

**Purification and Characterization of Native and
Recombinant Dipeptidyl Aminopeptidase 1 of
*Plasmodium falciparum***

Flora Yinglai-Hua Wang

Thesis submitted to the faculty of the
Virginia Polytechnic Institute and State University
In partial fulfillment of the requirements for the degree of

Master of Science
In
Life Science
Department of Biochemistry

Michael W. Klemba, Committee Chair
Jianyong Li, Committee Member
Dharmendar Rathore, Committee Member

May 16, 2008

Blacksburg, Virginia

Key Words: *Plasmodium falciparum*, DPAP1, protein purification, malaria

Purification and Characterization of Native and Recombinant Dipeptidyl Aminopeptidase1 of *Plasmodium falciparum*

Flora Y. Wang

Abstract

Plasmodium falciparum dipeptidyl aminopeptidase 1 (DPAP1) contributes to the degradation of hemoglobin by releasing dipeptides from globin oligopeptides in the food vacuole. The lack of success at DPAP1 gene disruption suggests that this exopeptidase is important for efficient growth during the erythrocytic asexual stage. DPAP1 is therefore an attractive target for the development of anti-malarial drugs that block the catabolism of hemoglobin. To guide the design of selective, potent DPAP1 inhibitors, it is necessary to characterize the substrate specificity of this enzyme along with its human homolog cathepsin C. Although native purification of DPAP1 is possible, the amount of purified enzyme obtained is insufficient for extensive biochemical characterization. To overcome this obstacle, a strategy was developed for the recombinant expression of soluble DPAP1 in the bacterium *Escherichia coli* and for its activation *in vitro*. The production of active recombinant DPAP1 presents three challenges: 1) expression of the protein in soluble form, 2) generation of the native N-terminus, and 3) cleavage of the pro-domain. Soluble expression of DPAP1 was achieved by fusing it to the C-terminus of maltose-binding protein (MBP). A linker sequence encoding a tobacco etch virus protease (TEVp) cleavage site was introduced between MBP and DPAP1 such that TEVp cleavage would generate the presumed native N-terminus of DPAP1. Incubation of the MBP-DPAP1 fusion with TEVp resulted in the release of free DPAP1 which hydrolyzed the fluorogenic

substrate prolyl-arginyl-7-amido-4 methyl coumarin (Pro-Arg-AMC). Various proteases were tested for the ability to excise the pro-region. Treatment with both trypsin and papain removed the pro-region and increased DPAP1 activity two to three fold. When assayed with Pro-Arg-AMC, trypsin-treated DPAP1 had kinetic properties similar to native enzyme whereas papain-treated DPAP1 deviated from Michaelis-Menten kinetics. Using a combinational dipeptidyl substrate library, the substrate specificities of native and recombinant (trypsin-activated) DPAP1, as well as of human cathepsin C were profiled. We find that both DPAP1 and human cathepsin C accept a wide spectrum of amino acid side chains at the substrate P1 and P2 positions. Interestingly, several P2 residues show high selectivity for DPAP1 or cathepsin C. The collected data point to the feasibility of designing inhibitors that are specific for DPAP1 over cathepsin C.

Acknowledgments

First and foremost I would like to thank my research advisor Dr. Klemba for his guidance and support. He has played a key role in the development of my research and the completion of my Master's degree. I am also grateful to my committee members, Jianyong Li and Dharmendar Rathore, for taking the time to read my thesis and attending my committee meetings and thesis defense.

I would like to thank the following people whose contributions have made this project possible: Brittney Bibb for cloning of MBP-DPAP1 into the expression vector, and the optimization of expression of fusion protein in *Escherichia coli*; Dr. Seema Dalal for adding the His₆ tag at the C-terminus of the MBP-DPAP1 construct, and Monica Alvarez for recombinant expression of DPAP1 in yeast.

Thanks also go to Dr. José Arnau of Prozymex for his generous gift of the cathepsin C enzyme; Dr. Edgar Deu Sandoval of Stanford (Dr. Matthew Bogyo' Lab) for the synthesis of dipeptide-AMC substrates; Dr. N. Thornberry from Merck Research Labs for providing us with the dipeptidyl-ACC positional scanning substrate library and Dr. Florian Schubot for providing us with TEV protease.

I would also like to thank my parents for their support and love. Finally, I would like to thank my family Joseph, Adeline and Andrew, whose love and support have been the biggest contributor to all my success.

Table of Contents

Acknowledgments.....	iv
List of Figures.....	vii
List of Tables.....	ix
Chapter 1: Introduction.....	1
1.1: Background.....	1
1.1.1: History of Malaria.....	1
1.1.2: Classification of Malaria and the Vector.....	1
1.1.3: Burden of Malaria.....	2
1.1.4: Life cycle of <i>Plasmodium</i> Parasites.....	3
1.1.5: Hemoglobin Degradation.....	4
1.1.6: Food Vacuole Peptidases as Antimalarial Drug Targets.....	6
1.1.7: Dipeptidyl Aminopeptidase 1.....	6
1.2: Objective and Thesis Overview.....	9
1.2.1: Objective.....	9
1.2.2: Thesis Overview.....	9
Chapter 2: Methodology.....	10
2.1: Methodology.....	10
2.1.1: Purification of Native DPAP1 from Trophozoite Extract.....	10
2.1.2: Expression of Recombinant DPAP1 in <i>E.coli</i>	10
2.1.3: Profiling of Substrate Specificity of DPAP1.....	14

2.2: Materials and Methods	16
2.2.1: Partial Purification of Native DPAP1	16
2.2.2: Purification of Recombinant DPAP1.....	17
2.2.3: Position Scanning Dipeptidyl-ACC Library	21
2.2.4: Immunoblotting and Silver Staining.....	22
Chapter 3: Results	24
3.1: Purification of Native DPAP1.....	24
3.2: Purification of Recombinant DPAP1	25
3.3: Dipeptidyl-ACC Positional Scanning Library.....	34
3.3.1: Substrate Specificity of DPAP1 and cathepsin C at P2 Subsite.....	36
3.3.2: Substrate Specificity of DPAP1 and cathepsin C at P1 Subsite.....	38
Chapter 4: Discussion	43
4.1: Summary.....	43
4.2: Short-term Goals	45
4.3: Long-term Goals	46
Bibliography	47

List of Figures

Figure 1.1: World distribution of malaria [3].	2
Figure 1.2: Life cycle of malaria causing <i>Plasmodium</i> Parasite [10].	5
Figure 1.3: Hemoglobin degradation in the food vacuole [17].	6
Figure 2.1: Structure of cathepsin C active site with bound inhibitor [23].	12
Figure 2.2: The three main challenges for the recombinant purification of DPAP1.	14
Figure 2.3: Schechter and Berger Nomenclature for substrate-enzyme interaction.	15
Figure 3.1: % Recovery of native purification of DPAP1.	25
Figure 3.2: Coomassie gel of MBP-DPAP1 fusion protein eluted from IMAC.	26
Figure 3.3: Anti-DPAP1 immunoblot of isolated pro-DPAP1 before and after TEV protease treatment.	28
Figure 3.4: Activity of recombinant DPAP1 upon treatment with TEV protease and A) Papain B) Trypsin.	29
Figure 3.5: Kinetics of Substrate Hydrolysis Rate vs. Pro-Arg-AMC concentrations.	30
Figure 3.6: Time course of trypsin digestion of recombinant DPAP1.	31
Figure 3.7: Silver stain of purified recombinant DPAP1.	32
Figure 3.8: Schematic diagram of recombinant purification of DPAP1.	33
Figure 3.9: Comparison of P1 and P2 substrate specificity between native and recombinant DPAP1.	38
Figure 3.10: Comparison of S1 and S2 Substrate Specificity between DPAP1 and Human cathepsin C.	39

Figure 3.11: Comparison of cathepsin C preferences at P2 position with published kinetic data[29]..... 41

List of Tables

Table 1.1: Recombinant DPAP1 Expression and Purification.....	8
Table 3.1: Novel Fluorogenic Dipeptidyl Substrates for Kinetic Analysis.....	42

Chapter 1: Introduction

1.1: Background

1.1.1: History of Malaria

The word malaria originated from the Italian: mala aria, which means “bad air”. Even though the symptoms of malaria, which are characterized by periodic fever, chills, enlarged spleen, were first documented in historical records more than 3,000 years ago, the true cause of malaria was not known until much later. In 1880, Dr. Alphonse Laveran, a French military physician, examined the blood droplets collected from the soldiers who suffered from “flu” like symptom, and documented that the true cause of intermittent fever is due to a type of protozoa, which he named *Oscillaria malariae*, later known as *Plasmodium falciparum*. In 1898, a British surgeon major named Ronald Ross discovered that the *Anopheline* mosquito was the vector that transmitted the malaria parasite to the hosts. About the same time, professor Giovanni Grassi at the University of Rome proved that malaria parasites (sporozoites) were introduced to human through the bite of a female mosquito.

1.1.2: Classification of Malaria and the Vector

Malaria parasite is classified as protozoan kingdom, of the class Sporozoa. Human malaria is caused by four species of the *Plasmodium* genus: *P. vivax*, *P. ovale*, *P. malariae*, and *P. falciparum*. Among these species, malaria caused by *P. falciparum* is particularly severe and often fatal in infants and young children.

The natural vectors of malaria parasites are female *Anopheles* mosquitoes. Among the approximately 400 *Anopheles* species, 40 transmit the *Plasmodial* parasite that causes human malaria [1].

1.1.3: Burden of Malaria

About 40% of the world's population is at risk of malaria [2]. Every year, more than 500 million are infected with malaria, and an estimated 2 million died from the effect of the disease, most of them were children under the age of five. 90% of malaria related fatalities occur in sub-Saharan Africa, where it has the highest malaria infection rate (Figure 1.1).



Figure 1.1: World distribution of malaria [3].

Malaria occurs in many locations of the tropical and the subtropical world. The majority of the world's malaria cases occur in Sub-Saharan Africa. The dark-red area indicates where malaria is most wide spread. The pink area indicates lower incidences of malaria. The grey area indicates no occurrence of malaria.

One of the challenges of malaria control is the increasing occurrence of resistance to available anti-malarial drugs. Over the years, strains of *Plasmodium falciparum* have developed resistance to most of the available antimalarials such as quinine, chloroquine, proguanil, sulfadoxine-pyrimethamine, mefloquine, and atovaquone [4]. To make matters more complicated, multi-drug resistance strains have also emerged. The loss of effectiveness of available and affordable drug such as those used in the Artemisinin based Combination Therapies (ACT) is just a matter of time. Since the development of a malaria vaccine is still at an early stage, there is an urgent need for the identification and characterization of new antimalarial targets.

1.1.4: Life cycle of *Plasmodium* Parasites

The malaria parasite *Plasmodium falciparum* has a complex life cycle, which involves two types of hosts: humans and female *Anopheles* mosquitoes [5]. Upon taking a blood meal from a malaria-infected individual, the female *Anopheles* mosquito ingests *Plasmodium* gametocytes and the sexual life cycle of *Plasmodium* begins inside the mosquito gut. Within the midgut of the mosquito, male and female gametes fertilize to form a diploid zygote. Development of the zygote leads to the formation of an ookinete. The resulting ookinete penetrates the mosquito gut wall and exists on the exterior of the gut wall as an oocyst. The oocyst then ruptures, releasing hundreds of sporozoite into the mosquito body cavity where these sporozoites eventually migrate to the salivary gland of the mosquito. When an infected *Anopheles* mosquito is taking a subsequent blood meal, the sporozoites from the salivary gland are released into the blood stream of the human host. After traveling to the liver via the blood stream, the sporozoites invade the liver cells where they proliferate into thousands of merozoites. The merozoites rupture the

hepatocytes and enter into the blood stream. The released merozoites invade red blood cells (RBCs) and begin an asexual life cycle. The asexual life cycle can be generally viewed as having three stages. During the erythrocytic phase of the asexual life cycle, internalized merozoites differentiate from the ring stage to the trophozoite stage and replicate during the schizont stage to produce daughter merozoites. The accumulation of merozoites leads to the rupture of the infected erythrocytes and release of the daughter merozoites into the blood stream. These merozoites then initiate another cycle by invading erythrocytes (Figure 1.2). Thus, during the erythrocytic stage of the *Plasmodium* life cycle, the parasite undergoes repeated rounds of multiplication and this replicative cycle is the cause of all malarial associated pathology.

