IV. Therefore, the evolutionary trends of the M1 aminopeptidase appear to show that the formation of smaller β -sheets in domain III coupled with less contacts between module I and domain II results in a protein that can open more easily.

The opening and closing event of the M1 aminopeptidases must involve changes to interactions between residues from domain II and IV. Investigation into these interactions showed differences in the electrostatic potential of the interactions between the various species. In the mammalian and TIFF3 enzymes that open more readily, the interactions between domains II and IV are mostly hydrophobic. In the lower organisms, the interactions are hydrophilic. The interactions between domains II and IV likely control or contribute to the open and closed states and need to change to allow the protein to open and close. Therefore, more hydrophilic interactions between the two domains is likely to keep the interactions intact, keeping the protein in a closed state. Hydrophobic interactions might decrease the attraction between the two domains and allow the protein to open more easily. This finding might explain why M1 aminopeptidases in the higher organisms are more readily able to sample the open conformation. Additionally, for TIFF3, domain IV is loosely packed with shorter α helices but longer loops, which also favors a ready switch between conformations.

One important caveat of my study is that the enzymes examined in this thesis were investigated in their monomeric form, however, for the mammalian enzymes, there is evidence of functional dimers (in addition to monomers)(147). Analysis of the dimer interface of *hAPA*, *hERAP1*, *hAPN* and *pAPN* suggests that the dimer interactions are conserved and mediated primarily at the solvent exposed pore of domain IV (Fig 6.11). Dimers have never been reported in bacterial or parasite M1 aminopeptidases and inspection of their domain IV shows that their longer α 19 likely prevents formation of a dimer as observed for the mammalian species (*hAPN*, *pAPN*, and *hAPA*). In *EcAPN* and *PfA-M1*, the protrusion of the long α 19 likely prevents the dimer formation. TIFF3 has short α -helices in domain IV and does not share the conserved dimer interactions with higher organisms, therefore, TIFF3 is also less likely to form dimers. Conserved glutamine residues were identified at the dimer interface of *hAPN*, *hAPA* and *pAPN*, and analysis of their RMSF throughout simulation showed little turbulence (RMSFs between 0.4 – 1.1 Å). This suggests that the motion observed in my simulation may not interfere with a dimer interaction.

8.3. Insight into the dynamics in *PfA-M1* can guide future antimalarial drug design.

PfA-M1 is an essential hemolysin in *P. falciparum*. Its ectodomain consists of four separate domains with the active site located in the catalytic domain II. Prior to this study, no open form of *PfA-M1* had been observed and it was assumed to not sample an open conformation (4). This meant that the active site was effectively buried from bulk solvent and asked the question of how a peptide substrate and small molecule inhibitors accessed the active site (4). Studies performed in this thesis aimed to shed light on possible intramolecular routes that peptide substrates or inhibitors may take to access the active site. A combination of computational biology, including MD and SMD, as well as experimental biology was used to (a) show that the design of small molecule inhibitors need to consider the dynamics of the substrate binding pockets, (b) show that a six amino acid hemoglobin derived peptide can be cleaved by *PfA-M1*; (c) show that despite earlier evidence, *PfA-M1* can indeed sample an open conformation.

To investigate potential ligand dynamics and diffusion in and around the active site at more relevant biological timescales, I chose to use bestatin as a model system with which to study the binding dynamics between *PfA-M1* and small molecule inhibitors. The dissociation of bestatin within *PfA-M1* pocket was interrogated by a series of MD simulations, producing a total simulation time of 2.3 μ s. This sample size provided a large ensemble of conformations (~230,000 snapshots) that I used to perform time-structure independent component analysis (tICA). The tICA permitted the generation of a Markov state model (MSM) to identify conformational states of protein and ligand, and transitions between these states. The obtained HMM states also provided a model for the association / dissociation of bestatin from *PfA-M1*. The results of this study show that a change in size of the S1 pocket has profound effects on the position of a P1 inhibitor moiety and that the dynamics of the S1 pocket also appear to dictate inhibitor egress from the active site (and possibly entrance). I identified key residues to target to improve potency by stabilizing both the inhibitor and target. The individual dynamics and control of substrate movement could be exploited to produce selective inhibitors that can differentiate between similar members of the M1 aminopeptidase superfamily.

The study on the dissociation of bestatin from *PfA-M1* arose from a total of 2.3 μ s of MD simulation. During all of this time, the dynamics of the protein were found to be very stable and maintained a closed conformation. In contrast, my other study of the unbound *PfA-M1*, that provided a total of 1.5 μ s, identified an open conformation of the protein. The open form of *PfA-M1* was similar to the published open crystal structures of *hERAP1*, *pAPN* and *TIFF3* (1, 149, 216) (Fig 3.10). Analysis of the free energy of the trajectories suggests that there is a significant energy barrier (4.53 kT) between the open and closed form of *PfA-M1*. The closed form of *PfA-M1* has a low free energy and this barrier is the likely reason that the open form of wild type *PfA-M1* has not been captured in crystallographic studies to date.

To provide experimental validation of the computational result, I attempted to identify the 'hinge' residues that were involved in the motion to open *PfA-M1*. Potential hinge residues were mutated and mutant proteins analyzed for structure and function. The mutants produced in this study showed compromised aminopeptidase activity, however, their crystal structures provided no direct evidence for any conformational change. In a separate study, mutagenesis of residues involved in a key salt bridge between domains II and IV did result in the resolution of a structure of *PfA-M1* that was in the open conformation (McGowan lab, unpublished). This mutant, *PfA-M1* (E319P), was catalytically inactive and the structure obtained was of low resolution (3.5 Å). Key elements of the active site (e.g. GAMEN motif) were missing in electron density. This is consistent with other studies of the open forms of M1 aminopeptidases that show that the active site / substrate pockets are poorly organized in the open conformation (149, 295). Comparison of the structure of *PfA-M1* (E319P) and my MD model showed that both open crystal structure and the open form in simulation data rotates domain I up and leaving huge gap between domain I and domain IV, however, the open *PfA-M1* crystal structure opened a gap between domains II and IV as observed in other open form crystals, while the simulation structure did not sample this conformation (1, 149, 216). The simulation data is in a less open form compared to the crystal structure. This suggests that simulations need a longer time scale to sample the complete dynamics in relation to the opening between domains II and IV.

The identification of an open form of *PfA-M1* changes the way we would view substrate access to the active site. The intramolecular routes that have been proposed from the bestatin study may still be possible. Although my *PfA-M1*-bestatin simulations still needs QM/MM approaches to evaluate the possibility of bestatin escape routes, the long classical simulations indicate that small inhibitors such as bestatin might stop M1 from opening. Moreover, it is now far more likely that the substrate and inhibitors access the active site whilst the protein is open. The protein then closes, forming the catalytic machinery required for proteolysis. This asks the question of the role of the C-terminal domain IV pore that is of significant size and was previously proposed to be the route of entry for peptide substrates (4). In my study, I showed that *PfA-M1* was able to cleave a six amino acid peptide called Hb6. This is the first time a peptide of this size has been shown to be cleaved by *PfA-M1*. I was unable to produce the co-crystal structure of *PfA-M1* bound to Hb6, despite production of catalytic mutants and extensive crystallographic trials. I therefore modelled the substrate into the protein and analyzed these models. The modelled *PfA-M1*-Hb6 suggests that residues P₃ - F₆ of Hb6 are located in the domain IV pore / channel. This position is consistent with the position of substrate in *pAPN*, *hAPN* and *hERAP1* (1, 144, 216). *hERAP1* is able to process very long peptide substrates (up to 16 residues with a preference for > 8) and the domain IV has been proposed to correctly position the long tail (216). A SMD study supported the hypothesis that the most likely route of substrate entrance to the active site is directly through the open form.

To date, a detailed study of substrate pockets in *PfA-M1* has been limited to only the P1 and P1' pockets. In this study, I propose the position of substrate pockets P2', P3', P4' and P5', where residues A468, T492, V493, R489, S542, and H543 are involved in forming these pockets. However, future experiments are needed to work out the maximum length of the peptide that *PfA-M1* can cut and if the activity is processive or not. These new substrate pockets provide more detailed knowledge of the activity of *PfA-M1* and potential new druggable sites that could be exploited in future inhibitor design.

The culmination of my studies into *PfA-M1* have proposed a new mechanism of action for the protease. The substrate likely enters the active site through an open form of

PfA-M1 and is positioned correctly by the movement of S1 pocket helix (residues 570-575). The move to an overall closed form to stabilize the binding of substrate, and Y580 helps to locate the peptide bond in the right place for proteolysis. Here, I postulate that the energy released from hydrolysis might aid the protein to overcome the energy barrier to open again for continuation of the process. The new fundamental mechanistic data produced in this thesis has implications for all M1 aminopeptidases and will be useful for future rational drug design to produce not only new antimalarials but other inhibitors targeting other members of the M1 aminopeptidase superfamily.

