

## INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning  
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA  
800-521-0600

**UMI<sup>®</sup>**



**UNIVERSITY OF ALBERTA**

**STRUCTURAL STUDIES OF MALARIAL ASPARTIC  
PROTEINASE ACTIVATION**

by

**NINA KHAZANOVICH BERNSTEIN**



**A THESIS**

**SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH  
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY**

**DEPARTMENT OF BIOCHEMISTRY**

**EDMONTON, ALBERTA  
SPRING, 2000**



National Library  
of Canada  
Acquisitions and  
Bibliographic Services  
Bibliographic Services

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque nationale  
du Canada  
Acquisitions et  
services bibliographiques  
services bibliographiques

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*

*Our file* *Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-59937-X

**Canada**

# UNIVERSITY OF ALBERTA

## LIBRARY RELEASE FORM

NAME OF AUTHOR: Nina Khazanovich  
Bernstein

TITLE OF THESIS: Structural Studies of  
Malarial Aspartic  
Proteinase Activation

DEGREE: Doctor of Philosophy

YEAR THIS DEGREE GRANTED: 2000

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.



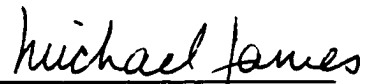
Department of Biochemistry  
University of Alberta  
Edmonton, Alberta  
CANADA  
T6G 2H7

Date: April 11/2000

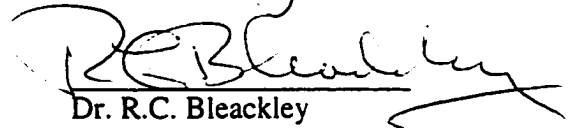
University of Alberta

Faculty of Graduate Studies and Research

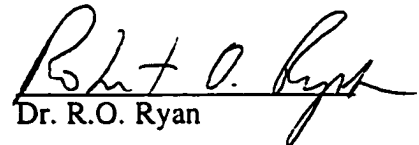
The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Structural studies of malarial aspartic proteinase activation submitted by Nina Khazanovich Bernstein in partial fulfillment of the requirements for the degree of Doctor of Philosophy.



Dr. M.N.G. James



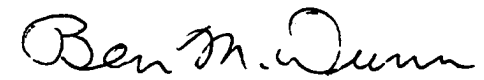
Dr. R.C. Bleackley



Dr. R.O. Ryan



Dr. O. Hindsgaul



Dr. B.M. Dunn

April 10, 2000

*Dedicated to my parents.*

## Abstract

Malaria is a complex and devastating disease that is a major public health problem in many areas of the world. During the blood stage of the malaria infection, the malaria parasite, *Plasmodium*, lives in the human host's erythrocytes, where it digests large quantities of hemoglobin as a source of nutrients. Hemoglobin digestion is an ordered and highly efficient process that is carried out in a specialized acidic compartment by several of the parasite's proteolytic enzymes. These enzymes, the hemoglobinases, include a cysteine proteinase, a metalloproteinase and at least two aspartic proteinases. The aspartic proteinases, plasmepsins, are the subject of this thesis.

Like all known eukaryotic aspartic proteinases, the plasmepsins are produced as inactive precursors, or zymogens, called proplasmepsins. Proteolytic removal of the amino-terminal extension, or prosegment, at acidic pH converts the inactive zymogen into the active proteinase. We determined the X-ray crystal structure of proplasmepsin II from the most dangerous human malaria parasite *Plasmodium falciparum*.

Comparison of the structure of this zymogen to the structure of the corresponding mature enzyme, plasmepsin II, revealed the basis of inactivity of proplasmepsin II. The prosegment and the first 15 amino acids of the mature sequence enforce a major domain rotation in the zymogen relative to the mature enzyme. This domain shift severely distorts the active site in proplasmepsin II relative to plasmepsin II. At this "immature" active site, the catalytic machinery is in the wrong orientation to carry out proteolysis. The crystal structure of proplasmepsin II was used to propose a possible pathway of autocatalytic activation that occurs at low pH *in vitro*, and to rationalize the need for acidic pH for maturase-assisted activation *in vivo*.

Until recently, our understanding of aspartic proteinase activation had been limited, on a structural level, to the gastric aspartic proteinase zymogens, human and porcine pepsinogen and human progastricsin. In all of these zymogens, the prosegment



prevents activity at neutral pH by blocking a pre-assembled active site. Disruption of several salt bridges at acidic pH clears the prosegment out of the proteinase's active site, leading to autocatalyzed activation. In proplasmepsin II, the method of inactivation is fundamentally different. In this case, the active site is accessible, but incompletely formed, and therefore unable to catalyze peptide bond hydrolysis.

We also investigated the activation of plasmepsin in another human malaria parasite, *P. vivax*, by determining the crystal structures of plasmepsin and proplasmepsin. *P. vivax* plasmepsin has the structure of a typical aspartic proteinase. *P. vivax* proplasmepsin possesses many of the structural features that were observed in *P. falciparum* proplasmepsin II, including the same method of inactivation. Thus, the domain shift that prevents the formation of a functional active site in the zymogen is established as a common mode of inactivation for the proplasmepsins.

The hemoglobinasases of *Plasmodium* have been recognized as targets for anti-malarial drug design, since inhibition of these enzymes kills the parasites in culture. Crystal structures of the proplasmepsins suggest that their unusual active site architecture may also render these zymogens potential drug design targets.

## Acknowledgements

I would like to thank my supervisor, Michael James, for his guidance and support during my graduate studies. It was wonderful to have him share his extensive knowledge of protein structure and function, and it has been a pleasure to learn from him.

I would like to acknowledge the generous contributions of our collaborators, Hansruedi Loetscher and Robert Ridley at Hoffmann La Roche who provided *P. falciparum* proplasmepsin II, and Charles Yowell and John Dame at the University of Florida, Gainesville, who provided *P. vivax* proplasmepsin. I also appreciate the many helpful discussions that I have had with John Dame and Daniel Bur (Hoffmann La Roche).

A very special note of thanks to Maia Cherney, not only for all her hard work in growing the crystals that I used in my experiments, but also for always sharing great enthusiasm for the projects.

I appreciate all the help that I received from the members of our group and from Randy Read's group. Marie, Masao, Jens, Katherine, Ernst, Steve, Craig, Ken, Natalie, Stan, Marc, Hong, Bart and Jonathan have helped me learn various aspects of protein crystallography and biochemistry. I would also like to thank Mae Wylie for help with various administrative tasks (which were usually sprung on her with little or no warning), and Janet Wright for all her help in organizing my parental leave and defense.

I would like to thank my supervisory and examination committee, Dr. Bleackley, Dr. Sykes and Dr. Ryan, as well as the external examiners Dr. Hindsgaul and Dr. Dunn.

I thank my friends, especially Kathy, Jo and Gregory who have been able to make grad school ever so much more interesting.

I am grateful to the University of Alberta, NSERC and AHFMR for financial support, and I especially appreciate AHFMR's wonderfully progressive policy of paid parental leaves.

Finally, and very importantly, I would like to thank my family: Alex for all his support, Nathan for never failing to make me smile, and my parents for all their hard work and patience.

## Table of Contents

CHAPTER 1: GENERAL INTRODUCTION .....	1
1.1 MALARIA - PAST AND PRESENT.....	1
1.2 BIOLOGY OF MALARIA .....	6
1.2.1 LIFE CYCLE OF THE PLASMODIUM PARASITE .....	7
1.2.2 HEMOGLOBIN DEGRADATION BY PLASMODIUM.....	9
1.2.3 PROTEINASES OF <i>PLASMODIUM</i> .....	10
1.3 REFERENCES .....	14
CHAPTER 2: CRYSTAL STRUCTURE OF <i>P. VIVAX</i> PLASMEPSIN.....	20
2.1 INTRODUCTION.....	20
2.1.1 BIOLOGICAL ROLE OF ASPARTIC PROTEINASES .....	20
2.1.2 STRUCTURE AND MECHANISM OF ASPARTIC PROTEINASES.....	21
2.1.3 FLEXIBILITY IN ASPARTIC PROTEINASES .....	31
2.1.4 MALARIAL PLASMEPSINS.....	35
2.2 METHODS .....	39
2.2.1 ACTIVATION AND CRYSTALLIZATION.....	39
2.2.2 DATA COLLECTION.....	40
2.2.3 STRUCTURE DETERMINATION AND REFINEMENT .....	40
2.2.4 MODELLING OF <i>P. OVALE</i> AND <i>P. MALARIAE</i> PLASMEPSINS.....	45
2.3 RESULTS AND DISCUSSION .....	45
2.3.1 DESCRIPTION OF STRUCTURE.....	45
2.3.2 COMPARISON OF THE NCS-RELATED MOLECULES.....	49
2.3.3 PEPSTATIN A BINDING .....	53
2.3.4 COMPARISON TO PFPMII .....	55
2.3.5 SUBSTRATE SPECIFICITIES OF THE PLASMEPSINS .....	65
2.3.6 COMPARISON WITH HUMAN APS .....	66
2.4 CONCLUSION .....	71
2.5 REFERENCES .....	72
CHAPTER 3: CRYSTAL STRUCTURE OF <i>P. FALCIPARUM</i> PROPLASMEPSIN II, A NOVEL ASPARTIC PROTEINASE ZYMOGEN.....	79
3.1 INTRODUCTION.....	79
3.1.1 GASTRIC AP ZYMOGENS.....	80
3.1.2 PROPHYTEPSIN .....	85
3.1.3 PROPLASMEPSINS .....	86
3.1.3.1 <i>Plasmepsin production in vivo</i> .....	88

3.1.3.2 Recombinant plasmepsins and truncated proplasmepsins .....	90
3.2 METHODS .....	91
3.2.1 CRYSTALLIZATION AND DATA COLLECTION .....	91
3.2.2 SELF-ROTATION SEARCH.....	92
3.2.3 STRUCTURE DETERMINATION AND REFINEMENT .....	92
3.2.4 MODELLING OF <i>P. FALCIPARUM</i> PROPLASMEPSIN I.....	95
3.3 RESULTS AND DISCUSSION .....	100
3.3.1 DESCRIPTION OF STRUCTURE.....	100
3.3.2 COMPARISONS AMONG NCS-RELATED MOLECULES.....	106
3.3.3 DIMERIZATION OF ZYMOGEN.....	108
3.3.4 COMPARISON TO PLASMEPSIN II .....	108
3.3.4.1 Rearrangement of the mature N-terminus.....	109
3.3.4.2 The domain shift.....	115
3.3.4.3 Inhibition of catalytic activity in the zymogen.....	115
3.3.4.4 Loop rearrangements.....	121
3.3.5 COMPARISON WITH GASTRIC AP ZYMOGENS AND PROPHYTEPSIN.....	124
3.3.6 EFFECT OF LOWERING THE PH.....	126
3.3.7 AUTOACTIVATION OF <i>P. FALCIPARUM</i> PROPLASMEPSIN I.....	128
3.3.8 A MODEL FOR THE ENTIRE PROPLASMEPSIN II .....	131
3.4 CONCLUSION .....	131
3.5 REFERENCES .....	134
CHAPTER 4: CRYSTAL STRUCTURE OF <i>P. VIVAX</i> PROPLASMEPSIN .....	140
4.1 INTRODUCTION.....	140
4.2 METHODS .....	141
4.2.1 CRYSTALLIZATION AND DATA COLLECTION .....	141
4.2.2 STRUCTURE DETERMINATION AND REFINEMENT .....	143
4.3 RESULTS AND DISCUSSION .....	148
4.3.1 QUALITY OF THE MODEL.....	148
4.3.2 DESCRIPTION OF STRUCTURE.....	152
4.3.3 COMPARISON OF NCS-RELATED MOLECULES.....	160
4.3.4 COMPARISON TO <i>P. VIVAX</i> PLASMEPSIN .....	162
4.3.5 COMPARISON TO <i>P. FALCIPARUM</i> PROPLASMEPSIN II .....	167
4.4 CONCLUSION .....	170
4.5 REFERENCES .....	172
CHAPTER 5: LOOKING BEYOND THE STRUCTURES.....	174
5.1 PROPLASMEPSINS AS DRUG DESIGN TARGETS .....	174

5.2 MEMBRANE ATTACHMENT.....	176
5.3 DO THE PROPLASMEPSINS ASSIST IN THEIR OWN TRAFFICKING?.....	179
5.4 REFERENCES .....	182

## List of Figures

1.1	Worldwide distribution of malaria.	2
1.2	Life cycle of the <i>Plasmodium</i> parasite.	8
1.3	Hemoglobin degradation in <i>P. falciparum</i> .	11
2.1	Sequence alignment of aspartic proteinases.	22
2.2	The structure of pepstatin.	23
2.3	The structure of an archetypal aspartic proteinase, pepsin.	25
2.4	Psi loops of pepsin.	26
2.5	Hydrogen bonding pattern at the active site of pepsin.	27
2.6	Catalytic mechanism of aspartic proteinases.	29
2.7	Loop flexibility in aspartic proteinases.	32,33
2.8	Subdomain flexibility in aspartic proteinases.	34
2.9	Refinement progress for PvPM.	42,43
2.10	Ramachandran plot for the PvPM crystal structure.	44
2.11	The structure of PvPM.	46
2.12	Hydrogen bonding interactions at the active site of PvPM.	48
2.13	DDM comparing the two NCS-related PvPM molecules.	51
2.14	Superposition of the two NCS-related molecules of PvPM.	52
2.15	Interactions of pepstatin with PvPM.	54
2.16	Superposition of pepstatin in two NCS-related molecules.	56
2.17	The dimer in the asymmetric unit of the PfPMII crystal.	58,59
2.18	The dimer in the asymmetric unit of the PvPM crystal.	60,61
2.19	Superposition of the active site clefts of PvPM and PfPMII.	62
2.20	Modelled substrate in the S2' pockets of PvPM and PfPMII.	67,68
2.21	Superposition of the active site clefts of PvPM and cathepsin D.	70
3.1	The structures of pepsinogen and prophytepsin.	81,82
3.2	Blocked active sites of pepsinogen and prophytepsin.	83,84
3.3	Sequence alignment of AP zymogen prosegments.	87
3.4	Trafficking of proplasmepsin to the digestive vacuole in <i>P. falciparum</i> .	89
3.5	Refinement progress for pPfPMII.	96-99
3.6	Structure of pPfPMII.	101
3.7	Interactions of the prosegment with the mature portion of pPfPMII.	103

3.8	Activation of PfPMII.	104
3.9	The pro-mature junction in pPfPMII.	105
3.10	Alternate conformations of Ser 37 and Trp 41 in molecule 4 of pPfPMII.	107
3.11	DDM comparing the mature portion of pPfPMII to PfPMII.	110
3.12	N-terminal rearrangement of pPfPMII and PfPMII.	111
3.13	Interaction of the N-terminal residues 11 – 20 with the active site.	112,113
3.14	Superposition of residues 30 – 329 of pPfPMII and PfPMII.	116
3.15	The active site regions of pPfPMII and PfPMII.	117,118
3.16	Superposition of the 158 to 169 loop in pPfPMII and PfPMII.	123
3.17	Salt bridges involving prosegment residues in pPfPMII.	127
3.18	The role of the Lys110p to Val mutation in the autoactivation of PfPMI.	130
3.19	A model for the full-length pPfPMII.	132
4.1	Refinement progress for pPvPM.	145-147
4.2	Residue properties of pPvPM.	149
4.3	Distribution of temperature factors in pPvPM.	150,151
4.4	The structure of pPvPM.	154
4.5	Interactions of the prosegment with the mature portion of pPvPM.	155
4.6	The pro-mature junction of pPvPM.	167
4.7	The active site regions of pPvPM and PvPM.	158,159
4.8	Superposition of the two molecules in the pPvPM asymmetric unit.	161
4.9	DDM comparing the mature portion of pPvPM with PvPM.	163
4.10	The N-terminal rearrangement between pPvPM and PvPM.	164
4.11	Superposition of residues 22 to 327 of pPvPM and PvPM.	166
4.12	Superposition of pPvPM and pPfPMII.	169
5.1	Docking results for pPfPMII.	177
5.2	Schematic representation of membrane attachment in AP zymogens.	178
5.3	Sequence similarity of the cytoplasmic portion of pPfPMII and yeast $\delta$ -COP	181

## List of Tables

2.1	The currently known plasmepsins.	37
2.2	Sequence comparison (% identity) among plasmepsins and other selected APs.	38
2.3	Summary of data and refinement statistics for PvPM.	41
3.1	Summary of data and refinement statistics for pPfPMII.	93
3.2	Interactions of N-terminal residues 11 to 20 with the Psi loops in pPfPMII.	114
4.1	Summary of data and refinement statistics for pPvPM.	142
4.2	Molecular replacement results for pPvPM.	144



## Abbreviations

$ F_c $	calculated structure factor amplitude
$ F_o $	observed structure factor amplitude
AP	aspartic proteinase
DDM	difference distance matrix
HAP	histo-aspartic protein
MR	molecular replacement
NCS	non-crystallographic symmetry
PbPM	<i>P. berghei</i> plasmepsin
PfPMI - IV	<i>P. falciparum</i> plasmepsins I - IV
PmPM	<i>P. malariae</i> plasmepsin
PoPM	<i>P. ovale</i> plasmepsin
pPfPMI - IV	<i>P. falciparum</i> proplasmepsins I - IV
pPvPM	<i>P. vivax</i> proplasmepsin
PSI	plant-specific insert
PvPM	<i>P. vivax</i> plasmepsin
rmsd	root mean squared deviation
$V_m$	Mathews coefficient [(volume of unit cell) / (protein molecular mass) * (number of asymmetric units in the unit cell)]
WHO	The World Health Organization

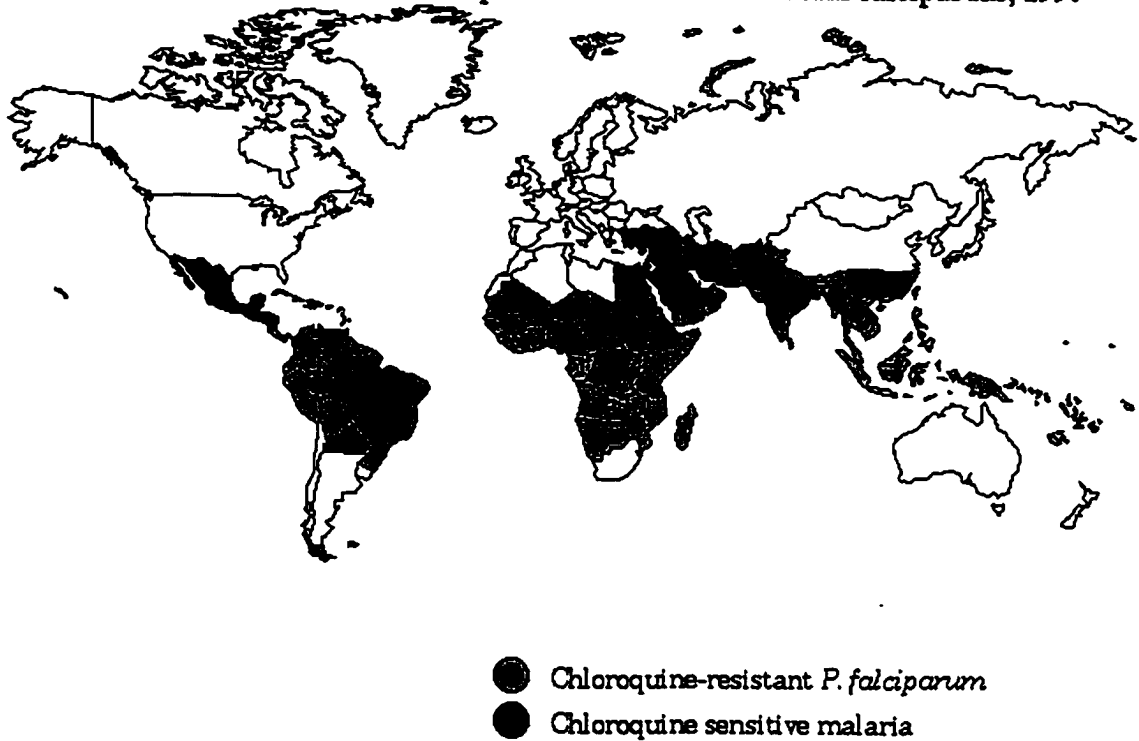
# Chapter 1: GENERAL INTRODUCTION

## 1.1 Malaria - past and present

Malaria is an ancient disease that, despite massive efforts aimed at its eradication and control, is still wreaking havoc in many areas of the world (Figure 1.1). Written references to malaria date as far back as 2700 BC. Malaria has had a major impact on human history, deciding outcomes of battles and influencing political and economic decisions. For instance, as recently as during World War II, 500 000 cases of malaria were recorded in the US army. General Douglas McArthur poignantly described the power of malaria: "This will be a long war, if for every division I have facing the enemy, I must count on a second division in the hospital with malaria, and a third division convalescing from this debilitating disease" [Russell *et al.*, 1946]. Malaria hampered the early development of the British colonies in North America. The construction of the Panama Canal was, for a long time, halted by a huge death toll from malaria and yellow fever. The spread of malaria is closely linked to ecology, and therefore, to economy. As various countries moved toward organized farming, the clearing of forests in favor of fields and rice paddies created new breeding ground for the malaria-carrying mosquitoes, sparking many epidemics.

Malaria has even caused extensive modification of the human genome. For instance, people who lack the Duffy blood group antigen, are resistant to infection with the malaria parasite *P. vivax* [Bruce-Chwatt, 1985]. Abnormalities in hemoglobin, such as the sickle-cell trait or  $\beta$ -thalassemia, render people less vulnerable to *P. falciparum*, the most dangerous malaria pathogen [Bruce-Chwatt, 1985]. A deficiency in glucose-6-phosphate dehydrogenase also affords partial protection against malaria [Hunt & Stocker, 1990]. The geographical distribution of these protective mutations parallels the worldwide distribution of malaria (Figure 1.1) [Bruce-Chwatt, 1985]. The Duffy antigen is absent in native West African populations. The sickle cell trait is found most commonly in Africa, or in people of African origin.  $\beta$ -thalassemia frequency is increased in parts of the Mediterranean where malaria existed in the past. Glucose-6-phosphate deficiency occurs in tropical Africa, the Mediterranean, Southwest Asia and India. It is likely that these

**Distribution of Malaria and Chloroquine- resistant *Plasmodium falciparum*, 1996**



**Figure 1.1** Worldwide distribution of malaria. (Division of Quarantine, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA.)

mutations have persisted in malarious areas due to the selective advantages that they provide against this disease.

The ancient Greeks and Romans realized that recurring fevers were somehow connected with stagnant bodies of water, such as swamps or marshes. The term "malaria" originated from the Italian "mal'aria", or "bad air", a name that was given to such intermittent fevers in the sixteenth century. In the late 1800s, extensive research by French, Italian and British scientists resulted in the identification of the etiological agent, *Plasmodium* parasites, and the transmission vector, *Anopheles* mosquitoes [Laveran, 1880; Ross, 1897; Grassi *et al.*, 1899]. Prior to this time, the only known remedies against malaria were quinine, derived from the bark of the Peruvian *Cinchona succirubra* tree, and the Chinese plant Qinghaosu or *Artemisia annua*, a source of artemisinin. In the 1930s and 1940s, new anti-malarial tools, the drug chloroquine (a safer derivative of quinine) and the insecticide DDT, were developed.

The availability of DDT, a cheap and powerful means of eliminating malaria transmission, stimulated the establishment of malaria eradication programs [Pampana & Russell, 1955; WHO, 1958]. Eradication achieved great success in some parts of the world, where a favorable combination of climate and socioeconomic conditions existed. In the United States, for instance, malaria virtually disappeared after five years of controlled indoor spraying with DDT [Institute of Medicine, 1991]. In other parts of the world, usually poor, politically unstable and tropical countries, eradication efforts proved unsuccessful. In yet other parts, Sub-Saharan Africa, the rate of transmission of malaria was so high, that eradication was not even attempted. Eventually, it was realized that where eradication was not working, malaria control (a less spectacular and less politically attractive concept) might be more effective [Gramiccia & Beales, 1988]. Responsibility for malaria control was gradually transferred to individual nations, many of whom lacked the technical and financial resources for this task. Furthermore, as the developed countries (the major sources of funding for eradication programs) no longer perceived malaria as a direct threat to them, international interest and financial support for the fight against malaria waned.

Far from disappearing, however, the problem was becoming more serious. One of the greatest difficulties was the development of resistance to drugs and insecticides. Today the problem of drug resistance is so widespread, that chloroquine is useless in most

of the malaria-affected regions (Figure 1.1) [Butler *et al.*, 1997a]. Multi-drug resistant strains of *P. falciparum* and chloroquine-resistant *P. vivax* have emerged [WHO, 1997]. Many of the drugs that have been developed to replace chloroquine are meeting the same fate [White, 1998]. A number of other drugs have extremely toxic side effects [Rosenthal, 1998]. DDT use and production are severely restricted due to environmental concerns [Institute of Medicine, 1991]. Political unrest and poverty are forcing population movements that place people at greater risk of malaria infection. In addition, war and civil strife have destroyed the health care infrastructures of a number of countries, facilitating the resurgence of malaria [WHO, 1997]. Economic pressures are leading to the same environmental disruption that has proved so favorable to the spread of malaria in the past. Logging, mining in forests, and the building of new settlements, are exposing more people to malaria infection [WHO, 1997]. Furthermore, changes in parasite distribution in favor of the deadly *P. falciparum* have been observed [WHO, 1997]. Indeed, the malaria situation today is dire.

Currently, over 40 % of the world's population lives in regions where malaria is endemic. Due to a lack of a standardized global monitoring system, it is difficult to give a precise quantitative description of the incidence of malaria. The World Health Organization (WHO) estimates that malaria infects 300 - 500 million people every year, and kills 1.5 - 2.7 million [WHO, 1997]. While malaria is still largely perceived as a third-world disease, it is having some direct impact on the developed countries, posing a potential danger to tourists, workers and military personnel going to tropical countries. In some cases, malaria has been transmitted by "stowaway mosquitoes", who inadvertently arrived on airplanes from malaria-endemic countries. Finally, several cases of autochthonous (locally transmitted) malaria have been reported in the US [Zucker, 1996]. Humanitarian concerns and a recognition of the potential renewed worldwide threat from this disease, have once again propelled malaria to the front of the global health care agenda. For example, the US National Institute of Allergy and Infectious Diseases and the UK's Wellcome Trust have steadily increased their funding of malaria research in the 1980s and 1990s [Butler *et al.*, 1997b]. The WHO and the Special Programme for Research and Training in Tropical Diseases (TDR) have recently established several programs specifically aimed at optimizing existing strategies and developing new methods of malaria control. These programs include the Roll Back Malaria project (RBM), the

Multilateral Initiative for Malaria in Africa (MIM) and the Medicines for Malaria Venture (MMV). Over the past few years, the pharmaceutical industry had withdrawn from malaria research, considering it economically unfeasible. The MMV, an unprecedented alliance between the pharmaceutical industry and the public sector, now aims to bring the expertise and technological capabilities of the pharmaceutical companies back into the malaria field. The aim of the MMV is to develop new anti-malarial drugs that will be both effective and affordable, and thus useful to the people who need them most [Ridley & Gutteridge, 1999].

The battle against malaria involves a complex interplay of factors. Aside from the consideration that must be given to drug resistance and toxicity and various socio-economic factors, one must take into account the immunological relationship between malaria and humans [Snow & Marsh, 1998]. In areas of intense and constant malaria transmission, the holoendemic areas, people develop immunity to malaria infection. In these areas malaria poses the greatest threat in early childhood, after the child loses the passive immunity inherited from the mother but before developing his or her own immunity. All children in these regions, mainly sub-Saharan Africa, become infected with malaria. The majority survive and acquire partial immunity. Still many, as many as a million per year, die [WHO, 1997]. In fact, these young children account for most of the malaria death toll. The immune individuals rarely suffer from severe malaria. However, in order to maintain immunity, an individual must be continually exposed to the infection. If the exposure is interrupted, due to a temporary absence from the malarious area, or due to a malaria control effort, the person loses immunity and again becomes highly susceptible to malaria [Molineaux, 1988]. The role of immunity must be taken into consideration in any malaria control measure to be implemented in holoendemic regions. For instance, it may be advantageous to treat the illness, rather than preventing infection completely. In areas of seasonal or occasional malaria transmission, where people have little or no immunity, the approach to malaria must be different. Due to inadequate immunity in their population, these areas are at a high risk for malaria epidemics. In these regions preventing infection, for example with the use of insecticide-impregnated bed nets, may be effective. Clearly, since malaria has such diverse effects in different parts of the world, and even on different individuals, no single method of malaria control will be universally applicable. Effective malaria control measures must be tailored to specific local situations,

and the optimal tools used in the fight against this disease are likely to vary in different cases.

The currently available choice of anti-malarial tools is becoming depleted, mainly due to the development of resistance, and new tools are urgently needed. The possibility of developing immunity to malaria has given hope for a malaria vaccine [Kwiatkowski & Marsh, 1997]. Unfortunately, the great deal of research devoted to this subject has, so far, achieved little success. Several new anti-malarial drugs are in the process of development, but they also have drawbacks. [Rosenthal, 1998]. For example, derivatives of artemisinin, an organic peroxide that likely damages the parasite's membrane proteins by free radical alkylation, appear promising [Meshnick *et al.*, 1996]. These substances act rapidly and are effective against drug-resistant malaria, but show rapid clearance from the body and recurrence of disease after the end of treatment [deVries & Dien, 1996]. Another drug, halofantrine, exhibits limited bioavailability [Bryson & Goa, 1992]. It is imperative to improve our understanding of various aspects of malaria, and to identify new targets for therapeutic intervention.

## 1.2 Biology of malaria

Malaria is caused by protozoan parasites of the genus *Plasmodium*, and affects a wide range of vertebrates, including birds, reptiles, primates and rodents [Knell, 1991]. Four species of *Plasmodium* infect humans, *P. falciparum*, *P. vivax*, *P. ovale* and *P. malariae*. The latter three species, *P. vivax*, *P. ovale* and *P. malariae*, cause the disease most commonly characterized by recurring fever. *P. vivax* and *P. ovale* cause tertian malaria, where the fever repeats every three days, whereas *P. malariae* causes quartan malaria, with the fever repeating every four days. *P. vivax* is the most widely distributed of the parasites, affecting temperate and tropical zones [Bruce-Chwatt, 1985]. *P. falciparum* is the most virulent, potentially causing fatal disease. *Falciparum* malaria is characterized by irregular tertian fever, and, if left untreated, can rapidly progress into coma and death in non-immune individuals. *P. ovale* and *P. malariae* are more rare. *P. falciparum* appears to have adapted to human hosts much more recently than the other three human malaria pathogens. Its lethal effects on the host attest to less than optimal adaptation. In addition, *P. falciparum* appears to be genetically closer to the avian and rodent pathogens than to *Plasmodium* species that infect primates [Knell, 1991].

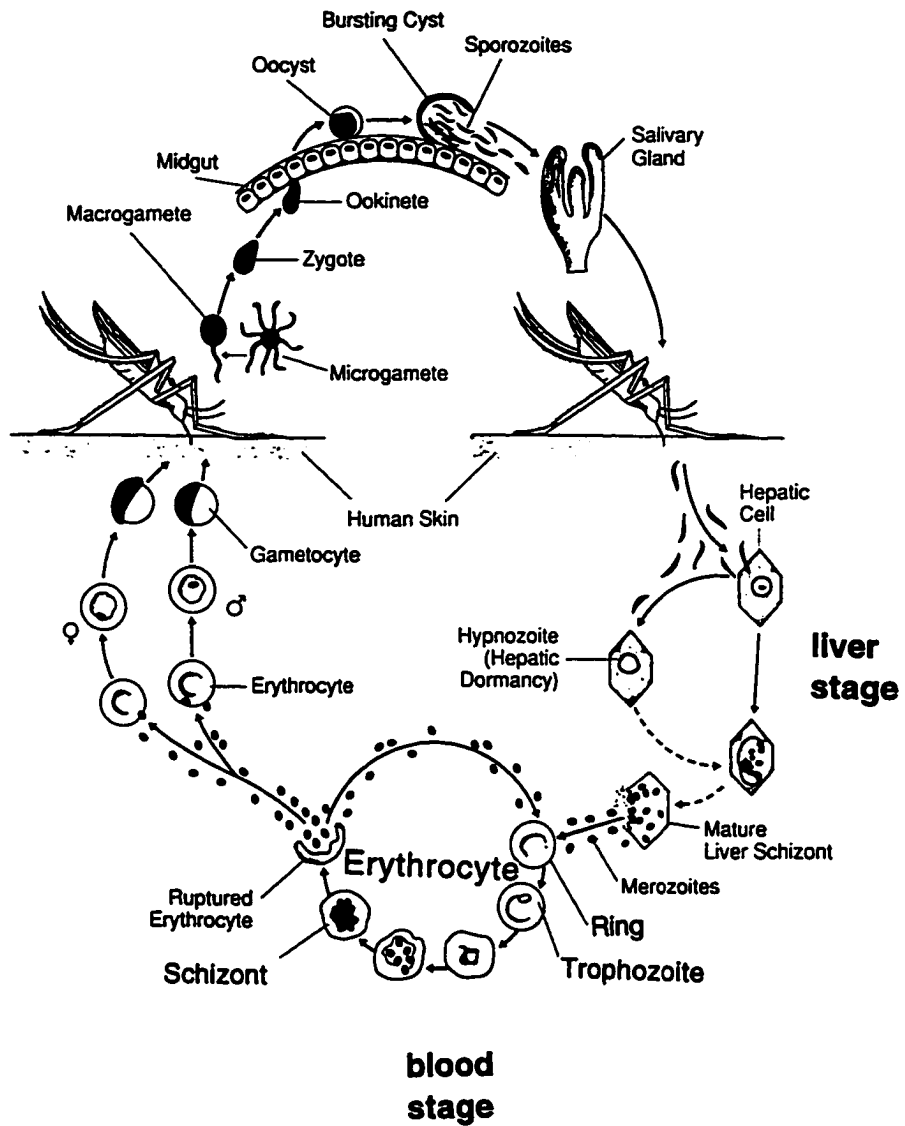
Malaria is transmitted by 50 - 60 species of *Anopheles* mosquitoes [Bruce-Chwatt, 1985]. Egg production by the female mosquito requires feeding on blood. When a female mosquito bites a human infected with malaria, she ingests the parasites along with the blood. After developing in the mosquito, the parasites are injected into another human during the mosquito's next blood meal. In hot and humid climates, the lifetime of the mosquito is long enough, and the development of the parasite is fast enough, to result in effective transmission of malaria. Although *Anopheles* mosquitoes can be found in cooler climates, such as Canada, their life span is shorter and the development of the parasite is significantly slower, decreasing the probability of malaria transmission [Molineaux, 1988]. The mosquito is called the "vector" in malaria transmission, and the human, or other vertebrate, is the "host". While it is the host who may become ill, the vector may also be affected by the malaria infection. Recent studies have shown that malaria infection modifies the behavior of mosquitoes, making them more aggressive toward humans [Koella *et al.*, 1998; Koella & Packer, 1996].

### 1.2.1 Life cycle of the Plasmodium parasite

The malaria parasite has a complex life cycle, transferring between its mosquito vector and human host, and adopting a variety of morphologically distinct forms (Figure 1.2) [Knell, 1991]. When a female mosquito bites a malaria-infected person, she ingests *Plasmodium* gametocytes. The parasite reproduces sexually within the mosquito, going through the zygote, ookinete and oocyst stages, finally ending up as sporozoites in the insect's salivary glands. The rate of sporogony, or parasite development in the mosquito, depends on the species of *Plasmodium* and on environmental factors, such as temperature and humidity. The most favorable conditions for sporogony are a temperature between 20 and 30°C and over 60 % humidity. For *P. vivax*, sporogony lasts 8 to 10 days at 28°C, and for *P. falciparum*, 9 to 11 days [Macdonald, 1957]. Sporogony usually cannot occur at temperatures below 15°C.

When the mosquito bites another person, the sporozoites are injected into the new host's blood. The parasites initially travel to the liver, where they develop into schizonts. This is the asymptomatic liver stage, or hepatic stage, of malaria, that lasts 9 - 16 days. The schizonts produce large numbers of merozoites, which leave the liver cells and invade red blood cells. The following step in parasite development is called the blood, or





**Figure 1.2** Life cycle of the *Plasmodium* parasite. (Adapted from [Institute of Medicine, 1991].)

erythrocytic, stage. This stage is further subdivided into the ring, trophozoite and schizont stages. Within the erythrocyte, the invading merozoite is enclosed by a single, erythrocyte-derived membrane in the parasitophorous vacuole. During the schizont stage, the parasite undergoes up to 5 rounds of asexual reproduction to yield new merozoites, that rupture the erythrocyte and invade other erythrocytes. The clinical symptoms of malaria, such as fever and anaemia, accompany the rupturing of red blood cells. The fever is caused by pyrogenic substances that are released into the plasma from the ruptured erythrocytes. The process of erythrocyte invasion and rupture repeats, leading to a periodic recurrence of fever. Eventually, differentiation of merozoites produces gametocytes, which can be ingested by a mosquito, completing the parasite's life cycle.

### **1.2.2 Hemoglobin degradation by plasmodium**

During the blood stage of malaria, the parasites imbibe the host erythrocyte's cytoplasm and digest its most abundant protein component, hemoglobin, as a source of metabolic energy. It has been estimated that the *Plasmodium* parasites digest as much as 80 % of the host cell's hemoglobin [Francis *et al.*, 1997]. Most of the hemoglobin degradation occurs during the metabolically active trophozoite stage [Francis *et al.*, 1997]. The erythrocyte's cytoplasm is ingested through a double membrane-enclosed structure called the cytostome. From there, hemoglobin-laden vesicles bud, eventually forming single membrane-enclosed digestive vacuoles. The digestive vacuole is the principal site of hemoglobin degradation [Goldberg *et al.*, 1990; Olliaro & Goldberg, 1995]. This is an acidic organelle, reminiscent of the lysosome, that contains the enzymes required for the breakdown of hemoglobin.