1.1.5: Hemoglobin Degradation

One of the remarkable features during the asexual development of the malarial parasite in the erythrocyte is the degradation of hemoglobin. During this time, 75% of the erythrocyte hemoglobin is internalized and degraded [6]. Hemoglobin degradation reaches its highest rate during the trophozoite stage. Studies have shown that *Plasmodium* incorporates amino acids derived from hemoglobin catabolism into its own protein [7, 8]. Hemoglobin degradation may also be necessary to prevent premature hemolysis of the erythrocytes [9].

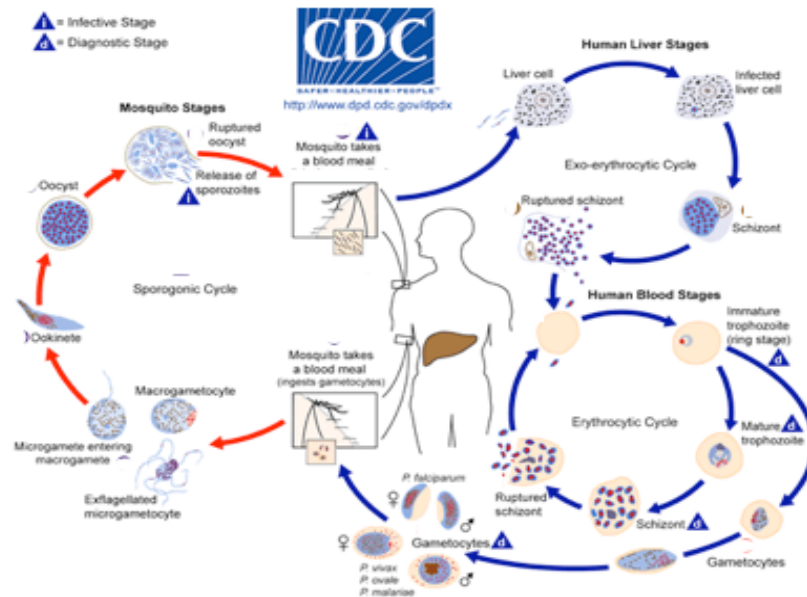


Figure 1.2: Life cycle of malaria causing *Plasmodium* Parasite [10].

The malaria parasite exhibits a complex life cycle involving a mosquito vector and a human host. The erythrocytic stage of malaria parasite's life cycle is responsible for all the pathology associated with malaria.

The process of hemoglobin degradation by malaria parasite occurs in an acidic organelle known as the food vacuole (FV). The *P. falciparum* food vacuole contains multiple proteases. The endopeptidases include the members of the aspartic protease (Plasmepsin I, II and IV), histo-aspartic protease [11-13], cysteine protease (Falcipain 2, 2' and 3) [14] and metalloprotease (falcilysin families)[15]. The enzymatic activities of these endopeptidases contribute to the dissociation of the hemoglobin tetramer and the cleavage of globin to oligopeptides. The food vacuole exopeptidases dipeptidyl amino peptidase 1 (DPAP1) [16], aminopeptidases P and aminopeptidase N [17] further degrade these oligopeptides to generate dipeptides and amino acids. Figure 1.3 is an illustration of hemoglobin degradation pathway in the food vacuole.

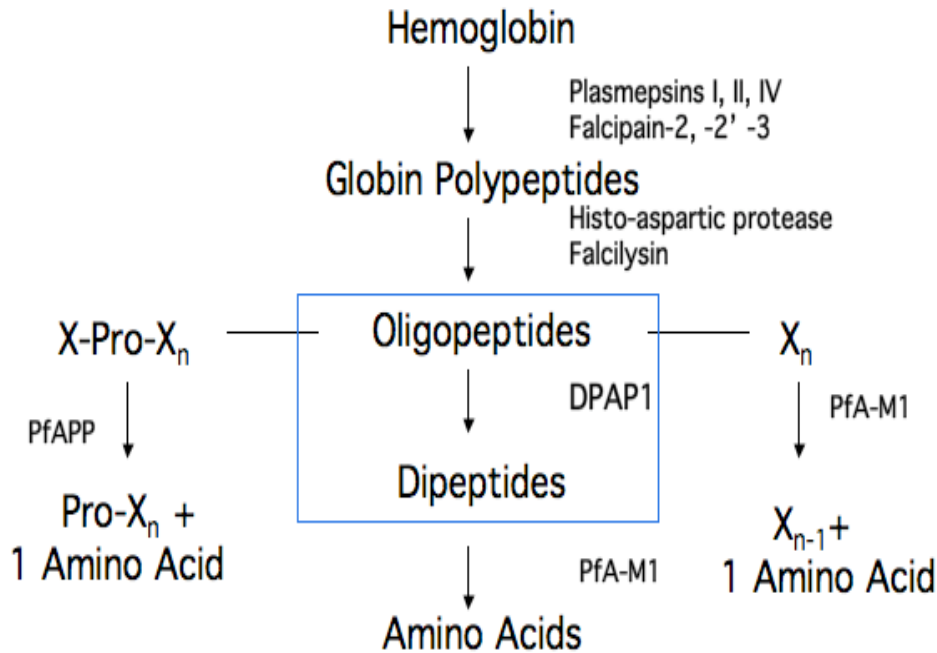


Figure 1.3: Hemoglobin degradation in the food vacuole [17].

The boxed area indicates the processing of oligopeptides by dipeptidyl aminopeptidase 1 (DPAP1), for the generation of dipeptides.

1.1.6: Food Vacuole Peptidases as Antimalarial Drug Targets

The catabolism of host cell hemoglobin is required for the intraerythrocytic growth of *Plasmodium* parasites. Cysteine protease inhibitors block hemoglobin hydrolysis, prevent parasite maturation, and kill parasites [18]. These results suggest that the proteases involved in hemoglobin degradation are potential anti-malaria targets [19]. These inhibitors might be good candidates for the development of antimalarials.

1.1.7: Dipeptidyl Aminopeptidase 1

Dipeptidyl aminopeptidase 1 (DPAP1) is another cysteine peptidase of the papain family that resides inside the food vacuole. DPAP1 is an addition to the list of potential

antimalarial drug target [16]. DPAP1 is a *Plasmodial* homolog of the human lysosomal exopeptidase cathepsin C. These two enzymes share 30% identity and 45% similarity. DPAP1 is involved in the later stage of hemoglobin degradation for the production of dipeptides (Figure 1.3). Localization studies of DPAP1 with a GFP tag at the C-terminus indicated that DPAP1 accumulates in the food vacuole during the trophozoite and schizont stages. Furthermore, attempts to disrupt the open reading frame encoding the DPAP1 were not successful [16]. These results suggested that DPAP1 plays a critical role in the degradation of hemoglobin and that inhibition of DPAP1 activity will prevent the growth of the parasite during the asexual stage. DPAP1 has two *Plasmodium* homologs DPAP2 and DPAP3. DPAP 2 is not involved in the process of hemoglobin degradation. Rather, it is expressed during the sexual stage [20]. DPAP3 appears to be involved in the release of the parasites from host erythrocytes[21]. Therefore, DPAP1 appears to be the only dipeptidyl aminopeptidase that is involved in the degradation of hemoglobin in the food vacuole. Taken together, these data suggest that DPAP1 is an attractive target for the development of anti-malarial drug.

To determine whether the inhibition of DPAP1 will block hemoglobin degradation in the host and kill the parasite, it is necessary to design selective and potent DPAP1 inhibitors. The development of DPAP1 specific inhibitors requires the characterization of the enzyme at the level of substrate specificity. Differences in substrate preference between the DPAP1 and its human homolog cathepsin C can be exploited for the design of DPAP1 inhibitors that are not toxic to the host.

The first step toward the characterization of DPAP1 is its purification. Even though native purification of DPAP1 is possible, but the amount of purified enzyme

obtained is insufficient for extensive biochemical characterization. Therefore, the expression of recombinant DPAP1 was attempted in eukaryotic and prokaryotic expression systems (Table 1.1). Recombinant DPAP1 could not be expressed in yeast (*Kluyveromyces lactis*). The expression of the full length DPAP1 in the bacteria (*E.coli*) resulted insoluble protein aggregates (data not shown). The expression of the poly- His₆ tagged version of DPAP1 at the C-terminus (DPAP1-His₆) in *Plasmodium falciparum* was also attempted. DPAP1 did not bind to an immobilized metal-ion affinity chromatography (IMAC) column, indicating that the C-terminal His-tag was cleaved by proteases in the food vacuole (data not shown). Therefore, the production of soluble and active recombinant DPAP1 became a major goal to facilitate the design of DPAP1 specific inhibitors.

Table 1.1: Recombinant DPAP1 Expression and Purification

Recombinant purification of DPAP1 has been attempted in different organisms. The method used and its outcome are listed.

Method	Outcome
Expressed in yeast (<i>Kluyveromyces lactis</i>)	No expression
Expressed in <i>Plasmodium falciparum</i> (His ₆ -tag)	His ₆ - tag was cleaved in FV
Expressed in bacteria (<i>Escherichia coli</i>) (full length)	Not soluble

1.2: Objective and Thesis Overview

1.2.1: Objective

The goal of this study was to develop a strategy for the recombinant purification and its subsequent activation of DPAP1. The successful expression of recombinant DPAP1 would enable a detailed comparison of the substrate specificity of purified DPAP1 and its human homolog cathepsin C. These results will provide a foundation for the design of potent and selective DPAP1 inhibitors.

1.2.2: Thesis Overview

Chapter 2 of this thesis provides a description of the strategies for recombinant expression of DPAP1 in *E.coli*. The approaches that were taken for the characterization of native DPAP1, recombinant DPAP1 as well as cathepsin C are also described. In Chapter 3, the results of each purification step toward the expression and activation of the recombinant DPAP1 are presented. The substrate preferences of DPAP1 and cathepsin C are also compared. In Chapter 4, conclusions are drawn, and directions for future research work are discussed.

Chapter 2: Methodology

2.1: Methodology

2.1.1: Purification of Native DPAP1 from Trophozoite Extract

Native purification of DPAP1 from soluble trophozoite extract was attempted using an established protocol[16]. DPAP1 was enriched through the sequential steps of hydrophobic interaction (HIC), anion exchange (monoQ) and gel filtration chromatography (GF). The purity of the enzyme was examined by loading a fraction of the collected sample onto a 12% sodium dodecyl sulfate (SDS)-polyacrylamide gel, and stained the gel with silver nitrate. Although the enriched DPAP1 is highly purified, the activity of the enzyme had decreased, and the amount of purified DPAP1 was not enough for subsequent biochemical characterization of the enzyme.

In an attempt to stabilize the purified DPAP1, we revised the purification process, and moved gel filtration ahead of anion exchange chromatography. Figure 2.1 showed the % recovery of DPAP1 activity after the sequential steps of HIC, GF and monoQ. The 10% recovery yield of purified enzyme was insufficient for combinatorial library profiling. Therefore, a partially purified DPAP1 was used for the dipeptidyl-ACC library assay.

2.1.2: Expression of Recombinant DPAP1 in *E.coli*

The production of active recombinant DPAP1 presents three obstacles: the expression of soluble protein, the generation of the Asp 1 residue at the N-terminus, and the removal of the pro-region. Studies have shown that the use of a natural affinity tag as

a fusion partner increases the solubility of the recombinant protein. Among the list of commonly used fusion partners, maltose binding protein (MBP) was an attractive choice because it enhances the solubility of its fusion partner and acts as an affinity tag at the same time[22]. Thus, using MBP as a fusion partner would not only lead to the production of soluble DPAP1 but also provide a tool for its removal by Amylose Resin. To express soluble DPAP1, the PCR amplified DPAP1 coding region were inserted into the downstream of the pMAL-c2X vector which encodes MBP for cytoplasmic expression of the fusion protein. (Done by Brittney Bibb). A His₆-tag was also introduced at the C-terminus of the DPAP1 coding region for the isolation of the full-length fusion protein by immobilized metal ion affinity chromatography (IMAC) after its expression in *E.coli*. (carried out by Dr. Seema Dalal)

The activation of recombinant protein requires not only the cleavage of MBP from the fusion protein but also the generation of aspartic acid at the N-terminus of the DPAP1. A crystal structure of cathepsin C complexed with Gly-Phe-diazomethane (Gly-Phe-CHN₂) revealed that the side chain of the carboxyl group of a conserved aspartic acid at the enzyme's N-terminus form ion pair with the charged N-terminal amino group of the substrate [23](Figure 2.1). Protein sequence alignment of DPAP1 with the homolog cathepsin C indicated that this aspartic acid is conserved[16]. Judging from the presumed function of the aspartic acid in cathepsin C and its conservation in DPAP1, it is reasonable to assume that the retention of aspartic acid at the N-terminus is a requirement for the activity of recombinant DPAP1.