8.4 The protein dynamics of *PfA-M17*, implications for the M17 aminopeptidase superfamily.

The family of hexameric M17 aminopeptidases are conserved throughout all kingdoms of life, and possess a characteristic homohexameric arrangement described as a dimer of trimers. In *P. falciparum*, *PfA-M17* is postulated to have a role in Hb digestion as well as essential house-keeping functions (96). To understand the mechanistic implications of the hexameric assembly, I undertook a thorough investigation of the dynamics of *PfA-M17*, and how those motions contribute to its function. The comparison of the metal center behavior between the monomer and hexamer simulation sets provided evidence that the hexamer stabilizes the active site and maintains the position of catalytic water, which is pivotal to the enzymatic function.

The original crystal structure of *PfA-M17* showed six channels to the central catalytic cavity and identified a ~20 Å flexible loop (residues 246-265, L9) that sits at the entrance to each of the channels (81). Initial modelling and energy minimisation of the starting loop conformations resulted in almost complete occlusion of the channel entrances. During the simulation of the hexamer, changes to the size and shape of these possible entrance pores was observed. Interestingly, the three pores at the front trimer were observed to open to a larger diameter, while the other three pores in the back trimer maintained their original pore sizes. The loop dynamics sampled during this movement supported previous hypotheses

that the L13 is very flexible. Moreover, it may be that substrate movement is controlled by the trimers in that the substrate may enter on one trimer face only.

Analysis of the active site of *PfA-M17* identified a novel loop movement that links the active site of the trimers. The detailed movement within the catalytic domain were analyzed. The catalytic loop L13 of subunit A² moved to the catalytic center of subunit A, and significant rearrangement of $\alpha 10$ and $\alpha 24$ in subunit A was required to facilitate such movement of L13. Experimental validation of this catalytic loop movement was provided by an X-ray crystal structure that had a similar conformation with occluded active sites due to the movement of this loop (McGowan laboratory, unpublished data). Comparison of my MD model and the inactive crystal structure showed that the results were consistent. The major difference observed was that the movement of the loop in the crystal structure was observed in all chains of the hexamer, whereas only one chain in the MD showed a similar movement. The reason for the discrepancy between the MD and the structural data likely results from the metal coordination in the active site and the limitations of our MD FF. The crystal structure showed that the zinc ions were rearranged and that a third and novel zinc position was observed. The rearrangement was necessary to facilitate the change in backbone structure that occurred when L13 moved. The nature of my MD simulation, specifically my hybrid 'bonded / non-bonded model' that was used to manage the quantum mechanics of cation simulation, required us to fix the positions of the active site zinc ions (212). This is a major limitation of using classical MD simulations for metallo-proteins as we cannot simulate the metals changing position / coordination. However, despite this significant limitation, the simulated backbone motions are remarkably similar to those observed in all active sites of the crystal structure. Future work might rely on the use of hybrid QM/MM simulation and increase the simulation length to micro seconds level to solve this problem.

Characterization of the function, and regulation of that function, described in this thesis for *PfA-M17* has implications for understanding how the large family of M17 enzymes functions, and provides insight into the vast range of functions attributed to the large and important enzyme family. The overall contribution of my study to the family of M17 aminopeptidases tells us that the arrangement protects the six active sites contained within

the central catalytic cavity and regulates access to the central cavity by key loops that flex to open and close access channels. A flexible loop links the active sites of each trimer to its neighbor. It is possible that this loop acts as a key regulatory mechanism to control the activity of *PfA-M17* in different environmental conditions.

8.5 CONCLUSIONS

My findings suggest that all M1 aminopeptidase enzymes can occupy an open or closed conformation, however, the open form of M1 aminopeptidase is more unstable and possibly sampled less in bacteria and parasites. The model complex structure of *PfA-M1-Hb6* provides predictions as to the potential substrate binding pockets that have not previously been explored and maps the protein dynamics within *PfA-M1* that are important to ligand binding. The study of the dynamics in *PfA-M17* indicate that the trimer communications are important for both substrate guidance and the catalytic mechanism, and that the hexamer is important to stabilize the catalytic machinery.

REFERENCES

1. Wong AH, Zhou D, Rini JM. The X-ray crystal structure of human aminopeptidase N reveals a novel dimer and the basis for peptide processing. *J Biol Chem.* 2012;287(44):36804-13.
2. Skinner-Adams TS, Stack CM, Trenholme KR, Brown CL, Grembecka J, Lowther J, et al. Plasmodium falciparum neutral aminopeptidases: new targets for anti-malarials. *Trends in biochemical sciences.* 2010;35(1):53-61.
3. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 2016;44(W1):W344-50.
4. McGowan S, Porter CJ, Lowther J, Stack CM, Golding SJ, Skinner-Adams TS, et al. Structural basis for the inhibition of the essential Plasmodium falciparum M1 neutral aminopeptidase. *Proceedings of the National Academy of Sciences of the United States of America.* 2009;106(8):2537-42.
5. Kochan G, Krojer T, Harvey D, Fischer R, Chen L, Vollmar M, et al. Crystal structures of the endoplasmic reticulum aminopeptidase-1 (ERAP1) reveal the molecular basis for N-terminal peptide trimming. *Proceedings of the National Academy of Sciences of the United States of America.* 2011;108(19):7745-50.
6. Thunnissen MM, Nordlund P, Haeggström JZ. Crystal structure of human leukotriene A4 hydrolase, a bifunctional enzyme in inflammation. *Nature Structural & Molecular Biology.* 2001;8(2):131-5.
7. Drinkwater N, Lee J, Yang W, Malcolm TR, McGowan S. M1 aminopeptidases as drug targets: broad applications or therapeutic niche? *Febs J.* 2017;284(10):1473-88.
8. Reiter P. From Shakespeare to Defoe: malaria in England in the Little Ice Age. *Emerging infectious diseases.* 2000;6(1):1-11.
9. Organization WH. World malaria report 2015.: World Health Organization; 2015.
10. Bartoloni A, Zammarchi L. Clinical aspects of uncomplicated and severe malaria. *Mediterranean journal of hematology and infectious diseases.* 2012;4(1):e2012026.
11. Athuman M, Kabanywany AM, Rohwer AC. Intermittent preventive antimalarial treatment for children with anaemia. *The Cochrane database of systematic reviews.* 2015;1:CD010767.
12. Nadjm B, Behrens RH. Malaria: an update for physicians. *Infectious disease clinics of North America.* 2012;26(2):243-59.
13. Mueller I, Zimmerman PA, Reeder JC. Plasmodium malariae and Plasmodium ovale--the "bashful" malaria parasites. *Trends in parasitology.* 2007;23(6):278-83.
14. Collins WE. Plasmodium knowlesi: a malaria parasite of monkeys and humans. *Annual review of entomology.* 2012;57:107-21.
15. Nosten F, Rogerson SJ, Beeson JG, McGready R, Mutabingwa TK, Brabin B. Malaria in pregnancy and the endemicity spectrum: what can we learn? *Trends in parasitology.* 2004;20(9):425-32.
16. Dev V, Phookan S, Sharma VP, Dash AP, Anand SP. Malaria parasite burden and treatment seeking behavior in ethnic communities of Assam, Northeastern India. *The Journal of infection.* 2006;52(2):131-9.
17. Wipasa J, Elliott S, Xu H, Good MF. Immunity to asexual blood stage malaria and vaccine approaches. *Immunology and cell biology.* 2002;80(5):401-14.
18. Miller LH, Baruch DI, Marsh K, Doumbo OK. The pathogenic basis of malaria. *Nature.* 2002;415(6872):673-9.
19. Kaufman TS, Ruveda EA. The quest for quinine: those who won the battles and those who won the war. *Angew Chem Int Edit.* 2005;44(6):854-85.
20. Kyle RA, Shampe MA. Discoverers of quinine. *Jama.* 1974;229(4):462.
21. Jacquerioz FA, Croft AM. Drugs for preventing malaria in travellers. *The Cochrane database of systematic reviews.* 2009;4(4):CD006491.
22. Uhlemann A-C, Krishna S. Antimalarial multi-drug resistance in Asia: mechanisms and assessment. *Malaria: Drugs, Disease and Post-genomic Biology: Springer; 2005. p. 39-53.*