Hydrolysis of hemoglobin produces short peptides and heme [Francis *et al.*, 1997]. The peptides are exported into the parasite's cytoplasm, where they are apparently further degraded by an aminopeptidase [Kolakovich *et al.*, 1997]. The resultant free amino acids are used by the parasite for protein synthesis. Free heme, a potentially toxic byproduct of hemoglobin digestion, is detoxified by polymerization to hemozoin, or malaria pigment. Hemozoin is a crystalline polymer consisting of chains of heme moieties linked by interaction of the propionate group of one heme with the iron atom of the next [Slater *et al.*, 1991]. Chloroquine is believed to function by inhibiting heme polymerization [Slater, 1993; Bray *et al.*, 1998]. In addition to sequestering the dangerous

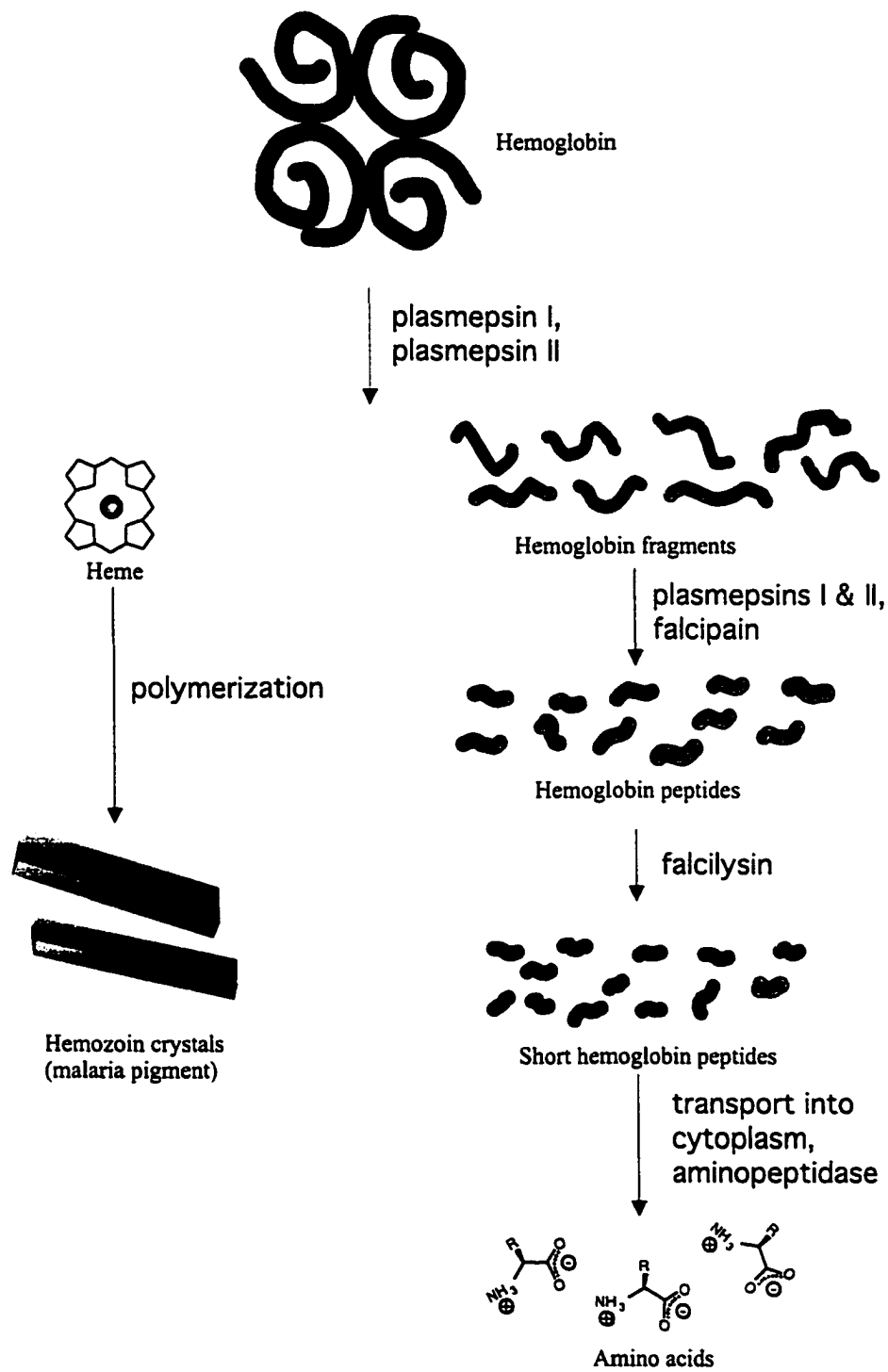
free heme, hemozoin has been shown to interact with the host's immune system [Arese & Schwarzer, 1997; Pichyangkul *et al.*, 1994; Schwarzer *et al.*, 1992]. When heme is released from hemoglobin, the oxidation state of its iron changes from +2 to +3, with a concomitant reduction of molecular oxygen to various reactive species, such as  $O_2^-$ ,  $OH^-$ , and  $H_2O_2$ . Superoxide dismutase and catalase, most likely obtained from the host during hemoglobin ingestion, contain the oxidative damage that these species can cause to the parasite [Fairfield *et al.*, 1988; Fairfield *et al.*, 1983].

Hemoglobin degradation is necessary for the parasite's survival, and the inhibition of this process, as described below, kills the parasites in culture [Francis *et al.*, 1994; Silva *et al.*, 1996; Moon *et al.*, 1997; Rosenthal, 1998; Berry, 1999]. Thus the hemoglobin degradation pathway of *Plasmodium* represents a useful point for chemotherapeutic intervention. In fact, the WHO has recognized the hemoglobin-degrading proteinases of *P. falciparum* as new anti-malarial drug design targets [Tyas *et al.*, 1999].

### 1.2.3 Proteinases of *Plasmodium*

The *Plasmodium* parasite has many proteolytic enzymes that perform a variety of functions. During the erythrocytic cycle, for instance, proteinases are required for red blood cell invasion, processing of merozoite surface antigens, hemoglobin digestion and merozoite release [Schrevel *et al.*, 1990].

Four proteinases have been purified from the digestive vacuoles of *P. falciparum*, and shown to participate in hemoglobin degradation. These hemoglobinses are two aspartic proteinases, plasmepsins I and II, a cysteine proteinase, falcipain, and a metalloproteinase, falcilysin [Goldberg *et al.*, 1991; Gluzman *et al.*, 1994; Francis *et al.*, 1994; Salas *et al.*, 1995; Francis *et al.*, 1996; Eggleston *et al.*, 1999]. The proteinases digest hemoglobin in an ordered fashion (Figure 1.3) [Gluzman *et al.*, 1994; Francis *et al.*, 1997; Eggleston *et al.*, 1999]. The initial cleavage of hemoglobin is made by the plasmepsins at the Phe 33 - Leu 34 bond of the  $\alpha$  chain. Residues 33 and 34 are located at the C-terminus of helix B, near the structurally important hinge region, which is found at the interface between the  $\alpha 1$  and  $\beta 1$  subunits of hemoglobin. Some local unfolding at low pH likely exposes this bond to the proteinase. Cleavage of the Phe 33 - Leu 34 bond may then cause further disruption of the hemoglobin tertiary structure, exposing other scissile bonds to proteolysis. After the initial cleavage, the plasmepsins and falcipain



**Figure 13** Hemoglobin degradation in *P. falciparum*.

hydrolyze the denatured hemoglobin to peptides. The peptides, which are 11 - 20 amino acids in length, can then be degraded by falcilysin to shorter peptides. The short peptides are transported out of the digestive vacuole, possibly by the *P. falciparum* P-glycoprotein homologue 1, a transporter protein that has been implicated in chloroquine resistance [Kolakovich *et al.*, 1997]. Aminopeptidase activity has been found in *Plasmodium*, and it appears to function in the cytoplasm, breaking down the short peptides from the digestive vacuole to individual amino acids [Florent *et al.*, 1998].

There is some debate in the literature regarding the exact role of the plasmepsins and falcipain in hemoglobin degradation. For instance, there are conflicting reports about falcipain being able to hydrolyze native hemoglobin [Gluzman *et al.*, 1994; Salas *et al.*, 1995; Francis *et al.*, 1996]. Questions about plasmepsin II center on the fact that a plasmepsin I-specific inhibitor blocks hemoglobin degradation, and that plasmepsin II cannot compensate for the loss of plasmepsin I activity [Francis *et al.*, 1994]. In addition, it has recently been shown that plasmepsin II can degrade proteins of the erythrocyte cytoskeleton at neutral pH, suggesting a possible auxiliary role in erythrocyte rupture [Le Bonniec *et al.*, 1999].

It has been shown that inhibitors of both aspartic and cysteine proteinases block hemoglobin digestion in *Plasmodium* and kill the parasite in culture [Francis *et al.*, 1994; Silva *et al.*, 1996; Moon *et al.*, 1997; Rosenthal, 1998; Berry, 1999]. Inhibitors of these proteinases may, therefore, have applications as anti-malarial drugs. All these hemoglobin-degrading proteinases are initially produced by the organism as inactive precursors, or zymogens. The zymogens become converted to mature, active proteinases when their hydrolytic function is required. The interruption of these conversion, or activation, processes may also yield novel anti-malarial therapies.

Any inhibitor design program is likely to benefit from the detailed molecular information that can be provided by a three-dimensional structure of the target protein. This thesis presents the structural studies of the conversion of two malarial aspartic proteinase zymogens, *P. falciparum* proplasmepsin II and *P. vivax* proplasmepsin, to the active enzymes.

The crystal structure of *P. falciparum* plasmepsin II (PfPMII) was solved earlier in complex with the inhibitor pepstatin A by A.M. Silva and coworkers [Silva *et al.*, 1996]. Extensive use has been made of this structural information, not only to rationalize the

substrate preferences of this and other plasmepsins, but to design a variety of inhibitors that may have potential as novel anti-malarial drugs [Silva *et al.*, 1996; Haque *et al.*, 1999; Westling *et al.*, 1997].

Chapter 2 of this thesis presents the X-ray crystal structure of a second malarial proteinase, plasmepsin from *Plasmodium vivax* (PvPM). While this enzyme is generally very similar to PfPMII, there are some subtle differences, which may be relevant to the function of these enzymes. Chapter 2 also serves as an introduction to aspartic proteinases, which will be helpful later, in our discussion of aspartic proteinase zymogen activation.

Chapters 3 and 4 deal with the activation of plasmepsins. In Chapter 3, the crystal structure of *P. falciparum* proplasmepsin II (pPfPMII) is described. Prior to the elucidation of the pPfPMII structure, our understanding of AP activation was limited, on a structural level, to the gastric enzymes pepsin and gastricsin [Sielecki *et al.*, 1991; Hartsuck *et al.*, 1992; Moore *et al.*, 1995; Khan *et al.*, 1997; Bateman *et al.*, 1998]. Soon after the publication of the pPfPMII structure [Bernstein *et al.*, 1999], the crystal structure of the barley AP zymogen, prophytepsin, was reported [Kervinen *et al.*, 1999]. While there are some important differences between prophytepsin and the gastric zymogens, all these precursors utilize similar mechanisms of inactivation. Proplasmepsin II, on the other hand, uses a completely different and unexpected way to prevent proteolytic activity, which not only makes for a fascinating biological story, but also provides a potential new target for anti-malarial drug design. Chapter 4 describes the crystal structure of proplasmepsin from *P. vivax*, which reproduces most of the unusual structural features seen in pPfPMII, and allows us to formulate some general conclusions about the mode of inactivation of the proplasmepsins, and possibly the mechanism of their conversion to the active enzymes.

Chapter 5 concludes the thesis with a discussion of proplasmepsins in the context of the more recently characterized APs, and with some ideas on the potential usefulness of proplasmepsins in the design of new anti-malarial drugs.

### 1.3 References

- Arese P, Schwarzzer E (1997) Malarial pigment (haemozoin): a very active 'inert' substance. *Ann Trop Med Parasitol* **91**: 501-16.
- Bateman KS, Chernaia MM, Tarasova NI, James MNG. (1998) Crystal structure of human pepsinogen A. *Adv Exp Med Biol* **436**: 259-63.
- Bernstein NK, Cherney MM, Loetscher H, Ridley RG, James MNG (1999) Crystal structure of the novel aspartic proteinase zymogen proplasmepsin II from *Plasmodium falciparum*. *Nat Struct Biol* **6**:32-7.
- Berry, C (1999) Proteases as drug targets for the treatment of malaria. In *Proteases of Infectious Agents*. (Dunn BM, ed. Academic Press, San Diego), pp. 165-188.
- Bray PG, Mungthin M, Ridley RG, Ward SA (1998) Access to hemozoin: the basis of chloroquine resistance. *Mol Pharmacol* **54**: 170-9.
- Bruce-Chwatt LJ (1985) *Essential malariology*. John Wiley & Sons, New York.
- Bryson HM, Goa KL (1992) Halofantrine. A review of its antimalarial activity, pharmacokinetic properties and therapeutic potential. *Drugs* **43**: 236-58.
- Butler D, Maurice J, O'Brien C (1997a) Time to put malaria control on the global agenda. *Nature* **386**: 535-536.
- Butler D, Maurice J, O'Brien C (1997b) Can industry be wooed back into the act? *Nature* **386**: 540.
- de Vries PJ, Dien TK (1996) Clinical pharmacology and therapeutic potential of artemisinin and its derivatives in the treatment of malaria. *Drugs* **52**: 818-36.
- Eggleson KK, Duffin KL, Goldberg DE (1999) Identification and characterization of falcilysin, a metalloproteinase in hemoglobin catabolism within the malaria parasite *Plasmodium falciparum*. *J Biol Chem* **274**: 32411-7.
- Fairfield AS, Abosch A, Ranz A, Eaton JW, Meshnick SR (1988) Oxidant defense enzymes of *Plasmodium falciparum*. *Mol Biochem Parasitol* **30**: 77-82.

Fairfield AS, Meshnick SR, Eaton JW (1983) Malaria parasites adopt host cell superoxide dismutase. *Science* **221**: 764-6.

Florent I, Derhy Z, Allary M, Monsigny M, Mayer R, Schrevel J (1998) A *Plasmodium falciparum* aminopeptidase gene belonging to the M1 family of zinc-metallopeptidases is expressed in erythrocytic stages. *Mol Biochem Parasitol* **97**: 149-60.

Francis SE, Banerjee R, Goldberg DE (1997) Biosynthesis and maturation of the malaria aspartic hemoglobinases plasmepsins I and II. *J Biol Chem* **272**: 14961-8.

Francis SE, Gluzman IY, Oksman A, Banerjee D, Goldberg DE (1996) Characterization of native falcipain, an enzyme involved in *Plasmodium falciparum* hemoglobin degradation. *Mol Biochem Parasitol* **83**: 189-200.

Francis SE, Gluzman IY, Oksman A, Knickerbocker A, Mueller R, Bryant ML, Sherman DR, Russell DG, Goldberg DE (1994) Molecular characterization and inhibition of a *Plasmodium falciparum* aspartic hemoglobinase. *EMBO J* **13**: 306-17.

Fusek M, Vetvicka V (1995) *Aspartic proteinases: physiology and pathology*. (CRC Press Inc., Boca Raton).

Gluzman IY, Francis SE, Oksman A, Smith CE, Duffin KL, Goldberg DE (1994) Order and specificity of the *Plasmodium falciparum* hemoglobin pathway. *J Clin Invest* **93**: 1602-8.

Goldberg DE, Slater AF, Beavis R, Chait B, Cerami A, Henderson GB (1991) Hemoglobin degradation in the human malaria pathogen *Plasmodium falciparum*: a catabolic pathway initiated by a specific aspartic protease. *J Exp Med* **173**: 961-9.

Goldberg DE, Slater AF, Cerami A, Henderson GB (1990) Hemoglobin degradation in the malaria parasite *Plasmodium falciparum*: an ordered process in a unique organelle. *Proc Natl Acad Sci U S A* **87**: 2931-5.

Gramiccia G, Beales PF (1988) The recent history of malaria control and eradication. In *Malaria – principles and practice of malariology*, Vol. 2 (Wernsdorfer WH & McGregor I, eds. Churchill Livingstone, London), pp. 1335-1378.

Grassi B, Bignami A, Bastianelli G (1899) Ulteriori ricerche sulla malaria. *Rendiconti della Reale Accademia dei Lincei, Roma* **VIII**: 434-438.



Haque TS, Skillman AG, Lee CE, Habashita H, Gluzman IY, Ewing TJ, Goldberg DE, Kuntz ID, Ellman JA (1999) Potent, low-molecular-weight non-peptide inhibitors of malarial aspartyl protease plasmepsin II. *J Med Chem* **42**: 1428-40.

Hartsuck JA, Koelsch G, Remington SJ (1992) The high-resolution crystal structure of porcine pepsinogen. *Proteins* **13**: 1-25.

Hunt NH, Stocker R (1990) Oxidative stress and the redox status of malaria-infected erythrocytes. *Blood Cells* **16**:499-526.

Institute of Medicine (U.S.) (1991) *Malaria: obstacles and opportunities: a report of the Committee for the Study on Malaria Prevention and Control: Status Review and Alternative Strategies, Division of International Health*. (Institute of Medicine / Oaks Jr SC. *et al.*, eds. National Academy Press, Washington, D.C.).

James MNG, Sielecki AR (1986) Molecular structure of an aspartic proteinase zymogen, porcine pepsinogen, at 1.8 Å resolution. *Nature* **319**: 33-8.

Kervinen J, Tobin GJ, Costa J, Waugh DS, Wlodawer A, Zdanov A (1999) Crystal structure of plant aspartic proteinase prophytepsin: inactivation and vacuolar targeting. *EMBO J* **18**: 3947-55.

Khan AR, Cherney MM, Tarasova NI, James MNG (1997) Structural characterization of activation 'intermediate 2' on the pathway to human gastricsin. *Nat Struct Biol* **4**: 1010-5.

Knell AJ (1991) *Malaria*. Oxford University Press, Oxford.

Koella JC, Packer MJ (1996) Malaria parasites enhance blood-feeding of their naturally infected vector *Anopheles punctulatus*. *Parasitology* **113**: 105-109.

Koella JC, Sorensen FL, Anderson RA (1998) The malaria parasite, *Plasmodium falciparum*, increases the frequency of multiple feeding of its mosquito vector, *Anopheles gambiae*. *Proc R Soc Lond B Biol Sci* **265**: 763-768.

Kolakovich KA, Gluzman IY, Duffin KL, Goldberg DE (1997) Generation of hemoglobin peptides in the acidic digestive vacuole of *Plasmodium falciparum* implicates peptide transport in amino acid production. *Mol Biochem Parasitol* **87**: 123-35.

Kwiatkowski D, Marsh K (1997) Development of a malaria vaccine. *Lancet* **350**: 1696-1701.

Laveran A (1880) Note sur un nouveau parasite trouvé dans le sang de plusieurs malades atteints de fièvre palustre. *Bulletin de l'Académie Médecine*, 2<sup>nd</sup> Series, 9: 1235-1236.

Le Bonniec S, Deregnacourt C, Redeker V, Banerjee R, Grellier P, Goldberg DE, Schrevel J (1999) Plasmepsin II, an acidic hemoglobinase from the *Plasmodium falciparum*. *J Biol Chem* 274: 14218-23.

Macdonald G (1957) *The epidemiology and control of malaria*. Oxford University Press, London.

Meshnick SR, Taylor TE, Kamchonwongpaisan S (1996) Artemisinin and the antimalarial endoperoxides: from herbal remedy to targeted chemotherapy. *Microbiol Rev* 60: 301-315.

Molineaux L (1988) The epidemiology of human malaria as an explanation of its distribution, including some implications for its control. In *Malaria – principles and practice of malariology*, Vol. 2 (Wernsdorfer WH, McGregor I, eds., Churchill Livingstone, London), pp. 913 – 998.

Moon RP, Tyas L, Certa U, Rupp K, Bur D, Jacquet C, Matile H, Loetscher H, Grueninger-Leitch F, Kay J, Dunn BM, Berry C, Ridley RG (1997) Expression and characterisation of plasmepsin I from *Plasmodium falciparum*. *Eur J Biochem* 244: 552-60.

Moore SA, Sielecki AR, Chernaia MM, Tarasova NI, James MNG (1995) Crystal and molecular structures of human progastricsin at 1.62 Å resolution. *J Mol Biol* 247: 466-85.

Olliaro PL & Goldberg DE (1995) The *Plasmodium* digestive vacuole: metabolic headquarters and choice drug target. *Parasitol Today* 11: 294-97.

Pampana EJ, Russell PF (1955) Malaria – a world problem. *World Health Organization Chronicle* 9: 31–96

Pichyangkul S, Saengkrai P, Webster HK (1994) *Plasmodium falciparum* pigment induces monocytes to release high levels of tumor necrosis factor-alpha and interleukin-1 beta. *Am J Trop Med Hyg* 51: 430-5.

Ridley RG, Gutteridge WE (1999) Medicines for Malaria Venture. Paper presented at the Médecins sans Frontières Symposium: Drugs for Communicable Diseases - Stimulating Development and Availability; Paris, October 15th 1999.

Rosenthal PJ (1998) Proteases of malaria parasites: new targets for chemotherapy. *Emerg Infect Dis* 4: 49-57.

Ross R (1897) On some peculiar pigmented cells found in two mosquitoes fed on malarial blood. *British Medical Journal*, 2: 1786-1788.

Russell PF, West CS, Manwell RD (1946) *Practical malariology*. W.B. Saunders Company, Philadelphia.

Salas F, Fichmann J, Lee GK, Scott MD, Rosenthal PJ (1995) Functional expression of falcipain, a *Plasmodium falciparum* cysteine proteinase, its role as a malarial hemoglobinase. *Infect Immun* 63: 2120-5.

Schrevel J, Deguercy A, Mayer R, Monsigny M (1990) Proteases in malaria-infected red blood cells. *Blood Cells* 16: 563-584.

Schwarzer E, Turrini F, Ulliers D, Giribaldi G, Ginsburg H, Arese P (1992) Impairment of macrophage functions after ingestion of *Plasmodium falciparum*-infected erythrocytes or isolated malarial pigment. *J Exp Med* 176: 1033-41.

Sielecki AR, Fujinaga M, Read RJ, James MNG (1991) Refined structure of porcine pepsinogen at 1.8 Å resolution. *J Mol Biol* 219: 671-92.

Silva AM, Lee AY, Gulnik SV, Maier P, Collins J, Bhat TN, Collins PJ, Cachau RE, Luker KE, Gluzman IY, Francis SE, Oksman A, Goldberg DE, Erickson JW (1996) Structure and inhibition of plasmepsin II, a hemoglobin-degrading enzyme from *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* 93: 10034-9.

Slater AF (1993) Chloroquine: mechanism of drug action and resistance in *Plasmodium falciparum*. *Pharmacol Ther* 57: 203-35.

Slater AF, Swiggard WJ, Orton BR, Flitter WD, Goldberg DE, Cerami A, Henderson GB (1991) An iron-carboxylate bond links the heme units of malaria pigment. *Proc Natl Acad Sci U S A* 88: 325-9.

Snow RW, Marsh K (1998) New insights into the epidemiology of malaria relevant for disease control. *Br Med Bull* 54: 293-309.

Tyas L, Gluzman I, Moon RP, Rupp K, Westling J, Ridley RG, Kay J, Goldberg DE, Berry C (1999) Naturally-occurring and recombinant forms of the aspartic proteinases plasmepsins I and II from the human malaria parasite *Plasmodium falciparum*. *FEBS Lett* 454: 210-4.

Westling J, Yowell CA, Majer P, Erickson JW, Dame JB, Dunn BM (1997) *Plasmodium falciparum*, *P. vivax*, and *P. malariae*: a comparison of the active site properties of plasmepsins cloned and expressed from three different species of the malaria parasite. *Exp Parasitol* 87: 185-93.

White, NJ (1998) Drug resistance in malaria *Br Med Bull* 54: 703-712.

World Health Organization (1958) *The first ten years of the World Health Organization*. World Health Organization, Geneva.

World Health Organization (1997) World malaria situation in 1994, part I. *Wkly Epidemiol Rec* 72: 269-276.

Zucker JR (1996) Changing patterns of autochthonous malaria transmission in the United States: a review of recent outbreaks. *Emerg Infect Dis* 2: 37-43.

## Chapter 2: CRYSTAL STRUCTURE OF *P. vivax* PLASMEPSIN

### 2.1 Introduction

#### 2.1.1 Biological role of aspartic proteinases

Aspartic proteinases (APs) comprise one of the four families into which proteolytic enzymes are classified on the basis of their catalytic mechanism. The other three families are the serine proteinases, cysteine proteinases and metalloproteinases. Aspartic proteinases are found in many organisms, including fungi, plants, animals, parasites and retroviruses; they carry out a wide variety of biological functions. These range from rather unspecific tasks such as digestion of dietary protein in the stomachs of vertebrates by the gastric APs (pepsin, gastricsin), to the extremely precise cleavage of angiotensinogen into angiotensin I by renin, an important step in the regulation of blood pressure. In addition, APs carry out intracellular protein catabolism (cathepsin D), antigen processing (cathepsins D and E) and viral polyprotein processing (HIV proteinase) [Fusek & Vetvicka, 1995]. The APs chymosin, cardosin and various microbial aspartic proteinases are used as milk clotting agents in cheese production [Faro *et al.*, 1999; Aikawa *et al.*, 1992]. A number of APs are also linked to disease. For instance, serum levels of pepsinogen may be used as an indicator for stomach cancer, and procathepsin D is secreted by certain breast tumors [Yoshihara *et al.*, 1998; Vetvicka *et al.*, 1997]. Direct involvement of APs in pathogenesis includes hemoglobin degradation by malarial plasmepsins, and the action of *Candida albicans* and *C. tropicalis* aspartic proteinases [Francis *et al.*, 1994; Hube *et al.*, 1996; Togni *et al.*, 1994]. Due to their association with disease and metabolic regulation, many APs are attractive targets for pharmaceutical intervention. For instance, HIV proteinase inhibitors are currently in use as drugs against AIDS, inhibitors of renin can be used to treat hypertension, and plasmepsin inhibitors have the potential to become anti-malarial drugs (see Chapter 1).

To avoid to the potential dangers of uncontrolled proteolysis, the activity of APs is tightly regulated. All known APs are produced as inactive precursors, zymogens, which

become converted to active (mature) proteinases when their function is required. Activation of APs will be discussed in more detail in Chapters 3 and 4 of this thesis, which are devoted to the activation of plasmepsins.

### 2.1.2 Structure and mechanism of aspartic proteinases

Eukaryotic APs, such as pepsin or cathepsin D, typically contain 325 to 350 amino acids, and are either monomeric or consist of more than one protein chain generated by proteolytic cleavage of a single chain. APs from retroviruses, such as HIV proteinase, are much smaller, with only about 100 to 125 amino acids, but function as homodimers. The APs share considerable sequence similarity, including several invariant or highly conserved regions (Figure 2.1). Most notable among the conserved regions are two D-T/S-G-T/S motifs in the eukaryotic APs. AP residue numbering is traditionally compared to porcine pepsin, where these motifs are numbered 32 to 35 and 215 to 218 (34 to 37 and 214 to 217 in plasmepsin numbering). A single similar sequence is conserved in the retroviral APs: DTGA (25 to 28 in HIV-1 proteinase). Since an active retroviral AP is a dimer, it contains two of these motifs. The invariant aspartates in these motifs are the catalytic residues of the AP. Asps 32 and 215 are susceptible to specific labeling by covalent inhibitors diazoacetyl norleucine methylester (DAN) and 1,2-epoxy-3-(*p*-nitrophenoxy)propane (EPNP), respectively [Rajagopalan *et al.*, 1966; Tang, 1971]. This pattern of reactivity indicates not only that the two aspartates are important for catalysis, but that they are chemically different. Furthermore, mutagenesis experiments confirmed the requirement of the two aspartates for catalytic activity [Lin *et al.*, 1989]. APs are also characterized by specific inhibition by pepstatin A, a hexapeptide, produced by *Streptomyces*, that contains the unusual amino acid statine (Sta) (Figure 2.2). As will be described below, statine mimics the transition state in AP catalysis.

The three-dimensional structures of a number of APs from various sources have been determined, and, together with results from biochemical studies, have yielded a wealth of information regarding these enzymes. Aspartic proteinases have a bilobal, or crescent, shape. The two lobes of eukaryotic APs contain an approximate two-fold internal symmetry, that is consistent with the similarity between the N- and C-terminal portions of their sequences and is believed to be the result of gene duplication [Tang *et al.*, 1978]. Retroviral APs, being homodimers, contain exact two-fold internal symmetry.

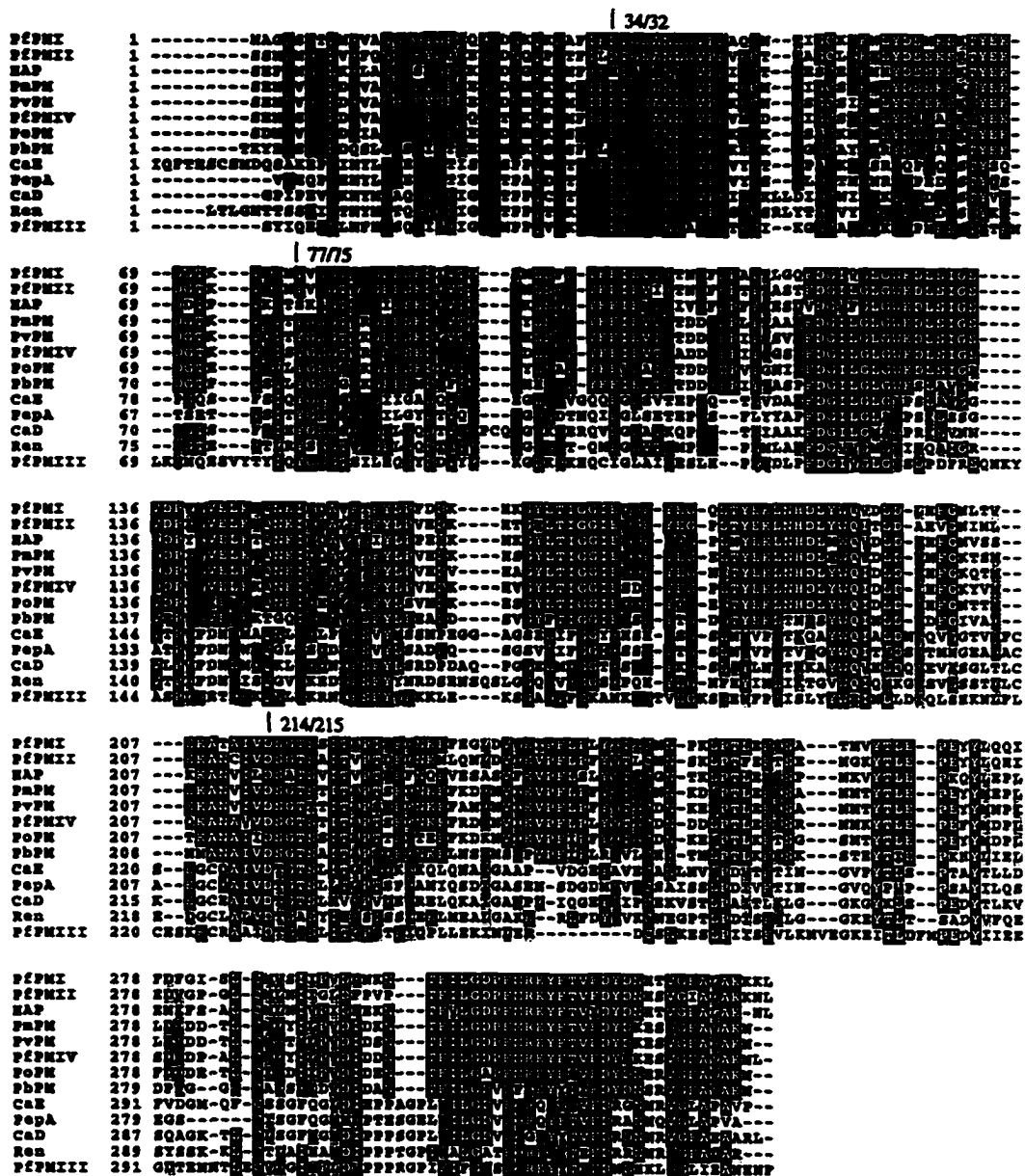
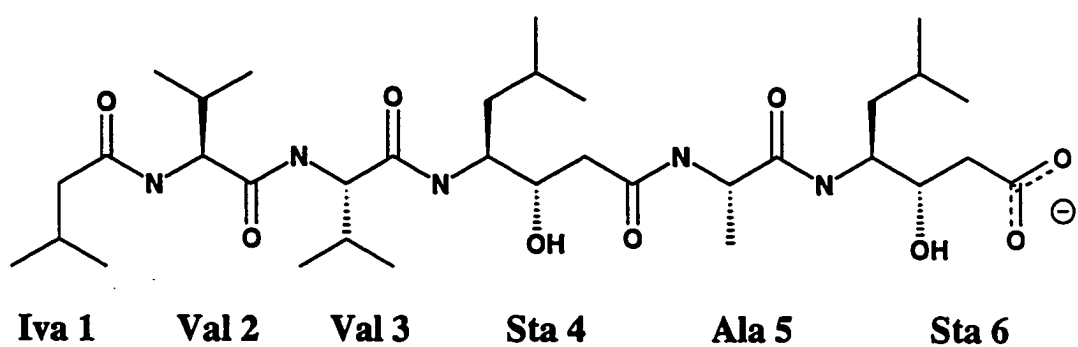


Figure 2.1 Sequence alignment of aspartic proteinases. The sequences were aligned with the program ClustalW(1.7), and BOXSHADE was used to highlight conserved sequence. Abbreviations are PpPMI-IV = *P. falciparum* plasmepsins I-IV, HAP = histo-aspartic protein from *P. falciparum*, PmPM = *P. malariae* plasmepsin, PvPM = *P. vivax* plasmepsin, PoPM = *P. ovale* plasmepsin, PbPM = *P. berghei* plasmepsin, CaE = human cathepsin E, CaD = human cathepsin D, PepA = human pepsin A, Ren = human renin. The numbers above the sequence indicate the catalytic aspartate residues and the conserved tyrosine in the flap (plasmepsin numbering/pepsin numbering).

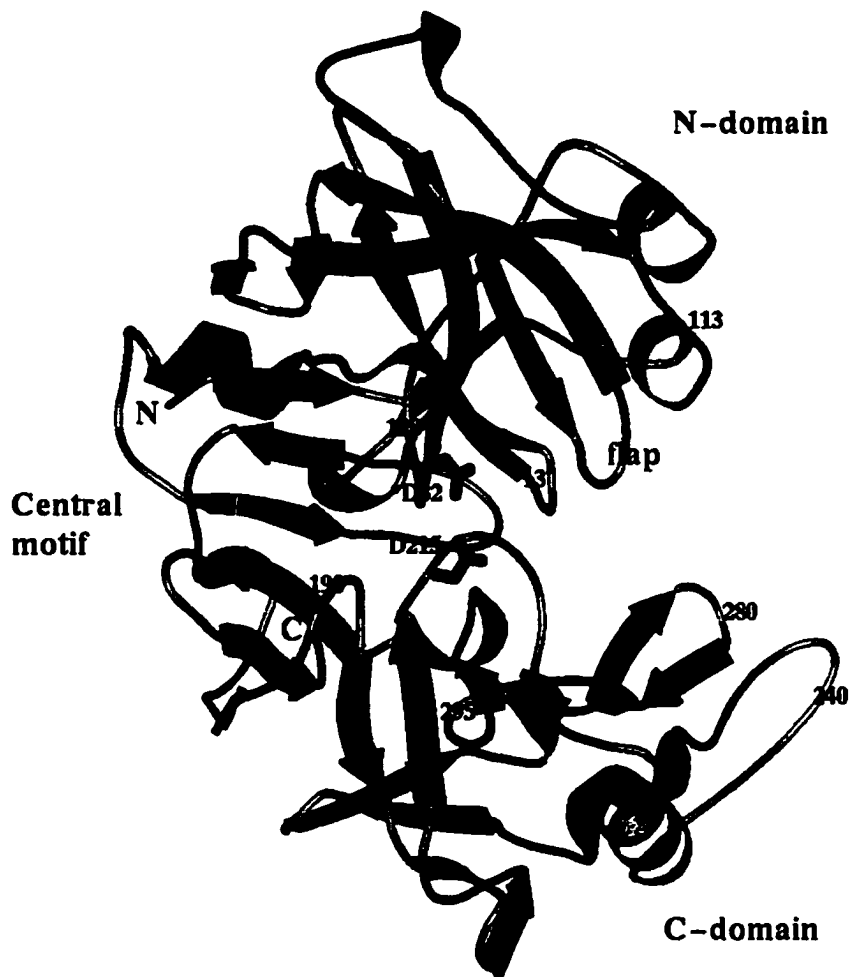


**Figure 2.2** The structure of pepstatin A.

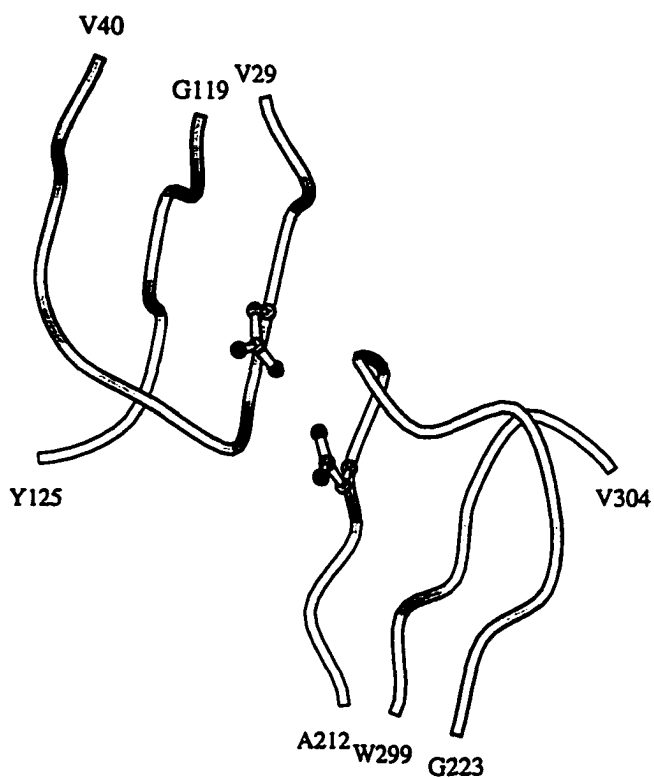


The structure of the eukaryotic APs may be divided into three domains: the N- and C-terminal domains and the central motif (Figure 2.3). The N- and C-domains consist predominantly of  $\beta$ -sheets packed in 3 and 2 layers, respectively. These two domains are attached to the central motif, a six-stranded anti-parallel  $\beta$ -sheet that is located at the back of the molecule. Between the two lobes of an AP is a long and deep cleft that contains the active site (the substrate-binding cleft or the active site cleft). Protein surface loops form the edges of the substrate-binding cleft, and the structural variability of these loops is largely responsible for the differences in substrate specificity among the APs. Loops surrounding the active site cleft are (in plasmepsin numbering): 12 to 17, 75 to 82, 109 to 120, 128 to 134, 190 to 195, 238 to 245, 275 to 286 and 291 to 298. Residues 67 to 93 form the flap, a  $\beta$ -hairpin that extends over the active site cleft and forms part of substrate-binding pockets S3, S2, S1 and S2'. (Throughout the thesis, the nomenclature of Schechter and Berger [1967] will be used, where substrate residues are labeled Pn and Pn', and the corresponding substrate-binding pockets on the enzyme are labeled Sn and Sn'. The numbering is away from the scissile bond, "n" in the direction of the substrate's N-terminus and "n'", in the direction of its C-terminus.) The flap is generally a flexible region in APs. Amino acids comprising the flap tend to have elevated temperature factors in structures of native APs, and significantly lower temperature factors in enzyme-inhibitor complexes [Fujinaga *et al.*, 1995; Suguna *et al.*, 1987]. Flexibility of the flap is important to its function, since the flap must open to admit substrate to the active site, close over the substrate during catalysis, and then open again to release the products.