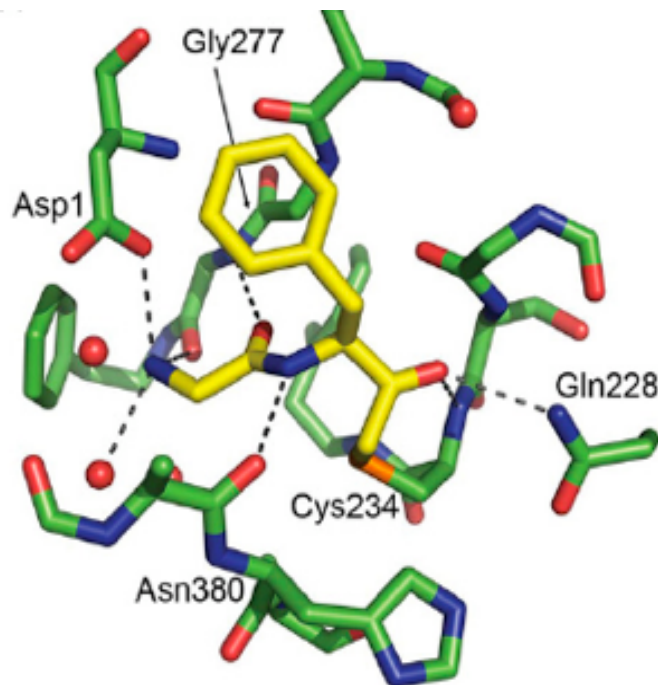


Figure 2.1: Structure of cathepsin C active site with bound inhibitor [23].

The side chain of the carboxyl group of a conserved aspartic acid at the enzyme's N-terminus (Asp 1) forms an ion pair with the charged N-terminal amino group of the substrate.

To generate the N-terminal aspartic acid of DPAP1, a linker sequence that encodes the sequence ENLYFQD was introduced between MBP and DPAP1. This linker sequence provides a cleavage site for tobacco etch virus (TEV) protease, which would cleave between the glutamine (Q) and the aspartic (D) residues with reasonable efficiency[22]. Cleavage made by TEV would not only separate the MBP from DPAP1 but also produce an aspartic residue at the N-terminus of DPAP1.

The recombinant MBP- ENLYFQD-DPAP1-His₆ clones were transformed into the Rosetta Gami strain of *E. coli* (Novagen). This strain is chosen because of its ability to enhance the formation of disulfide bonds in the bacterial cytoplasm and thereby promote the proper folding of protein in the bacterial cytoplasm.

Production of fully active DPAP1 is also likely to require removal of the pro-region. Activation of cathepsin C proenzyme requires the removal of a propeptide (pro-region) separating the N-terminal exclusion domain from the papain-like catalytic domain[24]. The structural determinations for cathepsin C further showed that the exclusion domain remains non-covalently associated with the catalytic domain. These results lead to the proposed model for the processing of proDPAP1 to the active, mature form of DPAP1[16]. In the parasite, proDPAP1 is presumably activated by other food vacuole proteases. To mimic the processing of the native DPAP1, different proteases were tested for the ability to generate mature recombinant DPAP1.

The three main challenges that needed to be addressed during recombinant purification of DPAP1 are illustrated in Figure 2.2. We anticipated that by expressing the recombinant DPAP1 as MBP-DPAP1 fusion protein would result soluble expression of DPAP1. The generation of the native like N-terminus by TEV cleavage and the subsequent removal of the pro-region of the enzyme would lead to the production of active recombinant DPAP1.

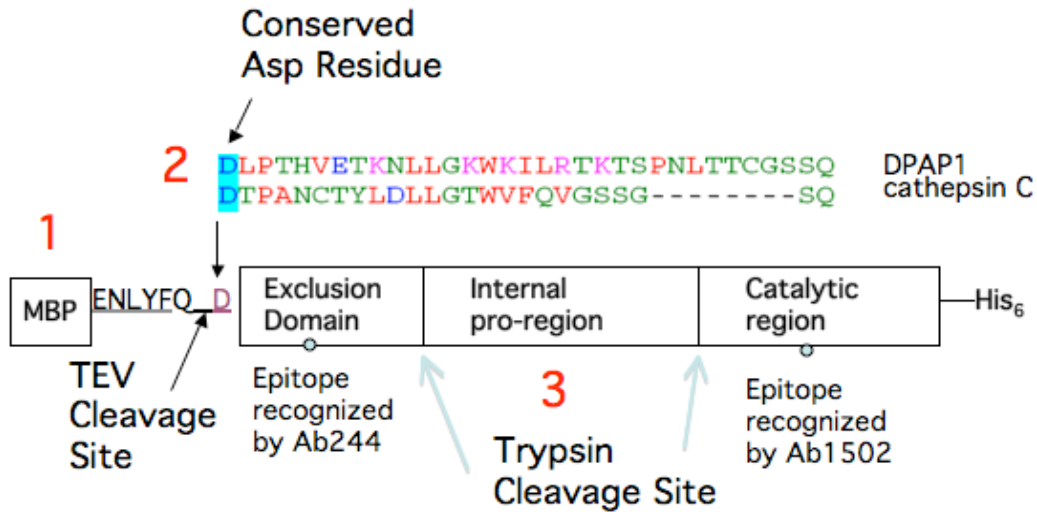


Figure 2.2: The three main challenges for the recombinant purification of DPAP1.

Challenges are indicated by red numbers: #1: expression of soluble DPAP1 fusion protein. #2: generation of native like N-terminus (Asp 1). #3: removal of pro-region. The protein sequence of cathepsin C was obtained from EMBL-EBI. The protein sequence of DPAP1 was obtained from PlasmoDB. Sequence alignment was generated using ClustalW. The model for the processing of DPAP1 is derived from Klemba et al [16].

2.1.3: Profiling of Substrate Specificity of DPAP1

A combinatorial peptide library consists of a systematic combination of amino acids and is a powerful tool for the study of enzyme substrate preferences[25]. We applied a combinatorial dipeptidyl ACC library to the study of DPAP1, because the enzyme cleaves dipeptides from the N-terminus. The interaction of DPAP1 with substrates may be described by using the general nomenclature developed by Schechter and Berger[26]. In this system of nomenclature, the substrate amino acid residues are called P, whereas the enzyme subsites on the protease that interact with the substrate are called S. The amino acid residues on the amino-terminal of the cleavage site (scissile

bond) are designated with P1, P2, P3 etc., whereas the amino acids residues on the carboxyl-terminal are designated with P1', P2' P3' etc (Figure 2.3)

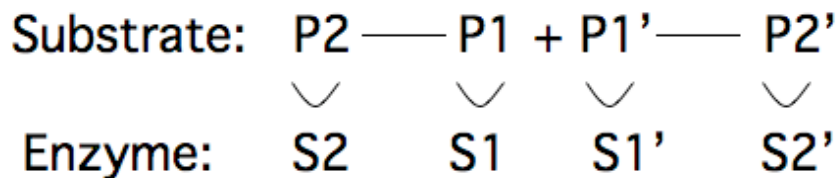


Figure 2.3: Schechter and Berger Nomenclature for substrate-enzyme interaction.

The “+” represents the scissile bond. The half circles represent enzyme’s subsite.

The dipeptidyl-ACC positional scanning library is a library of substrates that contains two sub-libraries. Using the Schechter and Berger system of nomenclature, the P1 sub-library (Z-AA+ACC) contains a defined amino acid (AA) at its P1 position and an equimolar mixture of all 20 amino acids (Z) at the P2 position. Conversely, the P2 sub-library (AA-Z+ACC) contains a defined amino acid (AA) at its P2 position and an equimolar mixture of all 20 amino acids (Z) at the P1 position. The P1' position of both sub-libraries, which is at the C-terminus of the scissile bond (+), is occupied by the ACC fluorophore (7-amino-4-carbamoylmethyl coumarin). The cleavage of dipeptide – fluorophore amide bond liberates the ACC fluoromore, resulting in an increase in fluorescence. For the profiling of DPAP1, (both natively purified and recombinantly expressed,) a quantity of purified DPAP1 that has similar rates for the cleavage of 10 μ M PR-AMC (best substrate that is commercially available) was used to cleave member of each substrate in the combinatorial peptide library. The dipeptidyl-ACC library was used

to profile partially purified native DPAP1, purified recombinant DPAP1, and purified recombinant cathepsin C.

One of the disadvantages of the native purification of the DPAP1 is the low yield of purified enzyme. Therefore, partially purified DPAP1 was used (HIC followed by monoQ) for the profiling of substrate specificity at S1 and S2 position. The background rate of hydrolysis was established by adding PR-FMK, an irreversible DPAP1 inhibitor, to the partially purified enzyme. More detail will be described in section 2.2.3A.

2.2: Materials and Methods

2.2.1: Partial Purification of Native DPAP1

Plasmodium falciparum clone 3D7 was cultured in human O+ erythrocytes in RPMI 1640 medium supplemented with 27 mM sodium bicarbonate, 11 mM glucose, 0.37 mM hypoxanthine, 10µg/ml gentamycin, and 5 g/liter Albumax. 420 mls of synchronized 3D7 trophozoites, (2% hematocrit, 12% parasitemia) were saponin treated (1 mg/ml) and harvested by centrifugation at 800 x g, 15 min at 4°C. The collected pellet was resuspended with 3 ml of lysis buffer (50 mM sodium malate, pH 5.2, 1 M (NH₄)₂SO₄, 1 mM EDTA, 1 µM pepstatin, and 2 µM leupeptin), and lysed by gentle sonication. Removal of cell debris was achieved by centrifugation at 800 x g, 10 min. at 4°C, followed by subsequent centrifugation of collected supernatant at 100,000 x g, 1hr. at 4°C. The cleared supernatant was filtered through a 0.45 µM syringe filter (Millipore) and loaded onto a phenyl sepharose HP column, (Amersham Biosciences) equilibrated

with loading buffer (50 mM sodium malate, pH 5.2, 1 M (NH₄)₂SO₄, 1 mM EDTA). Elution of DPAP1 was achieved with a linear gradient of (NH₄)₂SO₄ from 1M to 0M.

The active DPAP1 peak fractions were pooled together and dialyzed overnight against buffer composed of 50 mM bis-tris-HCl, pH 6.0, 1mM EDTA. The dialyzed sample was loaded onto a Mono Q HR 5/5 column equilibrated with loading buffer (50 mM bis-tris-HCl, pH 6.0, 1mM EDTA). The active DPAP1 was eluted with a linear gradient from 0 to 1 M NaCl. The peak fractions were tested for aminopeptidase, falcipain and DPAP1 activity. Fractions with the highest DPAP1 activity and negligible or no falcipain and aminopeptidase activity were pooled.

2.2.2: Purification of Recombinant DPAP1

A: Cloning and Expression of MBP-DPAP-His₆

The cloning of MBP-DPAP1-His₆ was done by Seema Dalal, and the optimization of the expression of MBP-DPAP1 fusion protein was done by Brittney Bibb. The DPAP1 coding region was PCR amplified using the forward primer 5'GCAC-GGAATTCGAAAACCTGTATTTTCAGGATTTACCAACCCATGTAGAAAC; and the reverse primer 5'GCACGCTGCAGTTAATGATGATGATGATGATGATTCCTA-ATTCCTTTTGCATTT. The amplified DPAP1 coding region was inserted in the pMAL-c2X vector (New England Biolabs) downstream of the malE coding sequence at the EcoRI and PstI restriction endonuclease sites. The clone was transformed into the Rosetta-gami strain of *E. coli*.

The optimal expression of MBP-DPAP1 fusion protein was achieved by adding 0.3 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) at 25°C for 6 hours. The induced cells were harvested by centrifugation (20,000 x g), and the pellet was stored at -80°C.