23. Miller LH, Su X. Artemisinin: discovery from the Chinese herbal garden. *Cell*. 2011;146(6):855-8.
24. Hsu E. Reflections on the 'discovery' of the antimalarial qinghao. *Brit J Clin Pharmacol*. 2006;61(6):666-70.
25. Organization WH. World Malaria Report 2015. WHO, Geneva. 2015.
26. Duffy PE, Mutabingwa TK. Artemisinin combination therapies. *Lancet*. 2006;367(9528):2037-9.
27. Davis TM, Karunajeewa HA, Ilett KF. Artemisinin-based combination therapies for uncomplicated malaria. *The Medical journal of Australia*. 2005;182(4):181-5.
28. Reyburn H. New WHO guidelines for the treatment of malaria. *Bmj*. 2010;340:c2637.
29. Howitt P, Darzi A, Yang GZ, Ashrafian H, Atun R, Barlow J, et al. Technologies for global health. *Lancet*. 2012;380(9840):507-35.
30. Malenga G, Palmer A, Staedke S, Kazadi W, Mutabingwa T, Ansah E, et al. Antimalarial treatment with artemisinin combination therapy in Africa. *Bmj*. 2005;331(7519):706-7.
31. Kokwaro G. Ongoing challenges in the management of malaria. *Malaria journal*. 2009;8 Suppl 1(Suppl 1):S2.
32. O'Brien C, Henrich PP, Passi N, Fidock DA. Recent clinical and molecular insights into emerging artemisinin resistance in *Plasmodium falciparum*. *Current opinion in infectious diseases*. 2011;24(6):570.
33. Fairhurst RM, Nayyar GM, Breman JG, Hallett R, Vennerstrom JL, Duong S, et al. Artemisinin-resistant malaria: research challenges, opportunities, and public health implications. *The American journal of tropical medicine and hygiene*. 2012;87(2):231-41.
34. Rogers WO, Sem R, Tero T, Chim P, Lim P, Muth S, et al. Failure of artesunate-mefloquine combination therapy for uncomplicated *Plasmodium falciparum* malaria in southern Cambodia. *Malaria journal*. 2009;8(1):10.
35. Noedl H, Se Y, Schaecher K, Smith BL, Socheat D, Fukuda MM, et al. Evidence of artemisinin-resistant malaria in western Cambodia. *The New England journal of medicine*. 2008;359(24):2619-20.
36. Neurath H, Walsh KA. Role of proteolytic enzymes in biological regulation (a review). *Proceedings of the National Academy of Sciences of the United States of America*. 1976;73(11):3825-32.
37. Bergmann M, Ross WF. On proteolytic enzymes X. The enzymes of papain and their activation. *J Biol Chem*. 1936;114(3):717-26.
38. Schechter I, Berger A. On the size of the active site in proteases. I. Papain. *Biochemical and biophysical research communications*. 1967;27(2):157-62.
39. Schechter I, Berger A. On the active site of proteases. III. Mapping the active site of papain; specific peptide inhibitors of papain. *Biochemical and biophysical research communications*. 1968;32(5):898-902.
40. Bond JS, Barrett AJ. Degradation of fructose-1,6-bisphosphate aldolase by cathepsin B. *The Biochemical journal*. 1980;189(1):17-25.
41. McKay MJ, Offermann MK, Barrett AJ, Bond JS. Action of human liver cathepsin B on the oxidized insulin B chain. *The Biochemical journal*. 1983;213(2):467-71.
42. Barrett AJ, McDonald JK. Nomenclature: protease, proteinase and peptidase. *The Biochemical journal*. 1986;237(3):935.
43. Bergkvist R. The proteolytic enzymes of *Aspergillus oryzae*. *Acta chem scand*. 1963;17(0):8.
44. Oda K. New families of carboxyl peptidases: serine-carboxyl peptidases and glutamic peptidases. *Journal of biochemistry*. 2012;151(1):13-25.
45. Rawlings ND, Barrett AJ, Bateman A. Asparagine peptide lyases: a seventh catalytic type of proteolytic enzymes. *J Biol Chem*. 2011;286(44):38321-8.
46. Rawlings ND, Barrett AJ, Finn R. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic acids research*. 2015:gkv1118.
47. Koonin EV, Makarova KS, Rogozin IB, Davidovic L, Letellier M-C, Pellegrini L. The rhomboids: a nearly ubiquitous family of intramembrane serine proteases that probably evolved by multiple ancient horizontal gene transfers. *Genome Biol*. 2003;4(3):1.

48. Lemberg MK, Freeman M. Functional and evolutionary implications of enhanced genomic analysis of rhomboid intramembrane proteases. *Genome research*. 2007;17(11):1634-46.
49. Brown MS, Ye J, Rawson RB, Goldstein JL. Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. *Cell*. 2000;100(4):391-8.
50. Lewis AP, Thomas PJ. A novel clan of zinc metallopeptidases with possible intramembrane cleavage properties. *Protein science : a publication of the Protein Society*. 1999;8(2):439-42.
51. Rudner DZ, Fawcett P, Losick R. A family of membrane-embedded metalloproteases involved in regulated proteolysis of membrane-associated transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*. 1999;96(26):14765-70.
52. Ye J, Rawson RB, Komuro R, Chen X, Davé UP, Prywes R, et al. ER Stress Induces Cleavage of Membrane-Bound ATF6 by the Same Proteases that Process SREBPs. *Mol Cell*. 2000;6(6):1355-64.
53. Brannigan JA, Dodson G, Duggleby HJ, Moody PC, Smith JL, Tomchick DR, et al. A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature*. 1995;378(6555):416-9.
54. Cheng H, Grishin NV. DOM - fold: A structure with crossing loops found in DmpA, ornithine acetyltransferase, and molybdenum cofactor - binding domain. *Protein science*. 2005;14(7):1902-10.
55. Schaechter M. *Encyclopedia of microbiology*: Academic Press; 2009.
56. Erez E, Fass D, Bibi E. How intramembrane proteases bury hydrolytic reactions in the membrane. *Nature*. 2009;459(7245):371-8.
57. Southan C. A genomic perspective on human proteases as drug targets. *Drug discovery today*. 2001;6(13):681-8.
58. Verhelst SH. Intramembrane proteases as drug targets. *The FEBS Journal*. 2016.
59. Yabluchanskiy A, Li Y, Chilton RJ, Lindsey ML. Matrix metalloproteinases: drug targets for myocardial infarction. *Current drug targets*. 2013;14(3):276-86.
60. Siklos M, BenAissa M, Thatcher GR. Cysteine proteases as therapeutic targets: does selectivity matter? A systematic review of calpain and cathepsin inhibitors. *Acta Pharmaceutica Sinica B*. 2015;5(6):506-19.
61. Neurath H. Evolution of proteolytic enzymes. *Science*. 1984;224(4647):350-7.
62. Johnson DE. Noncaspase proteases in apoptosis. *Leukemia*. 2000;14(9):1695-703.
63. Joseph K, Ghebrehiwet B, Kaplan AP. Activation of the kinin-forming cascade on the surface of endothelial cells. *Biological chemistry*. 2001;382(1):71-5.
64. Coughlin SR. Thrombin signalling and protease-activated receptors. *Nature*. 2000;407(6801):258-64.
65. Barros C, Crosby JA, Moreno RD. Early steps of sperm-egg interactions during mammalian fertilization. *Cell biology international*. 1996;20(1):33-9.
66. Davidson DJ, Higgins DL, Castellino FJ. Plasminogen activator activities of equimolar complexes of streptokinase with variant recombinant plasminogens. *Biochemistry*. 1990;29(14):3585-90.
67. Sim RB, Laich A. Serine proteases of the complement system. *Biochemical Society transactions*. 2000;28(5):545-50.
68. Collen D, Lijnen HR. The fibrinolytic system in man. *Critical reviews in oncology/hematology*. 1986;4(3):249-301.
69. LeMosy EK, Hong CC, Hashimoto C. Signal transduction by a protease cascade. *Trends in cell biology*. 1999;9(3):102-7.
70. Van den Steen PE, Opdenakker G, Wormald MR, Dwek RA, Rudd PM. Matrix remodelling enzymes, the protease cascade and glycosylation. *Biochimica et biophysica acta*. 2001;1528(2-3):61-73.
71. Selvarajan S, Lund LR, Takeuchi T, Craik CS, Werb Z. A plasma kallikrein-dependent plasminogen cascade required for adipocyte differentiation. *Nature cell biology*. 2001;3(3):267-75.