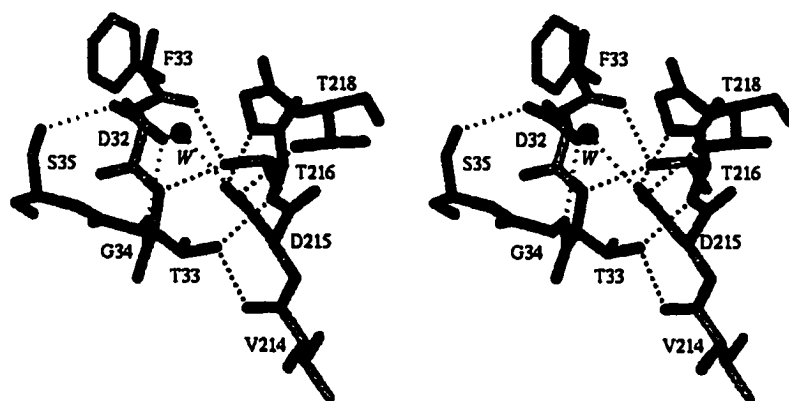
The catalytic machinery of an AP consists of two Asp residues on symmetrically-disposed active site loops (the Psi loops) contributed by the N- and C-domains (Figure 2.3). Each loop contains the conserved sequence Asp-Thr/Ser-Gly-Thr/Ser, and is bisected by a conserved hydrophobic-hydrophobic-Gly sequence (Figure 2.4), resembling the Greek letter Psi ( $\psi$ ). The catalytic aspartates are positioned in a symmetrical hydrogen bonding arrangement called the "fireman's grip" (Figure 2.5). In the fireman's grip configuration, the O $\gamma$ 1 (O $\gamma$ ) of the Thr (Ser) following each catalytic Asp, forms hydrogen bonds with the peptide O and NH of the residues preceding and following, respectively, the catalytic Asp on the opposite loop. In addition, within its Psi loop, each catalytic Asp forms hydrogen bonds through its carboxylate oxygens to the peptide NH of the conserved Gly and the hydroxyl oxygen of the second conserved Ser or Thr (Figure 2.5).



**Figure 2.3** The structure of an archetypal aspartic proteinase, human pepsin [Fujinaga *et al.*, 1995].  $\alpha$ -helices are colored violet;  $\beta$ -strands in the N- and C-domains, blue;  $\beta$ -strands in the central motif, green. The catalytic Asp residues are shown in red. Residue numbers identify loops that surround the active site cleft. The program Bobsript [Esnouf, 1997; Merritt & Murphy, 1994] was used to produce this figure, as well as figures 2.5, 2.7 - 2.9, 2.11, 2.12 and 2.14 - 2.21.



**Figure 2.4** The Psi loops of pepsin. The backbone trace of the loops is shown, and the catalytic Asp residues (Asp 32 and Asp 215) are shown in ball-and-stick representation. This figure was produced with MOLSCRIPT [Kraulis, 1991].

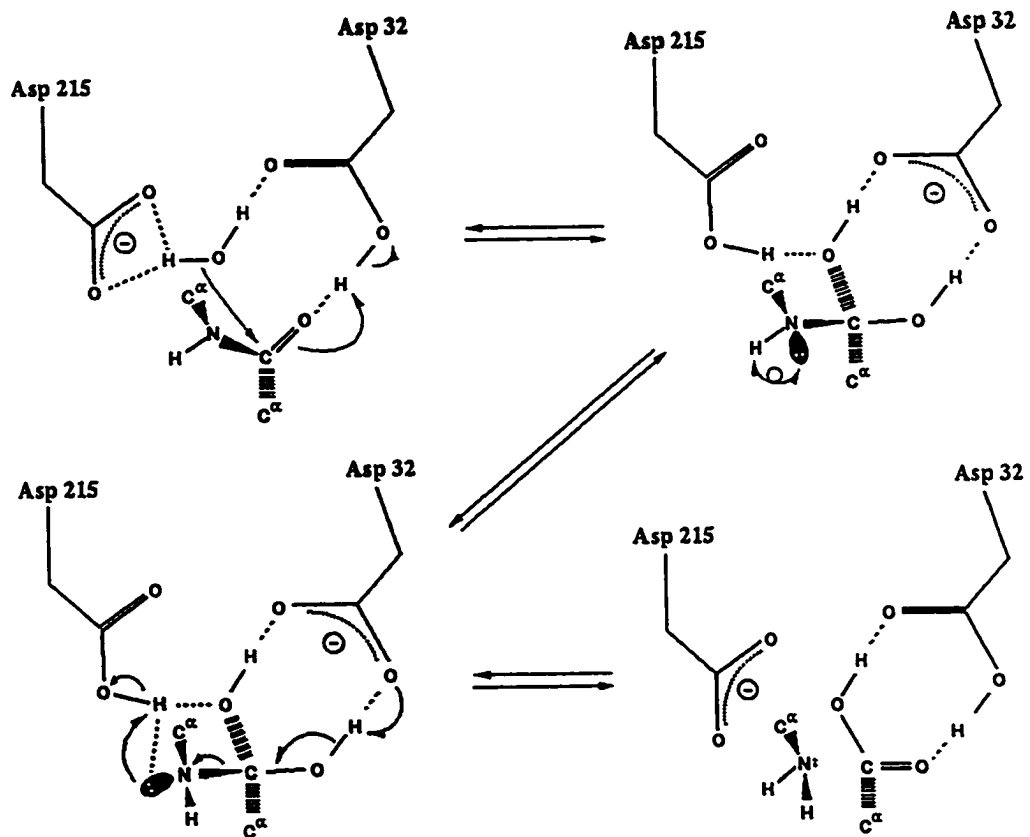


**Figure 2.5** The hydrogen bonding pattern at the active site of pepsin. Dashed lines indicate hydrogen bonds. The "fireman's grip" network is shown with black dashed lines.

The Asp side chains are therefore precisely positioned to interact with a water molecule located between the carboxyl groups of the two aspartates, the central water molecule. This water molecule is observed in the crystal structures of many native APs, and is replaced by an oxygen atom in a number of transition-state mimicking inhibitors, such as statine, difluorostatone, hydroxyethylene and phosphonate derivatives.

The catalytic mechanism of the APs has been the subject of a long debate in the literature. With the accumulation of a great deal of evidence, such as mutagenesis and kinetics data, as well as crystal structures of APs complexed with a variety of inhibitors, however, the diverse mechanistic proposals have converged to a common scheme (Figure 2.6) [Suguna *et al.*, 1987; James *et al.*, 1992; Veerapandian *et al.*, 1992]. The active site carries a charge of -1 and it appears that it is Asp 214 that is ionized, while Asp 34 is neutral, (protonated on its outer oxygen). This assignment of charge is deduced from the crystal structures of penicillopepsin and endothiapepsin in complex with difluorostatone inhibitors [James *et al.*, 1992; Veerapandian *et al.*, 1992]. In the presence of APs, difluorostatones become enzymatically hydrated to the corresponding gem-diols that are believed to mimic the tetrahedral intermediate of AP catalysis (see below). It should be noted that all the proposed mechanisms include assumptions about locations of protons. The assignment in figure 2.6 is based on the most favorable arrangement of protons that is derived from the difluorostatone complex structures. Furthermore, quantum mechanical calculations appear to be consistent with this arrangement [Beveridge, 1996; Beveridge & Heywood, 1993]. However, all the hydrogen atom locations are based on circumstantial evidence. A definitive assignment awaits direct observation of hydrogen atoms, either by neutron diffraction or by ultra-high resolution X-ray crystallography, as in the case of penicillopepsin, for which the crystal structure is currently being refined to 0.9 Å resolution (M.E. Fraser, personal communication).

As shown in figure 2.6, both Asps form hydrogen bonds to the central water molecule. When substrate binds, Asp 214 acts as a general base, deprotonating this water molecule and enabling it to carry out nucleophilic attack on the carbonyl carbon of the scissile peptide bond in the substrate. Nucleophilic attack is further assisted by the polarization of the P1 carbonyl bond. The carbonyl oxygen atom of the substrate is located in an electrophilic environment that polarizes the carbonyl bond and includes a hydrogen bond from Asp 34 Oδ1H and the partially positively charged edge of the Tyr 77



**Figure 2.6** The catalytic mechanism of aspartic proteinases. Adapted from [James *et al.*, 1992].

aromatic ring. During nucleophilic attack, Asp 34 O $\delta$ 1H becomes the general acid, protonating the carbonyl oxygen. Thus, nucleophilic attack by the water on the peptide carbonyl results in a gem-diol tetrahedral intermediate that is stabilized at the active site by hydrogen bonds to both catalytic Asps. The next step in the mechanism is inversion of configuration at the nitrogen of the scissile bond, which directs the nitrogen lone pair toward the O $\delta$ 2H of Asp 214. Protonation of the nitrogen by Asp 214 not only generates a good leaving group, but also weakens the stabilization of the gem-diol by eliminating the hydrogen bond between Asp 214 O $\delta$ 2H and the pro-R (statine-like) hydroxyl group of the diol. As the scissile bond nitrogen is protonated by Asp 214 O $\delta$ 2H, the pro-S hydroxyl becomes deprotonated by Asp 34 O $\delta$ 1, and the C-N bond breaks, leaving a free amine and a carboxylic acid, which forms a carboxylic acid dimer with Asp 34. The catalytic Asps are now in the same electronic state as at the beginning of the catalytic cycle. Diffusion of the products from the active site and the placement of a water molecule between the Asps regenerates the active enzyme.

Since the APs have a preference for long substrates (6 to 8 amino acids), it has been proposed that conformational strain in the substrate results from binding to the enzyme with the concomitant distortion of the scissile peptide bond from planarity, rendering it more susceptible to cleavage [Fruton, 1976; Pearl, 1984; Pearl, 1987]. This may indeed account for the secondary specificity of APs for amino acids distant from the scissile bond, *i.e.*, a dramatic improvement in  $k_{cat}$  without a change in  $K_m$  in substrates with a P3 residue [Fruton, 1976; Hofmann *et al.*, 1988]. The domain flexibility that is observed in structures of various APs is consistent with this hypothesis, since it may be easier for a flexible enzyme than for a rigid one to distort its bound substrate [Sali *et al.*, 1992].

It is also possible that strain in the substrate has an indirect effect on proteolysis, as has been postulated in the case of the maturation of glycosylasparaginase (GA), a process that involves intramolecular cleavage of a precursor. Crystal structures of two slowly-activating mutants of the GA precursor, determined to 1.9 Å resolution, revealed that the peptide bond following the scissile bond, between P1' and P2', is significantly non-planar ( $\omega = 153^\circ$ ) [Xu *et al.*, 1999]. It was thus proposed that the general relief of conformational strain may be one of the driving forces for proteolysis. The case may be similar with APs, but since no structures of functional APs with substrates are available, this proposal remains speculative. The crystal structure of the inactive mutant of a

retroviral AP, FIV proteinase, in complex with a modified substrate, determined to 2.0 Å resolution, shows no significant distortion of the scissile peptide bond, nor in the two surrounding peptide bonds [Laco *et al.*, 1997]. However, this structure does not rule out the possibility of peptide bond distortion in a natural substrate bound to a wild-type AP.

### 2.1.3 Flexibility in aspartic proteinases

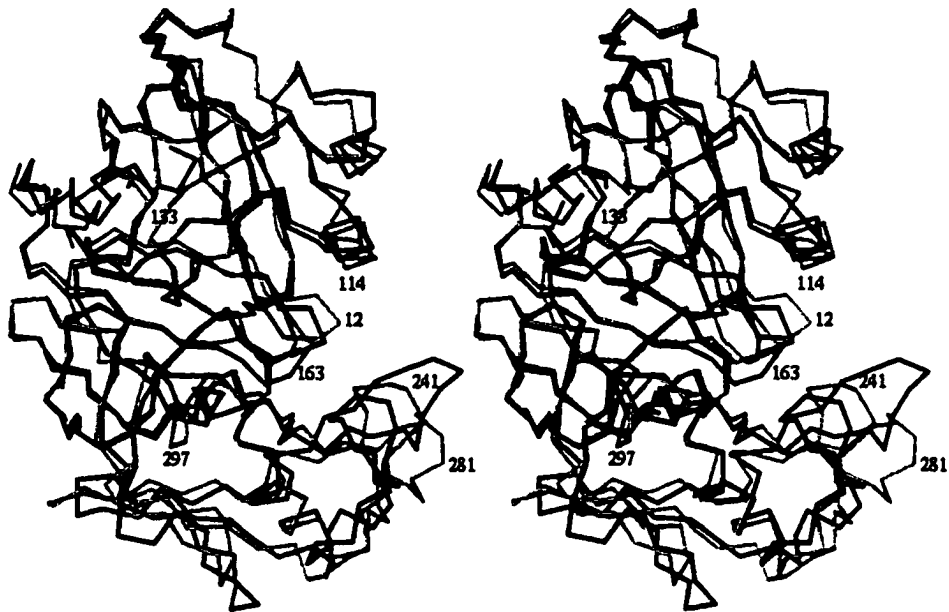
The many crystallographic studies of APs have not only elucidated their common features, such as the structurally conserved overall fold and active site, but have also shown that these molecules display quite a lot of flexibility. Commonly, a number of loops are either disordered or adopt different conformations in different molecules. In addition, a flexible subdomain has been identified in a number of structures [Sali *et al.*, 1992; Sali *et al.*, 1989; Abad-Zapatero *et al.*, 1990; Sielecki *et al.*, 1990; Fujinaga *et al.*, 1995; Silva *et al.*, 1996; Lee *et al.*, 1998]].

The flexible loops are found on the surface of the molecules, and many of them form the edges of the substrate-binding cleft (Figures 2.7a, b). These loops include (in plasmepsin numbering) residues 9 to 18, 108 to 119, 130 to 135, 159 to 166, 237 to 245, 275 to 288 and 294 to 298. The segments from 130 to 135 and from 294 to 298 are found on the Sn' side of the cleft, while the remaining ones are located on the Sn side. The flap, which also exhibits structural changes, covers the centre of the active site cleft.

The flexible subdomain is part of the C-domain, approximately residues 200 to 300 (Figure 2.8). For instance, rigid-body rotations of up to 4° of the C-domain (190 to 302) are found in complexes of endothiapepsin with various inhibitors. Pepsin and pepsinogen show a similar domain movement, where the rotation is approximately 6.5°. Comparisons between different aspartic proteinases and unrelated zymogens show much larger deviations, such as an 18° rotation between pepsinogen and some endothiapepsin-inhibitor complexes. Different cases involve different directions for the axis of rotation. In endothiapepsin, for instance, it is roughly parallel to the long axis of the molecule, passing through the tips of the two Psi loops [Sali *et al.*, 1992], whereas in porcine pepsin it is nearly perpendicular, passing approximately through the Cα atoms of residues 192 and 288 [Abad-Zapatero *et al.*, 1990]. Subdomain flexibility is also observed in the crystal structure of plasmepsin II complexed with pepstatin. In the latter case, the two

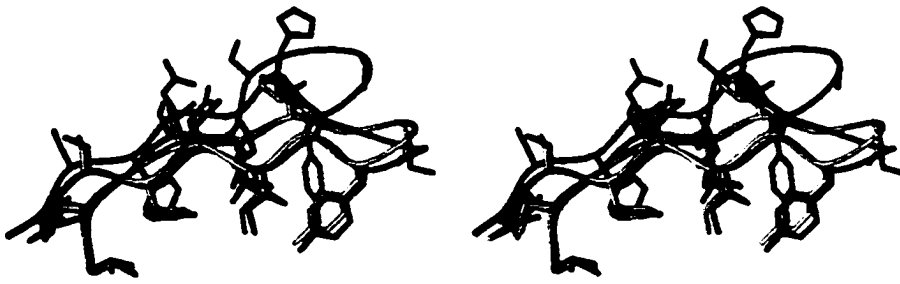


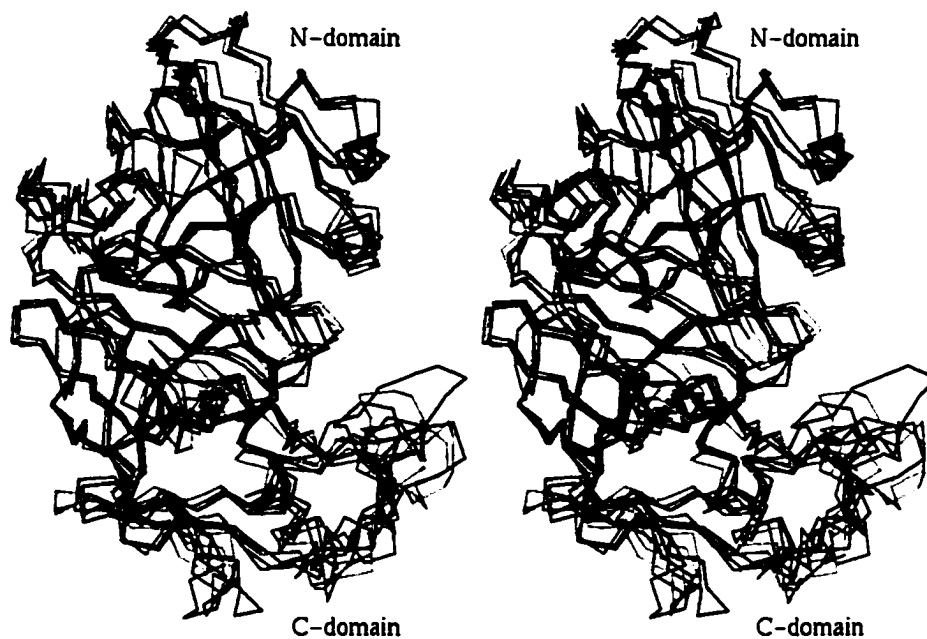
(a)



**Figure 2.7** Loop flexibility in aspartic proteinases. (a) Superposition of PvPM (cyan) and human cathepsin D (purple) illustrates structural differences in loops around the active site cleft. (b) The "flap" in three different structures of human cathepsin D: native cathepsin D (yellow), cathepsin D in complex with pepstatin A (cyan) and cathepsin D at pH 7.5 (pink) [Baldwin *et al.*, 1993; Lee *et al.*, 1999].

(b)





**Figure 2.8** Domain flexibility in aspartic proteinases. Superposition of PvPM (cyan, PDB ID 1QS8), PfPMII (yellow, 1SME), cathepsin D (dark blue, 1LYA), pepsin (magenta, 4PEP) and renin (aquamarine, 2REN).

molecules in the asymmetric unit show different orientations for the subdomain consisting of two polypeptide segments, 227 to 257 and 268 to 293, related by a rotation of approximately 5° [Silva *et al.*, 1996]. The crystal structure of cathepsin D at neutral pH, where it is reversibly inhibited by its own N-terminus, showed a 7° rotation of the C-subdomain [Lee *et al.*, 1998].

Neither the surface loop flexibility nor the C-domain movement in APs perturbs the geometry of the active site, which is standard for aspartic proteinases. Both types of flexibility appear to be functionally important, however, since they affect substrate recognition. Many of the loops in question contribute residues to the substrate-binding pockets, and their structure is likely to be important for substrate binding. The domain movement varies the size and shape of the substrate-binding cleft, most notably in the S3 pocket, thereby also influencing substrate recognition. Finally, the intrinsic flexibility of APs may be important for inducing strain in the substrate, which may be important to its catalytic mechanism [Pearl, 1984].

#### 2.1.4 Malarial plasmepsins

As described in Chapter 1, plasmepsins are the APs found in the protozoan organism, *Plasmodium*, the malarial parasite. To date, eight plasmepsins have been identified in 5 species of *Plasmodium* (Table 2.1). Most intensive research has been carried out with the *P. falciparum* enzymes, since this is the most dangerous of the malarial pathogens. It is nonetheless important to study the other species as well, because, while they do not directly cause mortality, they are costly in terms of morbidity. *P. vivax*, in particular, causes the largest number of malaria infections [Bruce-Chwatt, 1985]. Currently four APs are known in *P. falciparum* (PfPMI - IV), and one in each of *P. vivax* (PvPM), *P. ovale* (PoPM), *P. malariae* (PmPM) and *P. berghei*, the rodent malaria parasite (PbPM). PfPMIII and PfPMIV have been discovered only recently [Bowman *et al.*, 1999; Humphreys *et al.*, 1999], and their existence has brought to light the possibility that multiple plasmepsins exist in other species of *Plasmodium*. In addition, HAP, the histo-aspartic protein, highly similar in sequence to the plasmepsins, but with one of the catalytic Asps (32) replaced by a His, has also been found in *P. falciparum* [Berry *et al.*, 1999]. All the plasmepsins share a high degree of sequence similarity, with PvPM & PmPM being the most similar (87% identity) (Table 2.2). *P. falciparum* PMIII is distinct

from the other plasmepsins, sharing only about 25% sequence identity with them, and containing several insertions and deletions. PfPMIII actually shares more sequence similarity (42%) with the aspartic proteinase from another related parasite, *Eimeria acervulina* [Laurent *et al.*, 1993]. In addition to having a different organization, the gene for PfPMIII is located on chromosome 3, whereas the genes for the other plasmepsins are on chromosome 14. PfPMIII was discovered through genomic analysis, and it is quite likely that this enzyme acts at some other step of the parasite's life cycle than during the erythrocytic stage. *P. falciparum* PMIV has many features in common with PfPMI and PfPMII, which have been isolated from the trophozoite digestive vacuoles and shown to act as hemoglobinas. PfPMIV, however, appears to be more similar to plasmepsin from *P. vivax*, *P. ovale* and *P. malariae* than to PMI and PMII from *P. falciparum*. Therefore this enzyme may represent a common ortholog among the four species of the parasite that infect humans. Before the discovery of PfPMIV, it was believed that *P. vivax*, *P. ovale* and *P. malariae* each contained only one plasmepsin, and that this enzyme carried out the work of both PfPMI and PfPMII [Westling *et al.*, 1997]. When PfPMIV was found, however, it became evident that the situation is far more complex. It appears now that PvPM, PoPM, PmPM and PfPMIV are probably distinct in function from PfPMI and PfPMII. The discovery of PfPMIV also raises the possibility that there are more plasmepsins in *P. vivax*, *P. ovale* and *P. malariae*, i.e. enzymes corresponding to PfPMI and PfPMII. There is evidence that PvPM is expressed during the blood stage of parasite growth, since the corresponding mRNA has been amplified by RT-PCR from blood-stage RNA (J. Dame, personal communication). Likewise, PfPMIV has been cloned from cDNA confirming that the functional PfPMIV gene is expressed during the blood stage (J. Dame, personal communication). Preliminary immunolocalization experiments suggest that PvPM is present in the vicinity of the food vacuole in the parasite (J. Dame, personal communication). However, the biological role of this plasmepsin has not been elucidated, so it is not known if this enzyme actually functions as a hemoglobinase in the digestive vacuole, like PfPMI and PfPMII [Gluzman *et al.*, 1994], as a spectrinase, like PfPMII [Le Bonniec *et al* 1999], or whether it does something completely different. Since its function has not been characterized, the potential usefulness of this enzyme as a drug target is not yet known. Even if these enzymes do not prove to be viable drug targets, however, new information regarding all the plasmepsins, such as their three-dimensional

<b>Species</b>	<b>Name</b>	<b>Abbreviation</b>	<b>Reference</b>
<i>P. falciparum</i>	plasmepsin I	PfPMI	Francis <i>et al.</i> , 1994
	plasmepsin II	PfPMII	Dame <i>et al.</i> , 1994
	plasmepsin III	PfPMIII	Bowman <i>et al.</i> , 1999
	plasmepsin IV	PfPMIV	Humphreys <i>et al.</i> , 1999
	histo-aspartic protein	HAP	Berry <i>et al.</i> , 1999
<i>P. vivax</i>	plasmepsin	PvPM	
<i>P. malariae</i>	plasmepsin	PmPM	
<i>P. ovale</i>	plasmepsin	PoPM	
<i>P. berghei</i>	plasmepsin	PbPM	Humphreys <i>et al.</i> , 1999

**Table 2.1** The currently known plasmepsins.

	PfPM	PfPM	PfPM	PvPM	PmPM	PoPM	PbPM	HAP	CaD	CaE	Ren	PepA
	II	III	IV									
PfPMI	72	24	68	69	71	65	53	63	33	29	32	29
PfPMII		26	68	70	71	62	55	60	32	31	29	30
PfPMIII			24	27	26	27	25	22	29	30	28	29
PfPMIV				78	82	69	57	59	31	28	31	29
PvPM					87	75	56	59	30	28	29	27
PmPM						77	55	61	32	27	31	28
PoPM							56	56	32	30	30	29
PbPM								48	31	30	27	30
HAP									26	25	27	24
CaD										50	43	48
CaE											38	56
Ren												37

**Table 2.2** Sequence comparison (% identity) among plasmepsins and other selected APs. Abbreviations for the plasmepsins and HAP are as shown in Table 2.1. Other abbreviations are: CaD – human cathepsin D, CaE – human cathepsin E, Ren – human renin, PepA – human pepsin A.

structures, will help us gain a better understanding of the biology of malaria.

Extensive use has been made of the crystal structure of *P. falciparum* PMII in complex with pepstatin, determined earlier by Silva *et al.* [1996], not only in designing inhibitors, but also as a basis for modelling the structures of other plasmepsins in order to rationalize the differences in their kinetics and substrate specificity [Westling *et al.*, 1997].

We have recently determined the crystal structure of plasmepsin from *P. vivax* in complex with pepstatin. *P. vivax* plasmepsin contains 327 amino acids, and is initially produced as an inactive precursor, proplasmepsin. The conversion of proplasmepsin to PvPM is discussed in Chapter 4. PvPM shares 87% sequence identity with PmPM, about 80% with PoPM and PfPMIV, approximately 70% with PfPMI and PfPMII, and around 30% with the mammalian cellular APs cathepsins D and E (Table 2.2). Comparison of plasmepsins with the human APs is essential for the design of selective plasmepsin inhibitors, since plasmepsin inhibitors may cross-react with these enzymes and cause side effects.

## 2.2 Methods

### 2.2.1 Activation and crystallization

Mature *P. vivax* plasmepsin was generated by autoactivation of recombinant truncated *P. vivax* proplasmepsin at acidic pH. To 250  $\mu\text{L}$  of *P. vivax* proplasmepsin at a concentration of 0.13 mM (5 mg/mL), 25  $\mu\text{L}$  of 1M Na acetate, pH 4.5, was added. After incubation for 5 minutes at 37°C, the solution was stored at 4°C.

The complex of *P. vivax* plasmepsin with pepstatin was crystallized by vapor diffusion from the following mother liquor: 17% PEG 3000, 0.2M  $(\text{NH}_4)_2\text{SO}_4$ , 0.1M Na acetate pH 5.0, 3% MPD and 0.15%  $\beta$ -octyl glucoside. Plasmepsin (0.13 mM) in 20  $\mu\text{L}$  of the activation buffer was combined with 1  $\mu\text{L}$  of 5 mM solution of pepstatin in methanol giving a 1:2 protein to inhibitor ratio. Hanging or sitting drops were set up containing equal volumes of the plasmepsin:pepstatin complex and the reservoir solution described above. Seeding was required to produce diffraction-quality crystals. Resultant crystals belonged to the space group  $P4_12_12$ , with  $a = b = 145.02 \text{ \AA}$ ,  $c = 71.07 \text{ \AA}$  and contained 2 molecules in the asymmetric unit ( $V_m = 2.51 \text{ \AA}^3/\text{Da}$  for 2 molecules/asymmetric unit)



(Table 2.3) [Mathews, 1968]. The crystals proved to diffract rather poorly, and many crystals were screened before reasonably well-diffracting ones were found.

### 2.2.2 Data collection

X-ray diffraction data were collected on beamline X12B at the NSLS. The crystals were cryo-protected in 30% glycerol, 20% PEG 3000, 0.2 M  $(\text{NH}_4)_2\text{SO}_4$ , 0.1 M Na acetate buffer, pH 5.0, and flash-frozen in a cold  $\text{N}_2$  gas stream at 90K. 68 images ( $1^\circ$  rotations, 30 second exposure) were collected on a MAR image plate in the 18 cm scanning mode using a wavelength of 0.978 Å, with the crystal to detector distance of 210 mm. The data were integrated and scaled with Denzo and Scalepack [Otwinowski & Minor, 1996]. The redundancy of this data set was 4.7 and the R-merge was 8.4% for data between 25 and 2.5 Å (Table 2.3).

### 2.2.3 Structure determination and refinement

The structure was solved by molecular replacement in AMoRe [Navaza, 1994], using the coordinates of *P. falciparum* plasmepsin II (1sme, molecule 1 [Silva *et al.*, 1996]) as the search model. The molecular replacement search gave 2 solutions (correlation = 0.532, R = 45.3% for 10 - 3 Å data). Clear difference density ( $F_o - F_c$ ) was seen for a pepstatin molecule at the active site of each of the two plasmepsin molecules in the asymmetric unit (Figures 2.9a, b). The required sequence substitutions were built in O [Jones & Kjeldgaard, 1995], and the model was refined using X-PLOR and CNS using the maximum likelihood targets mlf and mli [Brunger, 1992; Brunger *et al.*, 1998; Pannu & Read, 1996]. The refinement included torsion angle simulated annealing, positional refinement and restrained individual temperature factor refinement. Anisotropic B-factor scaling and a bulk solvent correction were applied. 2-fold NCS restraints were initially applied and later dropped. The final model contained 655 amino acids (1 - 327 in molecule A and 1P - 327 in molecule B), 2 pepstatin A molecules, 2 acetate ions and 154 water molecules. The R- and free R-factors are 0.196 and 0.249, respectively, for all data between 25 and 2.5 Å. One amino acid, Asp B162, is found in the generously allowed region of the Ramachandran plot (Figure 2.10). This residue is located in a poorly-defined loop of the molecule.

**Table 2.3** Summary of data and refinement

Space Group	P4 <sub>1</sub> 2 <sub>1</sub> 2
Unit cell	$a = b = 145.02 \text{ \AA}$ $c = 71.07 \text{ \AA}$
Molecules/asymmetric unit	2
$V_m$ [Mathews, 1968]	$2.5 \text{ \AA}^3/\text{Da}$
Data completeness	98.3 [81.0] % (25 - 2.47 $\text{\AA}$ )
$F/\sigma F > 1$ , [ $F/\sigma F > 3$ ]	97.3 [53.7] % (2.51 - 2.47 $\text{\AA}$ )
$I/\sigma I$	12.97 [2.82 for 2.49 - 2.47 $\text{\AA}$ ]
Observed reflections	130851
Unique reflections	27347
$R_{\text{merge}}^1$	8.4 % [44.0% for 2.51 - 2.47 $\text{\AA}$ ]
<b>Final refinement statistics</b>	
R-factor <sup>2</sup> (25 - 2.50 $\text{\AA}$ , 26324 reflections)	19.7 %
R-free <sup>3</sup> (25 - 2.50 $\text{\AA}$ , 1353 reflections)	25.0 %
Number of protein atoms	5310
Number of acetate molecules	2
Number of water molecules	154
R.m.s. deviations from ideal geometry	
bond lengths	0.006 $\text{\AA}$
bond angles	1.5 °
improper angles	0.73 °
Temperature factors	
mean B (overall)	39.0 $\text{\AA}^2$
estimated from Wilson plot	40.3 $\text{\AA}^2$
Estimated cross-validated coordinate error	
from luzzati plot	0.37 $\text{\AA}$
from sigmaa	0.40 $\text{\AA}$
Bulk solvent model	
Ksol	0.328 $e/\text{\AA}^3$
Bsol	32.64 $\text{\AA}^2$
Ramachandran plot statistics for 581 non-proline and non-glycine residues [Laskowski <i>et al.</i> , 1993]	
residues in most favored regions	507 (87.3 %)
residues in additional allowed regions	73 (12.6 %)
residues in generously allowed regions	1 (0.2 %)
residues in disallowed regions	0 (0 %)

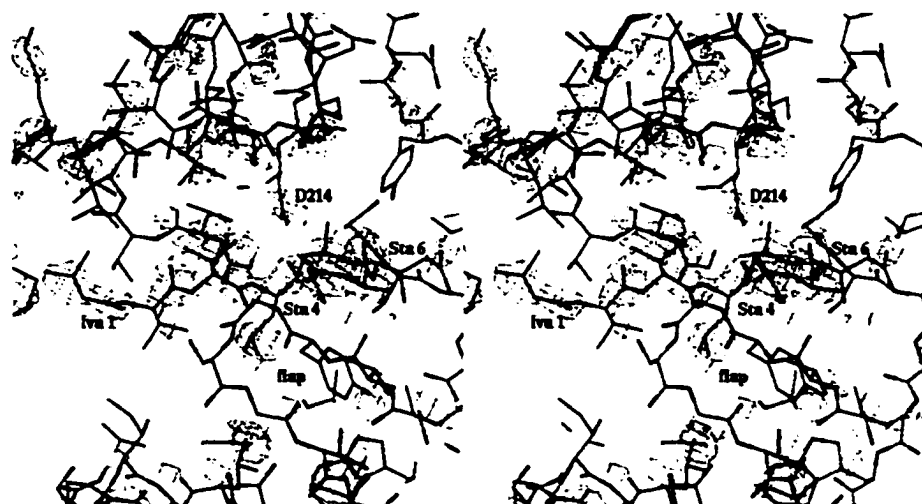
$$^1R_{\text{merge}} = \frac{\sum_{hklj} |I_{hklj} - \langle I_{hklj} \rangle|}{\sum_{hklj} \langle I_{hklj} \rangle}$$

$$^2R\text{-factor} = \frac{\sum |F_o| - |F_c|}{\sum |F_o|}$$

where  $|F_o|$  and  $|F_c|$  are the observed and calculated structure factor amplitudes, respectively.

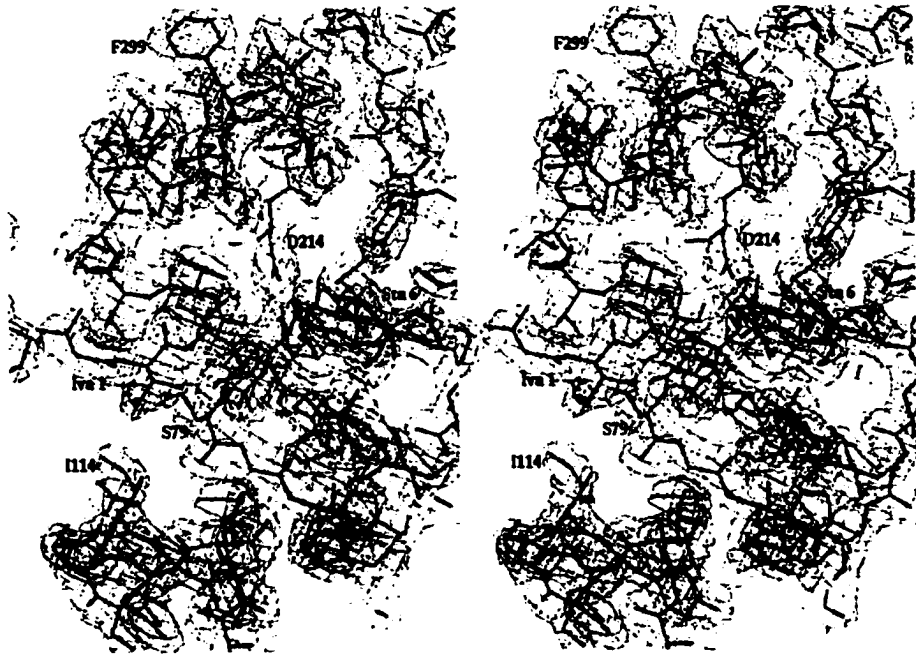
<sup>3</sup>R-free was calculated using a random set containing 2.1% of observations which were omitted during refinement [Brunger, 1992].

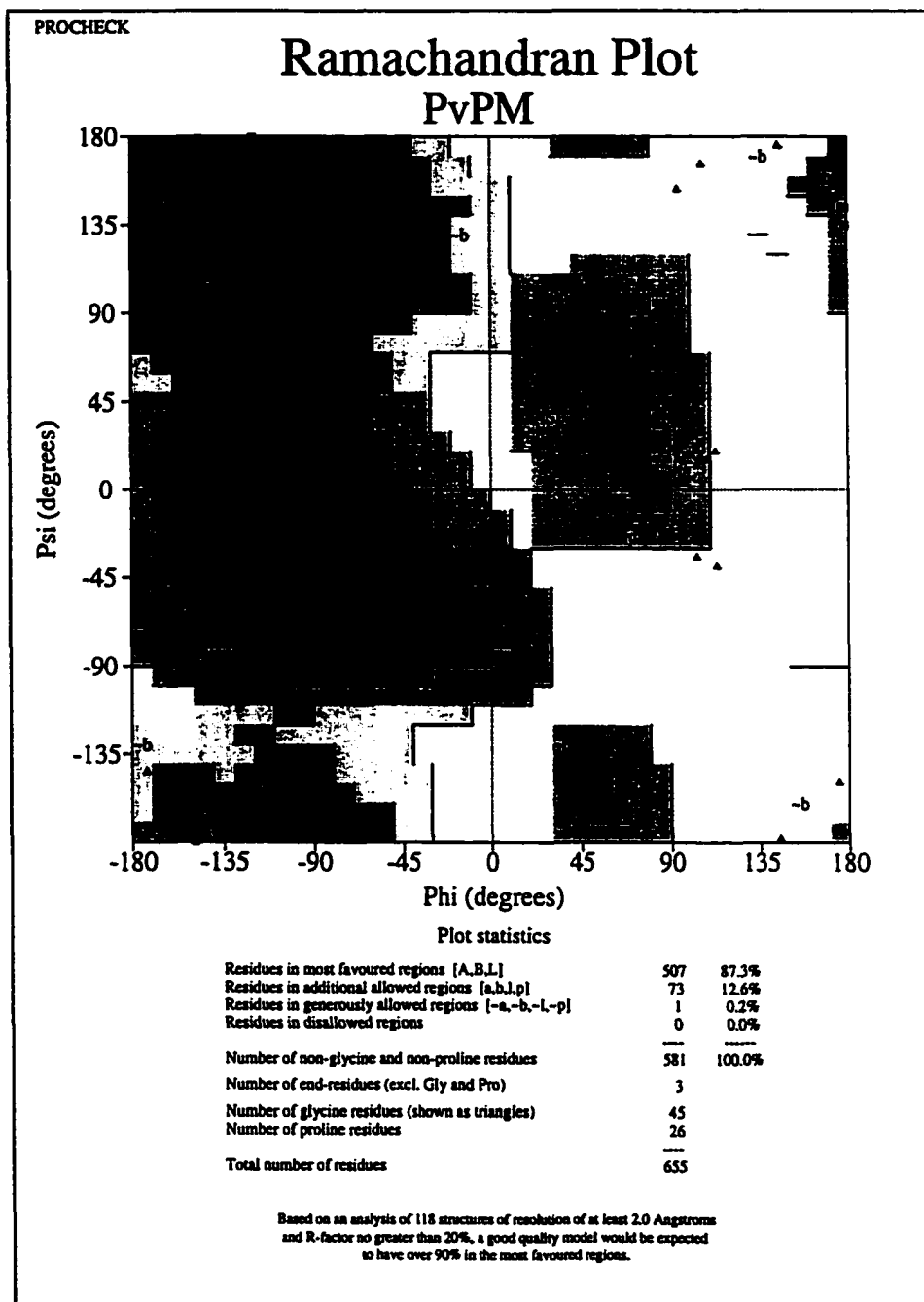
(a)



**Figure 2.9** Refinement progress for PvPM. (a) Difference ( $|F_O| - |F_C|$ ) electron density in the active site, contoured at  $2\sigma$ , superimposed on the final refined model, shown in ball-and-stick representation and colored by atom type. Pepstatin is shown with carbon atoms in cyan; PvPM carbons are in grey, oxygen in red, nitrogen in blue, and sulfur in yellow. (b) Final ( $2|F_O| - |F_C|$ ) map, contoured at  $1\sigma$ , superimposed on the final model.