B: Purification of MBP-TEV

The clone of MBP-TEV was purchased from Macromolecular Crystallography Laboratory. The expression and purification of MBP-TEV was achieved using an established protocol [27] with modification. *The E. coli* cells containing the MBP-TEV fusion protein were grown in Luria-Bertani (LB) media, to mid-log phase, and were induced with IPTG. The cells were harvested four hours after induction. The wet cell paste was collected by centrifugation, and resuspended with 15 mls lysis buffer (50 mM HEPES, pH 7.5, 200 mM NaCl, 1mM EDTA, 100 mM Pefabloc, and 1mg/ml lysozyme). After 30 minutes of incubation with the lysis buffer, the lysed cells were gently sonicated. Removal of cell debris was achieved by centrifugation at 15,000 x g, 10 min. at 4°C, followed by subsequent centrifugation of collected supernatant at 15,000 x g, 10 min. at 4°C. The collected supernatant was filtered with a 0.45 μ M syringe filter (Millipore), and loaded onto an amylose column equilibrated with loading buffer (50 mM HEPES, pH 7.5, 200 mM NaCl and 1mM EDTA). Elution of MBP-TEV was achieved with a linear gradient of α -methylglucopyranoside from 0 to 1M.

Fractions containing MBP-TEV were pooled and dialyzed against 50 mM HEPES pH 8.2, 5 mM dithiothreitol (DTT), 1 mM EDTA, 200 mM NaCl, and 10% glycerol overnight at 4°C. The aliquots of MBP-TEV were flash frozen in liquid nitrogen and stored at -80 °C.

C: Purification of MBP-DPAP1-His₆

The *E. coli* pellet containing the induced MBP-DPAP1 fusion protein was resuspended with lysis buffer (20 mM NaH₂PO₄ pH 7.4, 500 mM NaCl, 30 mM imidazole, 1mg/mL lysozyme). After 30 min. incubation on ice, the cells were subjected to gentle sonication, and then centrifuged at 20,000 x g for 40 minutes at 4°C. The collected supernatant was centrifuged again at 20,000 x g for 20 minutes at 4°C. The cleared supernatant was filtered through a 0.45 µm syringe filter (Millipore) and loaded onto an immobilized metal affinity chromatography (IMAC) column charged with Ni²⁺ and equilibrated with 20 mM NaH₂PO₄ pH 7.4, 500 mM NaCl, 30 mM imidazole. Elution of fusion protein was achieved by applying a linear gradient of imidazole from 30 mM to 500 mM. The fractions were loaded onto a 12% SDS-polyacrylamide gel and fractions containing the full length MBP-DPAP1 fusion protein were pooled and stored at 4°C.

D: Activation of Recombinant DPAP1

Cleavage of MBP from MBP-DPAP1-His₆ fusion protein was achieved by incubating the fusion protein with purified MBP-TEV (pre-dialyzed against 50 mM HEPES pH 8.2, 200 mM NaCl, 1mM EDTA, 1mM L - Glutathione, and 10% glycerol), in a final concentration of 10 mM Tris pH8, 960 ng/µl of MBP-TEV, 1mM GSSG/0.3mM GSH and 5mM EDTA overnight at room temperature. To separate the MBP-TEV, free MBP and any uncleaved MBP-DPAP1, the reaction mixture was incubated with amylose resin for 30 minutes on a rocker at room temperature. The supernatant contained the pro-DPAP1-His₆.

Further activation of DPAP1 required the removal of the pro-region. Initially, several different proteases were tested for the ability to increase DPAP1 activity. Among the proteases tested, treatment with both papain and trypsin lead to further activation of the enzyme. The supernatant collected from amylose resin was split into two aliquots. Half of the supernatant was cleaved with papain (1.5 ng of papain was added to every μl of pro-DPAP1-His₆) under acidic conditions (NaOAc, pH 5). The other half of the supernatant was cleaved with trypsin (1.5 ng of trypsin was added to every μl of pro-DPAP1-His₆) under the basic conditions (Tris, pH 8).

After 15 minutes incubation at 25 °C, the papain and trypsin activity was inhibited by the addition of 1 μM trans-epoxy succinyl-L-leucylamido-(4-guanido)butane (E64) or 0.5 mM 4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride (Pefabloc), respectively. Each of the reaction mixture was then loaded onto an IMAC column separately and the eluted peak fractions were pooled together and concentrated using Amicon Ultra-15 (Millipore). The concentrated sample was loaded onto the a Superdex S200 Gel Filtration column (Amersham Biosciences) equilibrated with 50 mM Tris pH8, 200 mM NaCl and 1 mM EDTA.

Fractions with the highest DPAP1 activity and negligible falcipain and aminopeptidase activities were pooled together. The pooled enzyme was mixed with 10% glycerol and 2mM DTT. The mixture was aliquoted, flash frozen in liquid nitrogen, and stored at -80°C. We found that the addition of 0.1% of Triton-X was added to the freshly thawed aliquots of DPAP1 was necessary to preserve its enzymatic activity.

2.2.3: Position Scanning Dipeptidyl-ACC Library

The combinatorial peptide library was used for the profiling of DPAP1 (partially purified and purified recombinant), and cathepsin C. Each substrate in the P1- and P2-sublibraries was diluted to 10 μ M with the assay buffer (50 mM 2-(N-morpholino) ethanesulfonic acid (NaMES), pH6, 30 mM NaCl, 2 mM DTT and 1mM EDTA), and was added to the Corning half area 96 well plate.

For native DPAP1, partially purified enzyme eluted from the monoQ column was incubated with 1 μ M E64, 1 μ M bestatin in the assay buffer (50 mM 2-(N-morpholino) ethanesulfonic acid (NaMES), pH 6, 30 mM NaCl, 2 mM DTT and 1mM EDTA), with or without 10 μ M PR-FMK, for 15 min, at room temperature. 50 μ L of the DPAP1-inhibitor mixture was then aliquoted to corresponding wells containing the substrate from the combinatorial peptide library.

For recombinant DPAP1, 2ul of purified enzyme, which had similar rates for the cleavage of 10 μ M PR-AMC as partially purified native DPAP1, was added to 10 μ M of substrate from P1 and P2 sub-library.

The enzyme cathepsin C (generously provided by Dr. José Arnau of Prozymex) was diluted with assay buffer (50 mM MES-HCl, 30 mM NaCl, 1mM EDTA and 2mM DTT). 0.5 ng of cathepsin C was added to 10 μ M of substrate in the P1 and P2 library diluted with assay buffer.

The hydrolysis of each substrate in the library by DPAP1 was monitored fluorometrically ($\lambda_{\text{ex}} = 380$, $\lambda_{\text{em}} = 460$) using a Perkin Elmer VICTOR³ 1420 microplate fluorometer. The initial hydrolysis rates were determined from fluorescence intensity versus time plots. The rates of hydrolysis of substrates were calculated by subtracting the

background rate (derived from the hydrolysis rate of each substrate with PR-FMK added) from the initial rate of hydrolysis. The corrected rates at the P1 position were compared as the percent maximal rate of the arginine (R) residue, which was designated as 100%. The hydrolysis rates of native and recombinant DPAP1 at the P2 position were compared as the percent maximal rate of the proline (P) residue, which was designated as 100%. The hydrolysis rates of DPAP1 with cathepsin C were compared as the percent maximal rate of the histidine (H), which was designated as 100%.

2.2.4: Immunoblotting and Silver Staining

DPAP1 was identified by western blotting using two DPAP1 specific antibodies: a monoclonal antibody 304.2.4.4 (abbreviated 244), which recognizes an epitope in the DPAP1 exclusion domain; and polyclonal antibody 1502, which recognizes an epitope in the catalytic domain[16]. The western blot was done according to an established protocol, as described briefly in the next paragraph.

Protein samples were mixed with 5x loading dye (250 mM Tris-HCl pH 6.8, 10% SDS, 50% glycerol, 25% β -mercaptoethanol and 0.5% bromophenol blue). The mixtures were boiled for 5 minutes, and loaded onto a 18% SDS-polyacrylamide gel, along with the protein standard (Bio-Rad). The resolved bands were transferred to nitrocellulose membrane, and subsequently blocked with 2% bovine serum albumin (BSA) in Tris-Buffered Saline Tween-20 (TBST) buffer. Immunoblotting was carried out with primary antibody 244 (1:200 dilution) & 1502 (1:1000 dilution), followed by secondary antibody (1:5,000 dilution). (Horseradish peroxidase conjugated anti-rabbit for Ab 244, anti-mouse for Ab 1502, GE Bio-sciences). Chemiluminescent signal was developed with Amersham ECL kit (GE Bio-sciences), and detected on a photographic film.

The purity of DPAP1 was assessed on a silver stained SDS polyacrylamide gel using an established protocol[16]. Purified DPAP1 protein was loaded onto a 18% SDS-polyacrylamide gel cross –linked with piperazine diacrylamide. The gel was then sensitized with 10% glutaraldehyde, followed by staining with silver diamine and development with formaldehyde, in 1% citric acid [28].

Chapter 3: Results

3.1: Purification of Native DPAP1

DPAP1 was purified from soluble trophozoite extract using the established protocol as mentioned in section 2.2.1. DPAP1 was enriched through the sequential steps of hydrophobic interaction, anion exchange and gel filtration chromatography. Repeatedly, we observed that highly purified DPAP1 activity was not stable. In an attempt to stabilize and concentrate the purified DPAP1, we switched the order of purification process such that gel filtration was applied before anion exchange chromatography. The percent recovery of the DPAP1 enzymatic activity was still very low. Figure 3.1 is a schematic diagram of the percent recovery of DPAP1 activity after each purification step. In order to have sufficient DPAP1 for the dipeptidyl-ACC library assay, we altered the purification protocol, such that DPAP1 was partially purified using HIC followed by monoQ. This partially purified enzyme was used for the profiling of DPAP1 substrate specificity.

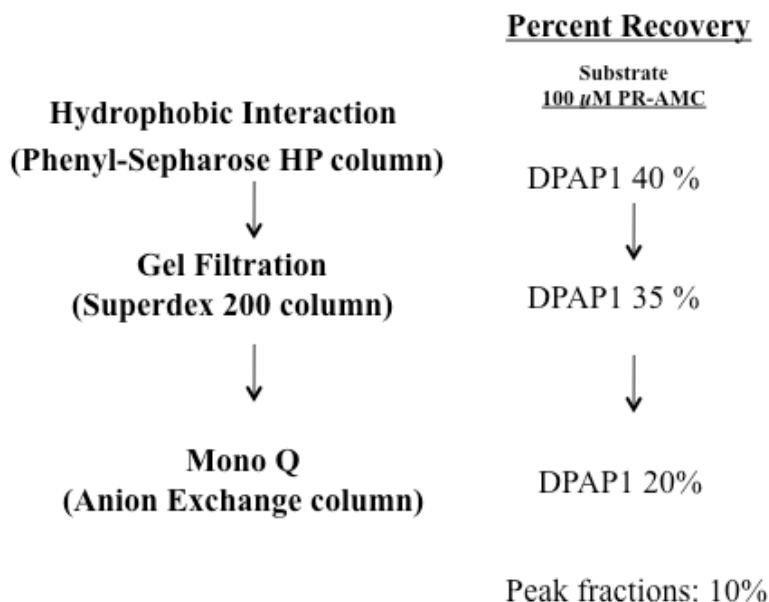


Figure 3.1: % Recovery of native purification of DPAP1.

The order of native purification of DPAP1 was revised from the established protocol. The purified fractions containing DPAP1 were assayed against 100 μ M PR-AMC.

3.2: Purification of Recombinant DPAP1

The cloning of the MBP-DPAP1-His₆ was done by Seema Dalal, and was described in section 2.2.2A. Expression vector pMAL-c2X, which would express DPAP1 in the cytoplasm, and vector pMAL-p4X, which would express DPAP1 in the periplasm, were used. Even though both vectors encode MBP, it was anticipated that periplasmic expression would be advantageous over cytoplasmic expression, because the oxidizing environment in the periplasm would favor disulfide bond formation. The optimization of MBP-DPAP1-His₆ expression in *Escherichia coli* cells was carried out by Brittney Bibb. Contrast to what we would expect, experimental results indicated that periplasmic expression of the fusion protein was mostly insoluble. A strain of *E. coli* (Rosetta-gami)

that would enable enhanced disulfide bond formations in the cytoplasm, which contains the pMAL-c2X vector, was chosen for further purification. For optimal expression, the Rosetta-gami cells were induced with 0.3mM IPTG at 25°C, and harvested 6 hours later.

The expressed MBP-DPAP1-His₆ fusion protein was purified by IMAC. The full-length fusion protein (126 kDa) were the predominant species (Figure 3.2). Impurities of various sizes were also observed.