72. Cheng XW, Huang Z, Kuzuya M, Okumura K, Murohara T. Cysteine protease cathepsins in atherosclerosis-based vascular disease and its complications. *Hypertension*. 2011;58(6):978-86.
73. Turk V, Turk B, Turk D. Lysosomal cysteine proteases: facts and opportunities. *The EMBO journal*. 2001;20(17):4629-33.
74. De Strooper B, Annaert W, Cupers P, Saftig P, Craessaerts K, Mumm JS, et al. A presenilin-1-dependent gamma-secretase-like protease mediates release of Notch intracellular domain. *Nature*. 1999;398(6727):518-22.
75. Struhl G, Greenwald I. Presenilin is required for activity and nuclear access of Notch in *Drosophila*. *Nature*. 1999;398(6727):522-5.
76. Ye Y, Lukinova N, Fortini ME. Neurogenic phenotypes and altered Notch processing in *Drosophila* Presenilin mutants. *Nature*. 1999;398(6727):525-9.
77. Lemberg MK, Bland FA, Weihofen A, Braud VM, Martoglio B. Intramembrane proteolysis of signal peptides: an essential step in the generation of HLA-E epitopes. *Journal of immunology*. 2001;167(11):6441-6.
78. McLauchlan J, Lemberg MK, Hope G, Martoglio B. Intramembrane proteolysis promotes trafficking of hepatitis C virus core protein to lipid droplets. *The EMBO journal*. 2002;21(15):3980-8.
79. Mucha A, Drag M, Dalton JP, Kafarski P. Metallo-aminopeptidase inhibitors. *Biochimie*. 2010;92(11):1509-29.
80. Kinch LN, Ginalski K, Grishin NV. Site-2 protease regulated intramembrane proteolysis: sequence homologs suggest an ancient signaling cascade. *Protein science : a publication of the Protein Society*. 2006;15(1):84-93.
81. McGowan S, Oellig CA, Birru WA, Caradoc-Davies TT, Stack CM, Lowther J, et al. Structure of the *Plasmodium falciparum* M17 aminopeptidase and significance for the design of drugs targeting the neutral exopeptidases. *Proceedings of the National Academy of Sciences of the United States of America*. 2010;107(6):2449-54.
82. Overall CM, Kleifeld O. Tumour microenvironment - opinion: validating matrix metalloproteinases as drug targets and anti-targets for cancer therapy. *Nat Rev Cancer*. 2006;6(3):227-39.
83. Sanz Y. Aminopeptidases. *Industrial Enzymes*: Springer; 2007. p. 243-60.
84. Barrett AJ, McDonald JK. *Mammalian proteases: a glossary and bibliography*: Academic Pr; 1986.
85. Taylor A. Aminopeptidases: structure and function. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 1993;7(2):290-8.
86. Christensen JE, Dudley EG, Pederson JA, Steele JL. Peptidases and amino acid catabolism in lactic acid bacteria. *Antonie van Leeuwenhoek*. 1999;76(1-4):217-46.
87. Scornik O, Botbol V. Bestatin as an Experimental Tool in Mammals. *Curr Drug Metab*. 2001;2(1):67-85.
88. Holz RC, Bzymek KP, Swierczek SI. Co-catalytic metallopeptidases as pharmaceutical targets. *Current opinion in chemical biology*. 2003;7(2):197-206.
89. Rigolet P, Xi XG, Rety S, Chich JF. The structural comparison of the bacterial PepX and human DPP-IV reveals sites for the design of inhibitors of PepX activity. *Febs J*. 2005;272(8):2050-9.
90. Inguibert N, Coric P, Dhotel H, Bonnard E, Llorens-Cortes C, Mota N, et al. Synthesis and in vitro activities of new non-peptidic APA inhibitors. *The journal of peptide research : official journal of the American Peptide Society*. 2005;65(2):175-88.
91. Seppo L, Jauhiainen T, Poussa T, Korpela R. A fermented milk high in bioactive peptides has a blood pressure-lowering effect in hypertensive subjects. *Am J Clin Nutr*. 2003;77(2):326-30.
92. FitzGerald RJ, O'Cuinn G. Enzymatic debittering of food protein hydrolysates. *Biotechnology advances*. 2006;24(2):234-7.
93. Maric S, Donnelly SM, Robinson MW, Skinner-Adams T, Trenholme KR, Gardiner DL, et al. The M17 leucine aminopeptidase of the malaria parasite *Plasmodium falciparum*: importance of active site metal ions in the binding of substrates and inhibitors. *Biochemistry*. 2009;48(23):5435-9.

94. Francis SE, Sullivan DJ, Jr., Goldberg DE. Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum*. *Annual review of microbiology*. 1997;51(1):97-123.
95. Klemba M, Gluzman I, Goldberg DE. A *Plasmodium falciparum* dipeptidyl aminopeptidase I participates in vacuolar hemoglobin degradation. *J Biol Chem*. 2004;279(41):43000-7.
96. Harbut MB, Velmourougane G, Dalal S, Reiss G, Whisstock JC, Onder O, et al. Bestatin-based chemical biology strategy reveals distinct roles for malaria M1- and M17-family aminopeptidases. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;108(34):E526-34.
97. R Trenholme K, L Brown C, S Skinner-Adams T, Stack C, Lowther J, To J, et al. Aminopeptidases of malaria parasites: new targets for chemotherapy. *Infectious Disorders-Drug Targets (Formerly Current Drug Targets-Infectious Disorders)*. 2010;10(3):217-25.
98. Stack CM, Lowther J, Cunningham E, Donnelly S, Gardiner DL, Trenholme KR, et al. Characterization of the *Plasmodium falciparum* M17 leucyl aminopeptidase. A protease involved in amino acid regulation with potential for antimalarial drug development. *J Biol Chem*. 2007;282(3):2069-80.
99. Lowther WT, Matthews BW. Structure and function of the methionine aminopeptidases. *Biochimica et biophysica acta*. 2000;1477(1-2):157-67.
100. Poreba M, McGowan S, Skinner-Adams TS, Trenholme KR, Gardiner DL, Whisstock JC, et al. Fingerprinting the substrate specificity of M1 and M17 aminopeptidases of human malaria, *Plasmodium falciparum*. *PloS one*. 2012;7(2):e31938.
101. Sivaraman KK, Oellig CA, Huynh K, Atkinson SC, Poreba M, Perugini MA, et al. X-ray crystal structure and specificity of the *Plasmodium falciparum* malaria aminopeptidase PfM18AAP. *Journal of molecular biology*. 2012;422(4):495-507.
102. Cunningham DF, O'Connor B. Proline specific peptidases. *Biochimica et biophysica acta*. 1997;1343(2):160-86.
103. Drinkwater N, Sivaraman KK, Bamert RS, Rut W, Mohamed K, Vinh NB, et al. Structure and substrate fingerprint of aminopeptidase P from *Plasmodium falciparum*. *The Biochemical journal*. 2016;473(19):3189-204.
104. Skinner-Adams TS, Lowther J, Teuscher F, Stack CM, Grembecka J, Mucha A, et al. Identification of phosphinate dipeptide analog inhibitors directed against the *Plasmodium falciparum* M17 leucine aminopeptidase as lead antimalarial compounds. *Journal of medicinal chemistry*. 2007;50(24):6024-31.
105. Skinner-Adams TS, Peatey CL, Anderson K, Trenholme KR, Krige D, Brown CL, et al. The aminopeptidase inhibitor CHR-2863 is an orally bioavailable inhibitor of murine malaria. *Antimicrobial agents and chemotherapy*. 2012;56(6):3244-9.
106. Velmourougane G, Harbut MB, Dalal S, McGowan S, Oellig CA, Meinhardt N, et al. Synthesis of new (-)-bestatin-based inhibitor libraries reveals a novel binding mode in the S1 pocket of the essential malaria M1 metalloaminopeptidase. *Journal of medicinal chemistry*. 2011;54(6):1655-66.
107. Kannan Sivaraman K, Paiardini A, Sienczyk M, Ruggeri C, Oellig CA, Dalton JP, et al. Synthesis and structure-activity relationships of phosphonic arginine mimetics as inhibitors of the M1 and M17 aminopeptidases from *Plasmodium falciparum*. *Journal of medicinal chemistry*. 2013;56(12):5213-7.
108. Paiardini A, Bamert RS, Kannan-Sivaraman K, Drinkwater N, Mistry SN, Scammells PJ, et al. Screening the Medicines for Malaria Venture "Malaria Box" against the *Plasmodium falciparum* aminopeptidases, M1, M17 and M18. *PloS one*. 2015;10(2):e0115859.
109. Ruggeri C, Drinkwater N, Sivaraman KK, Bamert RS, McGowan S, Paiardini A. Identification and Validation of a Potent Dual Inhibitor of the *P. falciparum* M1 and M17 Aminopeptidases Using Virtual Screening. *PloS one*. 2015;10(9):e0138957.
110. Sahi S, Rai S, Chaudhary M, Nain V. Modeling of human M1 aminopeptidases for in silico screening of potential *Plasmodium falciparum* alanine aminopeptidase (PfA-M1) specific inhibitors. *Bioinformatics*. 2014;10(8):518-25.
111. Sahi S, Raj U, Chaudhary M, Nain V. Modelling of human leucyl aminopeptidases for in silico off target binding analysis of potential *Plasmodium falciparum* leucine