(b)





**Figure 2.10** Ramachandran plot for the PvPM crystal structure. This figure was generated with PROCHECK [Laskowski *et al.*, 1993].

#### **2.2.4 Modelling of *P. ovale* and *P. malariae* plasmepsins**

Homology models of plasmepsins from *P. ovale* and *P. malariae* were constructed using the Homology module of Insight II (MSI, San Diego, CA). The crystal structure of *P. vivax* plasmepsin was used as the modelling template. Since the sequence alignments of these plasmepsins show no gaps or insertions, the models were built simply by introducing the required mutations into the *P. vivax* structure. The models were then subjected to 200 steps of energy minimization in CNS to relieve unfavorable contacts and to regularize the geometry.

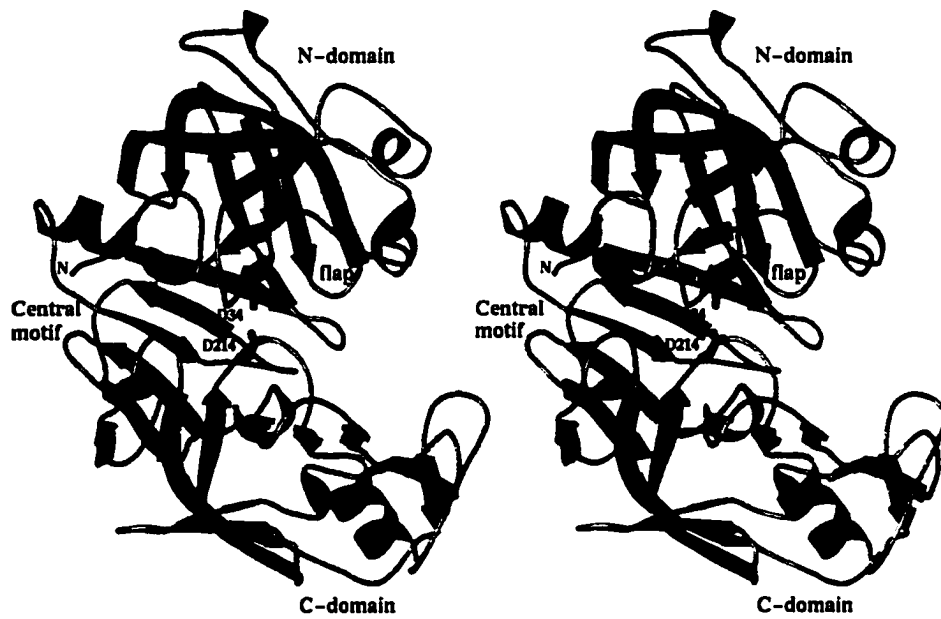
Substrates were modeled in the active site by modifying the structure of a phosphonate inhibitor in complex with penicillopepsin [Fraser *et al.*, 1992]. The required structural changes were built in O, and the geometry of the substrate peptide was regularized with the Refi\_zone command in O [Jones & Kjeldgaard, 1995].

Superpositions of molecules were performed with the program SUPPOS (Biomol package). Difference distance matrices were calculated with the program DDMP (P. Fleming, Center for Structural Biology at Yale University, New Haven, CT). Buried surface area was calculated using the program MS.

### **2.3 Results and discussion**

#### **2.3.1 Description of structure**

*P. vivax* plasmepsin (PvPM) has the typical bilobal shape characteristic of aspartic proteinases (Figure 2.11). Between the two lobes, which contain the N- and C-terminal domains, is the substrate-binding cleft. At the bottom of the cleft is the six-stranded anti-parallel  $\beta$ -sheet of the central motif. The secondary structure of PvPM consists predominantly of  $\beta$ -sheets. The central  $\beta$ -sheet is comprised of residues Asp 4 - Leu 8, Ser 153 - Tyr 157, Gly 166 - Ile 170, Thr 183 - Asp 186, Tyr 309 - Asp 314 and Ser 319 - Ile 324. The  $\beta$ -sheets in the N-domain are packed in 3 layers and contain residues Asp 9 - Val 11, Met 15 - Val 22, Gln 27 - Asp 34, Leu 40 - Pro 43, Glu 67 - Tyr 77, Gly 80 - Leu 93, Leu 96 - Asp 107 and Gly 122 - Gly 125. The C-domain contains a two-layer  $\beta$ -sheet that consists of residues Gln 194 - Phe 201, Gln 204 - Val 213, Ile 220 - Ala 222, Leu 257 - Lys 260, Thr 265 - Leu 268, Met 289 - Pro 291 and Thr 298 - Leu 301. In



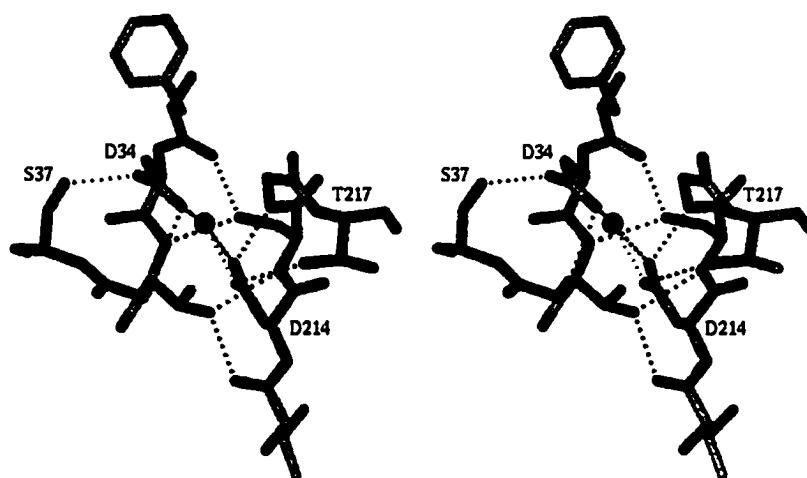
**Figure 2.11** The structure of PvPM. The coloring scheme is the same as in figure 2.3.

addition, the C-domain contains a small four-stranded anti-parallel  $\beta$ -sheet, residues Ile 237 - Lys 238, Tyr 245 - Thr 248, Met 274 - Pro 276 and Leu 284 - Ile 287. The structure contains 4  $\alpha$ -helices: Pro 113 - Ala 118 (preceded by one turn of  $3_{10}$  helix, Thr 108 - Glu 112), Pro 138 - Gln 146, Ile 223 - Ser 232 and Gly 302 - Lys 308; and another short  $3_{10}$  helix, Asp 59 - Ser 63.

The geometry of the active site of PvPM is typical of APs. The catalytic aspartate residues (34 and 214) form the same interactions as observed in all AP structures and described in the introduction (Figures 2.5, 2.12). In addition to their main chains' involvement in the "fireman's grip" hydrogen bonding pattern, the side chains of the aspartates interact with the neighbouring residues. O $\delta$ 1 of Asp 34 forms a hydrogen bond to Ser 37 O $\gamma$  (2.7 Å), and O $\delta$ 2 of Asp 34 forms a weak hydrogen bond with Gly 36 N (3.0 Å). Asp 214 O $\delta$ 1 and O $\delta$ 2 form hydrogen bonds with Thr 217 O $\gamma$ 1 (2.8 Å) and Gly 216 N (2.6 Å), respectively. As a result, the carboxylate groups of the aspartates are nearly coplanar with each other (both  $\chi_2$ s close to 180°), with their inner oxygens 2.7 Å apart. The diagonal distances between the inner and outer oxygens of the two Asps are different. The distance between O $\delta$ 1 of Asp 214 and O $\delta$ 2 of Asp 34 is 3.9 Å, while O $\delta$ 1 of Asp 34 and O $\delta$ 2 of Asp 214 are separated by 4.7 Å. The difference of 0.8 Å is significantly greater than the estimated coordinate error in the structure of PvPM (0.4 Å, see Table 2.3), and is therefore unlikely to be the result of an experimental error. Similar differences in the diagonal distances between the carboxylate oxygens have been observed in the crystal structures of other APs, generally at higher resolution [Pearl & Blundell, 1984; Sielecki *et al.*, 1990]. These differences have been used in quantum mechanical calculations to determine the most favorable hydrogen bonding arrangements and to rationalize the chemical inequivalence, i.e. protonation state, of the two catalytic aspartates [Beveridge, 1996]. In addition to their interactions with protein atoms, the outer oxygens of the Asps are in close contact with the statine OH of pepstatin A (see below).

All the plasmepsins currently known, with the exception of PfPMIII, have a serine in position 215, i.e. following the second catalytic aspartate. In most other known eukaryotic APs this residue is a threonine. As the crystal structures of both PvPM and PfPMII [Silva *et al.*, 1996] show, however, this substitution does not affect the structure of the active site. The hydroxyl group of Ser 215 superimposes with the hydroxyl group of Thr in other structures, donating a hydrogen bond to Asp 214. Placement of a Met at





**Figure 2.12** Hydrogen bonding interactions at the active site of PvPM.

position 306, which is occupied by a conserved, and one methylene shorter, Ile in other APs, compensates for the missing methyl group of Ser 215.

PvPM contains two disulfide bridges, between Cys 47 and Cys 52 and between Cys 249 and Cys 285. PfPMII also has a cysteine at position 211, which, in a structural superposition with pepsin, is located two residues away from Cys210. Cys 210 in pepsin participates in a disulfide bond with Cys 206. This disulfide is found in many APs, including vertebrate and plant enzymes, but is absent in the plasmepsins and the fungal APs. In PfPMII, Cys 211 is found, presumably in the reduced (SH) form, in a hydrophobic pocket, and in PvPM residue 211 is valine.

PvPM contains one *cis*-proline, residue 113, which is located in a turn between a  $3_{10}$ -helix and an  $\alpha$ -helix. A *cis*-Pro is also found at position 113 in PfPMII.

### 2.3.2 Comparison of the NCS-related molecules

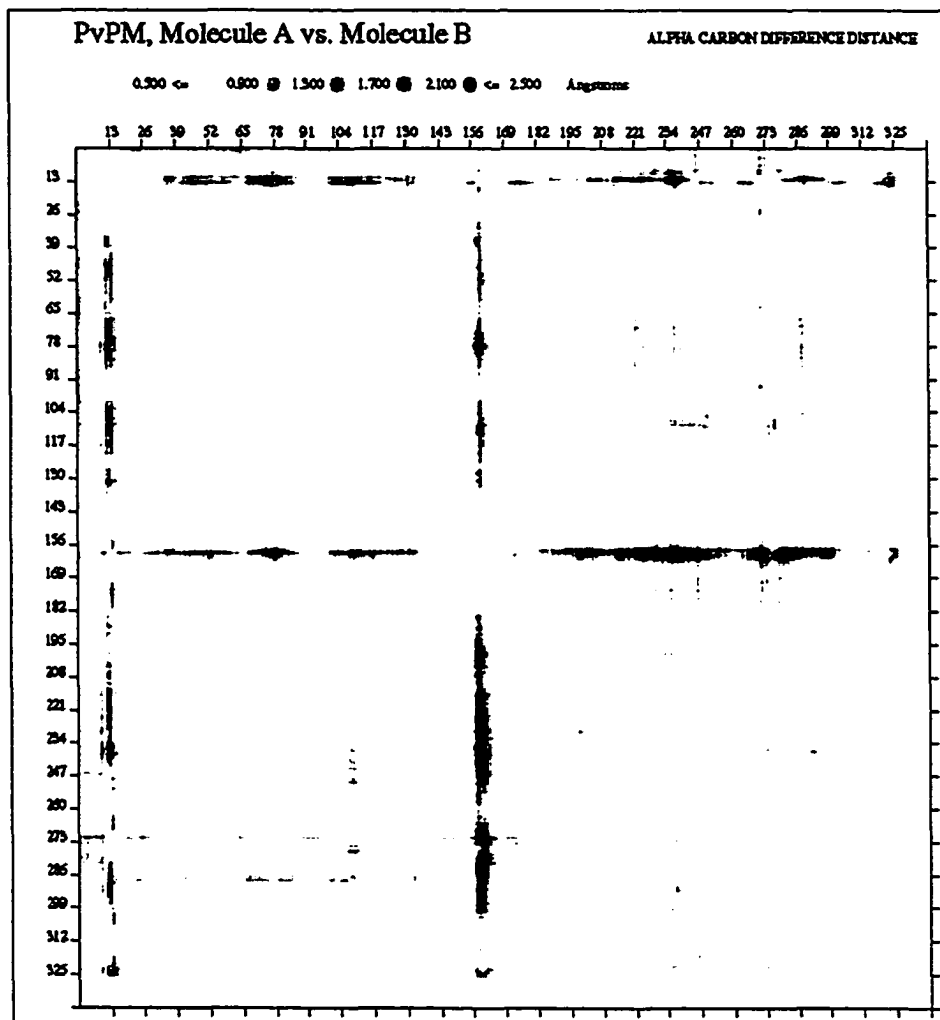
The asymmetric unit of the PvPM crystals consists of two molecules of PvPM, that are related by local, or non-crystallographic, symmetry (NCS). The presence of NCS in a crystal can make structure determination either easier or more difficult. NCS can be extremely helpful in solving the crystal structure, by allowing the use of averaging to improve electron density, for example [Bricogne, 1974; Kleywegt & Read, 1997]. Inclusion of NCS information in refinement is also a powerful means of improving the observation to parameter ratio, which is often less than optimal in crystallographic studies. On the other hand, the presence of multiple molecules in the asymmetric unit prolongs the map examination and manual rebuilding stage of refinement. Ultimately, while multiple molecules in the asymmetric unit make structure analysis somewhat more laborious, comparisons of these crystallographically independent molecules can yield valuable insights into the structure.

The two molecules of PvPM in the asymmetric unit are generally very similar in structure. Examination of the pairwise deviations, however, reveals some differences between the two molecules. The difference distance matrix (DDM) is a common means of analyzing structural differences between related molecules [Richards & Kundrot, 1988]. This matrix tabulates the differences in distances between all pairs of C $\alpha$  atoms within the two molecules and shows graphically local refolding or domain movements in one molecule relative to the other. Shaded areas of the DDM indicate changes in distance

between pairs of C $\alpha$  atoms in the two structures. The DDM for PvPM molecule A vs. molecule B (Figure 2.13) shows that two loops, around residue 14 and around residue 160, and a portion of the C-domain experience a conformational change between molecules A and B of PvPM. The program SUPPOS provided a more precise identification of the residues experiencing conformational changes: Asp 9 - Phe 16, Leu 158 - Gly 166, and Pro 223 - Asn 297.

The rmsd for the superposition of all 327 C $\alpha$  atoms is 0.64 Å. Superposing the C $\alpha$  atoms of only the 222 residues whose positions match to within 3 $\sigma$  improves the rmsd to 0.23 Å. Residues 223 - 297 constitute a subdomain that differs slightly in orientation, by a 2° rotation, between the two molecules (Figure 2.14). The maximal deviation is for Phe 241, with 1.8 Å between the C $\alpha$ s. Superposition of residues 220 - 300, gives an rmsd of 0.40 Å. As discussed in the introduction, flexibility of this region of APs (within the C-domain) is commonly observed. In a rather dramatic example, in PfPMII, the same subdomain differs in orientation by 5° between the two NCS-related molecules. Although the domain shift in the PvPM crystal structure is smaller than in PfPMII, it has an effect on the substrate-binding cleft. The segment from Pro 291 to Asp 295, which forms one rim of the cleft, is slightly shifted into the cleft in molecule A relative to molecule B. Thus Val 292 and Ile 294 are, respectively, 0.9 and 0.7 Å closer to the N-terminal lobe in molecule A than in B.

The surface loops Asp 9 to Phe 16 and Leu 158 to Gly 166 make a concerted shift due to different interactions with symmetry-related molecules. The loop from 158 to 166 is often poorly ordered in AP structures. While this loop does not contact the substrate-binding cleft directly, it interacts with the 9 - 16 loop, which, in turn, forms part of the S1 and S3 pockets. In molecule A of PvPM, Phe 241 from a symmetry-related molecule protrudes into the substrate-binding cleft close to Iva 1 (isovaline) of pepstatin, between the side chains of Ile 14 and Met 15, and pushes loop 9 - 16 slightly out of the cleft. To accommodate this movement, loop 158 - 166 must also move out of the way (Figure 2.14). In molecule B, on the other hand, His 168 Ne2 is involved in a hydrogen bond with Glu 253 of a symmetry-related molecule. As a result of this rearrangement, the shape of the S1/S3 pocket differs in the two molecules. In molecule A, the pocket is a little larger, since in molecule B, Met 15 moves deeper into the pocket. The C $\alpha$ s of Met 15 in the two molecules are 1.4 Å apart in the superposition, and their C $\epsilon$ s are 2.6 Å apart, while there is



**Figure 2.13** DDM comparing the two NCS-related molecules of PvPM. Dark shading indicates regions where the two molecules differ in conformation.



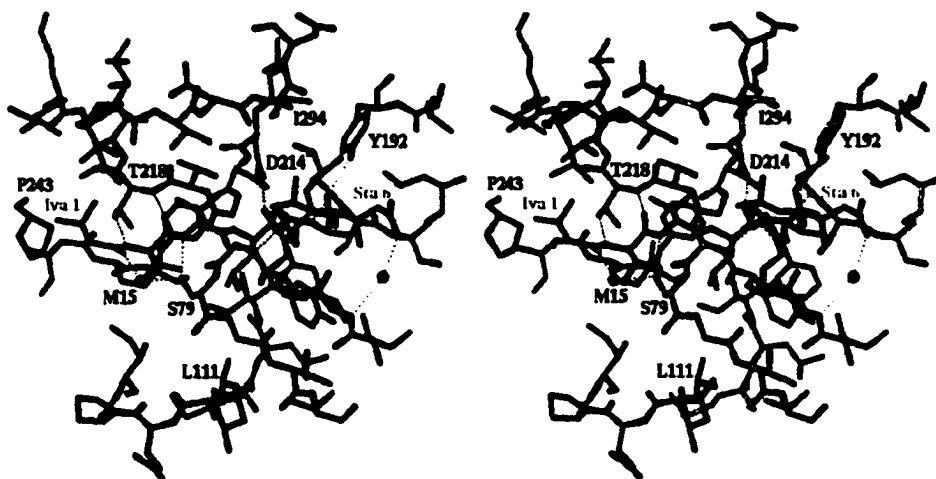
**Figure 2.14** Superposition of the two NCS-related molecules of PvPM. Molecule A is colored cyan and molecule B, dark blue.

essentially no difference in the orientation of Val 2 and Sta 4 of pepstatin in the two molecules. Recent mutagenesis studies on PfPMII identified Met 15 as an important determinant of S3 specificity [Westling *et al.*, 1999]. It is interesting to see that this residue is so mobile in PvPM. Furthermore, because the position of Met 15 in molecule A is strongly influenced by a crystal contact, molecule B may give a more accurate representation of the S1 and S3 pockets in solution. This example illustrates that when analyzing crystal structures, it is very important to consider symmetry contacts that may be affecting conformations of regions of the molecule under consideration. Finally, it should be noted that both of these loops are rather poorly ordered in the crystal, as can be judged by the poor quality of the associated electron density and by their elevated temperature factors.

### 2.3.3 Pepstatin A binding

Pepstatin A is a strong (generally with  $K_i$  in the nM range [Fusek & Vetvicka, 1995]) peptide-based inhibitor of aspartic proteinases (Figure 2.2). Pepstatin binds in an extended conformation in the substrate-binding cleft of plasmepsin. The OH of the statine moiety is located between the carboxylates of the catalytic aspartates, replacing the central water molecule that is found at this position in native APs. In this manner statine mimics the tetrahedral transition state in the nucleophilic attack of the water molecule on the scissile peptide bond. The leucine-like side chain of statine fits into the S1 pocket of plasmepsin. The isovaleryl and the two valine side chains of pepstatin occupy the S4, S3 and S2 subsites, respectively. Thus, pepstatin is a good structural model for the peptide substrate on the N-terminal (P<sub>n</sub>) side of the scissile bond. Since statine possesses two "extra" backbone atoms relative to an  $\alpha$ -amino acid, the structural correspondence between pepstatin and the substrate on the C-terminal (P<sub>n'</sub>) side of the scissile bond is more obscure. Pepstatin does not place anything into the S1' pocket. Ala fits approximately into the S2' pocket, and the side chain of the C-terminal statine is approximately in the S3' pocket of plasmepsin.

Pepstatin makes a number of interactions with plasmepsin, both through its side-chain and main-chain atoms (Figure 2.15). Most of the peptide O and NH atoms are involved in hydrogen bonds with the protein. The only exception is the NH of Sta 4, which does not form hydrogen bonds. This NH is 3.6 Å away from Thr 217 O $\gamma$ 1, and in



**Figure 2.15** Interactions of pepstatin with PvPM. Nitrogen atoms are colored blue, oxygen atoms, red and sulfur atoms, yellow. The carbon atoms of PvPM and pepstatin are in grey and cyan, respectively.

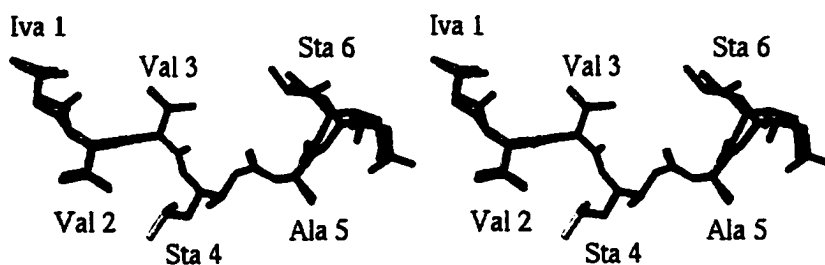
the substrate a hydrogen bond may exist between the two atoms. Furthermore, formation of a strong hydrogen bond in the transition state would facilitate the reaction. O of Iva 1 and O1 of Sta 6 form water-mediated hydrogen bonds to Ser 79 O $\gamma$  and Thr 76 NH, respectively. Most of the pepstatin side chains do not fill the substrate-binding pockets completely, meaning that a molecule that has a better fit to the plasmepsin active site should be a more potent inhibitor for this enzyme. If PvPM does prove to be a useful drug target, analysis of PvPM's interactions with pepstatin suggests that optimization of statine-based inhibitors may be a viable strategy for inhibitor design.

Superposition of the two pepstatin molecules in the asymmetric unit can give us some idea of how well ordered pepstatin binding is in PvPM. The central parts of the two molecules correspond closely, but the termini appear "frayed" (Figure 2.16). That is, the orientations of Iva 1 and Sta 6 differ between the two molecules, with Sta 6 exhibiting a particularly large conformational change. This observation is not surprising, considering that the conformation of a substrate is most critical close to the scissile bond. Due to the preference of APs for long peptide substrates, and the potential involvement of torsional strain in catalysis, however, one may expect that the binding of the substrate to the AP should also be well defined at subsites more distant from the scissile bond. In this case, pepstatin may simply not be an adequate model of the substrate, since the positions of its side chains do not correspond to those in the substrate on the C-terminal (Pn') side. Thus, the conformational variability seen for Sta 6 may not be present in the substrate. All 48 atoms of pepstatin superimpose with a rather high rmsd, 0.91 Å. When we only consider the 32 atoms that superimpose within 0.4 Å, which is the estimated coordinate error of this structure, the rmsd improves to 0.15 Å (Figure 2.16).

#### 2.3.4 Comparison to PfPMII

The crystal structures of PfPMII and PvPM showed that both proteins exhibit a movement of a common subdomain [Silva *et al.*, 1996]. The magnitude of the PfPMII movement, however, is greater than that of PvPM. The domain shift difference has a noticeable effect on the active site clefts in the two molecules, which may be relevant to their function. Before comparing the structures, however, we should consider the possible effects of crystal packing on the conformations of the molecules.





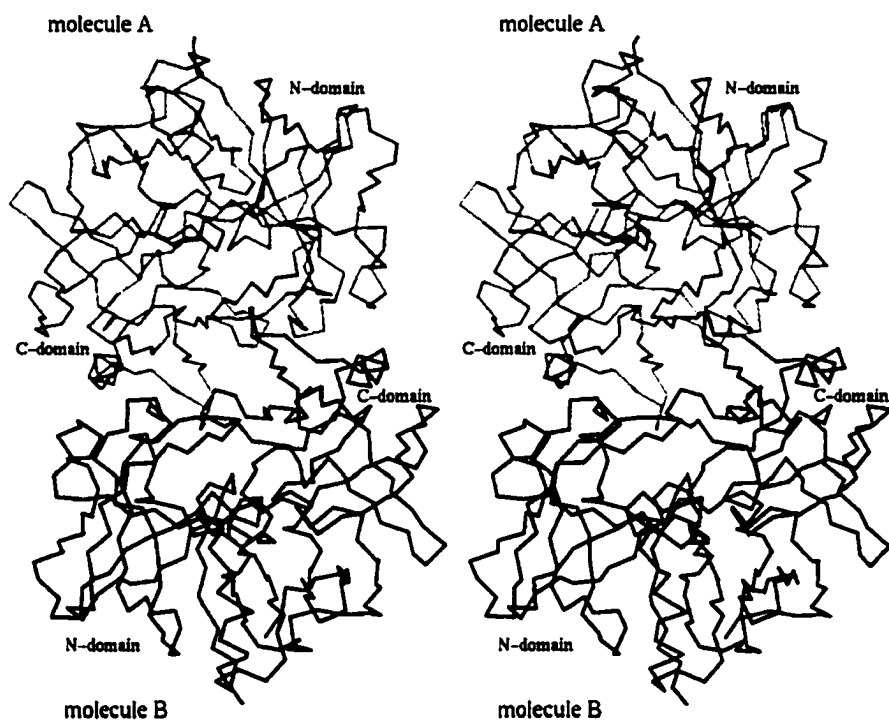
**Figure 2.16** Superposition of pepstatin in the two NCS-related molecules. Pairs of atoms separated by more than  $3\sigma$ , that were excluded from the final superposition, are colored magenta. The remaining atoms in molecule A are in yellow, and in molecule B, in cyan.

Both PvPM and PfPMII form dimers in the crystallographic asymmetric unit. The dimers, however, are different. In the PfPMII crystal, the two molecules pack quite symmetrically, interacting at their substrate-binding clefts. The PfPMII dimer is shown in figure 2.17a, looking down an approximate two-fold axis. In this arrangement, each molecule fits its 237 to 243 loop into the cleft of its partner, placing Phe 241 in the vicinity of the other molecule's S4 pocket. Superposition of the PfPMII dimer on itself, i.e. molecule B on molecule A, reveals that this dimer is not perfectly symmetric (Figure 2.17b). This can be expected, keeping in mind the 5° rotation by which the C-domains differ in molecules A and B. Overall, however, both molecules in the PfPMII asymmetric unit experience similar crystal packing environments. Thus the domain reorientation that they exhibit is most likely independent of the constraints of the crystal lattice. The dimer formed by PvPM is clearly asymmetric (Figure 2.18a), which is further confirmed by the superposition of the dimer on itself (Figure 2.18b). Molecule B points its Phe 241 into the S4 pocket of molecule A, but molecule A does not intrude into the cleft of molecule B. As discussed in a previous section, this packing difference causes some local conformational alteration between PvPM molecules A and B. On a global level, however, the two molecules of PvPM resemble each other more closely than do the two PfPMII molecules. Paradoxically, the pair of molecules that differ more, PfPMII, pack into a more symmetric dimer. This observation is encouraging, because it suggests that the difference in the magnitude of domain shifts in the PvPM and PfPMII crystal structures likely reflects the intrinsic flexibility of these enzymes.

Formation of the two dimers causes different amounts of accessible surface area to be buried. Dimerization of PfPMII buries 1232 Å<sup>2</sup> of accessible surface area, of which 85% (1046 Å<sup>2</sup>) is hydrophobic. In PvPM, dimerization buries a significantly smaller surface area, 829 Å<sup>2</sup>, which includes 75% (618 Å<sup>2</sup>) of hydrophobic area.

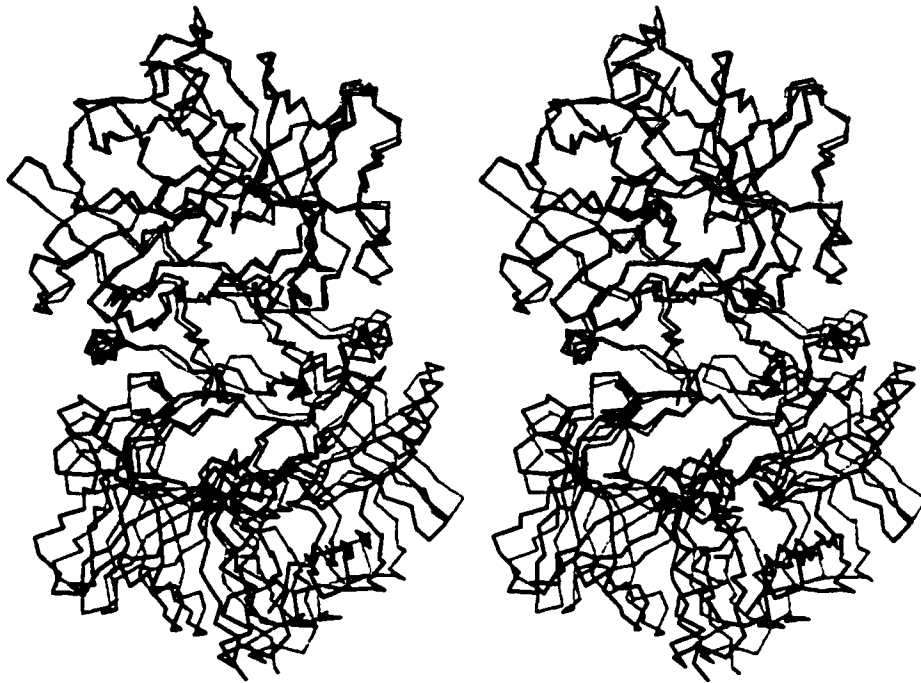
In PvPM, the loop 237 to 246 and the segment from 290 to 294, which form one edge of the cleft, are moved forward, i.e. into the cleft relative to PfPMII (Figure 2.19). Both of these effects decrease the size of the S4 pocket in PvPM, making it smaller than in PfPMII. Pro 243 is 3.6 Å closer to the N-lobe in PvPM than in PfPMII, and Leu 290 is 0.9 Å forward relative to Ile 290 in PfPMII.

(a)

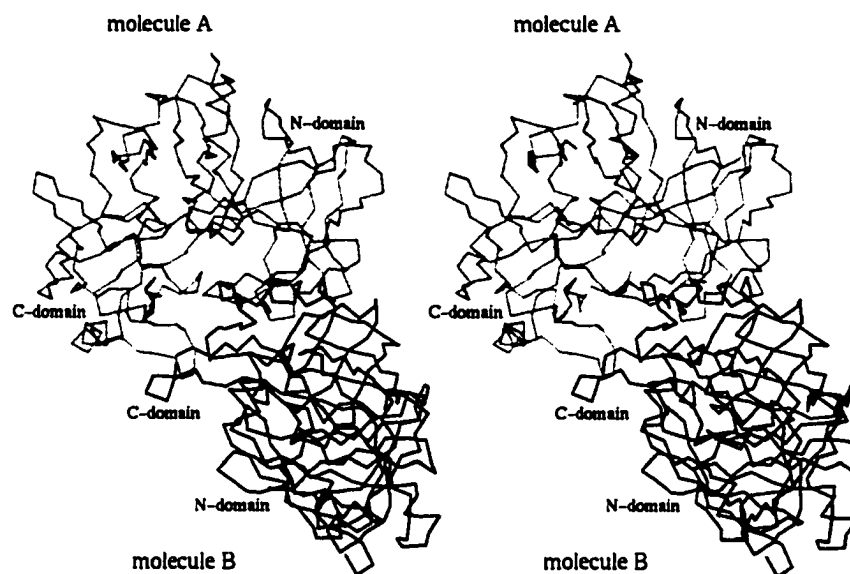


**Figure 2.17** The dimer in the asymmetric unit of the PfPMII crystal [Silva *et al.*, 1996]. Molecule A is colored cyan and molecule B is dark blue. (a) The PfPMII dimer. (b) The dimer superposed on itself (molecule A on molecule B).

(b)

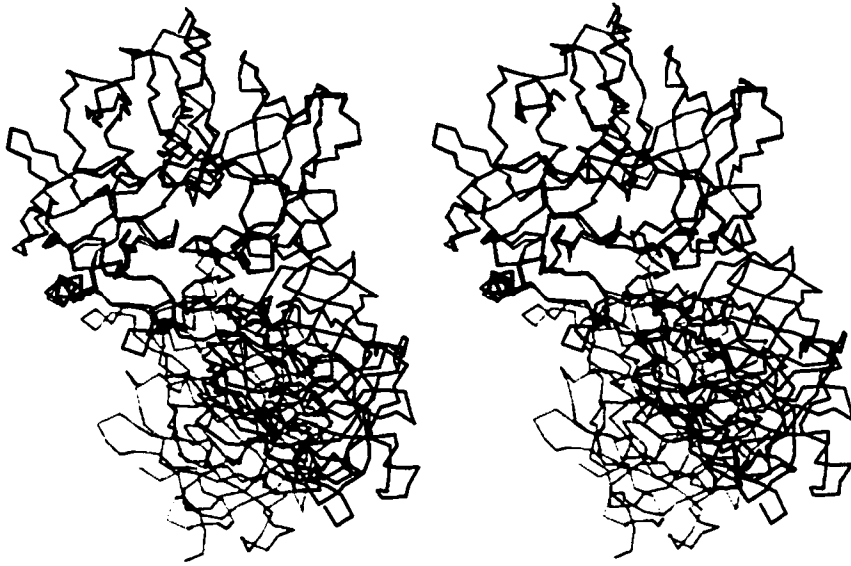


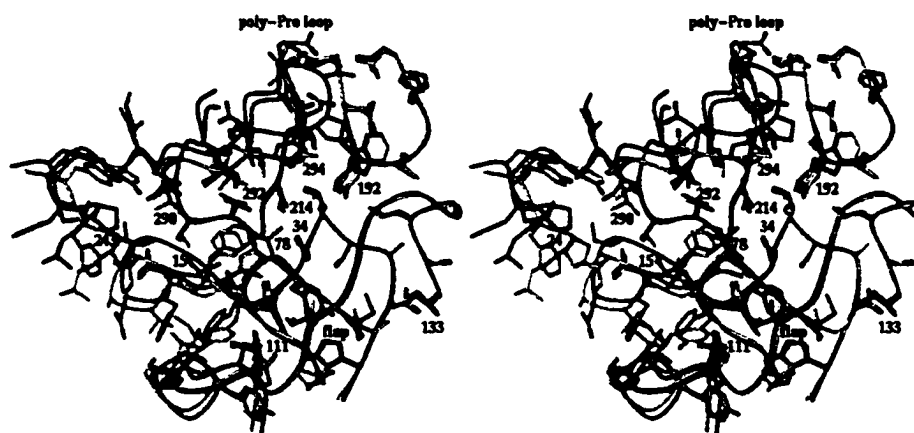
(a)



**Figure 2.18** The dimer in the asymmetric unit of the PvPM crystal. Molecule A is colored cyan and molecule B is dark blue. (a) The PvPM dimer. (b) The dimer superposed on itself (molecule A on molecule B).

(b)





**Figure 2.19** Superposition of the active site clefts of PvPM (yellow) and PfPMII (cyan). Side chains that point out of the cleft have been omitted for clarity.

A number of residues lining the substrate-binding cleft differ in amino acid type between PvPM and PfPMII, but the effect of the sequence substitutions is modulated by the difference in backbone positions due to the domain shift. Residue 292 is a Leu in PfPMII and a Val in PvPM. In previous studies, homology modelling of PvPM based on the structure of PfPMII, suggested that, since Val is smaller than Leu, the S2 pocket in PvPM is larger than in PfPMII [Westling *et al.*, 1997]. However, in the crystal structure of PvPM, the movement of the peptide backbone into the cleft shifts Val 292 forward so that its C $\gamma$ 1 corresponds to C $\delta$ 1 of Leu in PfPMII (Figure 2.19). The size of the S2 pocket is then approximately the same in the two enzymes. A similar argument goes for residue 294, forming part of the S1' pocket. This residue is an Ile in PvPM and a Phe in PfPMII.

Although Ile is smaller than Phe, the shift of the backbone forward in PvPM compensates for the size difference, leaving the S1' pocket sterically similar in the two structures (Figure 2.19).

One unusual feature of PfPMI and PfPMII is the substitution of Val at position 78 at the tip of the flap, for the highly conserved Gly. The only other known AP that does not have a Gly at this position is renin, where the corresponding residue is Ser 84. HAP, which is probably not a functional proteinase, but has a similar sequence to the plasmepsins, has a Lys, and several PAGs (pregnancy-associated glycoproteins, whose sequences also resemble APs) have His or Thr at this position [Berry *et al.*, 1999]. In PvPM and in all the other known plasmepsins, this residue is a Gly. The Val side chain causes a slight displacement of the tip of the flap (residues 78 and 79) in PfPMII relative to PvPM, as well as some difference in the conformation of Sta 6 of pepstatin. The distance from the tip of the flap across the cleft is shorter in PfPMII than in PvPM. In PfPMII, Val 78 C $\gamma$ 1 and Leu 292 C $\delta$ 2 are separated by 4.5 Å, whereas in PvPM, Gly 78 C $\alpha$  and Val 292 C $\delta$ 1 are separated by 6.5 Å. It appears that in PfPMII the entrance to the active site cleft is narrower and the scissile bond of the substrate is better sequestered from the solvent than in PvPM. The presence of inhibitor in both structures suggests that the observed differences are real.