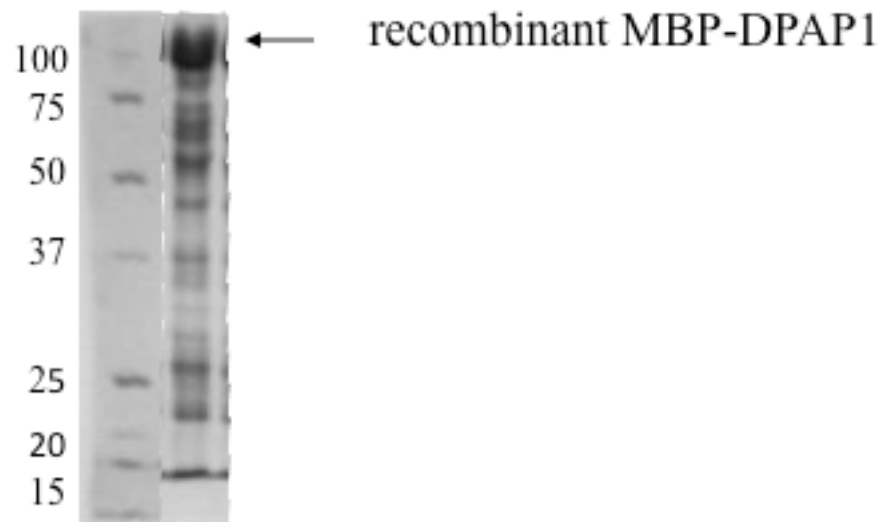


Figure 3.2: Coomassie gel of MBP-DPAP1 fusion protein eluted from IMAC.

Markers are labeled in kDa. The arrow points to the soluble MBP-DPAP1-His₆ with the expected size of 126 kDa. Size of markers is indicated in kDa.

The fractions contained full-length fusion protein were pooled and subjected to MBP-TEV digestion. For trial experiments, different types of the tobacco etch virus (TEV) were used under various cleavage concentrations, at different temperature. Cleavage of the fusion protein using His₆-TEV was first attempted at 4°C and room temperature, at various protease concentrations and pH conditions (pH 6, pH 7 or pH 8),

with or without reducing agent (1mM DTT, 1 mM L-glutathione (GSH), or 1 mM L-glutathione reduced/0.3 mM glutathione oxidized (GSH/GSSG). The experiment results indicated that activation of DPAP1 would occur when 240 ng of His₆-TEV was used to cleave every μ l of full length DPAP1, with the addition of 1 mM GSH /0.3 mM GSSG overnight, at room temperature.

Although partial activation with His₆-TEV cleavage was possible, we noticed a trend that there was a drop of DPAP1 activity when concentration of His₆-TEV exceeded 240/ μ l. To further optimize the cleavage condition, another type of protease, MBP-TEV, was purified and tested under the optimal condition used for His₆-TEV cleavage. MBP-TEV was preferred over other choices of TEV, because removal of MBP-TEV could be achieved by loading the reaction mixture onto amylose resin in a later step. When purified MBP-TEV was used to cleave the full length DPAP1, activation of DPAP1 was not observed. We speculated that the storage buffer of the MBP-TEV enzyme, which contained DTT, might inhibit the activation of DPAP1. This may also provide an explanation of decreased DPAP1 activity with increasing concentration of His₆-TEV, since DTT is included in the storage buffer.

MBP-TEV was dialyzed against buffer composed of 50 mM HEPES pH 8.2, 200 mM NaCl, 1mM EDTA, 1mM GSH, and 10% glycerol, in an attempt to replace the 5mM DTT with 1mM GSH. Trial experiments were followed using different concentrations of dialyzed MBP-TEV. Experiment results indicated that the most efficient cleavage could be achieved when 960 ng of dialyzed MPB-TEV was used for every μ l of IMAC purified MPB-DPAP1, under the cleavage condition of 1 mM GSH/0.3 mM GSSG, 5 mM EDTA, and 10 mM Tris pH8 buffer for overnight digestion at room temperature.

The separation of cleaved MBP from the pro-DPAP1, along with the MBP-TEV, was achieved by incubation of the MBP-TEV digested product with amylose resin. Centrifugation of the amylose beads, after incubation, separated the MBP containing species from soluble DPAP1, which was retained in the supernatant. Immunoblotting experiment using a DPAP1 specific antibody 1502 confirmed the production of proDPAP1 (Figure 3.3). The removal of MBP from DPAP1 leads to the generation of the native N-terminus. DPAP1 activity was observed at this stage (Figure 3.4).

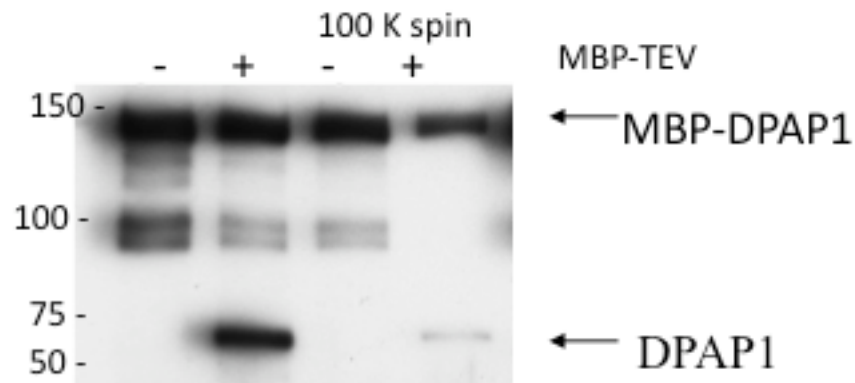


Figure 3.3: Anti-DPAP1 immunoblot of isolated pro-DPAP1 before and after TEV protease treatment.

The MBP-DPAP1 (126 kDa) was subjected to cleavage by MBP-TEV between Q and D residue of the linker sequence (ENLYFQD). The cleavage product was incubated with amylose, followed by centrifugation at 100,000 x g. The resulting DPAP1 in the supernatant was about 76 kDa as expected[16]. Size of markers is indicated in kDa.

The addition of protease to TEV cleaved DPAP1 should presumably lead to the removal of the pro-region from the partially activated enzyme. The preliminary results indicated that papain cleavage of DPAP1 increased activity by 3 folds (Figure 3.4A).

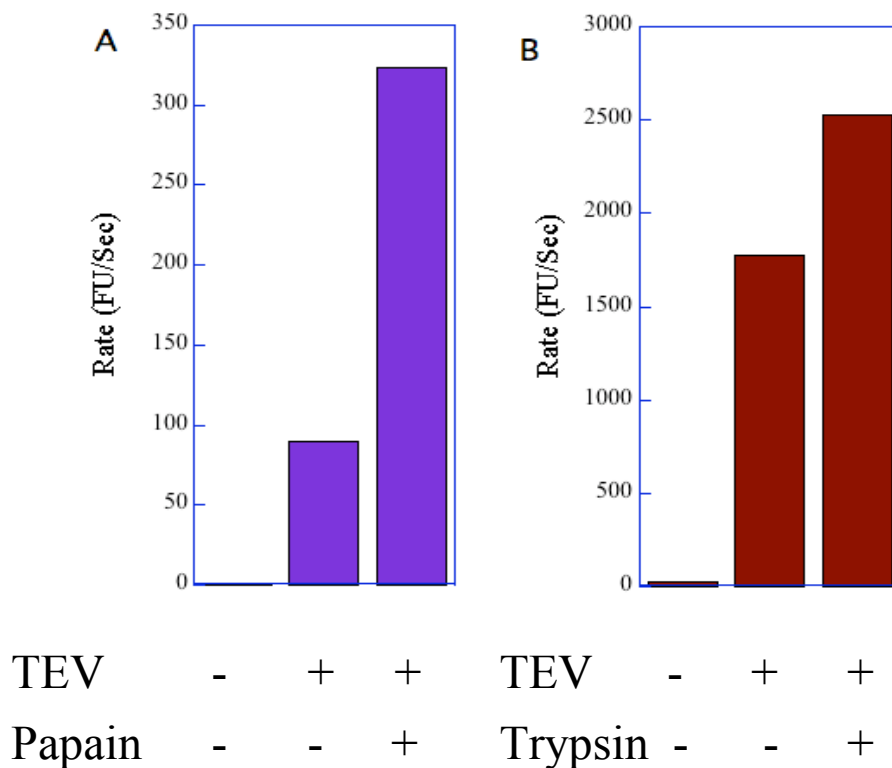


Figure 3.4: Activity of recombinant DPAP1 upon treatment with TEV protease and A) Papain B) Trypsin.

MBP-DPAP1 was subjected to TEV cleavage, followed by papain or trypsin cleavage. The generation of the native N-terminus, and subsequent removal of the pro-region lead to the activation of the DPAP1.

Although treatment with papain resulted in a three fold increase in DPAP1 activity, when the papain cleaved product was subsequently purified by gel filtration chromatography, papain activity co-migrated with DPAP1 (data not shown). In order to isolate DPAP1 from papain, the DPAP1 containing fractions eluted from the gel filtration column were pooled, and subsequently loaded onto an IMAC column (as referred in materials and method section). The collected DPAP1 containing fractions with negligible papain activity were pooled, and the K_m value of the papain treated DPAP1 was

determined. Kinetic analysis of purified and active recombinant DPAP1 indicated that papain treated recombinant enzyme deviated from the Michaelis-Menten curve (Figure 3.5 C).

In parallel experiments, different types of protease were used. Among those, trypsin treatment of TEV-cleaved DPAP1 resulted in an increase of DPAP1 enzymatic activity by 1.5 fold (Figure 3.4 B). The kinetic analysis of purified and active recombinant DPAP1 indicated that trypsin treated DPAP1 conformed to the Michaelis-Menten kinetics and showed a hyperbolic saturation curve similar to that of the native DPAP1. The K_m was close to that of the native enzyme. The experiments indicated that the K_m of native DPAP1 is 106 μM , and the K_m of trypsin- treated DPAP1 is 135 μM (Figure 3.5 A, B).

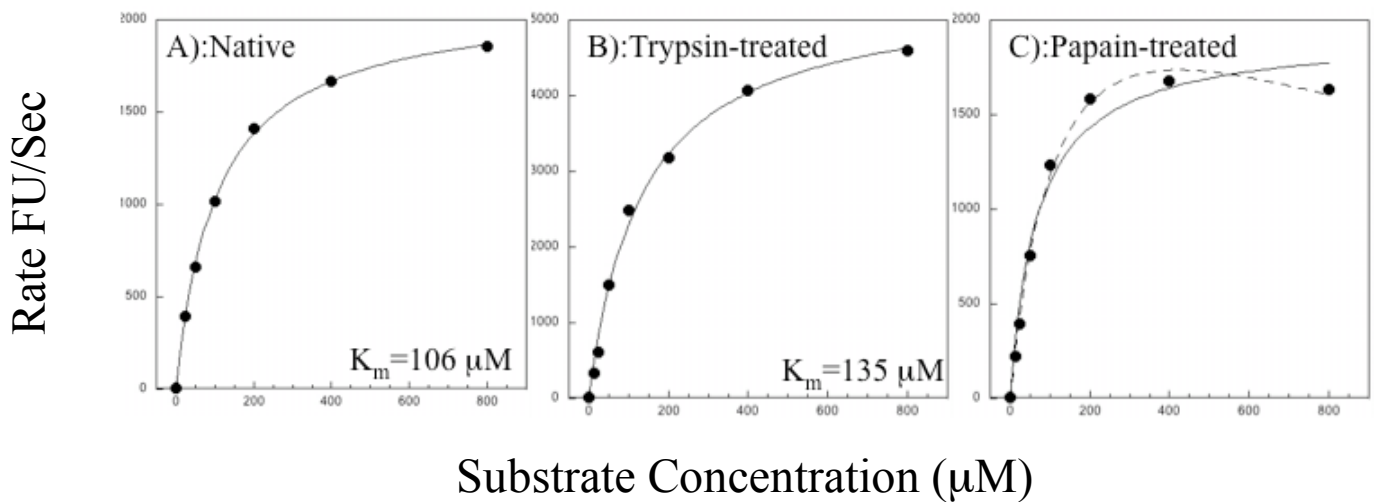


Figure 3.5: Kinetics of Substrate Hydrolysis Rate vs. Pro-Arg-AMC concentrations.