- aminopeptidase (PfA-M17) specific inhibitors. *Recent patents on endocrine, metabolic & immune drug discovery*. 2014;8(3):191-201.
112. Chaudhary M, Singh V, Anvikar AR, Sahi S. Screening and In Vitro Evaluation of Potential Plasmodium falciparum Leucyl Aminopeptidase Inhibitors. *Current computer-aided drug design*. 2016;12(4):282-93.
113. Mistry SN, Drinkwater N, Ruggeri C, Sivaraman KK, Loganathan S, Fletcher S, et al. Two-pronged attack: dual inhibition of Plasmodium falciparum M1 and M17 metalloaminopeptidases by a novel series of hydroxamic acid-based inhibitors. *Journal of medicinal chemistry*. 2014;57(21):9168-83.
114. Drinkwater N, Vinh NB, Mistry SN, Bamert RS, Ruggeri C, Holleran JP, et al. Potent dual inhibitors of Plasmodium falciparum M1 and M17 aminopeptidases through optimization of S1 pocket interactions. *European journal of medicinal chemistry*. 2016;110:43-64.
115. Rawlings N, Waller M, Barrett A, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic acids research*. 2013:gkt953.
116. Albiston A, Ye S, Chai S. Membrane Bound Members of the M1 Family: More Than Aminopeptidases. *Protein Peptide Lett*. 2004;11(5):491-500.
117. Hooper NM. Families of zinc metalloproteases. *FEBS letters*. 1994;354(1):1-6.
118. Aoyagi T, Suda H, Nagai M, Ogawa K, Suzuki J. Aminopeptidase activities on the surface of mammalian cells. *Biochimica et biophysica acta*. 1976;452(1):131-43.
119. Peer WA. The role of multifunctional M1 metallopeptidases in cell cycle progression. *Annals of botany*. 2011;107(7):1171-81.
120. Sanderink GJ, Artur Y, Siest G. Human aminopeptidases: a review of the literature. *Journal of clinical chemistry and clinical biochemistry Zeitschrift fur klinische Chemie und klinische Biochemie*. 1988;26(12):795-807.
121. Haeggstrom JZ, Funk CD. Lipxygenase and leukotriene pathways: biochemistry, biology, and roles in disease. *Chemical reviews*. 2011;111(10):5866-98.
122. Haeggström JZ, Rinaldo-Matthis A. Zinc Leukotriene A4 Hydrolase/Aminopeptidase Dual Activity. *Encyclopedia of Metalloproteins*: Springer; 2013. p. 2449-54.
123. Zini S, Fournie-Zaluski MC, Chauvel E, Roques BP, Corvol P, Llorens-Cortes C. Identification of metabolic pathways of brain angiotensin II and III using specific aminopeptidase inhibitors: predominant role of angiotensin III in the control of vasopressin release. *P Natl Acad Sci USA*. 1996;93(21):11968-73.
124. Bausback HH, Churchill L, Ward PE. Angiotensin metabolism by cerebral microvascular aminopeptidase A. *Biochemical pharmacology*. 1988;37(2):155-60.
125. Hashida H, Takabayashi A, Kanai M, Adachi M, Kondo K, Kohno N, et al. Aminopeptidase N is involved in cell motility and angiogenesis: Its clinical significance in human colon cancer. *Gastroenterology*. 2002;122(2):376-86.
126. Haraguchi N, Ishii H, Mimori K, Tanaka F, Ohkuma M, Kim HM, et al. CD13 is a therapeutic target in human liver cancer stem cells. *The Journal of clinical investigation*. 2010;120(9):3326-39.
127. Rogi T, Tsujimoto M, Nakazato H, Mizutani S, Tomoda Y. Human placental leucine aminopeptidase/oxytocinase. A new member of type II membrane-spanning zinc metallopeptidase family. *J Biol Chem*. 1996;271(1):56-61.
128. Herbst JJ, Ross SA, Scott HM, Bobin SA, Morris NJ, Lienhard GE, et al. Insulin stimulates cell surface aminopeptidase activity toward vasopressin in adipocytes. *The American journal of physiology*. 1997;272(4 Pt 1):E600-6.
129. Matsumoto H, Hattori A, Mizutani S, Tsujimoto M. *Cell-Surface Aminopeptidases: Basic and Clinical Aspects*. Elsevier Science BV, Amsterdam; 2001.
130. Lew RA, Mustafa T, Ye S, McDowall SG, Chai SY, Albiston AL. Angiotensin AT4 ligands are potent, competitive inhibitors of insulin regulated aminopeptidase (IRAP). *Journal of neurochemistry*. 2003;86(2):344-50.
131. Haroon N, Inman RD. Endoplasmic reticulum aminopeptidases: Biology and pathogenic potential. *Nature reviews Rheumatology*. 2010;6(8):461-7.
132. O'Cuinn G, O'Connor B, Elmore M. Degradation of thyrotropin-releasing hormone and luteinising hormone-releasing hormone by enzymes of brain tissue. *Journal of neurochemistry*. 1990;54(1):1-13.

133. Bauer K, Schomburg L, Heuer H, Schäfer MK-H. Thyrotropin releasing hormone (TRH), the TRH-receptor and the TRH-degrading ectoenzyme; three elements of a peptidergic signalling system. *Regulatory Peptides and Cognate Receptors*: Springer; 1999. p. 13-42.
134. Cruz C, Charli JL, Vargas MA, Joseph-Bravo P. Neuronal localization of pyroglutamate aminopeptidase II in primary cultures of fetal mouse brain. *Journal of neurochemistry*. 1991;56(5):1594-601.
135. Heuer H, Ehrchen J, Bauer K, Schafer MK. Region-specific expression of thyrotrophin-releasing hormone-degrading ectoenzyme in the rat central nervous system and pituitary gland. *The European journal of neuroscience*. 1998;10(4):1465-78.
136. Bauer K, Carmeliet P, Schulz M, Baes M, Deneef C. Regulation and cellular localization of the membrane-bound thyrotropin-releasing hormone-degrading enzyme in primary cultures of neuronal, glial and adenohypophyseal cells. *Endocrinology*. 1990;127(3):1224-33.
137. Fontés G, Lajoix A-D, Bergeron F, Cadel S, Prat A, Foulon T, et al. Miniglucagon (MG)-generating endopeptidase, which processes glucagon into MG, is composed of N-arginine dibasic convertase and aminopeptidase B. *Endocrinology*. 2005;146(2):702-12.
138. Pham VL, Cadel MS, Gouzy-Darmon C, Hanquez C, Beinfeld MC, Nicolas P, et al. Aminopeptidase B, a glucagon-processing enzyme: site directed mutagenesis of the Zn²⁺-binding motif and molecular modelling. *BMC biochemistry*. 2007;8(1):21.
139. Tholander F, Muroya A, Roques BP, Fournie-Zaluski MC, Thunnissen MM, Haeggstrom JZ. Structure-based dissection of the active site chemistry of leukotriene A4 hydrolase: implications for M1 aminopeptidases and inhibitor design. *Chemistry & biology*. 2008;15(9):920-9.
140. Yen C, Green L, Miller CG. Peptide accumulation during growth of peptidase deficient mutants. *Journal of molecular biology*. 1980;143(1):35-48.
141. Chandu D, Nandi D. PepN is the major aminopeptidase in *Escherichia coli*: insights on substrate specificity and role during sodium-salicylate-induced stress. *Microbiol-Sgm*. 2003;149(Pt 12):3437-47.
142. Tan PS, van Alen-Boerrigter IJ, Poolman B, Siezen RJ, de Vos WM, Konings WN. Characterization of the *Lactococcus lactis* pepN gene encoding an aminopeptidase homologous to mammalian aminopeptidase N. *FEBS letters*. 1992;306(1):9-16.
143. Tamura T, Tamura N, Cejka Z, Hegerl R, Lottspeich F, Baumeister W. Tricorn protease--the core of a modular proteolytic system. *Science*. 1996;274(5291):1385-9.
144. Chen L, Lin YL, Peng G, Li F. Structural basis for multifunctional roles of mammalian aminopeptidase N. *Proceedings of the National Academy of Sciences of the United States of America*. 2012;109(44):17966-71.
145. Mina-Osorio P. The moonlighting enzyme CD13: old and new functions to target. *Trends in molecular medicine*. 2008;14(8):361-71.
146. Hosein FN, Bandyopadhyay A, Peer WA, Murphy AS. The catalytic and protein-protein interaction domains are required for APM1 function. *Plant physiology*. 2010;152(4):2158-72.
147. Yang Y, Liu C, Lin YL, Li F. Structural insights into central hypertension regulation by human aminopeptidase A. *J Biol Chem*. 2013;288(35):25638-45.
148. Hermans SJ, Ascher DB, Hancock NC, Holien JK, Michell BJ, Chai SY, et al. Crystal structure of human insulin-regulated aminopeptidase with specificity for cyclic peptides. *Protein science : a publication of the Protein Society*. 2015;24(2):190-9.
149. Kyrieleis OJ, Goettig P, Kiefersauer R, Huber R, Brandstetter H. Crystal structures of the tricorn interacting factor F3 from *Thermoplasma acidophilum*, a zinc aminopeptidase in three different conformations. *Journal of molecular biology*. 2005;349(4):787-800.
150. Florent I, Derhy Z, Allary M, Monsigny M, Mayer R, Schrével J. *Plasmodium falciparum* aminopeptidase gene belonging to the M1 family of zinc-metalloproteinases is expressed in erythrocytic stages. *Molecular and biochemical parasitology*. 1998;97(1):149-60.
151. Aurrecochea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, et al. PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res*. 2009;37(Database issue):D539-43.