The helix next to the flap (108 - 118) is a major component of the substrate-binding cleft. Several residues on this helix differ between PvPM and PfPMII. Residue



111 is located at the bottom of the S1 pocket. This residue is a Phe in PfPMII and a Leu in PvPM. The smaller Leu results in a more spacious S1 pocket in PvPM. In addition, the electrostatic environment of this pocket must be different due to the substitution of an aliphatic side chain for a phenyl ring. Residue 110 is an Asp in PvPM and a Gly in PfPMII. In one molecule of PvPM, the side chain of Asp 110 accepts a hydrogen bond from Thr 81 NH, a residue in the flap. The absence of this hydrogen bond in PfPMII may affect the flexibility of the flap in that protein. Finally, the substitution of Ile in PvPM for Thr at position 114 probably has little effect on the S3 pocket, since both residues point their  $\gamma$ -methyl group into the pocket.

The loop between residues 294 and 298 has been shown to be important for substrate recognition and binding by renin [Dhanaraj *et al.*, 1992]. In many eukaryotic APs, this loop contains several, often sequential, proline residues, and is therefore called the "poly-proline loop". This loop differs in sequence between PvPM and PfPMII. In PfPMII, the sequence is Pro(295)-Val-Pro-Thr. In PvPM, this loop consists entirely of hydrophilic residues: Asp(295)-Ser-Asn-Thr. The conformation of this loop is also markedly different in the two structures (Figure 2.19), with PvPM presenting a more hydrophilic surface to an incoming substrate. The different composition and structure of this loop in the two enzymes may reflect preferences for different physiological substrates. Alternatively, the "poly-proline loop" may not be involved in substrate recognition in the plasmepsins [Luker *et al.*, 1996]. In both plasmepsins, the "poly-proline loop" is considerably shorter than in the vertebrate APs, such as cathepsin D and pepsin. The "poly-proline loop" does not affect S3' specificity in cathepsin D [Beyer & Dunn, 1998], suggesting that control of substrate recognition by this loop may only be necessary for a proteinase of exquisite specificity such as renin.

It is interesting to compare the structures of the flap in PvPM and PfPMII. In PvPM, a hydrogen bond between Asp 110 and Thr 81 NH likely provides stabilization to the flap, that is absent in PfPMII, where residue 110 is a Gly. In PvPM, the side chains of Thr 76 and Thr 81 form a hydrogen bond across the  $\beta$ -hairpin. In PfPMII, the corresponding hydrogen bond exists between Asn 76 and Thr 81. However, since asparagine has more degrees of freedom than threonine, it may be less energetically favorable to form this hydrogen bond in PfPMII. Another pair of residues that can interact across the hairpin is Asp 74 and Lys 83 in PvPM. No salt bridge is observed

between these surface residues in the crystal structure, but it is possible for them to interact electrostatically, providing some stabilization to the flap. In PfPMII, the corresponding residues are Glu 74 and Ser 83, whose potential interaction will probably be less stabilizing. Finally, residue 75 is an Ile in PvPM and a Met in PfPMII. The  $\beta$ -branched Ile is more conformationally restricted than the straight chain Met. As will be described in Chapters 3 and 4, the crystal structures of the zymogens of PvPM and PfPMII show more flexibility in the flap of *P. falciparum* proplasmepsin II. All of these factors may contribute to a more rigid flap in PvPM than in PfPMII.

PfPMII appears to be intrinsically more flexible than PvPM. A wider range of domain motion has been observed in PfPMII, and the flap may also be more flexible in PfPMII than in PvPM. Is there a reason for the increased flexibility of PfPMII? Additional flexibility may be necessary for PfPMII due to a narrower opening of the active site cleft near the tip of the flap, where Val 78 replaces the more commonly found Gly. In PvPM, the gap between the tip of the flap and the opposite side of the cleft is sufficiently wide, that a large rearrangement of the enzyme should not be required to admit substrate to the active site. Westling *et al.* studied the ability of recombinant PfPMII, PvPM and PmPM to hydrolyze a panel of systematically varied hemoglobin-based octapeptides [Westling *et al.*, 1997]. All the substrates tested had comparable  $K_m$  values for the three enzymes, but the  $k_{cat}$  values of PfPMII were markedly lower. Perhaps the reduced kinetic efficiency of PfPMII is due to a narrower opening of the active site cleft than in PvPM (and in PmPM, see below). Thus, the apparent difference in the flexibilities of the two enzymes may be related to their function. Of course, since we do not know the physiological substrate of PvPM, the experiments may not have reflected accurately the situation encountered *in vivo*, and we must treat this hypothesis with some caution.

### 2.3.5 Substrate specificities of the plasmepsins

A comparison of the substrate specificity of PfPMII, PvPM, PoPM and PmPM has been reported by Hung and Dunn [1999]. While the four enzymes showed similar preferences at subsites S3, S2 and S3', some differences were observed at subsite S2'. PfPMII and PoPM preferred Ala, and PvPM and PmPM preferred Arg at position P2'. Before the crystal structure of PvPM was available, plasmepsin kinetics were analyzed in terms of homology models built from the PfPMII structure of Silva *et al.* [1996]. As

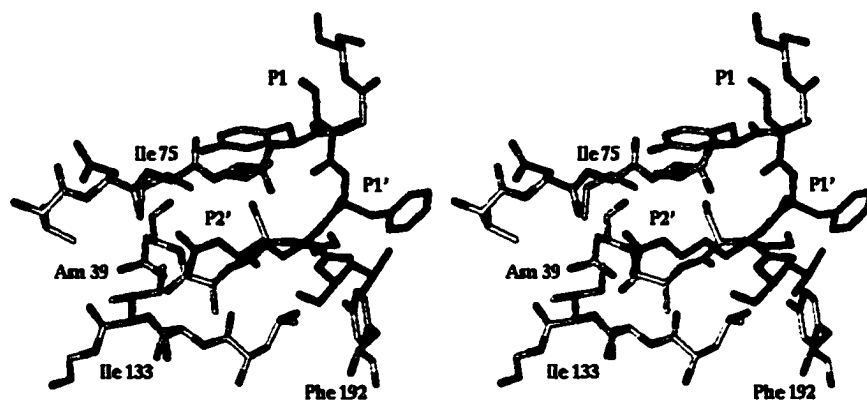
discussed in the previous section (Comparison to PfPMII), significant differences exist between the crystal structures of PfPMII and PvPM, suggesting that PfPMII was not an optimal model for PvPM. Because PvPM is closer in sequence to PoPM and PmPM than to PfPMII, PvPM should be a more suitable scaffold than PfPMII for constructing homology models of PoPM and PmPM. We therefore used our crystal structure of PvPM to build homology models of PoPM and PmPM in order to rationalize the differences in their substrate preferences at position P2'. A model of the substrate was built based on the coordinates of a phosphonate inhibitor of penicillopepsin [Fraser *et al.*, 1992].

Judging from the crystal structures of PfPMII and PvPM, and from the modeled structures of PoPM and PmPM, the difference in the S2' specificities of these enzymes may be due to the nature of residue 75. In both PvPM and PmPM, residue 75 is an Ile. In PfPMII it is a Met, and in PoPM it is a Leu. Being a  $\beta$ -branched amino acid, Ile has less conformational freedom than its non  $\beta$ -branched counterparts. It appears that the conformation that Ile 75 adopts ( $\chi_1 = -60^\circ$ ,  $\chi_2 = -60^\circ$ ) is the only one in which its side chain does not experience any steric clashes. Changing  $\chi_1$  to  $-180^\circ$  results in steric repulsion between Ile 75 C $\gamma_2$  and Val 82. Likewise, a  $\chi_1$  of  $60^\circ$  causes a severe clash of Ile 75 with Val 82 and Tyr 77. Therefore, Ile 75 is limited to having a  $\chi_1$  of  $-60^\circ$ . In this conformation, its C $\gamma_2$  is directed into the S2' pocket. In the PfPMII structure, the entire Met 75 side chain points into the S2' pocket. In the PoPM model, the Leu side chain points away from the S2' pocket, but it is also possible for it to have the same conformation as Met 75 in PfPMII, i.e., pointing into the S2' pocket. In that case, the side chains of Met and Leu may crowd the S2' pocket more than the much smaller methyl group (C $\gamma_2$ ) of Ile, leaving only enough space for an Ala of the substrate (Figure 2.20a). In PvPM and PmPM the S2' pocket appears to be larger, and better able to accommodate an Arg side chain at P2' (Figure 2.20b). It is also entirely possible that domain and loop flexibility may give rise to the differences in kinetic parameters of the plasmepsins.

### 2.3.6 Comparison with human APs

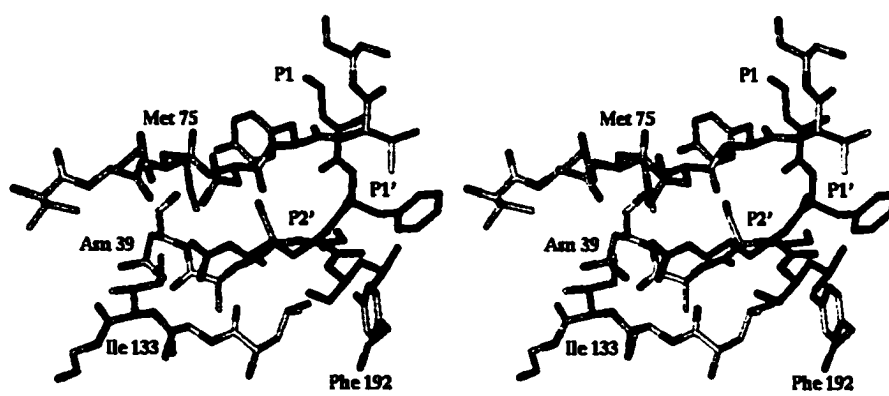
The human body contains many APs, such as the well-known gastric enzymes (pepsin and gastricsin), cathepsin D, cathepsin E and renin. In addition, several novel APs (napsins and memapsins) have recently been discovered [Tatnell *et al.*, 1998; Lin *et al.*,

(a)



**Figure 2.20** Modelled substrate in the S2' pockets of PvPM (a) and PfPMII (b). Substrate carbon atoms are colored cyan and enzyme carbon atoms are colored yellow.

(b)



2000]. If inhibitors of PvPM are to be developed into drugs, they must be selective for the parasite's proteinase and must not cross-react with human enzymes. Currently, structural information is available for pepsin, cathepsin D and renin.

Cross-reactivity of PvPM inhibitors with renin is probably a minor concern, since renin has a precise specificity for its single substrate, angiotensinogen. On a structural level, renin differs from PvPM at the tip of the flap, where the distance across the active site cleft, between Ser 84 O $\gamma$  and Met 303 C $\epsilon$ , is only 3.2 Å, compared to 6.5 Å for the corresponding residue pair (Gly 78 and Val 292) in PvPM. The S4 pocket differs in shape between renin and PvPM. Thr 219 and Leu 290 of PvPM are replaced by Tyr 231 and His 301, respectively, in renin. Met 15 in PvPM, a residue that influences P3 specificity, is replaced in renin by the polar Gln 19. Finally, the shape and amino acid composition of the poly-proline loop differs between the two enzymes. In renin this loop is longer, and protrudes more into the active site cleft.

Pepsin acts in the highly acidic environment of the human stomach (pH = 2) and undergoes irreversible denaturation above pH 5.4 [Foltmann, 1981]. PvPM most likely works at a higher pH than pepsin. Interactions of an inhibitor with these two enzymes at different pHs may differ. The loop from 237 to 245, forms part of the S4 pocket in PvPM, but points away from the pocket in pepsin. The poly-proline loop in pepsin is longer than in PvPM. Glu 13 in pepsin corresponds to Met 15 in PvPM. This sequence substitution affects the substrate specificity at P3, although at pH 2 Glu 13 is likely to be protonated, in which case the difference between this residue and Met will be less dramatic.

The lysosomal proteinase cathepsin D is the closest in sequence to PvPM. However, some structural differences can be detected between the two enzymes (Figure 2.21). The loop from 237 to 245 (part of the S4 pocket) is more hydrophilic in cathepsin D, with Glu 260B in cathepsin D corresponding to Pro 243 in PvPM. In cathepsin D, the distance across the cleft, between Gly 79A at the tip of the flap and Met 309B, is 4.7 Å, which is more similar to PfPMII (4.5 Å) than to PvPM (6.5 Å). The poly-proline loop is longer in cathepsin D and folded into the cleft, whereas in PvPM this loop points out of the cleft. In the S1 pocket, the aromatic ring of Phe 126B of cathepsin D is replaced by the side chain of Leu 111. The residue in cathepsin D that corresponds to Met 15 of PvPM is Gln 14A. The subtle steric and electronic differences in the S1/S3 pocket



**Figure 2.21** Superposition of the active site clefts of PvPM (yellow) and cathepsin D (pink). Residue numbering corresponds to PvPM.

between cathepsin D and PvPM can be helpful in endowing PvPM inhibitors with selectivity.

## 2.4 Conclusion

The crystal structure of PvPM is the first structure of the AP that is common to all four species of *Plasmodium* that infect humans. PvPM is, on the whole, a typical aspartic proteinase, exhibiting many structural features common to this class of enzymes.

Both of the currently known plasmepsin structures (PfPMII and PvPM) describe the enzyme in complex with pepstatin. APs whose three-dimensional structures have been elucidated with and without pepstatin, show small adjustments of the active site to enclose the inhibitor more effectively [Baldwin *et al.*, 1993; Fujinaga *et al.*, 1995]. An isothermal titration calorimetry study of the interaction of pepstatin with PfPMII indicated that a localized structural change (in the vicinity of His 164) occurs in the enzyme upon inhibitor binding [Xie *et al.*, 1997]. Crystal structures of uninhibited plasmepsins will be needed to explain this change. Structural information on the native plasmepsins should also be helpful for inhibitor design.

A comparison with the crystal structure of PfPMII suggests that PvPM possesses a lower degree of flexibility, which may support the idea that PvPM and related plasmepsins are distinct in function from PfPMI and PfPMII. The latter two enzymes have been characterized as hemoglobinases [Francis *et al.*, 1994; Gluzman *et al.*, 1994]. PfPMII has also been shown to act as a spectrinase [Le Bonniec *et al.*, 1999]. The function of PvPM, PmPM, PoPM and PfPMIV is possibly different from these. The spectrum of roles of this group of enzymes remains to be determined.



## 2.5 References

Abad-Zapatero C, Rydel TJ, Erickson J (1990) Revised 2.3 Å structure of porcine pepsin: evidence for a flexible subdomain. *Proteins* **8**: 62-81.

Baldwin ET, Bhat TN, Gulnik S, Hosur MV, Sowder RC, Cachau RE, Collins J, Silva AM, Erickson JW (1993) Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design. *Proc Natl Acad Sci U S A* **90**: 6796-800.

Berry C, Humphreys MJ, Matharu P, Granger R, Horrocks P, Moon RP, Certa U, Ridley RG, Bur D, Kay J (1999) A distinct member of the aspartic proteinase gene family from the human malaria parasite *Plasmodium falciparum*. *FEBS Lett* **447**: 149-54.

Beveridge A (1996) A theoretical study of torsional flexibility in the active site of aspartic proteinases: implications for catalysis. *Proteins* **24**: 322-34.

Beveridge AJ, Heywood GC (1993) A quantum mechanical study of the active site of aspartic proteinases. *Biochemistry* **32**: 3325-33.

Beyer BM, Dunn BM (1998) Prime region subsite specificity characterization of human cathepsin D: the dominant role of position 128. *Protein Sci* **7**: 88-95.

Bowman S, Lawson D, Basham D, Brown D, Chillingworth T, Churcher CM, Craig A, Davies RM, Devlin K, Feltwell T, Gentles S, Gwilliam R, Hamlin N, Harris D, Holroyd S, Hornsby T, Horrocks P, Jagels K, Jassal B, Kyes S, McLean J, Moule S, Mungall K, Murphy L, Barrell BG, *et al* (1999) The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* **400**: 532-538.

Bricogne G (1974) Geometric sources of redundancy in intensity data and their use for phase determination. *Acta Crystallogr* **A30**: 395-405.

Bruce-Chwatt LJ (1985) *Essential malariology*. John Wiley & Sons, New York.

Brunger AT (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**: 472-475.

Brunger AT (1993) *XPLOR: a system for X-ray crystallography and NMR* (Yale University Press, New Haven, Connecticut).

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54: 905-21.

Davies DR (1990) The structure and function of the aspartic proteinases. *Annu Rev Biophys Biophys Chem* 19: 189-215.

Dhanaraj V, Dealwis CG, Frazao C, Badasso M, Sibanda BL, Tickle IJ, Cooper JB, Driessen HP *et al.* (1992) X-ray analyses of peptide-inhibitor complexes define the structural basis of specificity for human and mouse renins. *Nature* 357: 466-72.

Esnouf RM (1997) An extensively modified version of MolScript that includes greatly enhanced colouring capabilities. *J Mol Graphics* 15: 133-138.

Faro C, Ramalho-Santos M, Vieira M, Mendes A, Simoes I, Andrade R, Verissimo P, Lin X, Tang J, Pires E (1999) Cloning and characterization of cDNA encoding cardosin A, an RGD-containing plant aspartic proteinase. *J Biol Chem* 274: 28724-9.

Foltmann B (1981) Gastric proteinases--structure, function, evolution and mechanism of action. *Essays Biochem* 17: 52-84.

Francis SE, Gluzman IY, Oksman A, Banerjee D, Goldberg DE (1996) Characterization of native falcipain, an enzyme involved in *Plasmodium falciparum* hemoglobin degradation. *Mol Biochem Parasitol* 83: 189-200.

Fraser ME, Strynadka NC, Bartlett PA, Hanson JE, James MNG (1992) Crystallographic analysis of transition-state mimics bound to penicillopepsin: phosphorus-containing peptide analogues. *Biochemistry* 31: 5201-14.

Fruton JS (1976) The mechanism of the catalytic action of pepsin and related acid proteinases. *Adv Enzymol Relat Areas Mol Biol* 44: 1-36.

Fujinaga M, Chernai MM, Tarasova NI, Mosimann SC, James MNG (1995) Crystal structure of human pepsin and its complex with pepstatin. *Protein Sci* 4: 960-72.

Fusek M, Vetvicka V (1995) *Aspartic proteinases: physiology and pathology*. (CRC Press Inc., Boca Raton).

- Gluzman IY, Francis SE, Oksman A, Smith CE, Duffin KL, Goldberg DE (1994) Order and specificity of the *Plasmodium falciparum* hemoglobin pathway. *J Clin Invest* **93**: 1602-8.
- Hofmann T, Allen B, Bendiner M, Blum M, Cunningham A (1988) Effect of secondary substrate binding in penicillopepsin: contributions of subsites S3 and S2' to kcat. *Biochemistry* **27**: 1140-6.
- Hube B (1996) *Candida albicans* secreted aspartyl proteinases. *Curr Top Med Mycol* **7**: 55-69.
- Humphreys MJ, Moon RP, Klinder A, Fowler SD, Rupp K, Bur D, Ridley RG, Berry C (1999) The aspartic proteinase from the rodent parasite *Plasmodium berghei* as a potential model for plasmepsins from the human malaria parasite, *Plasmodium falciparum*. *FEBS Lett* **463**: 43-8.
- Hung SH, Dunn BM (1999) The two sides of enzyme-substrate specificity: lessons from the classic enzymes. Presented at the VIIIth International Aspartic Proteinase Conference, Madeira, Portugal.
- James MNG, Sielecki AR, Hayakawa K, Gelb MH (1992) Crystallographic analysis of transition state mimics bound to penicillopepsin: difluorostatine- and difluorostatone-containing peptides. *Biochemistry* **31**: 3872-86.
- Jones TA, Kjeldgaard M (1995) *O the Manual, Version 5.11* (Uppsala, Sweden).
- Kleywegt GJ, Read RJ (1997) Not your average density. *Structure* **5**: 1557-69.
- Kraulis PJ (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Cryst* **24**: 946-950.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* **26**: 283-291.
- Laco GS, Schalk-Hihi S, Lubkowski J, Morris G, Zdanov A, Olson A, Elder JH, Wlodawer A, Gustchina A (1997) Crystal structures of the inactive D30N mutant of feline immunodeficiency virus protease complexed with a substrate and an inhibitor. *Biochemistry* **36**: 10696-708.

- Laurent F, Bourdieu C, Kaga M, Chilmonczyk S, Zgrzebski G, Yvone P, Pery P (1993) Cloning and characterization of an *Eimeria acervulina* sporozoite gene homologous to aspartyl proteinases. *Mol Biochem Parasitol* **62**: 303-12.
- Le Bonniec S, Deregnaucourt C, Redeker V, Banerjee R, Grellier P, Goldberg DE, Schrevel J (1999) Plasmepsin II, an acidic hemoglobinase from the *Plasmodium falciparum*. *J Biol Chem* **274**: 14218-23.
- Lee AY, Gulnik SV, Erickson JW (1998) Conformational switching in an aspartic proteinase. *Nat Struct Biol* **5**: 866-71.
- Lin X, Koelsch G, Wu S, Downs D, Dashti A, Tang J (2000) Human aspartic protease memapsin 2 cleaves the beta -secretase site of beta -amyloid precursor protein. *Proc Natl Acad Sci U S A* **97**: 1456-1460.
- Lin XL, Wong RN, Tang J (1989) Synthesis, purification, and active site mutagenesis of recombinant porcine pepsinogen. *J Biol Chem* **264**: 4482-9.
- Luker KE, Francis SE, Gluzman IY, Goldberg DE (1996) Kinetic analysis of plasmepsins I and II aspartic proteases of *Plasmodium falciparum* digestive vacuole. *Mol Biochem Parasitol* **79**: 71-78.
- Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* **33**: 491-497.
- Merritt EA, Murphy MEP (1994) Raster3D version 2.0: a program for photorealistic molecular graphics. *Acta Cryst D* **50**: 869-873.
- Navaza J (1994) AMoRe: an automated package for molecular replacement. *Acta Cryst, A* **50**: 157-163.
- Otwinowski Z, Minor W (1996) Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology*, **276**; (Carter Jr CW, Sweet R, eds., Academic Press, New York and London) pp. 307-326.
- Pannu NS, Read RJ (1996) Improved structure refinement through maximum likelihood. *Acta Cryst A* **52**: 659-668.
- Pearl L, Blundell T (1984) The active site of aspartic proteinases. *FEBS Lett* **174**: 96-101.

- Pearl LH (1984) The extended binding cleft of aspartic proteinases and its role in peptide hydrolysis. In *Aspartic proteinases and their inhibitors*. (Kostka V, ed., de Gruyter, Berlin). pp. 189-195.
- Pearl LH (1987) The catalytic mechanism of aspartic proteinases. *FEBS Lett* **214**: 8-12.
- Rajagopalan TG, Stein WH, Moore S (1966) The inactivation of pepsin by diazoacetyl norleucine methyl ester. *J Biol Chem* **241**: 4295-7.
- Richards FM, Kundrot CE (1988) Identification of structural motifs from protein coordinate data: secondary structure and first-level supersecondary structure. *Proteins* **3**: 71-84.
- Sali A, Veerapandian B, Cooper JB, Foundling SI, Hoover DJ, Blundell TL (1989) High-resolution X-ray diffraction study of the complex between endothiapepsin and an oligopeptide inhibitor: the analysis of the inhibitor binding and description of the rigid body shift in the enzyme. *EMBO J* **8**: 2179-88.
- Sali A, Veerapandian B, Cooper JB, Moss DS, Hofmann T, Blundell TL (1992) Domain flexibility in aspartic proteinases. *Proteins* **12**: 158-70.
- Schechter I, Berger A (1967) On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* **27**: 157-162.
- Sielecki AR, Fedorov AA, Boodhoo A, Andreeva NS, James MNG (1990) Molecular and crystal structures of monoclinic porcine pepsin refined at 1.8 Å resolution. *J Mol Biol* **1214**: 143-70.
- Silva AM, Lee AY, Gulnik SV, Maier P, Collins J, Bhat TN, Collins PJ, Cachau RE, Luker KE, Gluzman IY, Francis SE, Oksman A, Goldberg DE, Erickson JW (1996) Structure and inhibition of plasmepsin II, a hemoglobin-degrading enzyme from *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* **93**: 10034-9.
- Suguna K, Padlan EA, Smith CW, Carlson WD, Davies DR (1987) Binding of a reduced peptide inhibitor to the aspartic proteinase from *Rhizopus chinensis*: implications for a mechanism of action. *Proc Natl Acad Sci U S A* **84**: 7009-13.
- Tang J (1971) Specific and irreversible inactivation of pepsin by substrate-like epoxides. *J Biol Chem* **246**: 4510-7.

- Tang J, James MNG, Hsu IN, Jenkins JA, Blundell TL (1978) Structural evidence for gene duplication in the evolution of the acid proteases. *Nature* **271**: 618-21.
- Tatnell PJ, Powell DJ, Hill J, Smith TS, Tew DG, Kay J (1998) Napsins: new human aspartic proteinases. Distinction between two closely related genes. *FEBS Lett* **441**: 43-8.
- Togni G, Sanglard D, Monod M (1994) Acid proteinase secreted by *Candida tropicalis*: virulence in mice of a proteinase negative mutant. *J Med Vet Mycol* **32**: 257-65.
- Tyas L, Gluzman I, Moon RP, Rupp K, Westling J, Ridley RG, Kay J, Goldberg DE, Berry C (1999) Naturally-occurring and recombinant forms of the aspartic proteinases plasmepsins I and II from the human malaria parasite *Plasmodium falciparum*. *FEBS Lett* **454**: 210-4.
- Veerapandian B, Cooper JB, Sali A, Blundell TL, Rosati RL, Dominy BW, Damon DB, Hoover DJ (1992) Direct observation by X-ray analysis of the tetrahedral "intermediate" of aspartic proteinases. *Protein Sci* **1**: 322-8.
- Vetvicka V, Vetvickova J, Hilgert I, Voburka Z, Fusek M (1997) Analysis of the interaction of procathepsin D activation peptide with breast cancer cells. *Int J Cancer* **73**: 403-9.
- Westling J, Cipullo P, Hung SH, Saft H, Dame JB, Dunn BM (1999) Active site specificity of plasmepsin II. *Protein Sci* **8**: 2001-9.
- Westling J, Yowell CA, Majer P, Erickson JW, Dame JB, Dunn BM (1997) *Plasmodium falciparum*, *P. vivax*, and *P. malariae*: a comparison of the active site properties of plasmepsins cloned and expressed from three different species of the malaria parasite. *Exp Parasitol* **87**: 185-93.
- Xie D, Gulnik S, Collins L, Gustchina E, Suvorov L, Erickson JW (1997) Dissection of the pH dependence of inhibitor binding energetics for an aspartic protease: direct measurement of the protonation states of the catalytic aspartic acid residues. *Biochemistry* **36**: 16166-72.
- Xu Q, Buckley D, Guan C, Guo HC (1999) Structural insights into the mechanism of intramolecular proteolysis. *Cell* **98**:651 – 661.

Yoshihara M, Sumii K, Haruma K, Kiyohira K, Hattori N, Kitadai Y, Komoto K, Tanaka S, Kajiyama G (1998) Correlation of ratio of serum pepsinogen I and II with prevalence of gastric cancer and adenoma in Japanese subjects. *Am J Gastroenterol* 93: 1090-6.

# Chapter 3: CRYSTAL STRUCTURE OF PROPLASMEPSIN II FROM *P. falciparum*, A NOVEL ASPARTIC PROTEINASE ZYMOGEN

## 3.1 Introduction

All known aspartic proteinases are produced as inactive precursors that become converted to mature enzymes when their proteolytic function is required. The precursors, or zymogens, of eukaryotic APs contain an N-terminal extension (also called the prosegment or activation peptide), that interferes with catalytic activity and must be removed in the conversion to the active enzyme (activation). In many cases, additional proteolytic cleavages are required in the mature portion of the zymogen to generate the active proteinase. Retroviral APs, such as HIV proteinase, are synthesized as part of a polyprotein, where the segment N-terminal to the proteinase subunit may have inhibitory properties [Partin *et al.*, 1991]. An elaborate and precisely controlled process of activation is important for regulation of activity, correct folding and proper cellular sorting of the APs [Koelsch *et al.*, 1994]. Activation of aspartic proteinases has been extensively studied by biochemical and structural methods. The three-dimensional structures of many APs are known. Until last year, however, structural information was only available on the activation pathway of the gastric APs, enzymes that digest dietary protein in the gastric juice of vertebrates. Together with biochemical data, the crystal structures of porcine pepsinogen A, human progastricsin, human pepsinogen A and a progastricsin activation intermediate were used to explain, on a molecular level, why the gastric zymogens lack proteolytic activity and to describe the autocatalytic activation pathway at low pH [Sielecki *et al.*, 1991; Hartsuck *et al.*, 1992; Moore *et al.*, 1995; Bateman *et al.*, 1998; Khan *et al.*, 1997; Richter *et al.*, 1998; Khan *et al.*, 1999]. We have determined the first three-dimensional structure of a non-gastric aspartic proteinase zymogen, proplasmepsin II from the human malarial parasite *Plasmodium falciparum* [Dame *et al.*, 1994; Francis *et al.*, 1994; Bernstein *et al.*, 1999]. Recently, the crystal structure of a plant aspartic proteinase zymogen, prophytepsin from barley, was determined [Kervinen *et al.*, 1999].



Proplasmepsin II shows remarkable differences from both the gastric zymogens and prophytpsin, and represents a novel class of aspartic proteinase zymogen.

### 3.1.1 Gastric AP zymogens

The gastric APs include pepsin A, pepsin B, gastricsin and chymosin. Pepsin and gastricsin are found in adult vertebrates, and chymosin is present in the stomachs of neonate ruminants. Chymosin is also commercially important in the production of cheese, where it is used as a milk-clotting agent. The gastric APs are secreted proteins, initially synthesized in the neutral chief cells lining the stomach as preproenzymes, with the mature sequence preceded by a 40 - 50 amino acid long prosegment and a signal sequence [Fusek & Vetvicka, 1995]. Ingestion of food stimulates the release of the soluble zymogens into the lumen of the stomach. Upon exposure to the low pH environment of the stomach, the zymogens become converted to the mature enzymes. In most cases, the activation process is autocatalytic, involving one or more cleavages of the prosegment. The prosegments of gastric AP zymogens from different organisms show a high degree of sequence conservation [Richter *et al.*, 1998], specifically of a Lys - Tyr pair (36p - 37p in porcine pepsinogen). (Throughout the text, the suffix "p" in residue numbering refers to residues in the prosegment.) As crystal structures have shown, these two residues are crucial to the inhibition of catalytic activity in the zymogens [James & Sielecki, 1986; Sielecki *et al.*, 1991; Hartsuck *et al.*, 1992; Moore *et al.*, 1995].

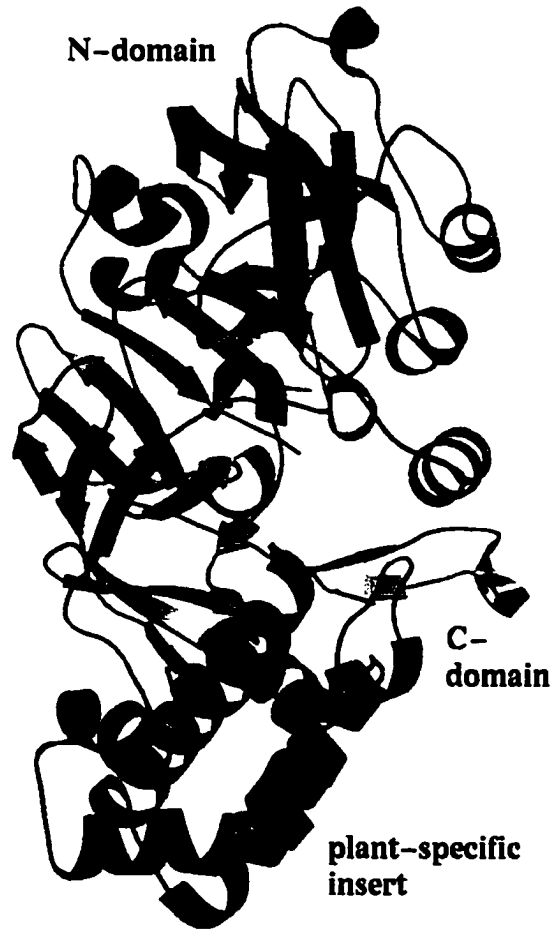
Porcine pepsinogen A, a representative of the gastric AP zymogens, is stabilized at neutral pH by the positively charged amino-terminal prosegment that occludes the active site cleft (Figure 3.1a)[Sielecki *et al.*, 1991; Hartsuck *et al.*, 1992]. The 44-residue prosegment is folded into a  $\beta$ -strand (Leu 1p to Leu 6p), 2  $\alpha$ -helices (Leu 12p to Lys 18p, Lys 21p to Thr 28p) and a  $3_{10}$ -helix (Pro 33p to Tyr 37p). The prosegment  $\beta$ -strand participates in the central six-stranded  $\beta$ -sheet, and becomes replaced by the mature N-terminus (Ile 1 to Leu 6) in pepsin. The helices lie in the substrate-binding cleft and block a pre-formed active site. The conserved lysine (36p) is directed into the active site, where it interacts with the two catalytic aspartates (32 and 215), neutralizing their potential mutual electrostatic repulsion at pH > 6.5. Two symmetrically disposed tyrosines (37p and 9) form hydrogen bonds to the catalytic aspartates, blocking the S1 and S1' substrate-binding pockets (Figure 3.2a). As the pH is lowered and acidic residues become protonated, the

(a)

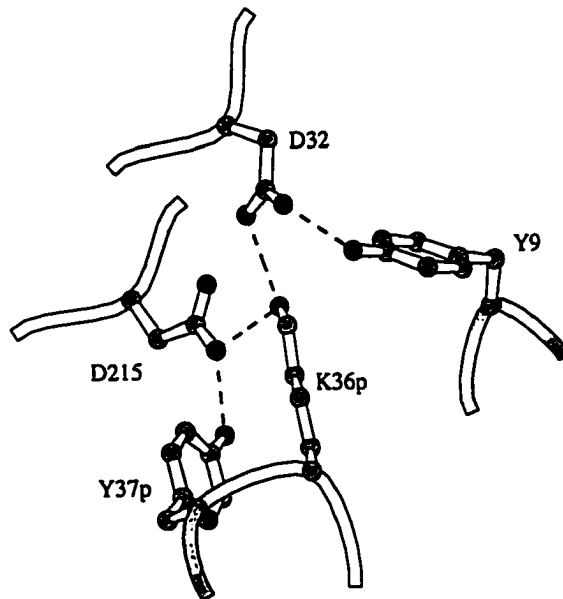


**Figure 3.1** The structures of porcine pepsinogen (a) and prophytepsin (b). The prosegment is colored magenta, the first 13 residues of the mature sequence are in cyan, and the remainder of the mature sequence is in yellow. The plant-specific insert in prophytepsin is shown in green. The side chains of the catalytic aspartates and the inhibitory lysine are shown in red and blue, respectively. The program Bobscript [Esnouf, 1997; Merritt & Murphy, 1994] was used to produce this figure, figures 3.5 – 3.10 and 3.12 – 3.18.

(b)

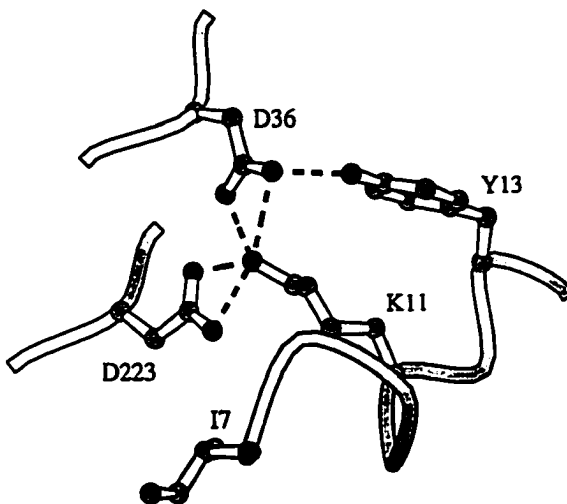


(a)



**Figure 3.2** The blocked active sites of pepsinogen and prophytepsin. Porcine pepsinogen (a) and prophytepsin (b) utilize somewhat different approaches to occlude their pre-formed active sites.

(b)



ionic interactions holding the prosegment helices in place disappear, and the prosegment unravels exposing a pre-formed active site. One or more specific cleavages, followed by dissociation of the prosegment and refolding of the mature N-terminus yield the active enzyme [Richter *et al.*, 1998]. The crystal structure of progastricsin intermediate II demonstrated that the prosegment  $\beta$ -strand remains non-covalently attached to the mature portion of the molecule, as part of the central  $\beta$ -sheet, after the C-terminal portion of the prosegment has vacated the active site cleft [Khan *et al.*, 1997]. This observation is not surprising, considering that while the prosegment helices interact with the mature portion primarily through ion pairing, the the prosegment  $\beta$ -strand forms numerous hydrogen bonds and hydrophobic interactions with the mature portion. Thus, a drop in pH is likely to disrupt the association of the prosegment helices with the mature portion, but should have little or no effect on the contacts made by the  $\beta$ -strand [Khan *et al.*, 1997]. Aside from the refolding of the N-terminus, only minor changes in conformation occur in the mature portion of pepsinogen upon conversion to pepsin.

The Lys36p-Tyr37p pair in pepsinogen, which is conserved in nearly all known gastric AP zymogens, procathepsins D and some fungal zymogens, is a likely indicator of the pepsinogen-like mode of inhibition.

### 3.1.2 Prophytepsin

Phyepsin is the aspartic proteinase found in barley, and is believed to be the plant homologue of the mammalian lysosomal processing proteinase, cathepsin D, and the yeast vacuolar proteinase A [Runeberg-Roos *et al.*, 1991]. Although its function in the plant is unknown, this enzyme has been implicated in protein turnover and, possibly, in apoptosis [Runeberg-Roos *et al.*, 1994; Runeberg-Roos *et al.*, 1998]. The zymogen, prophytepsin, contains a 41 amino acid N-terminal prosegment, a 338-residue mature AP sequence and a 104-residue plant-specific insert that is common to plant AP zymogens and may play a role in vacuolar targeting. The prosegment of prophytepsin shares limited sequence similarity with animal AP zymogens, such as pepsinogen and procathepsin D. Activation of phytepsin may be autocatalytic or may require a processing enzyme. It involves removal of the prosegment and several cleavages in the mature portion [Glathe *et al.*, 1998].