Kinetic analysis of: A) natively purified DPAP1 $K_m=106 \mu\text{M}$ B) recombinantly purified, trypsin treated DPAP1 $K_m=135 \mu\text{M}$, both conformed to Michaelis-Menten kinetics; whereas C) recombinantly purified, papain treated DPAP1 deviated from the Michaelis-Menten curve. Solid line: Michaelis-Menten curve; Dashed line in C): substrate inhibition curve.

To find out the end point of trypsin digestion of DPAP1, the gel filtration purified DPAP1, which had been treated with trypsin digestion for 15 minutes, was further digested with trypsin in a time course study. The cleaved DPAP1 resulted from each time point was identified by immunoblotting, using the DPAP1 specific antibodies for the catalytic region (Ab1502) and the exclusion region (Ab224) as depicted in Figure 3.6. The western Blot analysis indicated that the DPAP1 was cleaved into multiple polypeptides upon trypsin digestion. The results from time course study indicated that trypsin treatment of proDPAP1 lead to three major catalytic domain fragments and two major exclusion domain fragments. The processing of native DPAP1, presumably by other proteases in the food vacuole, resulted two major fragments at the catalytic domain, as shown by Klemba *et al.* The extra fragment observed upon trypsin digestion might be due to the cleavage of the loop region while at its active conformation.

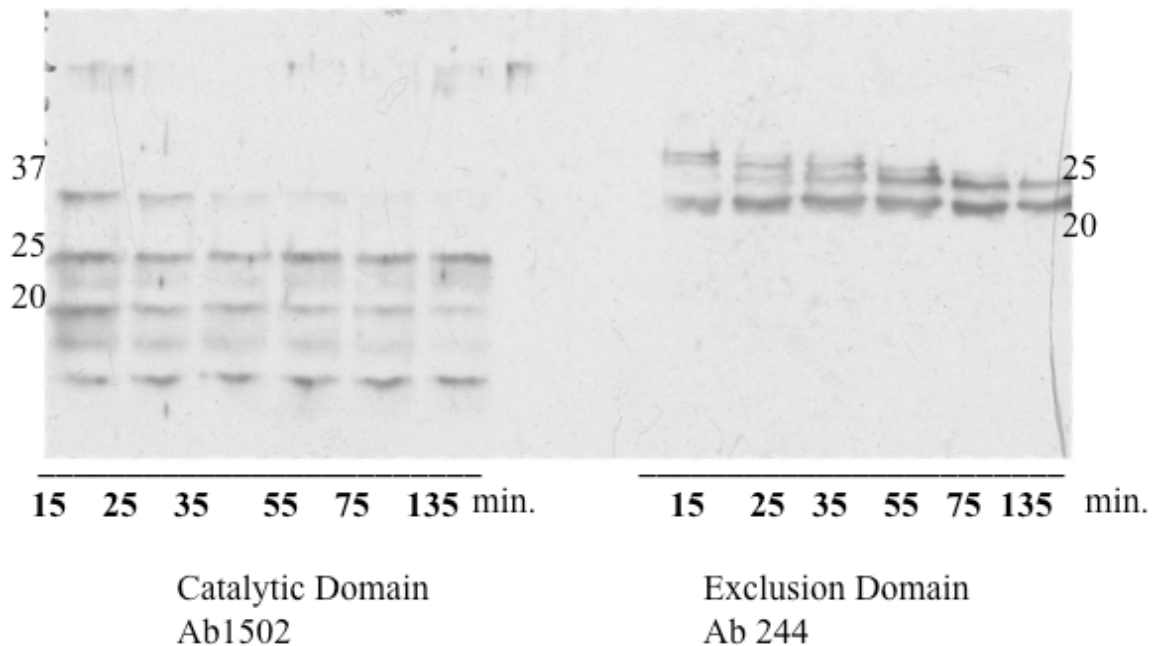


Figure 3.6: Time course of trypsin digestion of recombinant DPAP1.

Immunoblot of DPAP1 after trypsin digestion. The stability of the purified, active mature DPAP1 was followed over the course of 135 minutes. The processing of DPAP1 by trypsin digestion generated several polypeptides at the catalytic region.

Recombinant DPAP1 was tested for trypsin (using substrate 50 μ M z-FR-AMC in 50 mM N-2-Hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPE), pH 7.5, 100 mM NaCl) and aminopeptidase activity (using substrate 200 μ M L-AMC in 50 mM Tris-HCl pH 7.5). Result indicated that the contaminations of either trypsin or aminopeptidase activity were negligible.

The purity of the recombinantly purified DPAP1 was assessed on a silver stained SDS polyacrylamide gel (Figure 3.7). The processing of proDPAP1 to its mature form fragmented the pro-form into polypeptides of low molecular sizes.

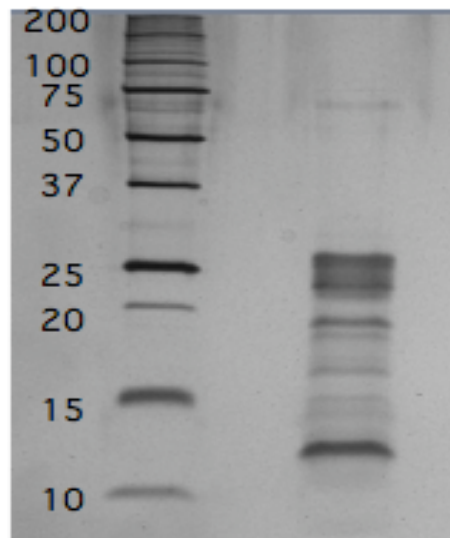


Figure 3.7: Silver stain of purified recombinant DPAP1.

The cleavage made by trypsin generated several polypeptides ranging from 25-12 kDa. Size of markers is indicated in kDa.

An optimized protocol for the recombinant purification of DPAP1 is summarized in Figure 3.8. The collected DPAP1 containing fractions (from gel filtration) were pooled and the enzyme was identified by immunoblotting, using the DPAP1 specific antibodies. The catalytic region was probed by antibody 1502, and the exclusion region was probed by antibody 244 as described in “materials and methods” section 2.2.4. Trypsin digestion of proDPAP1 produced polypeptides with low molecular weight comparing to the pro-DPAP1 (76 kDa) showed in Figure 3.3. The purity of DPAP1 was assessed on a silver stained gel was also described in “materials and methods” section 2.2.4.

Summary of Purification Strategy

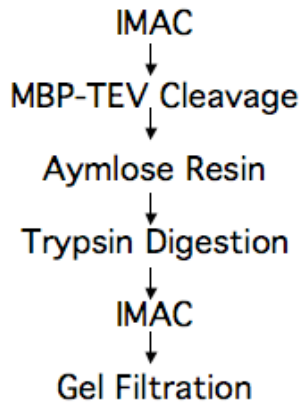


Figure 3.8: Schematic diagram of recombinant purification of DPAP1.

The development of the purification strategy requires the optimization of many steps, ranging from the choice of protease, the amount of protease needed for cleavage and the optimal buffer condition required for reaction to take place.

The final step involved in the purification process was to optimize the storage condition for purified DPAP1. When recombinantly purified DPAP1 was flash frozen in

gel filtration buffer alone, the enzyme activity decreased by 50% (data not shown). To stabilize the purified DPAP1 for long-term storage, different storage conditions were tested. Because DPAP1 requires a reduced cysteine for activity, the addition of reducing agent such as dithiothreitol (DTT) might stabilize the reduced form of the enzyme. Glycerol is known to stabilize protein structure. Therefore, different percentages of glycerol were each mixed with the purified DPAP1 with or without 2mM DTT, and were flash frozen in liquid nitrogen. The enzymatic activity of the thawed enzyme were compared with the unfrozen DPAP1, and the result indicated that 2 mM DTT, 10% glycerol was the best storage condition.

When the frozen recombinant DPAP1 was thawed for further analysis, it was discovered that the thawed DPAP1 lost its activity over time, at biological temperature (37 °C), 25 °C or on ice, even though the concentration of DTT was kept at 2mM in the reaction mixture. We speculated the enzyme might adsorb to the surface of the tube. To test this hypothesis, we added either BSA (1mg/ml), triton-X (0.1%), or both to the newly thawed enzyme, while kept the DTT concentration at 2mM. The activity of the freshly thawed DPAP1 were compared with those contained BSA and or triton-X, over 60 minute on ice, at 25 °C, or at 37 °C. Results indicated that addition of 0.1 % triton-X would stabilize the DPAP1 over time at 25 °C. Therefore, it was concluded that addition of 2mM DTT and 0.1% triton-X to the thawed recombinant DPAP1 would be necessary to preserve enzyme activity for analysis.

3.3: Dipeptidyl-ACC Positional Scanning Library

The dipeptidyl-ACC positional scanning library is composed of two sub-libraries. Each sub-library can be represented by using the following diagram:

P2 sub-library: AA - Z+ACC AA: Amino acid tested
 P2 P1 Z : Equimolar mixture of 20 amino acids
 P1 sub-library: Z - AA+ACC
 P2 P1 ACC: 7-amino-4 carbamoyl coumarin

+: Scissile bond, site where peptide cleavage occurred

Since the hydrolysis of PR-AMC, the best substrate commercially available for DPAP1, by both native and recombinant DPAP1, followed Michaelis-Menten kinetics (Figure 3.5), the catalytic efficiency (k_{cat}/K_m) of each substrate in the dipeptidyl-ACC library could be compared by determining this parameter. When assays are carried out at substrate concentrations well below the K_m of the substrates, the observed rate approximates $V[S]/K_m$. To determine the amount of substrate that should be used for the library assay, we used the K_m value of the best substrate commercially available for DPAP1, PR-AMC as a guide. Since the K_m value for both native and recombinant DPAP1 is about 100 μM (Figure 3.5), it was reasonable to assume that a substrate (in the dipeptide-ACC library) concentration of 10 μM is below the K_m . The relative catalytic efficiency of each substrate in the library could then be compared.

Because the amount of enzyme required for the combinatorial library exceeded our typical yields of highly purified DPAP1, a partial purification strategy, which gives higher yields, was adopted. Both E64 (cysteine protease inhibitor that would not inhibit DPAP1) and bestatin (aminopeptidase inhibitor) were added to the partially purified DPAP1, because preliminary result of pooled peak fractions eluted from the monoQ column indicated that aminopeptidase (using substrate 200 μM L-AMC in 50 mM Tris-

HCl pH 7.5) and falcipain activities (using substrate 50 μ M z-FR-AMC in 100 mM NaOAc, pH5.5, 100 mM DTT), although very low- migrated with the DPAP1 activity. The background activity was established by adding 10 μ M PR-FMK, a DPAP1 specific inhibitor, to the DPAP1 containing E64 and bestatin. The enzyme-inhibitor(s) complex was incubated for 15 minutes, and was mixed with 10 μ M of each substrate in the dipeptidyl-ACC library. The initial rate of hydrolysis was extracted from fluorescence vs. time plot, and the corrected reaction rate was calculated by subtracting the background rate. The relative hydrolysis rate was compared as % maximal rate, setting the proline as 100% for P2 sub-library, and arginine as 100% for P1 sub-library.

For the profiling of recombinant DPAP1, the amount of recombinant DPAP1 that would cleavage 10 μ M PR-AMC at a similar rate as the native DPAP1 was used for the library assay. The contamination of either trypsin or amino peptidase activity were negligible, therefore, it was not necessary to add inhibitors. The background was also established by adding 10 μ M PR-FMK to one set of reaction. The initial rate of hydrolysis and the corrected rate were calculated using the same protocol as mentioned above.

3.3.1: Substrate Specificity of DPAP1 and cathepsin C at P2 Subsite

The P1 and P2 subsite preferences of native DPAP1 are comparable to that of the recombinantly purified enzyme (Figure 3.9). DPAP1 did not cleave substrates that contain basic or acidic residues at the P2 site (Arg, Lys, Asp, Glu). Neither did DPAP1 prefer substrates with aromatic residues (Phe, Trp, Tyr) at the P2 position. On the other hand, aliphatic residues (Val, Ile, Nle (X), and Leu,) along with His, Gln, Ser, Thr, Ala,

and Met were all preferred by DPAP1 at the P2 site; whereas Gly, and Asn were disfavored (Fig. 3.9).