152. Allary M, Schrevel J, Florent I. Properties, stage-dependent expression and localization of Plasmodium falciparum M1 family zinc-aminopeptidase. *Parasitology*. 2002;125(Pt 1):1-10.
153. McGowan S. Working in concert: the metalloaminopeptidases from Plasmodium falciparum. *Current opinion in structural biology*. 2013;23(6):828-35.
154. Florent I, Derhy Z, Allary M, Monsigny M, Mayer R, Schrével J. A Plasmodium falciparum aminopeptidase gene belonging to the M1 family of zinc-metallopeptidases is expressed in erythrocytic stages. *Molecular and Biochemical Parasitology*. 1998;97(1-2):149-60.
155. Dalal S, Klemba M. Roles for two aminopeptidases in vacuolar hemoglobin catabolism in Plasmodium falciparum. *J Biol Chem*. 2007;282(49):35978-87.
156. Whisstock JC, McGowan S, Trenholme KR, Gardiner DL, Dalton JP. Reply to Klemba: Intracellular processing of the membrane-bound PfA-M1 neutral aminopeptidase, a target for new antimalarials. *P Natl Acad Sci USA*. 2009;106(22):E56-E.
157. Ragheb D, Dalal S, Bompiani KM, Ray WK, Klemba M. Distribution and biochemical properties of an M1-family aminopeptidase in Plasmodium falciparum indicate a role in vacuolar hemoglobin catabolism. *J Biol Chem*. 2011;286(31):27255-65.
158. Ito K, Nakajima Y, Onohara Y, Takeo M, Nakashima K, Matsubara F, et al. Crystal structure of aminopeptidase N (proteobacteria alanyl aminopeptidase) from Escherichia coli and conformational change of methionine 260 involved in substrate recognition. *J Biol Chem*. 2006;281(44):33664-76.
159. Addlagatta A, Gay L, Matthews BW. Structure of aminopeptidase N from Escherichia coli suggests a compartmentalized, gated active site. *Proceedings of the National Academy of Sciences of the United States of America*. 2006;103(36):13339-44.
160. Nocek B, Mulligan R, Bargassa M, Collart F, Joachimiak A. Crystal structure of aminopeptidase N from human pathogen Neisseria meningitidis. *Proteins*. 2008;70(1):273-9.
161. Grembecka J, Kafarski P. Leucine aminopeptidase as a target for inhibitor design. *Mini reviews in medicinal chemistry*. 2001;1(2):133-44.
162. Rawlings ND, Waller M, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res*. 2014;42(Database issue):D503-9.
163. York IA, Goldberg AL, Mo XY, Rock KL. Proteolysis and class I major histocompatibility complex antigen presentation. *Immunological reviews*. 1999;172(1):49-66.
164. Towne CF, York IA, Neijssen J, Karow ML, Murphy AJ, Valenzuela DM, et al. Leucine aminopeptidase is not essential for trimming peptides in the cytosol or generating epitopes for MHC class I antigen presentation. *Journal of immunology*. 2005;175(10):6605-14.
165. Beninga J, Rock KL, Goldberg AL. Interferon- γ can stimulate post-proteasomal trimming of the N terminus of an antigenic peptide by inducing leucine aminopeptidase. *J Biol Chem*. 1998;273(30):18734-42.
166. Taylor A, Daims M, Lee J, Surgenor T. Identification and quantification of leucine aminopeptidase in aged normal and cataractous human lenses and ability of bovine lens LAP to cleave bovine crystallins. *Current eye research*. 1982;2(1):47-56.
167. Cappiello M, Alterio V, Amodeo P, Del Corso A, Scaloni A, Pedone C, et al. Metal ion substitution in the catalytic site greatly affects the binding of sulfhydryl-containing compounds to leucyl aminopeptidase. *Biochemistry*. 2006;45(10):3226-34.
168. Hattori A, Matsumoto H, Mizutani S, Tsujimoto M. Molecular cloning of adipocyte-derived leucine aminopeptidase highly related to placental leucine aminopeptidase/oxytocinase. *Journal of biochemistry*. 1999;125(5):931-8.
169. Kazeto H, Nomura S, Ito N, Ito T, Watanabe Y, Kajiyama H, et al. Expression of adipocyte-derived leucine aminopeptidase in endometrial cancer. Association with tumor grade and CA-125. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*. 2003;24(4):203-8.
170. Matsui M, Fowler JH, Walling LL. Leucine aminopeptidases: diversity in structure and function. *Biological chemistry*. 2006;387(12):1535-44.

171. Grenier D, Michaud J. Selective growth inhibition of *Porphyromonas gingivalis* by bestatin. *FEMS microbiology letters*. 1994;123(1-2):193-9.
172. Modak JK, Rut W, Wijeyewickrema LC, Pike RN, Drag M, Roujeinikova A. Structural basis for substrate specificity of *Helicobacter pylori* M17 aminopeptidase. *Biochimie*. 2016;121:60-71.
173. Singh AK, Singh R, Tomar D, Pandya CD, Singh R. The leucine aminopeptidase of *Staphylococcus aureus* is secreted and contributes to biofilm formation. *Int J Infect Dis*. 2012;16(5):e375-81.
174. Strater N, Sherratt DJ, Colloms SD. X-ray structure of aminopeptidase A from *Escherichia coli* and a model for the nucleoprotein complex in Xer site-specific recombination. *The EMBO journal*. 1999;18(16):4513-22.
175. Kim H, Burley SK, Lipscomb WN. Re-refinement of the X-ray crystal structure of bovine lens leucine aminopeptidase complexed with bestatin. *Journal of molecular biology*. 1993;230(3):722-4.
176. Kim H, Lipscomb WN. X-ray crystallographic determination of the structure of bovine lens leucine aminopeptidase complexed with amastatin: formulation of a catalytic mechanism featuring a gem-diolate transition state. *Biochemistry*. 1993;32(33):8465-78.
177. Sträter N, Lipscomb W. Transition state analogue L-leucinephosphonic acid bound to bovine lens leucine aminopeptidase: X-ray structure at 1.65 Å resolution in a new crystal form. *Biochemistry*. 1995;34(28):9200-10.
178. Burley SK, David PR, Taylor A, Lipscomb WN. Molecular structure of leucine aminopeptidase at 2.7-Å resolution. *P Natl Acad Sci USA*. 1990;87(17):6878-82.
179. Straeter N, Lipscomb WN. Two-metal ion mechanism of bovine lens leucine aminopeptidase: active site solvent structure and binding mode of L-leucinal, a gem-diolate transition state analog, by X-ray crystallography. *Biochemistry*. 1995;34(45):14792-800.
180. Sträter N, Sun L, Kantrowitz E, Lipscomb WN. A bicarbonate ion as a general base in the mechanism of peptide hydrolysis by dizinc leucine aminopeptidase. *P Natl Acad Sci USA*. 1999;96(20):11151-5.
181. Burley SK, David PR, Sweet RM, Taylor A, Lipscomb WN. Structure determination and refinement of bovine lens leucine aminopeptidase and its complex with bestatin. *Journal of molecular biology*. 1992;224(1):113-40.
182. Bahl A, Brunk B, Crabtree J, Fraunholz MJ, Gajria B, Grant GR, et al. PlasmoDB: the Plasmodium genome resource. A database integrating experimental and computational data. *Nucleic Acids Res*. 2003;31(1):212-5.
183. Martin RE, Kirk K. Transport of the essential nutrient isoleucine in human erythrocytes infected with the malaria parasite *Plasmodium falciparum*. *Blood*. 2007;109(5):2217-24.
184. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nature Structural & Molecular Biology*. 2002;9(9):646-52.
185. Adcock SA, McCammon JA. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chemical reviews*. 2006;106(5):1589-615.
186. Verlet L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys Rev*. 1967;159(1):98-103.
187. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force - field parameter sets 53A5 and 53A6. *Journal of computational chemistry*. 2004;25(13):1656-76.
188. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, DeBolt S, et al. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput Phys Commun*. 1995;91(1-3):1-41.
189. MacKerell Jr AD, Bashford D, Bellott M, Dunbrack Jr RL, Evanseck JD, Field MJ, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry B*. 1998;102(18):3586-616.
190. Jorgensen WL, Tirado-Rives J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*. 1988;110(6):1657-66.