The crystal structure of prophytepsin establishes the conformation of only residues 6p to 26p of the prosegment. The remaining 15 prosegment residues as well as residue 1 of the mature sequence are disordered (Figure 3.1b)[Kervinen *et al.*, 1999]. Despite the missing parts, however, the structure is quite informative about the inhibition of catalytic activity. Similar to pepsinogen, the prosegment begins with a  $\beta$ -strand (Val 6p to Lys 12p), that serves as the first strand of the central  $\beta$ -sheet. The strand is followed by a loop (Arg 13p to Ile 15p) and an  $\alpha$ -helix (Asp 16p to Ser 26p) that enters the active site cleft. The active site is blocked by a portion of the mature N-terminus. Lys 11 forms ionic interactions with the catalytic aspartates (Asp 36 and Asp 223), and Tyr13 forms a hydrogen bond with Asp 36, blocking the S1 pocket. The S1' pocket is occluded by Ile 7 (Figure 3.2b). The N-terminal sequence is additionally stabilized in the active site cleft by main-chain interactions with residues 81 and 82 of the flap.

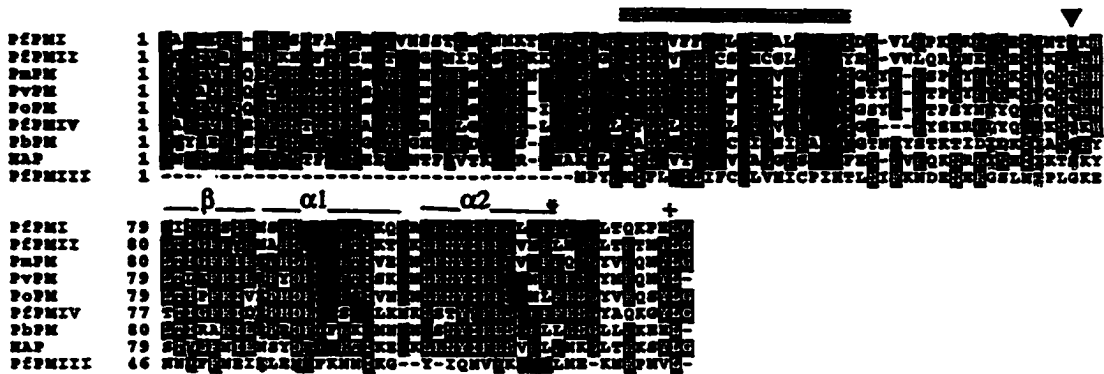
Thus prophytepsin uses a similar mechanism of inhibition as do the gastric AP zymogens, but the role of the inhibitory Lys 36p - Tyr 37p - Tyr 9 triad in pepsinogen is played in prophytepsin by Lys 11, Tyr 13 and Ile 7, all part of the mature enzyme sequence. In both types of zymogen the active site is pre-formed, and activation involves changes that make the active site cleft accessible to substrates, and refolding of the N-terminal residues to replace the prosegment  $\beta$ -strand in the central  $\beta$ -sheet.

The method of inactivation observed in prophytepsin is likely to be common among plant AP zymogens; they share extensive sequence homology in the prosegment and in the mature N-terminus. In addition, cathepsin D, which possesses both the Lys-Tyr pair in its prosegment and the Lys-Asn-Tyr sequence at its mature N-terminus (Figures 3.3, 2.1), may use either inhibition mechanism. A recent crystal structure of cathepsin D at pH 7.5 shows that the N-terminus re-enters the active site cleft and uses Lys 8 and Tyr 10 to block the active site in a similar fashion to that seen in prophytepsin [Lee *et al.*, 1998].

### 3.1.3 Proplasmepsins

The first malarial APs to be characterized in detail were *P. falciparum* plasmepsins I and II. When the amino acid sequences of their zymogens, proplasmepsins I and II (pPfPMI and pPfPMII), were analyzed, several unusual features were noticed in the prosegments (Figure 3.3) [Dame *et al.*, 1994; Francis *et al.* 1994]. These features

### Proplasmepsin prosegments



### Human AP zymogen prosegments

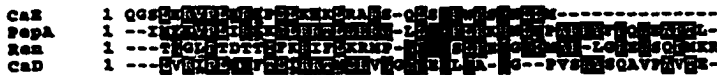


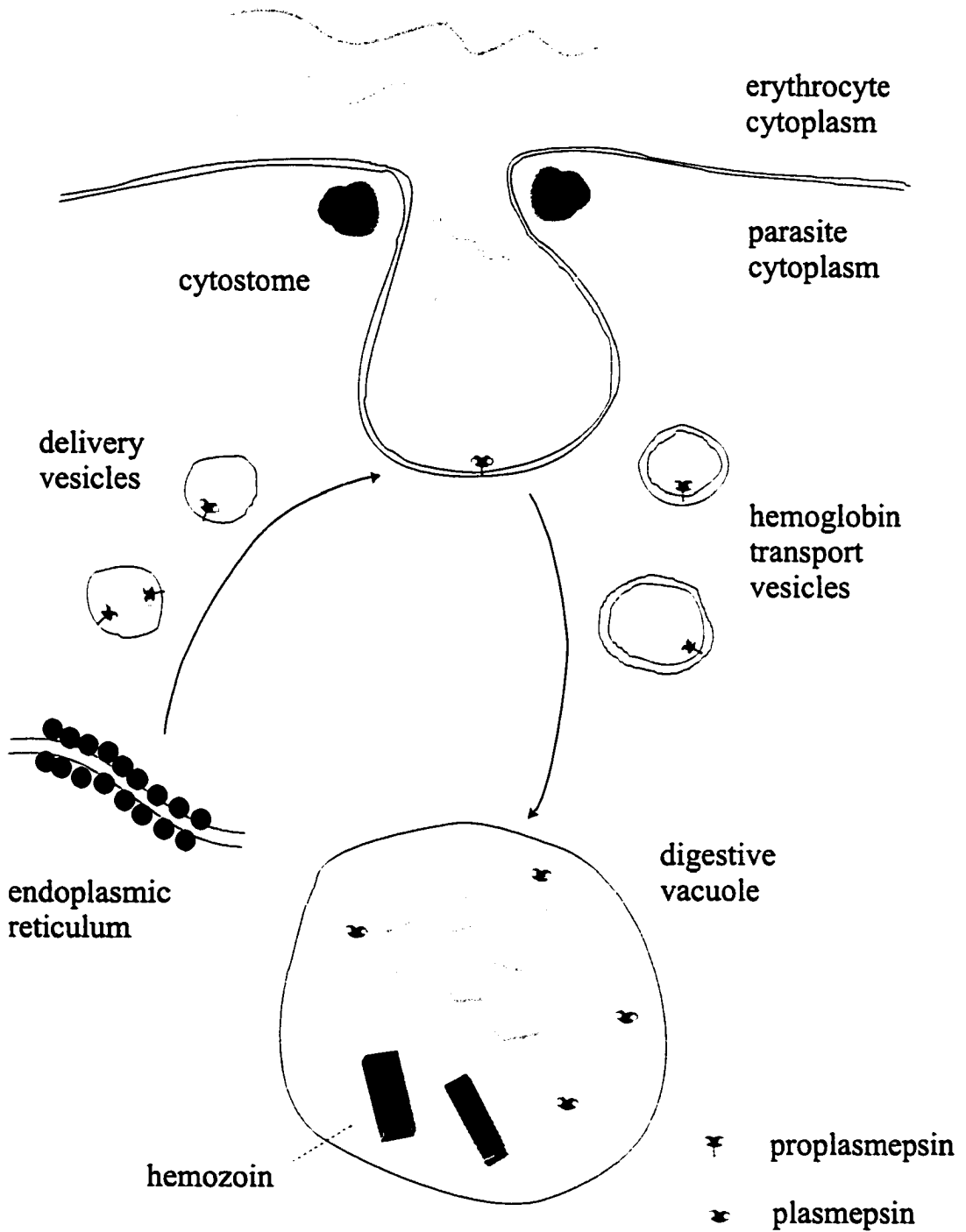
Figure 3.3 Sequence alignment of AP zymogen prosegments. (Abbreviations are as in figure 2.1.) Proplasmepsin prosegments and the human AP zymogen prosegments were aligned separately, *i.e.*, in two groups, using ClustalW(1.7). Secondary structure of the pPpMII prosegment is indicated above the proplasmepsin prosegment sequences. The bond cleaved in the autoactivation of pPpMII is shown by an asterisk (\*), and the bond cleaved in the autoactivation of pPpM is indicated by a (+). (▼) indicates the beginning of the truncated proplasmepsins. The line (====) indicates the signal anchor sequence.



were also later found in the other related APs, PfPMIV, PvPM, PmPM, PoPM and PbPM. The proplasmepsin prosegments do not share any appreciable sequence similarity with other AP zymogens, and are significantly longer than most AP zymogen prosegments. They contain about 124 amino acids compared to the gastric AP zymogen length of less than 50 amino acids. In contrast to other AP zymogens, proplasmepsins lack a signal sequence, but contain in the prosegment a stretch of 21 hydrophobic amino acids that was predicted to mediate membrane attachment. Indeed, it was shown later that PfPMI and PfPMII are type II membrane proteins. The prosegment hydrophobic region serves as a signal anchor sequence that tethers the zymogens to the membrane during transport from the ER to the digestive vacuole [Francis *et al.*;1997]. Proteolytic removal of the prosegment during activation, yields a soluble mature enzyme.

### 3.1.3.1 Plasmepsin production in vivo

Francis *et al.* [1997] investigated the biosynthesis and maturation of PfPMI and PfPMII. They determined that the two enzymes are expressed at different times during the erythrocytic cycle. The PfPMI mRNA level peaks early, in the ring stage, but the PfPMII message level reaches its maximum later, during the trophozoite stage. Earlier immunoelectron microscopy studies localized the plasmepsins (or their zymogens) at the parasite's surface and along the hemoglobin ingestion pathway [Francis *et al.*, 1994]. Although the secretory machinery of *Plasmodium* is poorly characterized, a rough model was proposed for plasmepsin trafficking (Figure 3.4). After synthesis in the ER, the membrane-bound zymogens travel, possibly *via* the Golgi apparatus, to the parasite's surface membrane, where they become associated with the cytosome. From there, they proceed in hemoglobin-laden vesicles to the digestive vacuole. Either *en route*, or in the digestive vacuole, the zymogens become converted to soluble, mature plasmepsins. Many questions remain about this mechanism. For instance, the disposition of proplasmepsin at the parasite surface and in the double membrane delimited hemoglobin-carrying vesicles is unclear. The zymogen may remain bound to the trophozoite membrane, as predicted by the fusion and vesiculation model of transport. In this case, the C-terminal (mature) portion of the zymogen would be located inside the parasitophorous vacuole, and plasmepsin would only come into contact with hemoglobin after the vesicle fuses with the single-membrane bounded digestive vacuole. Alternatively, proplasmepsin may somehow



**Figure 3.4** Trafficking of proplasmepsin to the digestive vacuole in *P. falciparum*. [Adapted from Francis *et al.*, 1997b.]

translocate across the parasite's membrane and insert into the parasitophorous vacuolar membrane, directing its C-terminus into the erythrocyte cytoplasm. Perhaps the process involves insertion of proplasmepsin into the membrane at sites of fusion between the parasite plasma membrane and the parasitophorous vacuole membrane [Francis *et al.*, 1997, Elmendorf & Haldar, 1993].

The experiments of Francis *et al.* showed that the activation of pPfPMI and pPfPMII is carried out *in vivo* by a maturase that requires acidic pH. The intriguing proteolytic activity was apparently only inhibited by the tripeptide aldehydes acetyl-leucyl-leucyl-norleucinal and acetyl-leucyl-leucyl-methional [Francis *et al.*, 1997]. These molecules also inhibit cysteine proteinases and the chymotryptic activity of proteasomes. Other inhibitors of cysteine proteinases or of proteasomes, however, had no effect on the maturase [Francis *et al.*, 1997]. Later results suggested that this maturase is actually an aspartic proteinase (D.E. Goldberg, personal communication).

### 3.1.3.2 Recombinant plasmepsins and truncated proplasmepsins

Plasmepsins are of interest as potential anti-malarial drug targets, since their inhibition kills *Plasmodium* in culture [Francis *et al.*, 1994; Silva *et al.*, 1996; Moon *et al.*, 1997; Rosenthal, 1998; Berry, 1999]. Due to widespread resistance to many anti-malarial therapies, there is an urgent need for new drugs to fight this disease [Institute of Medicine, 1991]. Since the isolation of plasmepsins in sufficient quantities from the parasites is impractical, recombinant enzymes were needed. Attempts to express a full-length proplasmepsin have been unsuccessful, giving either low expression levels or misinitiation of translation at an internal Met, yielding a shortened prosegment [Moon *et al.*, 1997; Luker *et al.*, 1996]. An efficient method of production of active recombinant plasmepsin II employed a truncated form of proplasmepsin II, in which the first 76 amino acids of the prosegment were missing (Figure 3.3) [Hill *et al.*, 1994]. This zymogen is comparable in size to the archetypal AP zymogens, is inactive at neutral pH and autoactivates at acidic pH (around 4.7) to yield a mature enzyme that is kinetically similar to the native enzyme purified laboriously from parasite digestive vacuoles [Tyas *et al.*, 1999]. The major peptide bond cleaved in autoactivation is Phe 112p - Leu 113p, involving the identical residues as the first bond cleaved in hemoglobin by the plasmepsins, Phe  $\alpha$ 33 - Leu  $\alpha$ 34 [Gluzman *et al.*, 1994; Goldberg *et al.*, 1991]. A minor site of cleavage in autoactivation is

Tyr 122p - Leu 123p [Hill *et al.*, 1994] (Figure 3.3). Autocatalytic cleavage at the major and minor sites results in a mature enzyme that is extended at the N-terminus by 12 and 2 amino acids, respectively. These extra residues do not appear to interfere with the recombinant enzyme's activity [Tyas *et al.*, 1999]. The bond cleaved during *in vivo* activation is that between Gly 124p and Ser 1 [Francis *et al.*, 1994; Gluzman *et al.*, 1994; Francis *et al.*, 1997].

The production of recombinant plasmepsin I (PfPMI) has proved much more difficult, with an analogous truncated zymogen being refractory to autoactivation unless a specific mutation is introduced near the cleavage site [Moon *et al.*, 1997]. Furthermore, the recombinant PfPMI produced from the engineered zymogen has significantly weaker catalytic properties than the native enzyme isolated from the parasite. Autoactivation of PfPMI is discussed in more detail below.

## 3.2 Methods

### 3.2.1 Crystallization and data collection

Proplasmepsin II from *Plasmodium falciparum* (expression clone obtained from C. Berry, University of Wales College of Cardiff, Cardiff, Wales, UK), expressed in *E. coli* and purified as described by Hill & *al.* [1994], was provided by Hansruedi Loetscher and Robert G. Ridley (Hoffmann La-Roche, Basel, Switzerland). To facilitate protein expression, in this construct of proplasmepsin II the prosegment was truncated to 48 amino acids (Figure 3.3), and 3 amino acids (GRG) were inadvertently added to the N-terminus as an artifact of the expression system.

Proplasmepsin II was crystallized by the vapor diffusion method. Sitting drops containing 2  $\mu\text{L}$  of protein solution (27 mg/mL) and 7  $\mu\text{L}$  reservoir solution were equilibrated against 1 mL reservoir solution (25% PEG4000, 0.18 M ammonium sulfate, 0.1 M TrisHCl, pH 7.5 - 8.0, 0.1 %  $\beta$ -octylglucoside). Seeding was required to obtain diffraction-quality crystals. The dominant crystal form at these conditions was triclinic, P1 ( $a = 52.02 \text{ \AA}$ ,  $b = 83.106 \text{ \AA}$ ,  $c = 98.42 \text{ \AA}$ ,  $\alpha = 98.55^\circ$ ,  $\beta = 97.13^\circ$ ,  $\gamma = 106.16^\circ$ ). However, one crystal, grown in the same conditions, was found to be monoclinic, P2<sub>1</sub> ( $a = 52.01 \text{ \AA}$ ,  $b = 133.22 \text{ \AA}$ ,  $c = 114.23 \text{ \AA}$ ,  $\beta = 98.45^\circ$ ). The P1 crystals had four molecules in the unit cell, and the P2<sub>1</sub> crystal form had four molecules in the asymmetric unit. The Matthews

coefficient [Mathews, 1968] was  $2.4 \text{ \AA}^3/\text{Da}$  and  $2.5 \text{ \AA}^3/\text{Da}$  with four molecules per asymmetric unit for the P1 and P2<sub>1</sub> crystal forms, respectively. For data collection, the crystals were transferred briefly to a cryo-protectant solution containing 30 % glycerol in reservoir solution, as described above, and flash-frozen in a cold nitrogen gas stream with a 600 Series Cryostream Cooler (Oxford Cryosystems, UK). Diffraction data for the P1 crystal were collected at 105K, on a DIP 2030H image plate detector (Mac Science Co., Ltd.) using double-mirror focussing optics and CuK $\alpha$  radiation generated with a Rigaku rotating anode generator RU-200BH operating at 45 kV and 75 mA. Diffraction data for the P2<sub>1</sub> crystal form were collected at 100K on beamline X12C at the National Synchrotron Light Source at BNL at a wavelength of 1.0  $\text{\AA}$ , using a MAR image plate detector. All the data sets were processed and reduced with Denzo and Scalepack [Otwinowski & Minor, 1996] (Table 3.1).

### 3.2.2 Self-rotation search

A self-rotation search was performed on the P1 data using the program POLARRFN [CCP4, 1994]. Three approximately mutually perpendicular non-crystallographic two-fold rotation axes were found. One was parallel to the **a**-axis, the second was 55° away from **c\*** in the plane normal to **a**, and the third was 93° away from the second axis in the same plane. For the P2<sub>1</sub> data, the native Patterson map revealed a peak at (0.43, 1/2, 0.61), which varied in height from 19 to 24% of the origin peak, depending on resolution. This indicated that a non-crystallographic 2-fold axis must be parallel to the crystallographic 2<sub>1</sub> screw axis (**b**). The presence of this axis was confirmed, together with 2-fold axes parallel to **a** and **c\***, in a self-rotation search with POLARRFN.

### 3.2.3 Structure determination and refinement

The structures were solved by molecular replacement using AMoRe [Navaza, 1994]. For the P1 crystal form, the initial search model was the structure of mature plasmepsin II [Silva *et al.*, 1996]. Four weak solutions consistent with the self-rotation results were obtained (corr. = 0.34, R = 51.5%). The electron density map showed that while the C-domain and central motif (139-329) had reasonable density, the N-domain (1-138) had extremely poor electron density, meaning that a domain shift had taken place

**Table 3.1** Summary of data and refinement

Space Group	P1	P2 <sub>1</sub>
Unit cell	$a = 52.02 \text{ \AA}$ $b = 83.96 \text{ \AA}$ $c = 98.42 \text{ \AA}$ $\alpha = 98.55^\circ$ $\beta = 97.13^\circ$ $\gamma = 106.16^\circ$	$a = 52.01 \text{ \AA}$ $b = 133.22 \text{ \AA}$ $c = 114.23 \text{ \AA}$ $\beta = 98.45^\circ$
Molecules/asymmetric unit	4	4
$V_m$ [Mathews, 1968]	$2.4 \text{ \AA}^3/\text{Da}$	$2.5 \text{ \AA}^3/\text{Da}$
Data completeness	89.0 % (20 - 2.1 \AA)	96.1 [80.6] % (20 - 1.85 \AA)
$F/\sigma F > 1$ , [ $F/\sigma F > 3$ ]	50.2 % (2.14 - 2.10 \AA)	89.6 [45.2] % (1.88 - 1.85 \AA)
$I/\sigma I$	10.5	10.33 [2.18 for 1.90 - 1.85 \AA]
Observed reflections	305261	421877
Unique reflections	80252	125752
$R_{\text{merge}}^1$	5.6 % [19.3% for 2.18 - 2.10 \AA]	8.4 % [57.0% for 1.92 - 1.85 \AA]
<b>Final refinement statistics for P2<sub>1</sub> data</b>		
R-factor <sup>2</sup> (20 - 1.85 \AA, 123053 reflections)		21.4 %
R-free <sup>3</sup> (20 - 1.85 \AA, 2699 reflections)		23.7 %
Number of protein atoms		11531
Number of water molecules		731
Number of glycerol molecules		6
R.m.s. deviations from ideal geometry		
bond lengths		0.006 \AA
bond angles		1.46 °
improper angles		0.68 °
Temperature factors		
mean B (overall)		23.0 \AA <sup>2</sup>
estimated from Wilson plot		17.7 \AA <sup>2</sup>
Estimated cross-validated coordinate error		
from luzzati plot		0.26 \AA
from sigmaa		0.22 \AA
Bulk solvent model		
Ksol		0.444 e/\AA <sup>3</sup>
Bsol		64.25 \AA <sup>2</sup>
Ramachandran plot statistics for 1275 non-proline and non-glycine residues [Laskowski <i>et al.</i> , 1993]		
residues in most favored regions		1150 (90.2 %)
residues in additional allowed regions		121 (9.5 %)
residues in generously allowed regions		2 (0.2 %)
residues in disallowed regions		2 (0.2 %)

$$^1 R_{\text{merge}} = \frac{\sum_{hklj} |I_{hklj} - \langle I_{hklj} \rangle|}{\sum_{hklj} \langle I_{hklj} \rangle}$$

<sup>2</sup>R-factor =  $\frac{\sum ||F_o| - |F_c||}{\sum |F_o|}$ , where  $|F_o|$  and  $|F_c|$  are the observed and calculated structure factor amplitudes, respectively.

<sup>3</sup>R-free was calculated using a random set containing 2.1% of observations which were omitted during refinement [Brunger, 1992].

between the mature enzyme and the zymogen. The MR search was therefore repeated using separate N- and C- domains (11-129, and 135-328, respectively). Residues 1-11 were omitted from the search model based on the rearrangement of these residues that is observed in the pepsinogen to pepsin conversion [Sielecki *et al.*, 1991]. The final molecular replacement solutions had a correlation of 0.61 and R-factor of 42.1% for data between 10 and 3 Å. The resulting model was refined in X-PLOR [Brunger, 1993] using the maximum likelihood target mlf2 [Pannu & Read, 1996]. Initially only data between 9 and 2.8 Å,  $|F| > 3\sigma$  were included in the refinement. Gradually more data were added until all data between 13 and 2.1 Å were included. NCS restraints were used, but the weight was gradually lowered. After 3 cycles of energy minimization, simulated annealing and B-factor refinement, alternating with manual model building in O [Jones & Kjeldgaard, 1995], the R-factor was 30.4%, and free R-factor was 35.0% (rmsd bonds = 0.011 Å, rmsd angles = 1.7° rmsd impropers = 1.44°). This model was missing 209 out of 1520 residues in the asymmetric unit. Refinement of the P1 structure was not continued beyond this stage, since at this point we switched to the P2<sub>1</sub> data, which extended to higher resolution and were of higher quality (Table 3.1).

Molecular replacement was carried out in AMoRe [Navaza, 1994] using 10 - 3 Å P2<sub>1</sub> data and the domain-shifted model from the P1 crystal form. Due to the non-crystallographic 2-fold rotation axis being parallel to the crystallographic 2<sub>1</sub> axis, only 3 rotation solutions were observed, and the fourth was generated by applying 2-fold rotations from the self-rotation search to the rotation function solutions and testing them in a translation search. As a result, four strong solutions were obtained (corr. = 0.652, R = 36.8%). The correctness of these solutions was further confirmed when electron density maps, calculated with the CCP4 program suite [CCP4, 1994], showed partial density for the prosegment, which had been absent from the MR search and from map phasing (Figures 3.5a, b). The model was initially refined in X-PLOR using the mlf2 target and data between 20 and 2.0 Å. Strong NCS restraints (groups = main chain atoms in 11-138, 139-328, ncs-weight = 300) were used initially, and gradually dropped. The starting R-factor was 44.9% with a free R-factor of 44.8%. After 3 cycles of energy minimization, simulated annealing and B-factor refinement in X-PLOR and manual model fitting in O, we switched to CNS [Brunger *et al.*, 1998], where we used energy minimization, torsion

angle dynamics and restrained individual B-factor refinement with the mlf and mli targets. The CNS maximum likelihood targets mlf and mli (which corresponds to the X-PLOR target mlf2) assume a Gaussian distribution of the measurement error in structure factor amplitudes ( $|F|$ ) and diffraction intensities ( $I$  or  $|F|^2$ ), respectively [Pannu & Read, 1996]. Anisotropic B-factor scaling and a bulk solvent correction were also applied. A check of the model with WHAT\_CHECK [Hooft *et al.*, 1996] near the end of the refinement revealed systematic distortions of the bond angles along cell edges, which was taken to indicate errors in the initial determination of cell parameters due to an inaccurate knowledge of the wavelength in the synchrotron data collection. Application of the cell deformation matrix suggested in WHAT\_CHECK to the cell dimensions and coordinates, followed by energy minimization in CNS, was repeated until no bond length directionality was detected by WHAT\_CHECK. The final model contains 1461 protein residues, 731 water molecules and 6 glycerol molecules from the cryosolvent. All 4 protein molecules are missing 4-5 residues at the N-terminus and 6 residues at the tip of the flap. Molecule 3 is missing 14 residues between the prosegment and the mature segment, and molecule 2 is missing 1 residue at the C-terminus. Eleven residues have been assigned alternate conformations: Val 42 and Val 246 in molecule 1, Val 42 in molecule 2, Leu 40, Val 42, Leu 124 and Thr 165 in molecule 3, and Ser 37, Trp 41, Gln 275 and Leu 292 in molecule 4. The final refined electron density is shown in figures 3.5b and c. The structure contains 4 outliers on the Ramachandran plot, all of which are located in a poorly defined loop (Figure 3.5d). The final R and free R-factors are 21.4% and 23.8%, respectively for all data to 1.85Å.

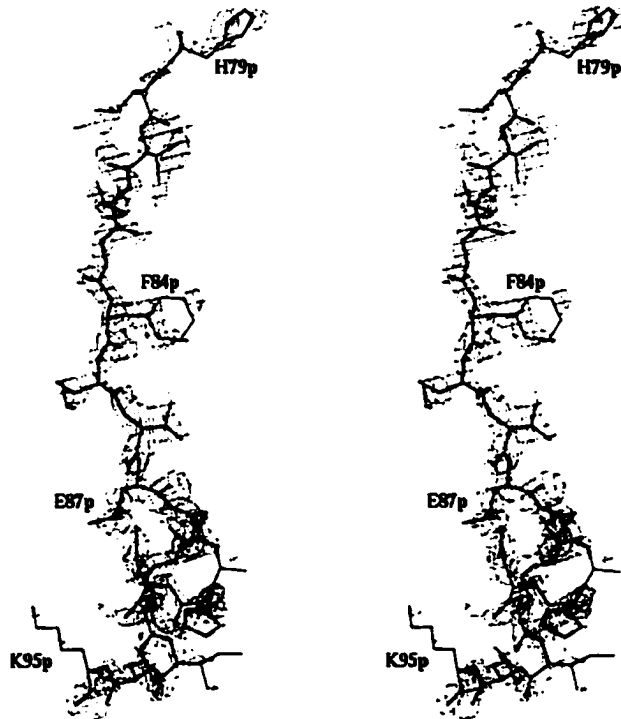
The coordinates and structure factors (for the monoclinic crystal form) have been deposited in the PDB, ID codes 1pfz and r1pfzsf, respectively.

### 3.2.4 Modelling of *P. falciparum* proplasmepsin I

A homology model of *P. falciparum* proplasmepsin I was built in O by introducing the required mutations to molecule 1 of the proplasmepsin II structure (crystal form P21). The model was subjected to energy minimization in CNS to relieve any strain introduced by the sequence changes.

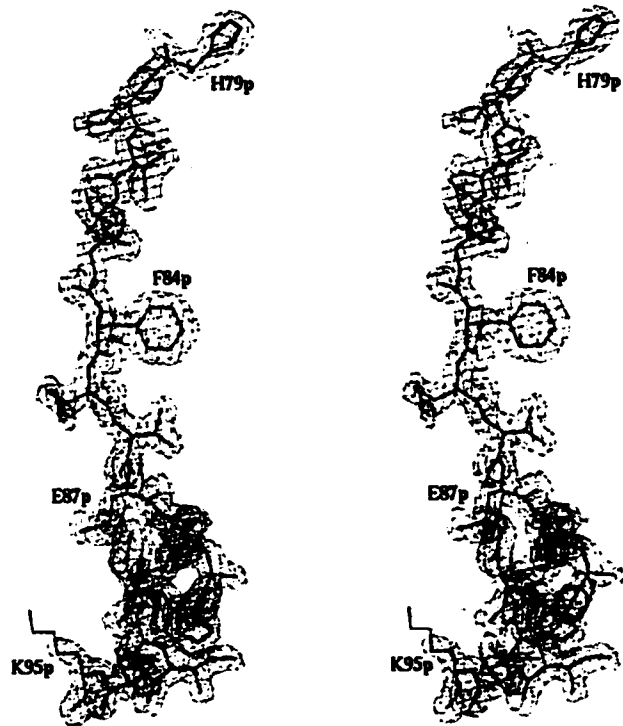


(a)

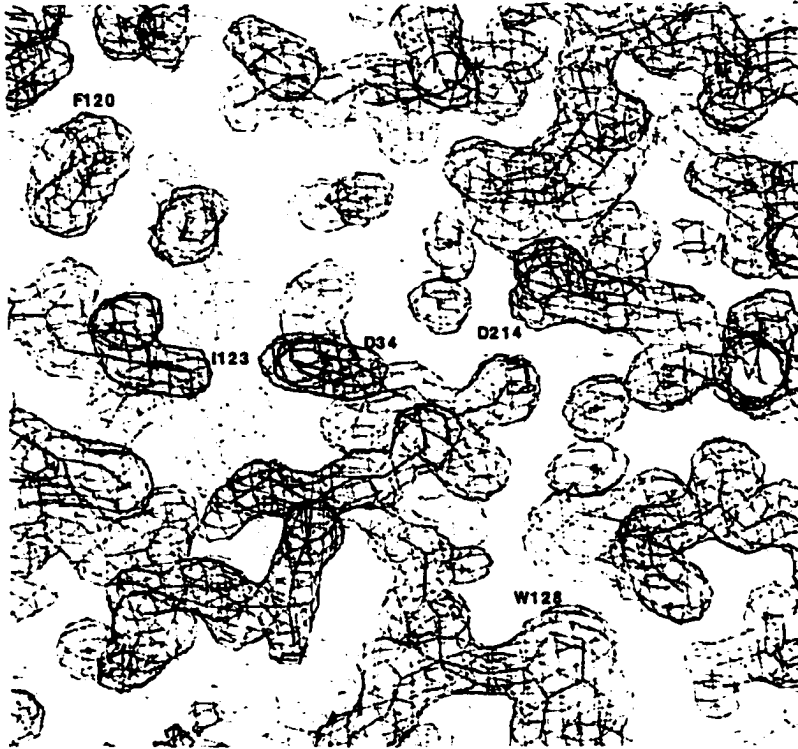


**Figure 3.5** Refinement progress for pPfPMII. (a) Difference ( $|F_O| - |F_C|$ ) map phased with the molecular replacement model, contoured at  $2\sigma$ , for part of the prosegment. (b) Final refined ( $2|F_O| - |F_C|$ ) map contoured at  $1\sigma$  about the model. (c) Final ( $2|F_O| - |F_C|$ ) map contoured at  $1\sigma$  at the active site region of molecule A. (d) Ramachandran plot.

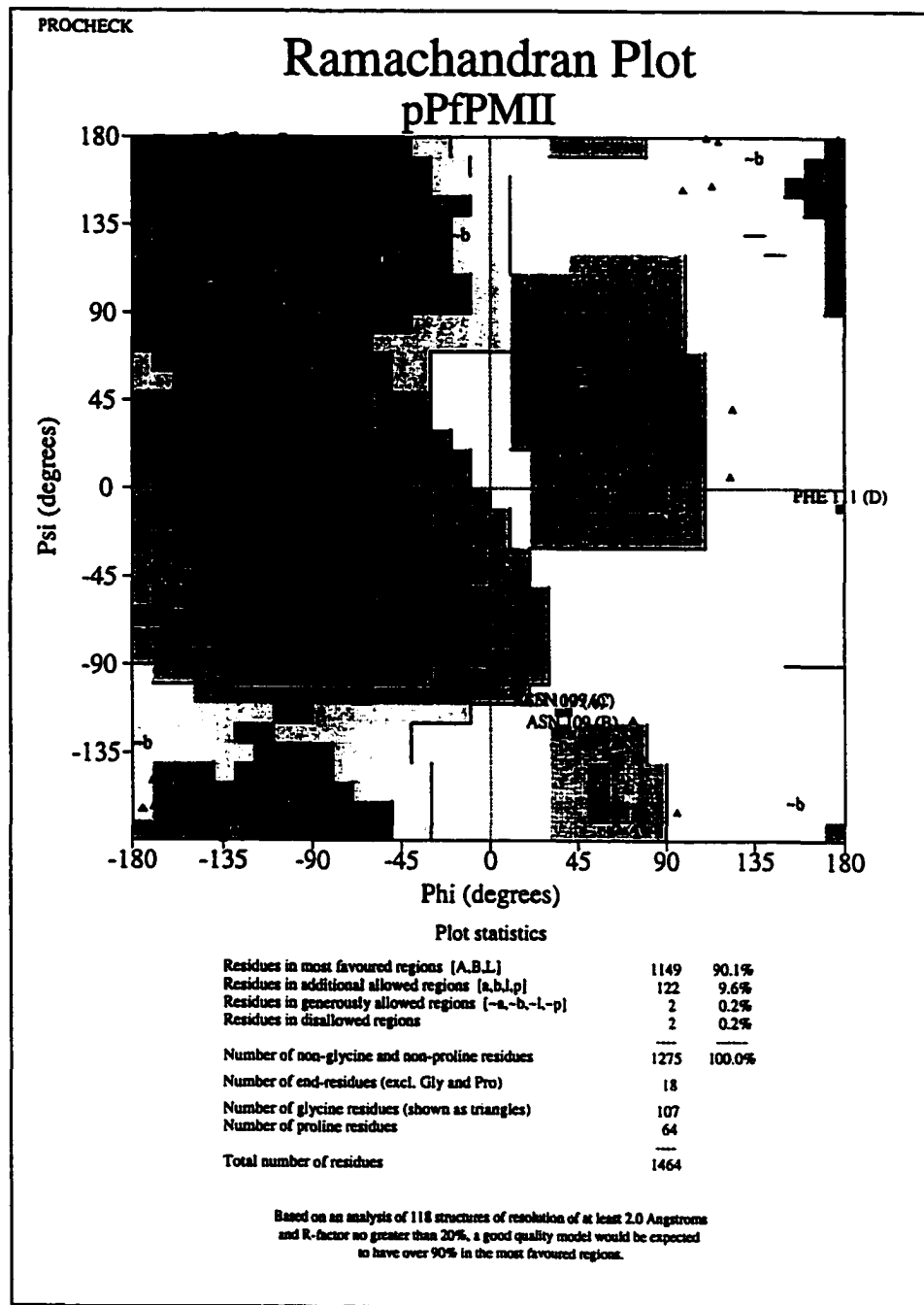
(b)



(c)



(d)

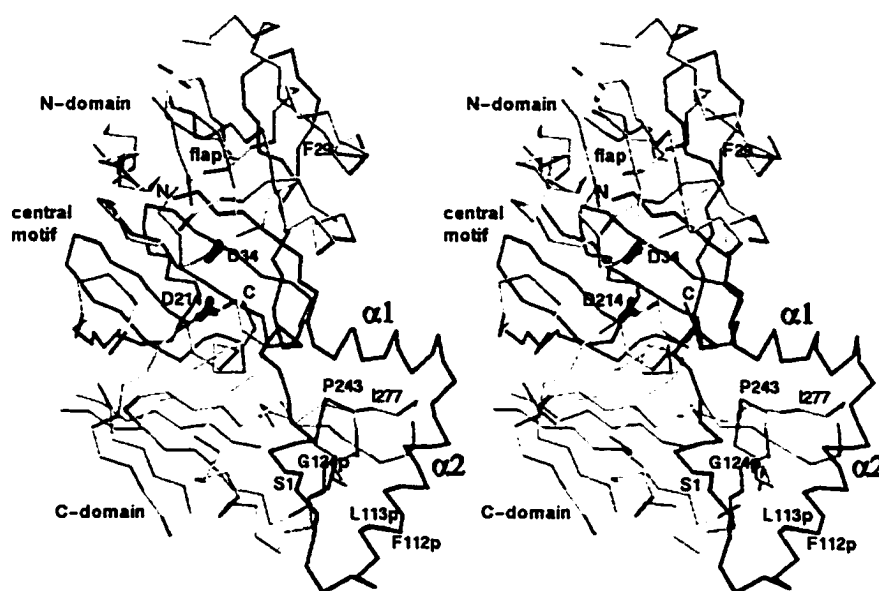


## 3.3 Results and Discussion

### 3.3.1 Description of structure

To allow efficient expression, the zymogen used in this investigation had the prosegment truncated to 48 residues (numbered 77p - 124p), approximately the same size as the prosegments of the archetypal aspartic proteinase zymogens [Hill *et al.*, 1994]. The structure of the truncated proplasmepsin II (from now on referred to as proplasmepsin II or pPfPMII) is shown in figure 3.6. This molecule is crescent-shaped, with one lobe containing the N-domain, and the other lobe, the C-domain. These two domains are attached to the central motif, a six-stranded anti-parallel  $\beta$ -sheet that forms the backbone of the molecule. It is the C-domain and central  $\beta$ -sheet that interact primarily with the prosegment. The secondary structure of the mature portion of proplasmepsin II, for the most part, is similar to that of plasmepsin II and other monomeric aspartic proteinases. The mature portion consists of a central 6-stranded  $\beta$ -sheet flanked by the N- and C-domains containing similarly folded, orthogonally arranged  $\beta$ -sheets packed in 3 and 2 layers, respectively. The residues that comprise the central motif are His 79p to Ile 86p, Ala 152 to Leu 187 and Phe 310 to Leu 329. The N-domain consists of Met 15 to Lys 129. The C-domain contains Asn 188 to Tyr 309, as well as the associated portions of the prosegment and mature N-terminus (Glu 87p to Ile 14). Residues Asp 130 to Asn 151, which fold into a short helix (Ile 139 to Asn 147) surrounded by two loops, form the connection between the N-domain and the central motif. Other interdomain connections occur at Ile 86p - Glu 87p and Leu 187 - Asn 188 (central motif to C-domain), Tyr 309 - Phe 310 (C-domain to central motif) and Ile 14 - Met 15 (C-domain to N-domain).