Substrate preferences of the DPAP1 homolog, cathepsin c were compared with DPAP1 at the P2 subsite (Fig.3.10). Like DPAP1, cathepsin C did not cleave substrates that contain basic residues (Lys, Arg) or acidic residues (Asp, Glu). Crystal structure of the DPAP1 homolog human cathepsin C revealed that the S2 binding site is a deep pocket, with the side chain carboxylate of the N-terminal of Asp1 positioned at the entrance[24]. It was speculated that the interaction of the carboxylate side chain of Asp with the N-terminal amino group of the substrate might be required for the formation of enzyme-substrate complex. This might provide an explanation why basic residues like Arg or Lys are not tolerated at P2 position. It was suggested that the side chains of basic residues, Arg with positively charged guanidino group, and Lys with positively charged ϵ -amino group, at pH 6, might interact with the side chain of Asp, negatively charged carboxylate ion. This interaction prevent the basic residue from entering S2 subsite, consequently, enzyme -substrate complex could not be formed[29].

Further comparison of cathepsin C with DPAP1 at the S2 subsite indicated that Pro, Ile were preferred by DPAP1, but not by cathepsin C, and Val was preferred by both enzymes.

Although aromatic residues (Phe, Tyr, Trp) were not preferred by DPAP1 at the S2 subsite, but Phe is well accepted by cathepsin C, which correlates with the crystal structure of the enzyme that showed the S2 subsite as a deep pocket that could accommodate amino acid with long side chains[23]. Furthermore, strait chain aliphatic residues (Ala, Met, Ser) were preferred by both DPAP1 and cathepsin C.

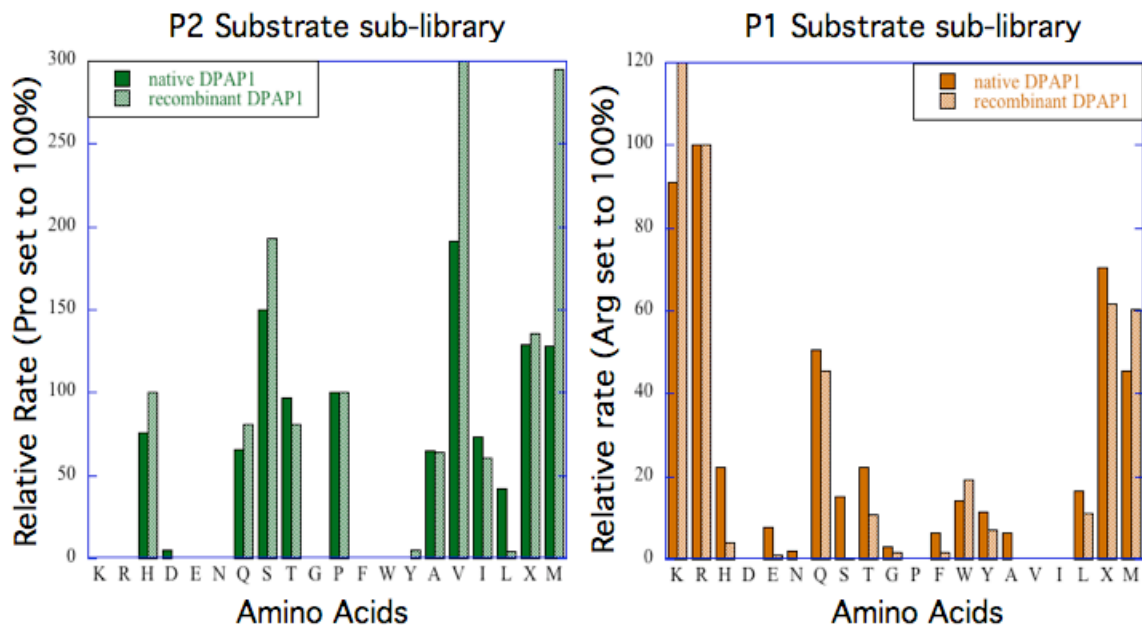


Figure 3.9: Comparison of P1 and P2 substrate specificity between native and recombinant DPAP1.

P2 sub-library: AA - Z+ACC, proline cleavage rate was set to 100%, P1 sub-library: Z - AA+ACC, arginine cleavage rate was set to 100%. AA: Amino acid tested, Z: Equal molar mixture of 20 amino acids, ACC: 7-amino-4 carbamoyl coumarin. Relative rate was calculated by subtracting the background rate (enzyme with PR-FMK) and the corrected rate for each amino acid was divided by the rate of the amino acid designated as 100% in each sub-library. Overall, substrate preferences at P1 and P2 position by DPAP1 are comparable.

3.3.2: Substrate Specificity of DPAP1 and cathepsin C at P1 Subsite

When DPAP1 substrate specificity was compared with cathepsin C, no obvious difference could be found at the P1 subsite (Fig. 3.10). Contrast to preferences made at the P2 subsite, basic residues (Arg and Lys) were strongly favored by DPAP1 and cathepsin C at P1 subsite. On the other hand, Pro, Val and Ile, which were favored at the P2 subsite, were not preferred at the P1 subsite (Fig 3.9). Both DPAP1 and cathepsin C enzymes would prefer Met, polar (Gln, Ser, Thr), aromatic (His, Phe, Tyr, Trp), aliphatic (Nle (X), Leu) and Gly the P1 position. The preferences for Glu, Asn, and Ala were

much lower for both enzymes at the P1 position. And neither enzyme would prefer Asp at P1.

When compared with published data (Tran *et al.* 2002) for the substrate specificity of cathepsin C, our data not only validated the substrate preferences tested in the literature, we also completed the list with additional amino acids at both P1 and P2 positions, which had not been studied before. For P2 subsite, we added Ile and Nle to the list. For P1 profiling, we added residues Asp, Asn, Thr, Val and Nle to the list.

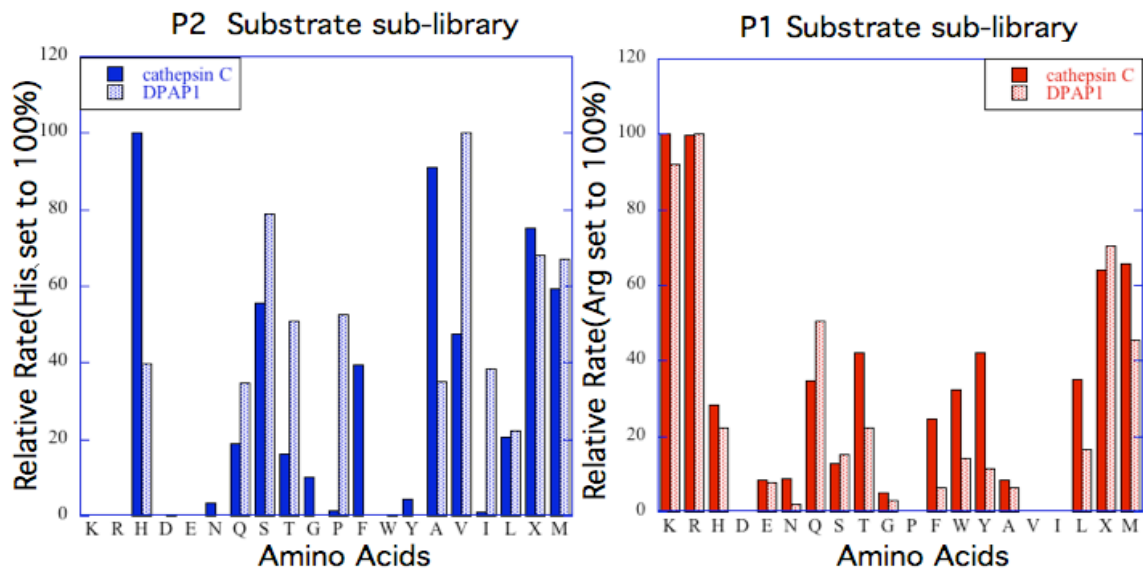


Figure 3.10: Comparison of S1 and S2 Substrate Specificity between DPAP1 and Human cathepsin C.

P2 sub-library: AA - Z+ACC, His cleavage rate was set to 100%, P1 sub-library: Z - AA+ACC, arginine cleavage rate was set to 100%. AA: Amino acid tested, Z: Equal molar mixture of 20 amino acids, ACC: 7-amino-4 carbamoyl coumarin. Corrected rate for DPAP1 was calculated by subtracting the background rate (enzyme with PR-FMK). Correct rate for cathepsin C, was calculated by subtracting 10, the estimated background, since inhibitor was not included in the assay. The relative rate was derived by dividing the corrected rates for each amino acid by the maximal cleavage of rate of amino acid in each sub-library.

Furthermore, it is worth noting that the cathepsin C substrate preferences data generated in this study correlated well with the kinetic data in the literature[29]. When the relative rate (% maximal hydrolysis) of a set of corresponding substrate was plotted against the k_{cat}/K_m of P2 specific dipeptide substrate (P1 is Phe, published by Tran *et al*, 2002), the resulting plot indicated that our data is comparable to the values published (Fig. 3.11) and the cleavage rate generated from the dipeptidyl-ACC library correlated well with the dipeptide substrate with the same residue at the P2 position.

Since the amount of substrate used for the library assay was below the K_m value for most, if not all 20 substrates in each of the sub-library, it could be assumed that the relative cleavage rate at each subsite would reflect the substrate affinity as well as the intrinsic catalytic rate. To further validate the results generated from the dipeptidyl-ACC library, it would be necessary to determine the kinetic parameters for the chosen dipeptidyl-ACC substrates. Because the differences between of DPAP1 and cathepsin C enzyme lie at the P2 subsite, we selected a set of representative residues at the P2 position that are mostly favored by DPAP1 (Pro, Ile) or cathepsin C (Phe). Valine, a residue preferred by both DPAP1 and cathepsin C, and Gly, a residue that was disfavored by neither were also chosen as candidates for the kinetic study. The P1 position will be fixed with Arg, because it was the most preferred residue for both DPAP1 and cathepsin C enzymes. Table 3.1 illustrates the dipeptide substrate that would be used for the kinetic analysis. Our collaborator Dr. Bogoy from Stanford University kindly provided us with the substrate listed in the table, and PR-AMC was purchased from Bachem. It would be anticipated that IR-AMC, which incorporated the highly preferred residue at P1 and P2 positions for DPAP1, would be cleaved with greater catalytic efficiency when compared

with substrate such as FR-AMC, which is preferred by cathepsin C. The opposite result would be expected should cathepsin C be used to cleave the same substrates. The kinetic analysis of the fluorogenic dipeptide substrates listed in the table will be used as a guide for the design of specific DPAP1 inhibitors.

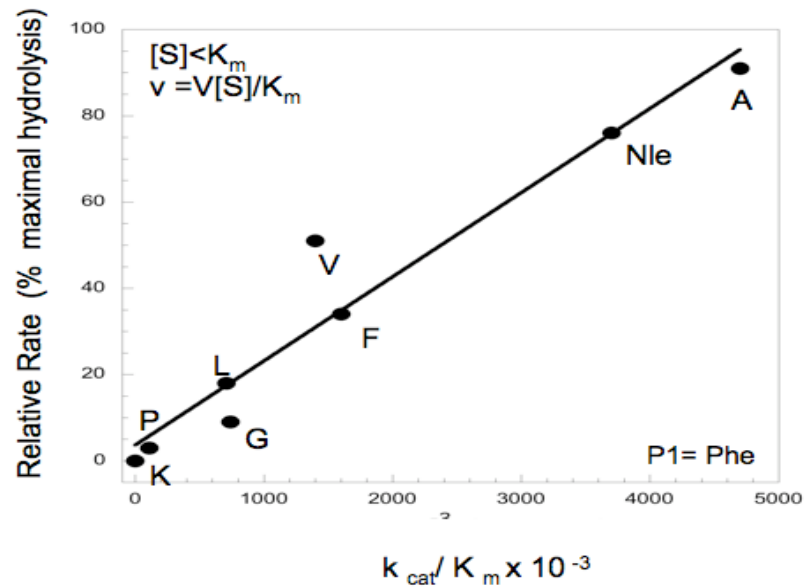


Figure 3.11: Comparison of cathepsin C preferences at P2 position with published kinetic data[29].

The relative % hydrolysis was plotted against the kinetic constant of the corresponding dipeptide with the same amino acid at the P2 position. The K_{cat} / K_m values were obtained from Tran *et al* (2002) Arch. Biochem. Biophys. 403 160-170. The correlation between cleavage efficiency and kinetic constant is evident.

Table 3.1: Novel Fluorogenic Dipeptidyl Substrates for Kinetic Analysis

The chosen substrates contain a combination of specific residues that will either be highly preferred or disfavored by the DPAP1 and/or cathepsin C enzyme. The substrates would be generously provided by Dr. M. Bogyo, other than GR-AMC and PR-AMC, which are commercially available from Becham.