191. Jorgensen WL, Tirado-Rives J. Development of the OPLS-AA force field for organic and biomolecular systems. *Abstr Pap Am Chem S.* 1998;216:U696-U.
192. Cheatham III TE, Kollman PA. Molecular dynamics simulation of nucleic acids. *Annual Review of Physical Chemistry.* 2000;51(1):435-71.
193. Norberg J, Nilsson L. Molecular dynamics applied to nucleic acids. *Accounts of chemical research.* 2002;35(6):465-72.
194. Moraitakis G, Purkiss AG, Goodfellow JM. Simulated dynamics and biological macromolecules. *Rep Prog Phys.* 2003;66(3):383-406.
195. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun.* 1995;91(1-3):43-56.
196. Hancock RD. Molecular mechanics calculations and metal ion recognition. *Accounts of chemical research.* 2002;23(8):253-7.
197. Hancock RD. Molecular Mechanics Calculations as a Tool in Coordination Chemistry. *Prog Inorg Chem.* 1989;37:187-291.
198. Aqvist J, Warshel A. Computer simulation of the initial proton transfer step in human carbonic anhydrase I. *Journal of molecular biology.* 1992;224(1):7-14.
199. Pang YP. Successful molecular dynamics simulation of two zinc complexes bridged by a hydroxide in phosphotriesterase using the cationic dummy atom method. *Proteins.* 2001;45(3):183-9.
200. Pang Y-P, Xu K, EL YAZAL J, Prendergast FG. Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach. *Protein Science.* 2000;9(10):1857-65.
201. Peters MB, Yang Y, Wang B, Fusti-Molnar L, Weaver MN, Merz KM, Jr. Structural Survey of Zinc Containing Proteins and the Development of the Zinc AMBER Force Field (ZAFF). *J Chem Theory Comput.* 2010;6(9):2935-47.
202. Hoops SC, Anderson KW, Merz KM. Force field design for metalloproteins. *Journal of the American Chemical Society.* 1991;113(22):8262-70.
203. Bayly CI, Cieplak P, Cornell W, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem-Us.* 1993;97(40):10269-80.
204. Åqvist J. Modelling of ion-ligand interactions in solutions and biomolecules. *Journal of Molecular Structure: THEOCHEM.* 1992;256:135-52.
205. Lin H, Truhlar DG. QM/MM: what have we learned, where are we, and where do we go from here? *Theoretical Chemistry Accounts.* 2007;117(2):185-99.
206. Senn HM, Thiel W. QM/MM methods for biomolecular systems. *Angew Chem Int Edit.* 2009;48(7):1198-229.
207. Gao J, Amara P, Alhambra C, Field MJ. A Generalized Hybrid Orbital (GHO) Method for the Treatment of Boundary Atoms in Combined QM/MM Calculations. *J Phys Chem A.* 1998;102(24):4714-21.
208. Jones PM, Robinson MW, Dalton JP, George AM. The Plasmodium falciparum Malaria M1 Alanine Aminopeptidase (Pf A-M1): Insights of Catalytic Mechanism and Function from MD Simulations. *PLoS one.* 2011;6(12):e28589.
209. Fiser A, Sali A. Modeller: generation and refinement of homology-based protein structure models. *Methods in enzymology.* 2003;374:461-91.
210. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry.* 2004;25(13):1605-12.
211. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, et al. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* 2007;35(Web Server issue):W522-5.
212. Yang W, Riley BT, Lei X, Porebski BT, Kass I, Buckle AM, et al. Generation of AMBER force field parameters for zinc centres of M1 and M17 family aminopeptidases. *Journal of biomolecular structure & dynamics.* 2017:1-10.
213. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics.* 1983;79(2):926-35.

214. Siegbahn PE, Borowski T. Modeling enzymatic reactions involving transition metals. *Accounts of chemical research*. 2006;39(10):729-38.
215. Merz Jr KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc*. 1995;117:5179-97.
216. Nguyen TT, Chang SC, Evnouchidou I, York IA, Zikos C, Rock KL, et al. Structural basis for antigenic peptide precursor processing by the endoplasmic reticulum aminopeptidase ERAP1. *Nat Struct Mol Biol*. 2011;18(5):604-13.
217. Case De, Darden T, Cheatham T, Simmerling C, Wang J, Duke R, et al. Amber 11. University of California; 2010.
218. Case D, Darden T, Cheatham III T, Simmerling C, Wang J, Duke R, et al. AMBER 12. University of California, San Francisco. 2012.
219. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*. 2006;65(3):712-25.
220. Wickstrom L, Okur A, Simmerling C. Evaluating the performance of the ff99SB force field based on NMR scalar coupling data. *Biophysical journal*. 2009;97(3):853-6.
221. D.A. Case TAD, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman (2015), AMBER 15, University of California, San Francisco.
222. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005;26(16):1781-802.
223. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*. 1995;103(19):8577-93.
224. Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Peptide Folding: When Simulation Meets Experiment. *Angewandte Chemie International Edition*. 1999;38(1-2):236-40.
225. Lindahl E, Hess B, Van Der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model*. 2001;7(8):306-17.
226. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem*. 2005;26(16):1701-18.
227. Turner P. XMGRACE, Version 5.1. 19. Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR. 2005.
228. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput*. 2008;4(3):435-47.
229. Bakan A, Meireles LM, Bahar I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics*. 2011;27(11):1575-7.
230. Altis A, Otten M, Nguyen PH, Hegger R, Stock G. Construction of the free energy landscape of biomolecules via dihedral angle principal component analysis. *J Chem Phys*. 2008;128(24):245102.
231. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. 2015.
232. Atkinson SC, Armistead JS, Mathias DK, Sandeu MM, Tao D, Borhani-Dizaji N, et al. The Anopheles-midgut APN1 structure reveals a new malaria transmission-blocking vaccine epitope. *Nat Struct Mol Biol*. 2015;22(7):532-9.
233. Grant BJ, Rodrigues AP, ElSawy KM, McCammon JA, Caves LS. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*. 2006;22(21):2695-6.
234. Studio R. RStudio: integrated development environment for R. RStudio Inc, Boston, Massachusetts. 2012.
235. Hayward S, Lee RA. Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J Mol Graph Model*. 2002;21(3):181-3.