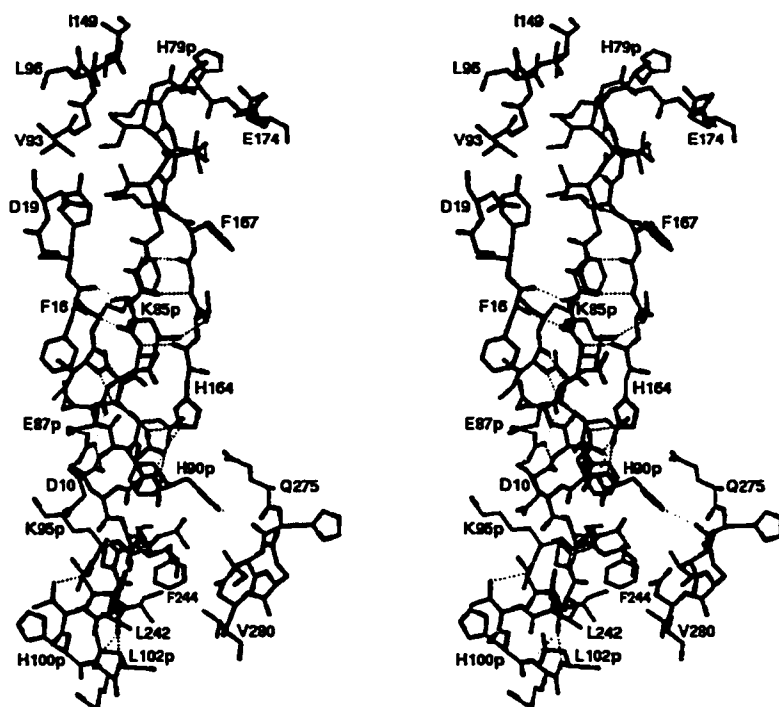
The structure shows why 48 amino acids may be an appropriate length for the truncated prosegment. Almost the entire truncated prosegment interacts with the mature portion of proplasmepsin, producing a compact crescent-shaped molecule. Judging from the structure, the remainder of a longer prosegment may have been disordered in the crystals, interfered with crystallization or lead to unwanted aggregation of the zymogen in solution. The prosegment has a clearly defined secondary structure: a  $\beta$ -strand followed by an  $\alpha$ -helix, a turn, a second  $\alpha$ -helix and a coil connection to the mature segment (Figures 3.3, 3.6).



**Figure 3.6** Structure of pPfPMII. The prosegment is in magenta, the first 29 residues of the mature sequence are in cyan, the mature portion of the central motif is in green, and the remaining portions of the N- and C-domains are in yellow. The side chains of the catalytic aspartates, 34 and 214, are shown in red. The first residue of this model is His 79p (N) and the last residue is Leu 329 (C).

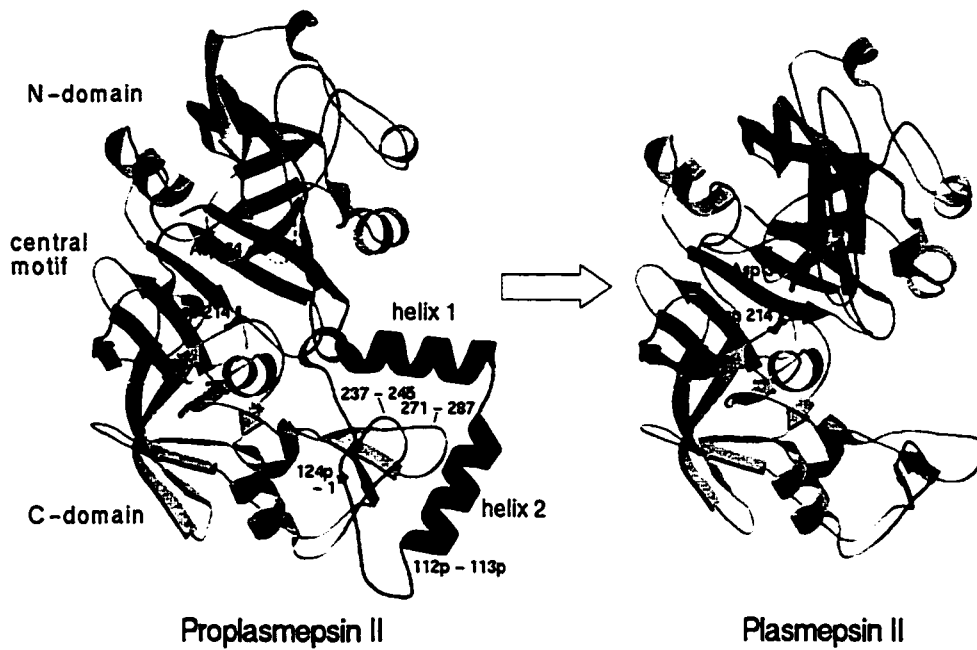
Residues 80p to 87p form the first strand in the central 6-stranded  $\beta$ -sheet (Figures 3.6, 3.7). Leu 80p to Val 86p are involved in anti-parallel  $\beta$ -sheet hydrogen bonds with Ile 170 to His 164, while Lys 85p to Glu 87p form anti-parallel  $\beta$ -sheet hydrogen bonds to the N-terminal residues Phe 16 to Ile 14. The side chains in the prosegment  $\beta$ -strand participate in a number of interactions. Thr 81p O $\gamma$ 1 forms a hydrogen bond with Thr 169 O $\gamma$ 1, and the buried hydrophobic residues, Leu 80p, Ile 82p, Phe 84p and Val 86p fit into hydrophobic pockets formed by the mature segment. The prosegment strand becomes replaced by residues 4 to 11 of the mature enzyme upon activation (Figure 3.8). Residues Asn 88p to Thr 99p form  $\alpha$ -helix 1, with Asn 88p serving as the N-cap (Figure 3.7). This helix runs along the C-domain, and makes hydrogen bonding and hydrophobic interactions with the mature portion. In addition to forming a salt bridge with Glu 87p, Arg 92p forms hydrogen bonds to the peptide oxygens of Asp 10 and Asn 12. Asp 91p forms a salt bridge to His 164, and His 90p N $\epsilon$ 2 makes a hydrogen bond with His 276 O. Ala 89p, Ile 93p, Leu 94p and Ile 97p all constitute the hydrophobic face of helix 1, which packs against the mature segment. The loop formed by residues 100p - 101p connects helix 1 to helix 2.  $\alpha$ -helix 2, consisting of residues Leu 102p to Leu 113p, lies in a groove between loops 237 to 245 and 271 to 287. This helix has fewer interactions with the mature portion than do the preceding helix and  $\beta$ -strand. The only hydrogen bond to the mature segment is between Tyr 105p OH and Pro 240 O. Hydrophobic interactions are made by the aromatic ring of Tyr 105p, Ile 106p, and Leu 113p. The Phe 112p - Leu 113p bond, the major site of cleavage during *in vitro* autoactivation, is located at the C-terminus of helix 2.

Residues 114p to 124p form a coil connection to the mature segment of the protein. The pro-mature junction, Gly 124p - Ser 1, is located in a compact surface loop delimited by Tyr 122p and Asp 4, the "Tyr-Asp" loop (Figure 3.9). Asp 4 is at the centre of the network maintaining the structure of this loop, and anchoring it to the body of the protein. The "Tyr-Asp" loop is closed off by a hydrogen bond between O $\delta$ 1 of Asp 4 and Tyr 122p OH. A hydrogen bond from Asp 4 NH to Ser 1 O $\gamma$  provides additional stabilization. The loop is attached to the C-domain by hydrogen bonds between Asp 4 carboxylate oxygens and the amide NH of Phe 241, and by a salt bridge between Asp 4 O $\delta$ 2 and Lys 238 N $\zeta$ .

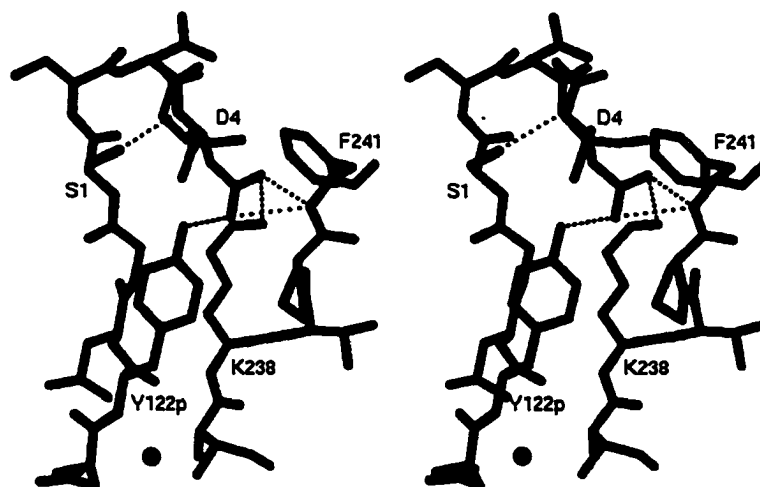


**Figure 3.7** Interactions of the prosegment with the mature portion of pPpPMII. The  $\beta$ -strand and  $\alpha$ -helix 1 of the prosegment (carbon atoms in cyan), and the surrounding mature portions (carbon atoms in grey) are shown.





**Figure 3.8** Activation of PfPMII. The coloring scheme is the same as in figure 3.6. Asterisks (\*) indicate scissile bonds in autoactivation (Phe 112p - Leu 113p) and in activation by a maturase (Gly 124p - Ser 1). The disordered tip of the flap in pPfPMII is indicated by a dashed line.



**Figure 3.9** The pro-mature junction in pPfPMII. Residues 237 - 241, which interact with the pro-mature junction, are shown with carbons colored cyan.

The flap in APs is the  $\beta$ -hairpin loop which covers the active site and contributes to the substrate-binding pockets S2, S1 and S2'. Comparisons of structures of zymogens and their respective mature proteinases, with and without bound inhibitors, show that the flap is quite flexible, with the position of its tip rearranging depending on the occupancy of the active site (Figure 2.7b). In the structure of proplasmepsin II, the tip of the flap (Asn 76 to Thr 81) is not defined in the electron density map and is probably disordered in all four molecules in the crystal.

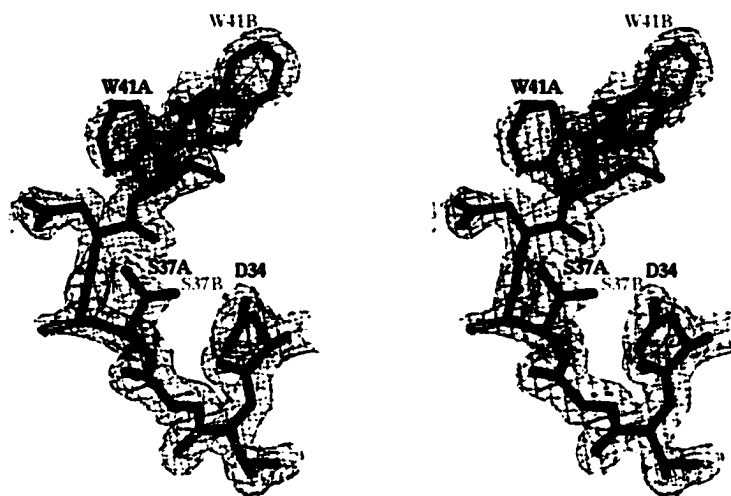
### 3.3.2 Comparisons among NCS-related molecules

The four molecules in the asymmetric unit of the proplasmepsin II crystal have essentially the same structures with a few minor differences. Molecule 3 is missing 14 residues bridging the prosegment and the mature segment, from residue 119p to residue 8. Loop 22 - 27 has two distinct conformations, one in molecules 1 and 3, and a different one in molecules 2 and 4. This variation is probably due to differences in crystal packing interactions.

The C-terminus of the prosegment helix 2, as well as the surrounding segments, Thr 221 to Leu 254 and Pro 270 to Thr 298, show some differences among the four molecules. It should be noted, however, that helix 2 of the prosegment makes relatively few contacts with the mature portion of the protein but is strongly influenced by crystal packing.

Molecule 4 has Ser 37 and Trp 41 in two prominent alternate conformations (Figure 3.10). In plasmepsin II complexed with pepstatin, Trp 41 ( $\chi_1 = 65^\circ$ ,  $\chi_2 = -103^\circ$ ) N $\epsilon$ 1 forms a hydrogen bond with Tyr 77 OH, which is located at the tip of the flap, while Ser 37 O $\gamma$  forms a hydrogen bond with Asp 34 O $\delta$ 2 [Silva *et al.*, 1996]. This pattern of hydrogen bonds is commonly observed in monomeric aspartic proteinases. Since the tip of the flap, including Tyr 77, is disordered in pPfPMII, in all four NCS-related molecules Trp 41 adopts a different conformation ( $\chi_1 = -51^\circ$ ,  $\chi_2 = 70^\circ$ ), its N $\epsilon$ 1 making a hydrogen bond with Ser 37 O $\gamma$ , whose  $\chi_1$  also changes accordingly. In molecule 4, Ser 37 and Trp 41 sample both conformations (Figure 3.10).

In a recent structure of renin in complex with a novel piperidine-based inhibitor, a similar reorientation of the corresponding residue, Trp 45, is observed as part of an



**Figure 3.10** Alternate conformations of Ser 37 and Trp 41 in molecule 4 of pPfPMII. Positions of side chains with carbon atoms colored cyan correspond to those found in PfPMII. The final refined  $2|F_o| - |F_c|$  map, contoured at  $1\sigma$  about the model, is shown.

induced fit due to inhibitor binding [Oefner *et al.*, 1999; Vieira *et al.*, 1999]. This rotation of the Trp 45 side chain opens a new binding pocket into which the inhibitor places an *o*-chlorobenzoate moiety. In addition, the tip of the flap turns up, and Tyr 83 (Tyr 77 in plasmepsin numbering) swings out by 120° to accommodate the inhibitor. The same compound has been reported to inhibit *P. falciparum* plasmepsins I and II, presumably by the same mechanism. In pPfPMII, with Trp 41 in its dominant conformation, the S1/S3 subsite extends into deep a pocket lined by Ile 32, Pro 43, Val 82, Val 105, Thr 108, Phe 111, Tyr 115, Phe 120 and Ile 123. Thus, the crystal structure shows that the plasmepsins may be pre-disposed to the conformational change that is required for the binding of this piperidine inhibitor. The novel binding mode may potentially be exploited in the design of anti-malarial drugs that act against the plasmepsins. Furthermore, the hydrophobic pocket revealed by the flipping of the Trp 41 side chain may be useful for designing molecules that inhibit conversion of proplasmepsin to the mature enzyme.

### 3.3.3 Dimerization of zymogen

Proplasmepsin II crystallizes in two space groups with 4 molecules in each asymmetric unit. In both cases, the asymmetric unit consists of two dimers in which proplasmepsin molecules interact through the interhelical loop and helix 2 of the prosegment and are related by an approximate two-fold rotation. The dimers are virtually the same in both space groups. A different packing arrangement of these dimers results in the different space groups. It is interesting to note that the truncated proplasmepsin II that was used in this study dimerizes in solution, whereas mature plasmepsin II is predominantly monomeric (H. Loetscher, personal communication). The dimers observed in the crystals of the zymogen are therefore likely to be the ones observed in solution. The dimerization interface consists mainly of hydrophobic interactions, and buries 1758 Å<sup>2</sup> of accessible surface area, including 461 Å<sup>2</sup> of hydrophilic and 1297 Å<sup>2</sup> of hydrophobic area.

### 3.3.4 Comparison to plasmepsin II

The crystal structure of plasmepsin II has been solved earlier in complex with the inhibitor pepstatin A by A. Silva *et al.* [1996]. A detailed structural comparison between the zymogen, proplasmepsin II, and the mature enzyme, plasmepsin II, should yield some

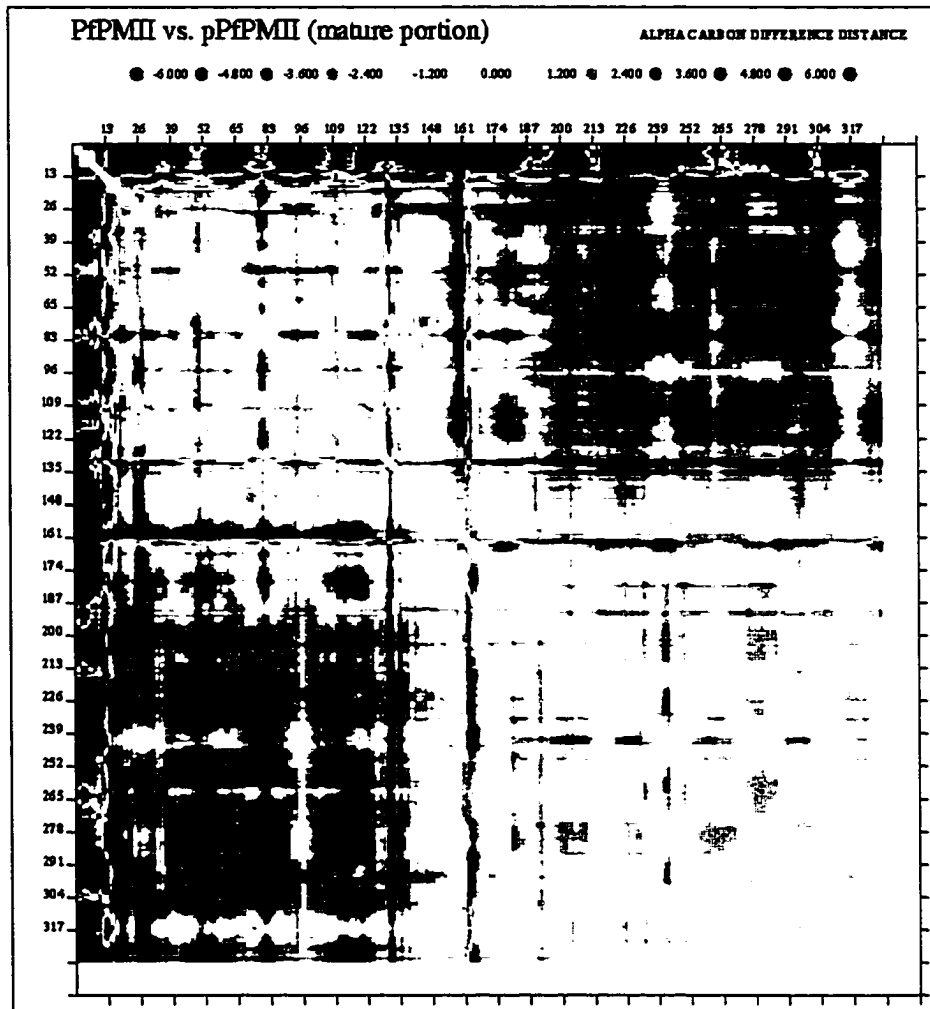
insights into the mechanism of inactivation in the zymogen, and possibly provide clues to the conversion pathway of the zymogen to the active enzyme.

Despite the overall similarity in secondary structure of proplasmepsin II to that of plasmepsin II [Silva *et al.*, 1996], there are a number of important differences between the two molecules, on both the global and local levels, as illustrated by the difference distance matrix for the mature portion of proplasmepsin II vs. plasmepsin II (Figure 3.11). The matrix indicates two major differences between the zymogen and enzyme: refolding of the N-terminus and a domain shift. In addition, some loop rearrangements occur between the two molecules.

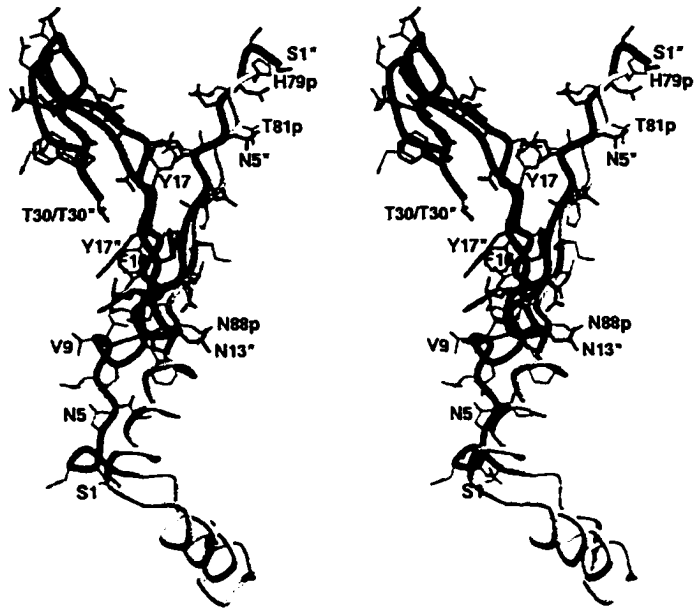
#### 3.3.4.1 Rearrangement of the mature N-terminus

The largest difference in secondary structure between plasmepsin II and proplasmepsin II is seen at the N-terminus of the mature sequence (Figure 3.12). Residues 1 to 14, undergo a severe conformational change upon activation, and residues 15 to 29 make a smaller rearrangement. Residues 4 to 11 must refold from their conformation in proplasmepsin II to replace residues 80p to 87p as the first strand of the central 6-stranded  $\beta$ -sheet. In the zymogen, Ser 1 to Asp 4 are involved in a Type IV  $\beta$ -turn in the "Tyr-Asp" loop [Hutchinson & Thornton, 1994]. Glu 7 to Phe 11 form a short 3<sub>10</sub> helix that protrudes into the active site cleft, partially obstructing it relative to the cleft of PfPMII in the region of the S3 and S4 substrate-binding pockets.

After Gln 12, whose peptide oxygen makes a hydrogen bond to Arg 92p N $\epsilon$ 2, the polypeptide chain adopts an extended conformation, interacting with the prosegment and the active site Psi loops. Residues from Phe 11 to Ala 20 form 13 hydrogen bonds with the Psi loops, 5 of them through water molecules (Figure 3.13a, Table 3.2). This segment is held in place by four additional anti-parallel  $\beta$ -sheet hydrogen bonds between Leu 14 to Phe 16 and Lys 85p to Glu 87p. In PfPMII, the conformation, and consequently the hydrogen bonding pattern, up to Met 15 is different. Residues Phe 11 - Met 15 are involved in a turn between two  $\beta$ -strands, and only Gly 18 and Ala 20 are available to form hydrogen bonds to Phe 31 and Phe 29, respectively (Figure 3.13b). At residue 14 (in pPfPMII), the proplasmepsin and plasmepsin polypeptide chains have similar orientations, but the sequence is out of register by one amino acid, *i.e.*, Ile 14 in proplasmepsin corresponds roughly to the position of Met 15 in plasmepsin (Figure 3.12). Such a shift in register is possible because of the hydrophobic character of the side chains in this



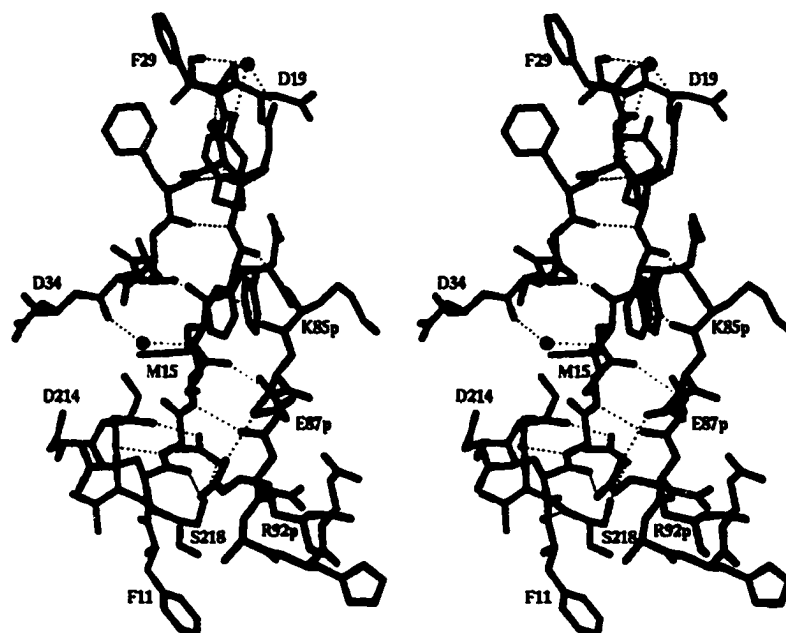
**Figure 3.11** DDM plot comparing the mature portion of pPfPMII to PfPMII. Regions colored blue indicate residues that are closer in PfPMII, and regions colored red indicate residues that are more distant from each other.



**Figure 3.12** N-terminal rearrangement of PfPMII upon activation. The prosegment (yellow) and first 30 amino acids (green) of pPfPMII are superimposed on the first 30 amino acids of PfPMII (red). Side chains of residues 89p - 123p have been omitted for clarity.

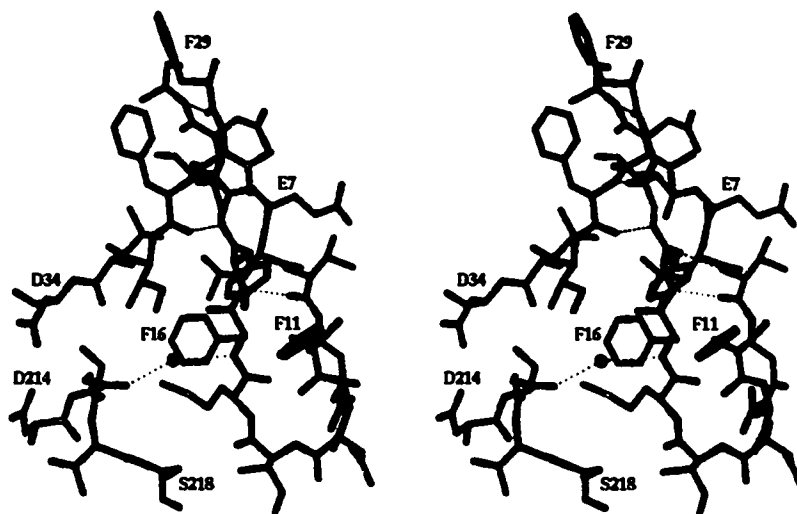


(a)



**Figure 3.13** Interaction of the N-terminal residues 11 - 20 with the active site. (a) In pPfPMII residues 11 - 20 (carbon atoms in cyan) are positioned by the prosegment (carbon atoms in magenta) to interact with both Psi loops (carbon atoms in grey). (b) In PfPMII, these residues interact only with the N-domain Psi loop.

(b)



Residues		Distance (Å)
Phe 11 O	Ser 218 N	2.8
Asn 13 N	Gly 216 O	2.8
Asn 13 Nδ2	Ser 215 O	2.7
Asn 13 Oδ1	Ser 218 Oγ	2.8
Met 15 O	Ile 33 N	2.9
Tyr 17 O	Phe 31 N	2.8
Tyr 17 N	Phe 31 O	2.7
Met 15 N	Wat 94	2.8
Wat 94	Ile 33 O	2.6
Tyr 17 O	Wat 192	2.8
Wat 192	Phe 29 O	2.7
Wat 192	Ala 20 N	2.6
Gly 18 O	Wat 361	2.7
Wat 361	Ala 20 O	2.6
Wat 361	Phe 29 N	3.0

**Table 3.2** Interactions of N-terminal residues 11 to 20 with the Psi loops in pPpPMII

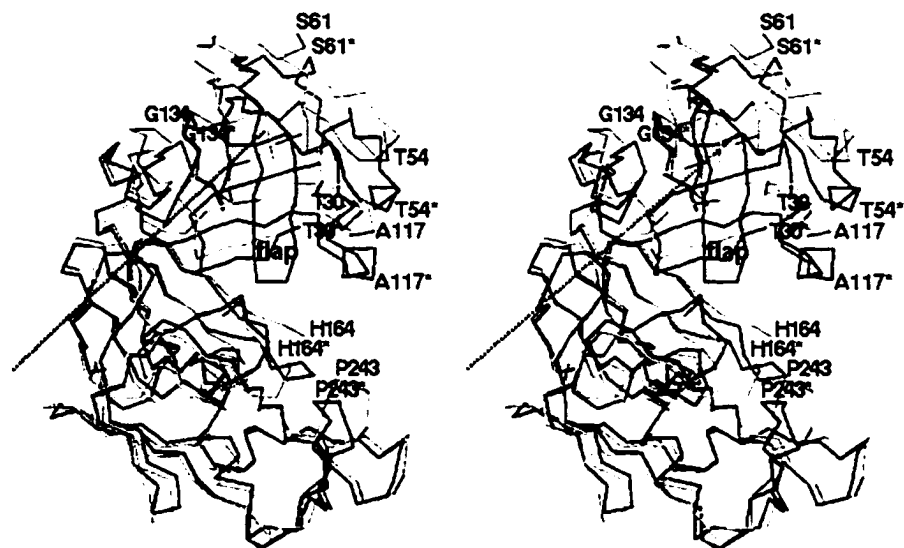
segment (14 - 18). This shift is not seen in pepsin nor in pepsinogen, possibly because it would switch the positions of hydrophobic residues Met 12 and Tyr 14 with the charged residue Glu 13 (Figure 3.3). A bulge at 18 and 19 brings the polypeptide chains back into register by residue 20. This bulge is stabilized by the side chain of Tyr 17, which fits into a pocket lined by the Asp 19 - Ala 20 peptide bond, Val 93 and Ile 82p (Figure 3.7). The bulge alters the main-chain hydrogen bonding pattern in proplasmepsin relative to plasmepsin and changes the conformation of the connecting loop (Gly 23 to Phe 29) (Figure 3.12).

#### 3.3.4.2 The domain shift

On a global level, the most prominent difference between PfPMII and pPfPMII is a large domain shift that renders the zymogen molecule more open than the active enzyme (Figure 3.14). We have divided the molecule into three domains, the N- and C-domains and the central motif, as proposed by Sali *et al.* [1992]. This domain designation is based on the observed rigid body shift, and is different from the canonical division of aspartic proteinases into N- and C- domains at the centre of the interlobe  $\beta$ -sheet at approximately residue 180 [Fusek & Vetvicka, 1995]. We will consider only residues 30 to 329, since the first 29 residues of the mature sequence are refolded between the enzyme and zymogen, as described in the preceding section. When the central motif and the C-domain (138 to 329) of the two molecules are superimposed, the N-domains (30 to 129) are related by a rotation of approximately  $14^\circ$  about the axis shown in figure 3.14. Since this domain shift was observed in eight independent molecules in two crystal forms (P1 and P2<sub>1</sub> with 4 molecules in each asymmetric unit), it was presumed to be a real phenomenon rather than an artifact of crystallization.

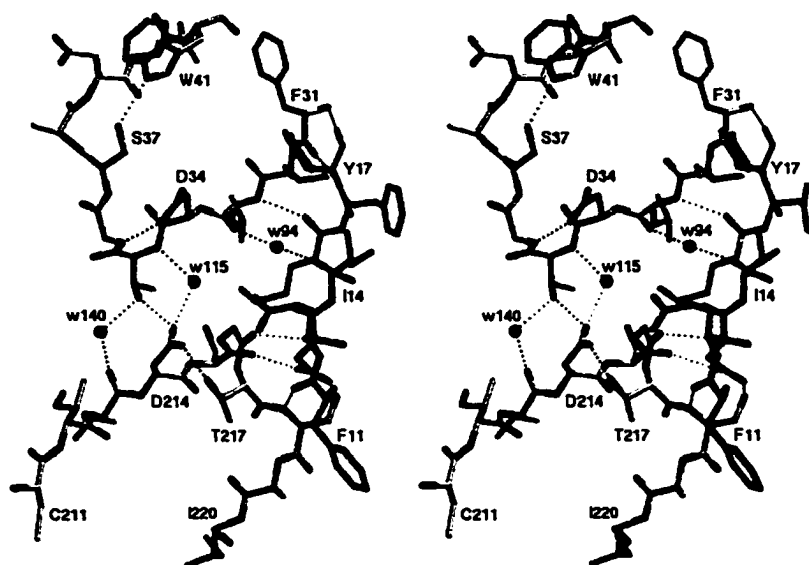
#### 3.3.4.3 Inhibition of catalytic activity in the zymogen

As a result of the domain opening in pPfPMII, the substrate-binding cleft is wider than in plasmepsin II, and the active site region is severely distorted relative to the mature enzyme (Figures 3.15a, b). This is in contrast to the gastric zymogens, where the aspartic proteinase active site is essentially fully formed. In pPfPMII, the active site Psi loops are much farther apart than in PfPMII, separating the catalytic Asp residues (34 and 214) by



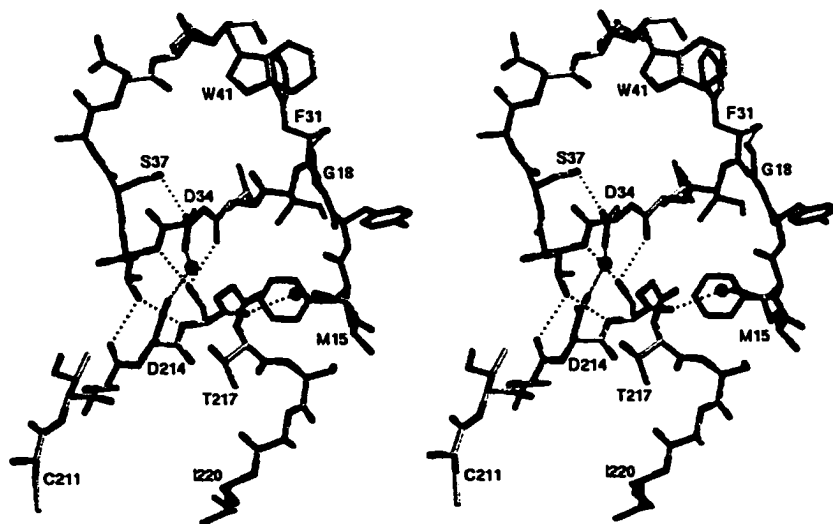
**Figure 3.14** Superposition of residues 30 - 329 of pPfPMII (yellow) and PfPMII (red). In this view, the C-domain and central motif have been superposed in O [Jones & Kjeldgaard, 1995]. The black dotted line indicates the axis of rotation that relates the N-domains in the two molecules. Labels with asterisks (\*) denote PfPMII residues.

(a)



**Figure 3.15** The active site regions of pPfPMII and PfPMII. (a) The "immature" active site of pPfPMII. (b) The active site of PfPMII. The N-terminal segment that interacts with the Psi loops is shown with carbon atoms in cyan. The magenta sphere in (b) is the statine OH from the PfPMII/pepstatin complex [Silva *et al.*, 1996] that mimics the nucleophilic water molecule at the active sites of native APs.

(b)



an additional 3.2 Å. The distance between the inner O $\delta$ s of Asp 34 and Asp 214 is increased from 2.9 Å in plasmepsin II to 6.1 Å in the zymogen. At this position, the Asp residues are not within contact distance, as they are in mature aspartic proteinases.

An extensive network of hydrogen bonds exists at the "immature" active site in pPfPMII; this network is completely different from the symmetrical "fireman's grip" holding together the active sites of aspartic proteinases [Fusek & Vetvicka, 1995]. In the structure of PfPMII complexed with pepstatin, Ser 215 O $\gamma$  forms hydrogen bonds to Leu 33 O, Thr 35 NH and Thr 35 O $\gamma$ 1, whereas Thr 35 O $\gamma$ 1 interacts with Val 213 O and Ser 215 N. Within the Psi loops, hydrogen bonds between the carboxylate oxygens of each catalytic Asp and neighboring residues fix the side chain conformation of the Asps, as discussed in Chapter 2. This intricate pattern of hydrogen bonds found at the mature AP active site enables the carboxylate groups of Asp 34 and Asp 214 to form hydrogen bonds with the OH of pepstatin, an inhibitor atom that mimics the central water molecule serving as the nucleophile in the first step of the AP catalytic mechanism.

In pPfPMII, Asp 34 hydrogen bonds through its O $\delta$ 1 to Gly 36 N. Asp 214 O $\delta$ 1 makes a hydrogen bond with Thr 217 O $\gamma$ 1. Asp 214 O $\delta$ 2, the "inner" oxygen, hydrogen bonds through water 115 to Thr 35 NH across the active site cleft. Asp 214 O $\delta$ 2 also interacts with Gly 216 NH and with Thr 35 O $\gamma$ 1, which, in turn, hydrogen bonds with Val 213 O through water 140. Finally, water 94 mediates the interaction between Leu 33 O and Met 15 N.

The lack of catalytic activity in proplasmepsin II is easily explained by the absence of a properly formed active site. The catalytic aspartic acid residues are too distant from one another in the zymogen to carry out the general base activation of a nucleophilic water molecule. In addition, some rearrangements must occur within the Psi loops to bring the catalytic aspartates into their active conformation. Both Asps must adjust their  $\chi$ 2 torsion angle by about 40°. Also, in the N-terminal Psi loop, Trp 41 and Ser 37 must change their side chain conformations, as described above.

We can identify several factors that may contribute to the domain separation in proplasmepsin II: electrostatic repulsion between the catalytic Asp residues at neutral pH, the well-defined structure of the prosegment and pro-mature junction, the interaction of the prosegment with the C-domain and the interaction of the mature N-terminus with the active site Psi loops. The  $\beta$ -strand anchors the prosegment securely to the body of the protein,



whereas the remainder of the prosegment wraps around the C-domain before connecting to the mature segment, forming a "harness" that maintains the molecule in an open conformation. A network of hydrogen bonds defines a rigid structure around the N-terminus of  $\alpha$ -helix 1 (Figure 3.7). This positions the helix-loop-helix unit with helix 1 packing against a hydrophobic face of the C-domain and helix 2 lying in the groove between loops 237 to 245 and 271 to 287, thus giving the "harness" a holding point on the C-domain (Figure 3.6). Six hydrogen bonds to Lys 85p, Glu 87p and Arg 92p orient the segment from Phe 11 to Tyr 17 enabling it to form hydrogen bonds to both Psi loops, plying them apart (Figure 3.13a). Insertion of the Met 15 side chain between the Psi loops may provide additional assistance in separating the loops (Figure 3.13a). In plasmepsin II, by contrast, the Phe 11 - Tyr 17 segment is in a different conformation and farther from the Psi loops, with only Gly 18 and Ala 20 forming hydrogen bonds to the N-domain Psi loop (Figure 3.13b). A number of steric clashes can also be foreseen between residues of the N-domain and the "harness" if the zymogen were to adopt the plasmepsin-like, domain-closed conformation. The  $3_{10}$  helix (Glu 7 - Phe 11) would collide with the flap, Glu 87p would clash with Ala 117 and Ser 118, and Ile 14 - Glu 21 would interfere with Ser 118 - Phe 120 and Gln 26 - Ile 32.

The domain shift in proplasmepsin II is quite different from the flexibility generally observed in APs or zymogens [Sali *et al.*, 1992; Abad-Zapatero *et al.*, 1990; Sielecki *et al.*, 1990; Silva *et al.*, 1996; Lee *et al.*, 1998]. As discussed in Chapter 2, AP domain movements commonly occur within the C-domain, and affect the size and shape of the substrate-binding pockets, but not the orientation of the catalytic machinery.