	Favored	Disfavored
DPAP1	PR-AMC IR-AMC	
Cathepsin C	FR-AMC	
DPAP1 & Cathepsin C	VR-AMC	GR-AMC

Chapter 4: Discussion

4.1: Summary

To combat the debilitating disease of malaria, and to alleviate the crisis of multiple drug-resistances recently emerged by the malaria-causing parasite, the need to identify new drug targets is urgent. Proteases involved in the hemoglobin degradation pathway are attractive targets for the development of anti-malaria drugs. Among those potential drug targets, the endopeptidases that are involved in the early stage of hemoglobin degradation have been studied extensively. A recent gene knockout study of the four digestive vacuole aspartic proteases (plasmepsins) indicated that all the four are not required for parasite survival. Rather, the quadruple DV plasmepsin knockout mutants would only slow down the growth rate during the asexual stage of *Plasmodium falciparum* compare to the wild type parasite *in vitro*[30]. Therefore, the aspartic plasmepsins are no longer considered as good targets for the development of antimalarials. The processing of the plasmepsin enzymes to their mature forms involved another well studied endopeptidases named the falcipain[31]. Although potent falcipain inhibitors have been reported [32], it is still questionable whether these inhibitors would be specific to the *Plasmodium* parasites, because there are up to 10 human cysteine endopeptidases existed in the host[33]. On the other hand, the recently discovered dipeptidyl aminopeptidase 1 (DPAP1) has only one homolog known as cathepsin C[16].

The discovery and identification of DPAP1 in the food vacuole of *Plasmodium falciparum* by Klemba *et al* provided the foundation for this study. Judging from the preliminary result which indicated that DPAP1 is important for parasite proliferation and

survival[16], we hypothesized that the inactivation of DPAP1 would impede its growth and kill the parasite. To prove this hypothesis would require the design of DPAP1 specific inhibitors that could inactivate the enzyme but at the same time would have minimum toxicity against to its human host. Therefore, the overall goal of the project is to answer the question: Is that possible to develop potent DPAP1 specific inhibitor(s)? What biological effect would be observed upon inhibition of DPAP1?

“Knowing your enemy is the way to win the war [34].” The design of specific and potent inhibitors of DPAP1 would require extensive knowledge of the enzyme at the level of substrate specificity, oligomerization state, and the type of warhead that can efficiently deliver the inhibitor to its destination- the food vacuole.

Extensive characterization of DPAP1 requires the availability of the purified enzyme in its active form. Although native purification of DPAP1 from *Plasmodium falciparum* could be achieved using established protocol, the amount of purified enzyme was limited. Consequently, a partially purified DPAP1 enzyme was used for the profiling of substrate specificity, using the dipeptidyl-ACC library. Appropriate inhibitors were added to the partially purified enzyme, and subsequent adjustment was made during the calculation of cleavage rates for the substrates in the library. This further illustrated the need for the generation of stable and active recombinant DPAP1.

Multiple attempts and modifications were made for the recombinant expression and purification of DPA1 enzyme. But the production of purified soluble and active DPAP1 was not successful. In this study, a strategy was developed, and the approach toward the design and execution of a feasible methodology, which lead to the successful purification of recombinant DPAP1.

During the purification of recombinant purification of DPAP1, the purity of the enzyme as well as the activity and stability of the purified enzyme was monitored using the best substrate available on the market (PR-AMC). Further validation came from the results generated from the dipeptidyl-ACC substrate specificity assay. When the recombinantly purified enzyme was compared with natively purified DPAP1 (partially purified) for its substrate specificity, the data indicated that both forms of DPAP1 enzyme are comparable to each other, and DPAP1 showed selectivity at each of the P1 and P2 subsites.

To differentiate DPAP1 from its human homolog cathepsin C, we also profiled the substrate specificity of cathepsin C. Comparison of these two enzymes at the P1 and P2 subsites indicated that preferences at P1 subsite were similar between DPAP1 and cathepsin C. On the contrary, at the P2 subsite, different preferences were observed between the two enzymes. These data suggested that there are difference between DPAP1 and cathepsin C at the level of substrate specificity, and the specificity is governed at the P2 subsite.

4.2: Short-term Goals

Further characterization and differentiation of DPAP1 and cathepsin C would require the validation of the preferences for the P2 subsite identified from the dipeptidyl-ACC library assay. Synthesized fluorogenic dipeptide substrate (by our collaborator) would be cleaved by DPAP1 and cathepsin C. The kinetic constants (k_{cat} and K_m) would be determined, and results will be compared, and correlation between higher cleavage efficiency and P2 subsite preference could be identified.

Cathepsin C is the only known tetramer among papain-family cysteine proteases[24]. Although this unique feature was not observed in DPAP1 during the purification of DPAP1, more detailed characterization of both natively purified DPAP1 and recombinantly purified enzyme would be necessary to gain insight into the processing of the enzyme to the mature, active form.

4.3: Long-term Goals

By exploiting the differences between DPAP1 and cathepsin C through their characterization, and incorporating the kinetic data obtained from the cleavage of fluorogenic dipeptide substrate, it would potentially lead to the design of potent and selective DPAP1 inhibitors.

One important aspect of inhibitor design is the choice of functional group that would be attached to the chosen peptides for the generation of peptide based inhibitor. Among the list of choices, vinyl sulfone (VS) –based inhibitor is preferred for its potency against various types of cysteine proteases [35]. Therefore, the synthesized dipeptidyl vinyl sulfone inhibitor would be used to evaluation of the potency (K_i), and selectivity (comparing to K_i of other proteases in the food vacuole) *in vitro*, and efficacy (IC_{50}) in cultured *Plasmodium* parasite.

Bibliography

1. Collins, F.H. and S.M. Paskewitz, *Malaria: current and future prospects for control*. Annu Rev Entomol, 1995. **40**: p. 195-219.
2. Snow, R.W., et al., *The global distribution of clinical episodes of Plasmodium falciparum malaria*. Nature, 2005. **434**(7030): p. 214-7.
3. Guerra, C.A., et al., *The limits and intensity of Plasmodium falciparum transmission: implications for malaria control and elimination worldwide*. PLoS Med, 2008. **5**(2): p. e38.
4. Wongsrichanalai, C., et al., *Epidemiology of drug-resistant malaria*. Lancet Infect Dis, 2002. **2**(4): p. 209-18.
5. Greenwood, B.M., et al., *Malaria: progress, perils, and prospects for eradication*. J Clin Invest, 2008. **118**(4): p. 1266-1276.
6. Loria, P., et al., *Inhibition of the peroxidative degradation of haem as the basis of action of chloroquine and other quinoline antimalarials*. Biochem. J., 1999. **339**: p. 363-70.
7. Francis, S.E., R. Banerjee, and D.E. Goldberg, *Biosynthesis and maturation of the malaria aspartic hemoglobinases plasmepsins I and II*. J. Biol. Chem., 1997. **272**(23): p. 14961-8.
8. Sherman, I.W. and L. Tanigoshi, *Incorporation of ¹⁴C-amino acids by malaria (Plasmodium lophurae). IV. In vivo utilization of host cell haemoglobin*. Int. J. Biochem., 1970. **1**: p. 635-637.
9. Allen, R.J. and K. Kirk, *Cell volume control in the Plasmodium-infected erythrocyte*. Trends Parasitol., 2004. **20**(1): p. 7-10.
10. CDC, http://www.cdc.gov/malaria/biology/life_cycle.htm.
11. Gluzman, I.Y., et al., *Order and specificity of the Plasmodium falciparum hemoglobin degradation pathway*. J. Clin. Invest., 1994. **93**(4): p. 1602-8.
12. Goldberg, D.E., et al., *Hemoglobin degradation in the human malaria pathogen Plasmodium falciparum: a catabolic pathway initiated by a specific aspartic protease*. J. Exp. Med., 1991. **173**(4): p. 961-9.
13. Wyatt, D.M. and C. Berry, *Activity and inhibition of plasmepsin IV, a new aspartic proteinase from the malaria parasite, Plasmodium falciparum*. FEBS Lett., 2002. **513**(2-3): p. 159-62.
14. Shenai, B.R., A.V. Semenov, and P.J. Rosenthal, *Stage-specific antimalarial activity of cysteine protease inhibitors*. Biol. Chem., 2002. **383**(5): p. 843-7.
15. Eggleston, K.K., K.L. Duffin, and D.E. Goldberg, *Identification and characterization of falcilysin, a metallopeptidase involved in hemoglobin catabolism within the malaria parasite Plasmodium falciparum*. J. Biol. Chem., 1999. **274**(45): p. 32411-7.
16. Klemba, M., I. Gluzman, and D.E. Goldberg, *A Plasmodium falciparum dipeptidyl aminopeptidase I participates in vacuolar hemoglobin degradation*. J. Biol. Chem., 2004. **279**(41): p. 43000-43007.

17. Dalal, S. and M. Klemba, *Roles for two aminopeptidases in vacuolar hemoglobin catabolism in Plasmodium falciparum*. J. Biol. Chem., 2007. **282**(49): p. 35978-87.
18. Rosenthal, P.J., *Proteases of malaria parasites: new targets for chemotherapy*. Emerg. Infect. Dis., 1998. **4**(1): p. 49-57.
19. Rosenthal, P.J., *Hydrolysis of erythrocyte proteins by proteases of malaria parasites*. Curr. Opin. Hematol., 2002. **9**(2): p. 140-5.
20. Klemba, M., *Unpublished Data*.
21. Arastu-Kapur, S., et al., *Chemically mapping protease pathways involved in the regulation of erythrocyte rupture by the human malaria parasite Plasmodium falciparum* Nat. Chem. Biol., 2008: p. in press.
22. Kapust, R.B., et al., *The PI' specificity of tobacco etch virus protease*. Biochem. Biophys. Res. Commun., 2002. **294**(5): p. 949-55.
23. Molgaard, A., et al., *The crystal structure of human dipeptidyl peptidase I (cathepsin C) in complex with the inhibitor Gly-Phe-CHN2*. Biochem J, 2007. **401**(3): p. 645-50.
24. Turk, D., et al., *Structure of human dipeptidyl peptidase I (cathepsin C): exclusion domain added to an endopeptidase framework creates the machine for activation of granular serine proteases*. EMBO J., 2001. **20**(23): p. 6570-82.
25. Leiting, B., et al., *Catalytic properties and inhibition of proline-specific dipeptidyl peptidases II, IV and VII*. Biochem. J., 2003. **371**(Pt 2): p. 525-32.
26. Schechter, I. and A. Berger, *On the size of the active site in proteases. I. Papain*. Biochem. Biophys. Res. Commun., 1967. **27**: p. 157-62.
27. Waugh D.S., C., S., Tropea, J., *Expression & Purification of MBP-TEV (S219V)-Arg5*. http://mc1.ncifcrf.gov/waugh_tech/protocols/pur_mbptev.pdf.
28. Rabilloud, T., *A comparison between low background silver diammine and silver nitrate protein stains*. Electrophoresis, 1992. **13**(7): p. 429-39.
29. Tran, T.V., et al., *Dipeptidyl peptidase I: importance of progranzyme activation sequences, other dipeptide sequences, and the N-terminal amino group of synthetic substrates for enzyme activity*. Arch. Biochem. Biophys., 2002. **403**(2): p. 160-70.
30. Bonilla, J.A., et al., *Critical roles for the digestive vacuole plasmepsins of Plasmodium falciparum in vacuolar function*. Mol. Microbiol., 2007. **E-pub**: p. doi: 10.1111/j.1365-2958.2007.05768.x.
31. Drew, M.E., et al., *Plasmodium food vacuole plasmepsins are activated by falcipains*. J Biol Chem, 2008. **283**(19): p. 12870-6.
32. Shenai, B.R., et al., *Structure-activity relationships for inhibition of cysteine protease activity and development of Plasmodium falciparum by peptidyl vinyl sulfones*. Antimicrob. Agents Chemother., 2003. **47**(1): p. 154-60.
33. Turk, B., D. Turk, and V. Turk, *Lysosomal cysteine proteases: more than scavengers*. Biochim. Biophys. Acta, 2000. **1477**(1-2): p. 98-111.
34. Sun, T., *The Art of War*. 6th century BC.
35. Palmer, J.T., et al., *Vinyl sulfones as mechanism-based cysteine protease inhibitors*. J. Med. Chem., 1995. **38**(17): p. 3193-6.