236. Wriggers W, Schulten K. Protein domain movements: detection of rigid domains and visualization of hinges in comparisons of atomic coordinates. *Proteins*. 1997;29(1):1-14.
237. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996;14(1):33-8, 27-8.
238. Mondelli M. Automatic Data Extraction from PDB and PDBePISA: an Extensive View on Protein Binding Surfaces with EPOS: Università di Siena; 2015.
239. Pei J, Grishin NV. PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods in molecular biology*. 2014;1079:263-71.
240. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta crystallographica Section D, Biological crystallography*. 2011;67(Pt 4):271-81.
241. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. *Acta crystallographica Section D, Biological crystallography*. 2011;67(Pt 4):235-42.
242. Brünger AT. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*. 1992;355:472-5.
243. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta crystallographica Section D, Biological crystallography*. 2010;66(Pt 4):486-501.
244. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*. 2007;35(Web Server issue):W375-83.
245. Rawlings ND, Morton FR, Barrett AJ. MEROPS: the peptidase database. *Nucleic Acids Res*. 2006;34(Database issue):D270-2.
246. Reguera J, Santiago C, Mudgal G, Ordone D, Enjuanes L, Casasnovas JM. Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. *PLoS Pathog*. 2012;8(8):e1002859.
247. Santiago C, Mudgal G, Reguera J, Recacha R, Albrecht S, Enjuanes L, et al. Allosteric inhibition of aminopeptidase N functions related to tumor growth and virus infection. *Sci Rep*. 2017;7:46045.
248. Dalal S, Ragheb DR, Schubot FD, Klemba M. A naturally variable residue in the S1 subsite of M1 family aminopeptidases modulates catalytic properties and promotes functional specialization. *J Biol Chem*. 2013;288(36):26004-12.
249. Lin F, Wang R. Systematic Derivation of AMBER Force Field Parameters Applicable to Zinc-Containing Systems. *J Chem Theory Comput*. 2010;6(6):1852-70.
250. Hu L, Ryde U. Comparison of Methods to Obtain Force-Field Parameters for Metal Sites. *J Chem Theory Comput*. 2011;7(8):2452-63.
251. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophysical journal*. 2001;80(1):505-15.
252. Mu Y, Nguyen PH, Stock G. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins*. 2005;58(1):45-52.
253. Girdlestone C, Hayward S. The DynDom3D Webserver for the Analysis of Domain Movements in Multimeric Proteins. *J Comput Biol*. 2016;23(1):21-6.
254. Hansson T, Oostenbrink C, van Gunsteren W. Molecular dynamics simulations. *Current opinion in structural biology*. 2002;12(2):190-6.
255. Fukunishi H, Watanabe O, Takada S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *The Journal of chemical physics*. 2002;116(20):9058-67.
256. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett*. 1999;314(1-2):141-51.
257. Laio A, Gervasio FL. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep Prog Phys*. 2008;71(12):126601.
258. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J Chem Phys*. 2004;120(24):11919-29.

259. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure*. 2008;16(5):673-83.
260. Rosati M, Dalal S, Klemba M. Two cap residues in the S1 subsite of a Plasmodium falciparum M1-family aminopeptidase promote broad specificity and enhance catalysis. *Mol Biochem Parasitol*. 2017;217:7-12.
261. Toxvaerd S, Heilmann OJ, Dyre JC. Energy conservation in molecular dynamics simulations of classical systems. *J Chem Phys*. 2012;136(22):224106.
262. Mpakali A, Saridakis E, Harlos K, Zhao Y, Papakyriakou A, Kokkala P, et al. Crystal Structure of Insulin-Regulated Aminopeptidase with Bound Substrate Analogue Provides Insight on Antigenic Epitope Precursor Recognition and Processing. *Journal of immunology*. 2015;195(6):2842-51.
263. Tame JR, Vallone B. The structures of deoxy human haemoglobin and the mutant Hb Tyr α 42His at 120 K. *Acta Crystallographica Section D: Biological Crystallography*. 2000;56(7):805-11.
264. Allen WJ, Bevan DR. Steered molecular dynamics simulations reveal important mechanisms in reversible monoamine oxidase B inhibition. *Biochemistry*. 2011;50(29):6441-54.
265. Patel JS, Berteotti A, Ronsisvalle S, Rocchia W, Cavalli A. Steered molecular dynamics simulations for studying protein-ligand interaction in cyclin-dependent kinase 5. *Journal of chemical information and modeling*. 2014;54(2):470-80.
266. Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *The Journal of chemical physics*. 2003;119(6):3559-66.
267. Izrailev S, Crofts AR, Berry EA, Schulten K. Steered molecular dynamics simulation of the Rieske subunit motion in the cytochrome bc(1) complex. *Biophysical journal*. 1999;77(4):1753-68.
268. Park S, Schulten K. Calculating potentials of mean force from steered molecular dynamics simulations. *J Chem Phys*. 2004;120(13):5946-61.
269. Laskowski RA. PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res*. 2001;29(1):221-2.
270. Gu YQ, Walling LL. Identification of residues critical for activity of the wound-induced leucine aminopeptidase (LAP-A) of tomato. *European journal of biochemistry*. 2002;269(6):1630-40.
271. Chen G, Edwards T, D'Souza V M, Holz RC. Mechanistic studies on the aminopeptidase from *Aeromonas proteolytica*: a two-metal ion mechanism for peptide hydrolysis. *Biochemistry*. 1997;36(14):4278-86.
272. Kitao A, Go N. Investigating protein dynamics in collective coordinate space. *Current opinion in structural biology*. 1999;9(2):164-9.
273. Amadei A, Linssen AB, Berendsen HJ. Essential dynamics of proteins. *Proteins*. 1993;17(4):412-25.
274. Hayward S, Go N. Collective variable description of native protein dynamics. *Annu Rev Phys Chem*. 1995;46(1):223-50.
275. Gerstein M, Chothia C. Packing at the protein-water interface. *P Natl Acad Sci USA*. 1996;93(19):10167-72.
276. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem*. 1997;48(1):545-600.
277. Dill KA, Chan HS. From Levinthal to pathways to funnels. *Nature structural biology*. 1997;4(1):10-9.
278. Wales DJ. The energy landscape as a unifying theme in molecular science. *Philos T Roy Soc A*. 2005;363(1827):357-75; discussion 75-7.
279. Kraft M, Schleberger C, Weckesser J, Schulz GE. Binding structure of the leucine aminopeptidase inhibitor microginin FR1. *FEBS letters*. 2006;580(30):6943-7.
280. Kale A, Pijning T, Sonke T, Dijkstra BW, Thunnissen AM. Crystal structure of the leucine aminopeptidase from *Pseudomonas putida* reveals the molecular basis for its enantioselectivity and broad substrate specificity. *Journal of molecular biology*. 2010;398(5):703-14.

281. Gu YQ, Walling LL. Specificity of the wound - induced leucine aminopeptidase (LAP - A) of tomato. *The FEBS Journal*. 2000;267(4):1178-87.
282. DuPrez KT, Scranton MA, Walling LL, Fan L. Structural insights into chaperone-activity enhancement by a K354E mutation in tomato acidic leucine aminopeptidase. *Acta Crystallographica Section D: Structural Biology*. 2016;72(5):694-702.
283. Mann G, Huo L, Adam S, Nardone B, Vendome J, Westwood NJ, et al. Structure and Substrate Recognition of the Botromycin Maturation Enzyme BotP. *ChemBiochem*. 2016;17(23):2286-92.
284. Timm J, Valente M, García-Caballero D, Wilson K, González-Pacanowska D. Structural Characterization of Acidic M17 Leucine Aminopeptidases from the TriTryps and Evaluation of Their Role in Nutrient Starvation in *Trypanosoma brucei*. *mSphere*. 2017;2(4):e00226-17.
285. Bauvois B, Dauzonne D. Aminopeptidase-N/CD13 (EC 3.4.11.2) inhibitors: chemistry, biological evaluations, and therapeutic prospects. *Medicinal research reviews*. 2006;26(1):88-130.
286. Hitzerd SM, Verbrugge SE, Ossenkoppele G, Jansen G, Peters GJ. Positioning of aminopeptidase inhibitors in next generation cancer therapy. *Amino Acids*. 2014;46(4):793-808.
287. Amadei A, Ceruso MA, Di Nola A. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins: Structure, Function, and Genetics*. 1999;36(4):419-24.
288. Gumpena R, Kishor C, Ganji RJ, Jain N, Addlagatta A. Glu121 - Lys319 salt bridge between catalytic and N - terminal domains is pivotal for the activity and stability of *Escherichia coli* aminopeptidase N. *Protein science*. 2012;21(5):727-36.
289. Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS computational biology*. 2012;8(10):e1002708.
290. Aqvist J, Warshel A. Free energy relationships in metalloenzyme-catalyzed reactions. Calculations of the effects of metal ion substitutions in staphylococcal nuclease. *Journal of the American Chemical Society*. 1990;112(8):2860-8.
291. Duarte F, Bauer P, Barrozo A, Amrein BA, Purg M, Aqvist J, et al. Force field independent metal parameters using a nonbonded dummy model. *The journal of physical chemistry B*. 2014;118(16):4351-62.
292. Li P, Merz Jr KM. Metal Ion Modeling Using Classical Mechanics. *Chemical reviews*. 2017.
293. Pang Y-P. Novel Zinc Protein Molecular Dynamics Simulations: Steps Toward Antiangiogenesis for Cancer Treatment. *J Mol Model*. 1999;5(10):196-202.
294. Stote RH, Karplus M. Zinc binding in proteins and solution: a simple but accurate nonbonded representation. *Proteins*. 1995;23(1):12-31.
295. Stamogiannos A, Maben Z, Papakyriakou A, Mpakali A, Kokkala P, Georgiadis D, et al. Critical Role of Interdomain Interactions in the Conformational Change and Catalytic Mechanism of Endoplasmic Reticulum Aminopeptidase 1. *Biochemistry*. 2017;56(10):1546-58.