The plasmepsin/proplasmepsin domain shift is unprecedented in terms of the division into rigid bodies, as well as the magnitude, direction and effect of the movement. In this system, the rigid bodies are the N-domain and the C-domain plus the central motif. The  $14^\circ$  rotation is much larger than any observed for a single enzyme. The most remarkable aspect of the plasmepsin/proplasmepsin domain shift is that it completely changes the architecture of the active site. The active site loops are far apart in the zymogen and must undergo a large movement to collapse to their mature form upon activation. Such a separation between the active site Psi loops has not been seen in any previously determined monomeric AP or zymogen structure. As will be described in the

next chapter, a similar rotation of the N-domain is observed in *P. vivax* proplasmepsin, suggesting that this is a general feature of proplasmepsin inactivation.

Finally, it is interesting to note that some of the C-domain flexibility common to aspartic proteinases is also observed in proplasmepsin II. The four molecules in the P2<sub>1</sub> asymmetric unit show subtle differences in the orientations of loops 221 to 254 and 270 to 298, which correspond to the flexible subdomain of plasmepsin II [Silva *et al.*, 1996].

#### 3.3.4.4 Loop rearrangements

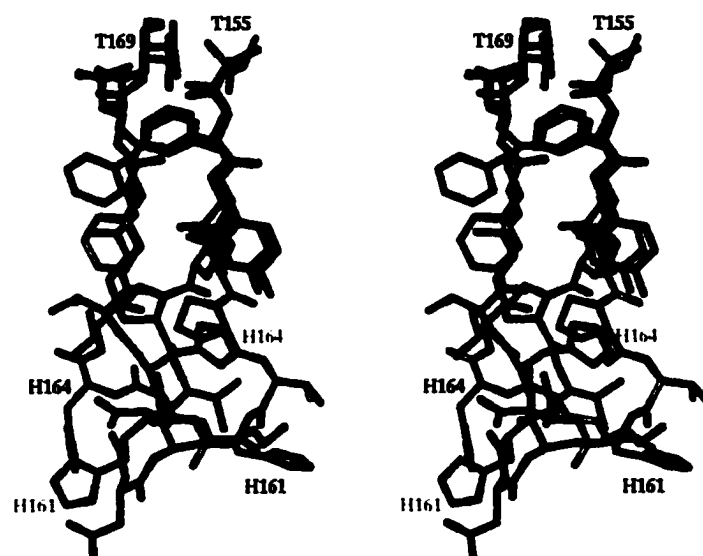
Four additional loops differ in conformation between plasmepsin and proplasmepsin (Figure 3.14). Two of these, 108 to 114 and 129 to 134, flank the disordered flap. In the remaining two, 158 to 169 and 237 to 245, the variation is due to interactions with the prosegment or the N-terminal region which refolds upon maturation.

Residues 109 to 113 have rather poor electron density. Simulated annealing omit maps helped to determine the path of this loop. For most of the residues in this loop, the side chain density is either very poor or absent. On the other hand, the main chain density is continuous everywhere except at Pro 113. This residue is in the *cis* conformation in plasmepsin II, but was modelled as *trans*-proline in proplasmepsin, which is more consistent with the electron density and overall loop conformation. Asn 109 is in a disallowed conformation, but it was modelled based only on main-chain electron density. It appears that the disorder of the flap in the zymogen is propagated to the surrounding structure.

The loop between Lys 129 and Ser 135 flanks the other side of the flap. This loop connects the two rigid bodies involved in the domain shift between the zymogen and the enzyme. While this region does not make any contacts with the flap, its conformation may be indirectly influenced by the difference in the flap conformations. In plasmepsin Tyr 77 OH at the tip of the flap forms a hydrogen bond to Trp 41 Nε1. In proplasmepsin, since the flap is disordered, this bond is not formed and the Trp 41 side chain points in a different direction, hydrogen bonding with Ser 37 Oγ and slightly displacing Ala 38 and Asp 39. In this position, Ala 38 would have a steric clash with Ser 132 if the 129 to 135 loop were in the plasmepsin conformation. The rearrangement of this loop alters the shape of the S2' pocket. In pPfPMII, this segment is pushed out of the pocket, creating a deeper cavity than in PfPMII.

The segment from Leu 158 to Thr 169 has the largest conformational change in the C-domain between proplasmepsin and plasmepsin (Figure 3.16). This loop is engaged in an extensive network of hydrogen bonds, with two residues making either direct or water-mediated contacts to the prosegment. Lys 163 peptide O forms a hydrogen bond to N $\delta$  of Asn 88p through water 29. His 164 Ne2 donates a hydrogen bond to Asp 91p O $\delta$ 2. This is one of the electrostatic interactions tethering helix 1 of the prosegment to the C-domain of proplasmepsin. The disruption of this salt bridge at low pH may contribute to the conformational change involved in plasmepsin activation. In proplasmepsin II, the 158 to 169 loop is roughly in the plane of the central  $\beta$ -sheet. In PfPMII, by contrast, a potential steric clash with Leu 8 - Phe 16, part of the re-folded N-terminus, pushes His 161 - Gly 166 out of the plane of the sheet. His 164 is an interesting residue, since it is proposed to be involved in a local conformational change in the mature enzyme upon binding of pepstatin [Xie *et al.*, 1997]. His 164 is believed to be solvent-exposed in the uninhibited plasmepsin, becoming buried, with a concomitant rise in pKa, when the inhibitor binds. It is possible that the local conformational change in question may be related to the subdomain movement observed in the PfPMII/pepstatin complex. On the other hand, the difference in the structure of loop 158 to 169 between the zymogen and enzyme, while not likely to be directly responsible for the change in the environment of His 164 during inhibitor binding to plasmepsin, shows that this region is inherently capable of dramatic changes in conformation.

Loop 237 to 245 is encircled by the prosegment and the mature N-terminus in pPfPMII. In general, its local conformation is approximately the same in the enzyme as in the zymogen, but this loop is displaced between the two structures (Figure 3.14). The loop makes a number of interactions with the prosegment and pro-mature junction on both sides (Figure 3.6). On one side, Phe 244 makes van der Waals contacts with Ile 93p and Ile 106p, Leu 242 contacts Tyr 105p and Ile 106p, and Val 239 contacts Tyr 105p. Tyr 105p OH also forms a hydrogen bond to Pro 240 O. On the other side of the loop, Tyr 245 contacts Ile 6. Lys 238 NH makes a hydrogen bond to Tyr 122p O, and its side chain folds back to form hydrogen bonds to Asp 4 O $\delta$ 1, as well as to Val 239 O and Leu 242 O. In addition, Val 236 O and Asn 233 NH form hydrogen bonds to the side chain N $\delta$ 2 and O $\delta$ 1 of Asn 121p.



**Figure 3.16** Superposition of the 158 to 169 loop in pPfPMII (carbon atoms in grey) and PfPMII (carbon atoms in cyan).

It should be noted that all the loops described in this section are either flexible or undergo distinct structural changes in other aspartic proteinases and zymogens. For example, the position of the 3<sub>10</sub> helix, Ser 110 to Tyr 114, varies between pepsin and pepsinogen [Sielecki *et al.*, 1991]. The loop corresponding to residues 124 to 135 is disordered in progastricsin, but in the activation intermediate it adopts a well-defined conformation, similar to that found in pepsin [Khan *et al.*, 1997]. Loop 157 to 161 is well-ordered in pepsin, but is disordered in pepsinogen [Hartsuck *et al.*, 1992], and the loop around residue 240 is highly mobile in both pepsin and pepsinogen [Hartsuck *et al.*, 1992].

### 3.3.5 Comparison with gastric AP zymogens and prophytepsin

When we compare the structure of proplasmepsin II to those of the gastric aspartic proteinase zymogens, porcine pepsinogen A and human progastricsin, and the barley AP zymogen, prophytepsin, we observe a striking difference in the mode of inhibition of catalytic activity.

In the gastric zymogens, the prosegment blocks the substrate-binding cleft, rendering the active site inaccessible to substrates (Figures 3.1a, 3.2a). A lysine residue of the prosegment protrudes into the active site, forming salt bridges with the two catalytic Asp residues. In addition, two tyrosine residues (37p and 9) form hydrogen bonds to the catalytic aspartates, preventing substrate access to the active site [Sielecki *et al.*, 1991]. Prophytepsin uses a somewhat different, but related, method of inhibition. In this zymogen, it is the mature N-terminus that provides the inhibitory lysine (Lys 11) (Figures 3.1b, 3.2b). The roles of the two inhibitory tyrosines (37p and 9 in pepsinogen) are played by Tyr13 and Ile 7, which block the S1 and S1' pockets, respectively [Kervinen *et al.*, 1999]. Thus, the inhibition mechanism of prophytepsin is quite similar to that of the gastric AP zymogens, but the role of the critical Lys36p - Tyr37p - Tyr9 triad in pepsinogen is played in prophytepsin by residues of the mature sequence, Lys11, Tyr13 and Ile7. In both pepsinogen and prophytepsin, the active site machinery is pre-assembled, and its structure undergoes only a minor change upon activation [Khan & James, 1997; Kervinen *et al.*, 1999].

In proplasmepsin, the prosegment does not enter the substrate-binding cleft. Rather, it associates with the C-domain of the protein, as part of a "harness" that separates the two lobes of the molecule (Figures 3.6, 3.8). A portion of the "harness", the  $3_{10}$  helix formed by Glu 7 - Phe 11, slightly occludes the cleft in the vicinity of the S3 and S4 substrate-binding pockets. Overall, however, the active site in proplasmepsin is substrate-accessible, but severely distorted and therefore not functional.

pPfPMII shares some similarity with prophytepsin and the gastric zymogens in the secondary structure of the prosegments. In all three types of zymogen the prosegment contributes the first strand to the central  $\beta$ -sheet, which must be replaced by the mature N-terminus upon activation. Since  $\beta$ -strand replacement is seen in such structurally diverse aspartic proteinase zymogens, it must be an intrinsic feature in the activation of eukaryotic aspartic proteinases. One possible reason for this is that insertion of the prosegment strand into the central  $\beta$ -sheet is important in the common folding pathway used by these zymogens. Another possibility is that anchoring the prosegment in the central  $\beta$ -sheet is an effective means of attaching it to the mature portion of the molecule.

The  $\beta$ -strand is followed by an  $\alpha$ -helix which runs in approximately the same direction in pPfPMII and pepsinogen, but is a little longer in proplasmepsin. The major structural difference appears after this helix, where the directions of the second helices differ by  $66^\circ$ . In pepsinogen, the second  $\alpha$ -helix points into the cleft, where it is followed by a short  $3_{10}$  helix containing Lys 36p. In proplasmepsin, the second  $\alpha$ -helix points away from the active site cleft, and lies between the loops 237 - 245 and 271 - 287. Since the prosegments follow different paths in these zymogens, the pro-mature junctions are found in different locations as well. The mature N-terminus occupies the substrate-binding cleft in the gastric zymogens, but in proplasmepsin it runs along the C-domain near the 237 to 245 loop. In prophytepsin, an  $\alpha$ -helix is also observed after the  $\beta$ -strand in the prosegment. This helix enters the active site cleft, but its direction differs from that of helix 1 in either pepsinogen or proplasmepsin. The pro-mature junction is not defined in the structure of prophytepsin, but its location can be inferred to be near the active site cleft.

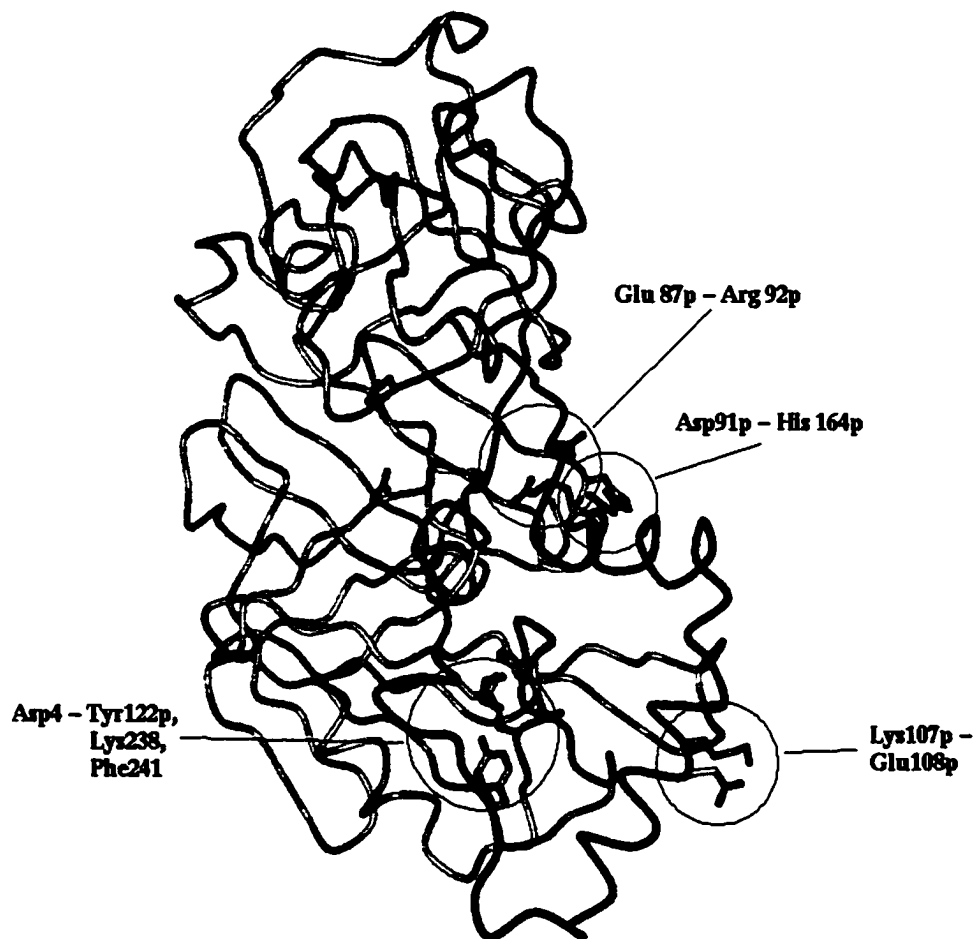
In all three types of zymogen, major refolding of the N-terminus is required during conversion to the active enzyme (residues 1 - 14 in pPfPMII and pepsinogen, and 1 - 17 in

prophytepsin). However, in proplasmepsin, residues 15 to 29 also need to be refolded slightly in order to assume their mature conformation.

In each type of zymogen the length of prosegment that interacts with the body of the molecule is 40 - 45 amino acids. In prophyepsin and the gastric zymogens, this is the entire prosegment, whereas in proplasmepsin, this is the truncated portion of the prosegment. Additionally, in proplasmepsin, as in the gastric zymogens, the prosegment is rich in positively-charged residues that help to stabilize the zymogen at neutral pH.

### 3.3.6 Effect of lowering the pH

Based on the structure of pPfPMII, we can propose some changes that might occur in the molecule as the pH is lowered, leading to either autocatalytic or proteinase-assisted activation. As in the low pH-induced conversion of the gastric aspartic proteinases, the first step is likely to be the disruption of several key salt bridges in the zymogen, which causes structural rearrangements. In proplasmepsin II several hydrogen bonds between residues of the prosegment and the mature segment involve Asp or Glu residues (Figure 3.17). The most prominent among these is the network of hydrogen bonds at the pro-mature junction centered around Asp 4 (Figure 3.9). The interactions made by the carboxylate oxygens of Asp 4 tether the junction to the 237 - 245 loop of the mature segment and close off the "Tyr-Asp" loop, fastening the prosegment around the C-domain of proplasmepsin. A drop in pH would result in the protonation of Asp 4 and disruption of the hydrogen bonds made by its side chain. As a result, the "Tyr-Asp" loop opens up, in effect lengthening the prosegment by 5 residues. The prosegment can no longer act as a tight "harness" to maintain the domain opening in proplasmepsin. Disruption of the Asp 91p to His 164 hydrogen bond would also weaken the association between the prosegment and the C-domain, and loss of the Glu 87p to Arg 92p and Lys 107p to Glu 108p salt bridges may assist further in destabilizing the prosegment. Destabilization of the prosegment helices may eliminate some of the interactions of residues 11 to 17 with the Psi loops. These changes should allow the molecule to assume the active, domain-closed, conformation. By analogy with the activation of the gastric aspartic proteinases [Richter *et al.*, 1998; Khan *et al.*, 1997], the prosegment  $\beta$ -strand probably still remains in its place, only exchanging for the mature N-terminus after the prosegment is cleaved. The bond cleaved in autoactivation is located near the C-terminus



**Figure 3.17** Salt bridges involving prosegment residues in pPpPMII. The coloring scheme for the backbone trace of pPpPMII is the same as in figure 3.6. Side chains for the catalytic aspartates and for the residues involved in salt bridges with the prosegment are shown.



of helix 2, which must be one of the early locations to unravel as the prosegment loses its native secondary structure. The scissile bond can then be presented to a plasmepsin active site for either uni- or bi-molecular cleavage.

Autoactivation of proplasmepsin II has been observed *in vitro* at a range of pH from 3.9 to 5.5, taking place more readily between pH 3.9 and 4.7 [Moon *et al.*, 1997]. The lower pH range covers the pKas of Asp and Glu carboxylates in proteins [Tandford, 1962], even taking into account some pKa depression which may be expected due to these residues' participation in salt bridges and hydrogen bonds. For instance, the involvement of Asp 4 in a number of hydrogen bonds and a salt bridge is likely to lower its side chain pKa relative to that of a solvent-exposed carboxylate. We should consider, however, that interactions of several Asp or Glu residues are disrupted, and therefore the transition may be cooperative.

A report on the potential role of PMII in the degradation of the erythrocyte cytoskeleton revealed that plasmepsin that has not been exposed to low pH, the standard procedure for autoactivation, was able to cleave spectrin at pH 6.8 after a prolonged (20 hour) incubation [Le Bonniec *et al.*, 1999]. SDS-PAGE analysis showed that a small proportion of the zymogen underwent autoactivation, implying that the proteolytic activity was due to the mature enzyme and not the zymogen (C Deregnacourt, personal communication). Similarly, autoactivation of gastric zymogens has been observed after extended storage at pH near neutrality [Foltmann, 1988].

The requirement for low pH for activation by a maturase, that is observed *in vivo*, is less conclusive based on the pPfPMII structure. The pro-mature junction (124p - 1) is located in the "Tyr-Asp" loop at the surface of the molecule, and should be accessible to the external maturase. Acidification may be necessary to loosen this loop in order for the scissile bond to assume an extended conformation that would be suitable for binding to a proteolytic enzyme's active site. Alternatively, low pH may be required if the maturase itself has an acidic pH optimum. Further characterization of the maturase will be needed to resolve this issue.

### **3.3.7 Autoactivation of *P. falciparum* proplasmepsin I**

The role of *P. falciparum* plasmepsin I in hemoglobin digestion has been clearly established, and this enzyme has been recognized as an anti-malarial drug target [Tyas *et*

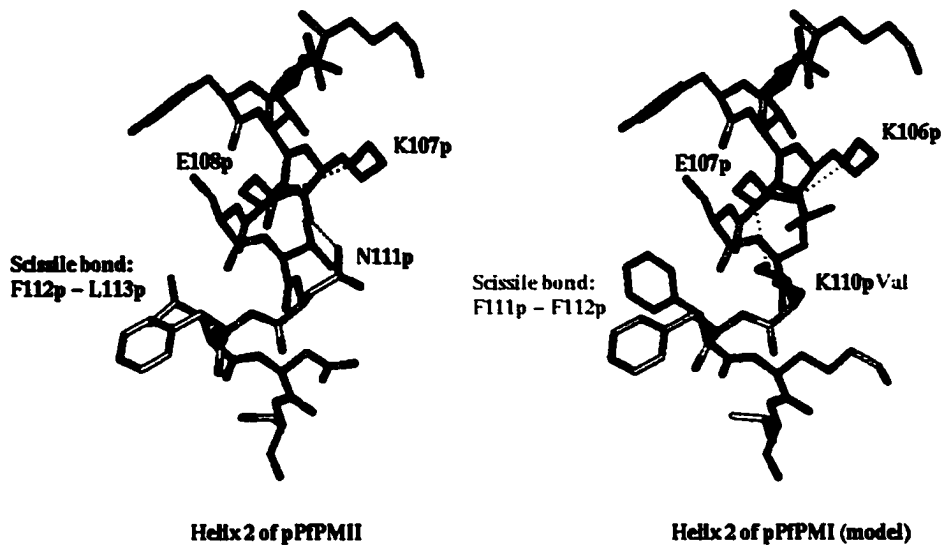
*al.*, 1999]. Unfortunately, production of recombinant PfPMI has been fraught with difficulties. Although high levels of expression of truncated pPfPMI were achieved, the zymogen could not be converted to mature enzyme [Moon *et al.*, 1997]. Various conditions were tested, but none proved conducive to autoactivation. A chimeric zymogen, containing the prosegment of pPfPMII and the mature portion of PfPMI, also resisted attempts at autoactivation.

A model of PfPMI, based on the crystal structure of PfPMII, showed that the sequence around the expected site of autocatalytic cleavage may have been preventing activation [Moon *et al.*, 1997]. The major bond cleaved in autoactivation of pPfPMII, Phe 112p - Leu 113p, is flanked by asparagines. In pPfPMI, the corresponding bond, Phe 111p - Phe 112p, is surrounded in sequence by lysines. Binding of this segment at the PfPMI active site would place Lys 110p into the S2 binding pocket, which was shown to be highly unfavorable. The model of PfPMI suggested that valine at this position would make the prosegment a more suitable substrate for the proteinase. Indeed, the K110pV mutant of pPfPMI, proved capable of autoactivation at acidic pH [Moon *et al.*, 1997].

The crystal structure of pPfPMII suggests a complementary explanation for the K110pV mutant of pPfPMI being amenable to autoactivation. A homology model of pPfPMI, based on the crystal structure of pPfPMII, shows that the Phe 111p - Phe 112p is located in helix 2 of the prosegment, and Lys 110p is involved in a salt bridge with Glu 107p (Figure 3.18). This interaction stabilizes the helix, and its removal, by replacing the Lys with a Val, may destabilize the helical structure. Thus, the K110pV mutation may help the autoproteolysis reaction to proceed by raising the energy of the ground state. In the mutant zymogen, autoactivation is actually observed at two sites: Phe 111p - Phe 112p, and to a smaller extent at Leu 116p - Thr 117p (Figure 3.3). The majority of both wild-type and mutant protein was highly aggregated, so it is difficult to judge which zymogen was better folded. However, the observation of cleavage at the Leu 116p - Thr 117p bond, where no surrounding sequence was altered, suggests a difference in folding between the wild-type and mutant forms of pPfPMI.

The activity of the recombinant PfPMI was significantly lower than that of PfPMI purified from the *P. falciparum* trophozoites [Moon *et al.*, 1997; Tyas *et al.*, 1999]. In assays for the hydrolysis of a series of hemoglobin-based peptides, the  $K_m$  values were similar between the native and recombinant PfPMI, while the  $k_{cat}$  values were up to 10

pPfPMII: ..LKNYIKESVNFLNSGLTKTNYLG...  
 pPfPMI: ..LKNYIKESLKFFKTGLTQKPHLG...  
                   V         ▲



**Figure 3.18** The role of the Lys 110p to Val mutation in the autoactivation of PfPMI.  $\alpha$ -helix 2 of the pPfPMII prosegment (left) is compared to the corresponding helix in the homology model of pPfPMI (right). The sequences of this region of pPfPMII and pPfPMI are shown at the top. Green labels indicate the K110pV mutation and the subsequent autoactivation. The red triangle above the pPfPMII sequence indicates the scissile bond in autoactivation. The green triangles below the pPfPMI sequence indicate bonds that are cleaved autocatalytically after the K110pV mutation is introduced.

times lower for the recombinant enzyme. Tyas *et al.* proposed that the reduced activity of recombinant PfPMI may be due to Lys 113p blocking the active site. A similar inhibitory mechanism was suggested earlier for pseudocathepsin D [Beyer & Dunn, 1996].

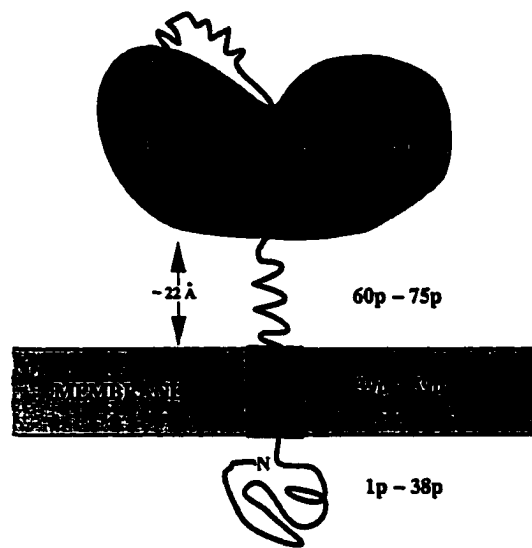
### 3.3.8 A model for the entire proplasmepsin II

The structure presented here is that of a truncated version of proplasmepsin II, and it is interesting to consider how the complete molecule, with 124 amino acids in the prosegment, may look. Due to the close association of the 48 residues of the truncated prosegment with the mature portion of the molecule, and from the location and direction of the N-terminus of the truncated proplasmepsin II, we can deduce that the remaining 76 residues of the prosegment likely protrude away from the body of the molecule. A stretch of approximately 17 residues connects the truncated prosegment to the hydrophobic segment, which has been predicted to form a trans-membrane helix [Francis *et al.*, 1994; Francis *et al.*, 1997]. Preceding this region are 38 residues, which presumably extend into solution on the opposite side of the membrane. Several methods of secondary structure prediction, which had various degrees of success with the truncated prosegment, propose an  $\alpha$ -helix for residues Glu 60p to Lys 75p. A linker consisting of a 15-residue  $\alpha$ -helix would place the mature portion of proplasmepsin II approximately 22 Å from the membrane. The predictions are less clear for the first 38 residues of the prosegment. A schematic diagram of the proposed complete proplasmepsin II model is shown in Figure 3.19.

The trans-membrane segment of proplasmepsin II serves as a signal anchor sequence [Francis *et al.*, 1997], instead of the signal sequences commonly found in other aspartic proteinase preproenzymes [Fusek & Vetvicka, 1995; Koelsch *et al.*, 1994]. Membrane attachment of proplasmepsin II enables this zymogen to be targeted through the secretory system of *P. falciparum*, from the ER to the surface of the parasite [Francis *et al.*, 1997]. From here, proplasmepsin is delivered to the digestive vacuole, where it is activated by a maturase.

## 3.4 Conclusion

The crystal structure of the truncated *P. falciparum* proplasmepsin II has shown us a novel type of aspartic proteinase zymogen and has added to our understanding of



**Figure 3.19** A model for the full-length pPfPMII.

plasmepsin activation. A recent survey of all known structures of proteinase zymogens [Khan & James, 1998] showed that while different mechanisms are employed, in all cases inhibition of catalytic activity is achieved by the prosegment and possibly parts of the mature sequence masking pre-assembled catalytic machinery. Proplasmepsin II, however, has made use of a different way to prevent proteolytic activity, an accessible but incompletely formed ("immature") active site. This resembles zymogen inactivation of the chymotrypsin-like serine proteinases, which do not have properly folded oxyanion holes [Steitz *et al.*, 1969; Freer *et al.*, 1970].

Much more remains to be learned about the proplasmepsin system. For instance, it is interesting to know the structure and function of the N-terminus of the prosegment. In addition, many questions exist about the targeting of proplasmepsin to the *Plasmodium* digestive vacuole. Finally, the characterization of the maturase and the elucidation of the activation mechanism are of major interest, since new anti-malarial drugs may be designed to target proplasmepsin activation. These drugs may be substances that inhibit the maturase, or molecules that fit into the distinctly-shaped active site cleft of proplasmepsin and prevent the domain closing that must occur upon activation.

### 3.5 References

- Abad-Zapatero C, Rydel TJ, Erickson J (1990) Revised 2.3 Å structure of porcine pepsin: evidence for a flexible subdomain. *Proteins* **8**: 62-81.
- Bateman KS, Chernaia MM, Tarasova NI, James MNG. (1998) Crystal structure of human pepsinogen A. *Adv Exp Med Biol* **436**: 259-63.
- Bernstein NK, Cherney MM, Loetscher H, Ridley RG, James MNG (1999) Crystal structure of the novel aspartic proteinase zymogen proplasmepsin II from *Plasmodium falciparum*. *Nat Struct Biol* **6**:32-7.
- Berry, C (1999) Proteases as drug targets for the treatment of malaria. In *Proteases of Infectious Agents*. (Dunn BM, ed. Academic Press, San Diego), pp. 165-188.
- Beyer BM, Dunn BM (1996) Self-activation of recombinant human lysosomal procathepsin D at a newly engineered cleavage junction, "short" pseudocathepsin D. *J Biol Chem* **271**: 15590-6.
- Bruce-Chwatt LJ (1985) *Essential malariology*. John Wiley & Sons, New York.
- Brunger AT (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**: 472-475.
- Brunger AT (1993) *XPLOR: a system for X-ray crystallography and NMR* (Yale University Press, New Haven, Connecticut).
- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**: 905-21.
- CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Cryst D* **50**: 760-763.
- Dame JB, Reddy GR, Yowell CA, Dunn BM, Kay J, Berry C (1994) Sequence, expression and modeled structure of an aspartic proteinase from the human malaria parasite *Plasmodium falciparum*. *Mol Biochem Parasitol* **64**: 177-90.

Elmendorf HG, Haldar K (1993) Secretory transport in *Plasmodium*. *Parasitol Today* **9**: 98-102.

Esnouf RM (1997) An extensively modified version of MolScript that includes greatly enhanced colouring capabilities. *J Mol Graphics* **15**: 133-138.

Foltmann B, (1988) Structure and function of propeptides in zymogens for aspartic proteinases. *Biol Chem Hoppe Seyler* **369** Suppl: 311-4.

Francis SE, Banerjee R, Goldberg DE (1997) Biosynthesis and maturation of the malaria aspartic hemoglobinases plasmeprins I and II. *J Biol Chem* **272**: 14961-8.

Francis SE, Gluzman IY, Oksman A, Knickerbocker A, Mueller R, Bryant ML, Sherman DR, Russell DG, Goldberg DE (1994) Molecular characterization and inhibition of a *Plasmodium falciparum* aspartic hemoglobinase. *EMBO J* **13**: 306-17.

Freer ST, Kraut J, Robertus JD, Wright HT, Xuong NH (1970) Chymotrypsinogen: 2.5-angstrom crystal structure, comparison with alpha-chymotrypsin, and implications for zymogen activation. *Biochemistry* **9**: 1997-2009.

Fusek M, Vetvicka V (1995) *Aspartic proteinases: physiology and pathology*. (CRC Press Inc., Boca Raton).

Glathe S, Kervinen J, Nimtz M, Li GH, Tobin GJ, Copeland TD, Ashford DA, Wlodawer A, Costa J (1998) Transport and activation of the vacuolar aspartic proteinase phytepsin in barley (*Hordeum vulgare* L.). *J Biol Chem* **273**: 31230-6.

Gluzman IY, Francis SE, Oksman A, Smith CE, Duffin KL, Goldberg DE (1994) Order and specificity of the *Plasmodium falciparum* hemoglobin pathway. *J Clin Invest* **93**: 1602-8.

Goldberg DE, Slater AF, Beavis R, Chait B, Cerami A, Henderson GB (1991) Hemoglobin degradation in the human malaria pathogen *Plasmodium falciparum*: a catabolic pathway initiated by a specific aspartic protease. *J Exp Med* **173**: 961-9.

Hartsuck JA, Koelsch G, Remington SJ (1992) The high-resolution crystal structure of porcine pepsinogen. *Proteins* **13**: 1-25.



- Hill J, Tyas L, Phylip LH, Kay J, Dunn BM, Berry C (1994) High level expression and characterisation of Plasmepsin II, an aspartic proteinase from *Plasmodium falciparum*. *FEBS Lett* **352**: 155-8.
- Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. *Nature* **381**: 272.
- Hutchinson EG, Thornton JM (1994) A revised set of potentials for beta-turn formation in proteins. *Protein Sci* **3**: 2207-16.
- Institute of Medicine (U.S.) (1991) *Malaria: obstacles and opportunities: a report of the Committee for the Study on Malaria Prevention and Control: Status Review and Alternative Strategies, Division of International Health*. (Institute of Medicine / Oaks Jr SC. *et al.*, eds. National Academy Press, Washington, D.C.).
- James MNG, Sielecki AR (1986) Molecular structure of an aspartic proteinase zymogen, porcine pepsinogen, at 1.8 Å resolution. *Nature* **319**: 33-8.
- Jones TA, Kjeldgaard M (1995) *O the Manual, Version 5.11* (Uppsala, Sweden).
- Kervinen J, Tobin GJ, Costa J, Waugh DS, Wlodawer A, Zdanov A (1999) Crystal structure of plant aspartic proteinase prophytepsin: inactivation and vacuolar targeting. *EMBO J* **18**: 3947-55.
- Khan AR, Cherney MM, Tarasova NI, James MNG (1997) Structural characterization of activation 'intermediate 2' on the pathway to human gastricsin. *Nat Struct Biol* **4**: 1010-5.
- Khan AR, James MNG (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci* **7**: 815-36.
- Khan AR, Khazanovich-Bernstein N, Bergmann EM, James MNG (1999) Structural aspects of activation pathways of aspartic protease zymogens and viral 3C protease precursors. *Proc Natl Acad Sci U S A* **96**: 10968-75.
- Koelsch G, Mares M, Metcalf P, Fusek M (1994) Multiple functions of pro-parts of aspartic proteinase zymogens. *FEBS Lett* **343**: 6-10.
- Kraulis PJ (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Cryst* **24**: 946-950.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**: 283-291.

Le Bonniec S, Deregnacourt C, Redeker V, Banerjee R, Grellier P, Goldberg DE, Schrevel J (1999) Plasmepsin II, an acidic hemoglobinase from the *Plasmodium falciparum*. *J Biol Chem* **274**: 14218-23.

Lee AY, Gulnik SV, Erickson JW (1998) Conformational switching in an aspartic proteinase. *Nat Struct Biol* **5**: 866-71.

Luker KE, Francis SE, Gluzman IY, Goldberg DE (1996) Kinetic analysis of plasmepsins I and II aspartic proteases of *Plasmodium falciparum* digestive vacuole. *Mol Biochem Parasitol* **79**: 71-78.

Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* **33**: 491-497.

Merritt EA, Murphy MEP (1994) Raster3D version 2.0: a program for photorealistic molecular graphics. *Acta Cryst D* **50**: 869-873.

Moon RP, Tyas L, Certa U, Rupp K, Bur D, Jacquet C, Matile H, Loetscher H, Grueninger-Leitch F, Kay J, Dunn BM, Berry C, Ridley RG (1997) Expression and characterisation of plasmepsin I from *Plasmodium falciparum*. *Eur J Biochem* **244**: 552-60.

Moore SA, Sielecki AR, Chernaiia MM, Tarasova NI, James MNG (1995) Crystal and molecular structures of human progastricsin at 1.62 Å resolution. *J Mol Biol* **247**: 466-85.

Navaza J (1994) AMoRe: an automated package for molecular replacement. *Acta Cryst.* **A50**: 157-163.

Oefner C, Binggeli A, Breu V, Bur D, Clozel JP, D'Arcy A, Dorn A, Fischli W, Gruninger F, Guller R, Hirth G, Marki H, Mathews S, Müller M, Ridley RG, Stadler H, Vieira E, Wilhelm M, Winkler F, Wostl W (1999) Renin inhibition by substituted piperidines: a novel paradigm for the inhibition of monomeric aspartic proteinases? *Chem Biol* **6**: 127-31.

- Otwinowski Z, Minor W (1996) Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology*, 276; (Carter Jr CW, Sweet R, eds., Academic Press, New York and London) pp. 307–326.
- Pannu NS, Read RJ (1996) Improved structure refinement through maximum likelihood. *Acta Cryst A* 52: 659-668.
- Partin K, Zybarth G, Ehrlich L, DeCrombrughe M, Wimmer E, Carter C (1991) Deletion of sequences upstream of the proteinase improves the proteolytic processing of human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A* 88:4 776-80.
- Richter C, Tanaka T, Yada RY (1998) Mechanism of activation of the gastric aspartic proteinases: pepsinogen, progastricsin and prochymosin. *Biochem J* 335: 481-90.
- Rosenthal PJ (1998) Proteases of malaria parasites: new targets for chemotherapy. *Emerg Infect Dis* 4: 49-57.
- Runeberg-Roos P, Kervinen J, Kovaleva V, Raikhel NV, Gal S (1994) The aspartic proteinase of barley is a vacuolar enzyme that processes probarley lectin in vitro. *Plant Physiol* 105: 321-9.
- Runeberg-Roos P, Saarma M (1998) Phytpepsin, a barley vacuolar aspartic proteinase, is highly expressed during autolysis of developing tracheary elements and sieve cells. *Plant J* 15: 139-45.
- Runeberg-Roos P, Tormakangas K, Ostman A (1991) Primary structure of a barley-grain aspartic proteinase. A plant aspartic proteinase resembling mammalian cathepsin D. *Eur J Biochem* 202: 1021-7.
- Sali A, Veerapandian B, Cooper JB, Moss DS, Hofmann T, Blundell TL (1992) Domain flexibility in aspartic proteinases. *Proteins* 12: 158-70.
- Sielecki AR, Fedorov AA, Boodhoo A, Andreeva NS, James MNG (1990) Molecular and crystal structures of monoclinic porcine pepsin refined at 1.8 Å resolution. *J Mol Biol* 214: 143-70.
- Sielecki AR, Fujinaga M, Read RJ, James MNG (1991) Refined structure of porcine pepsinogen at 1.8 Å resolution. *J Mol Biol* 219: 671-92.

Silva AM, Lee AY, Gulnik SV, Maier P, Collins J, Bhat TN, Collins PJ, Cachau RE, Luker KE, Gluzman IY, Francis SE, Oksman A, Goldberg DE, Erickson JW (1996) Structure and inhibition of plasmepsin II, a hemoglobin-degrading enzyme from *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* **93**: 10034-9.

Steitz TA, Henderson R, Blow DM (1969) Structure of crystalline alpha-chymotrypsin. 3. Crystallographic studies of substrates and inhibitors bound to the active site of alpha-chymotrypsin. *J Mol Biol* **46**: 337-48.

Tandford C (1962) The interpretation of hydrogen ion titration curves of proteins. *Adv Protein Chem* **17**: 69-165.

Tyas L, Gluzman I, Moon RP, Rupp K, Westling J, Ridley RG, Kay J, Goldberg DE, Berry C (1999) Naturally-occurring and recombinant forms of the aspartic proteinases plasmepsins I and II from the human malaria parasite *Plasmodium falciparum*. *FEBS Lett* **454**: 210-4.

Vieira E, Binggeli A, Breu V, Bur D, Fischli W, Guller R, Hirth G, Marki HP, Muller M, Oefner C, Scalone M, Stadler H, Wilhelm M, Wostl W (1999) Substituted piperidines--highly potent renin inhibitors due to induced fit adaptation of the active site. *Bioorg Med Chem Lett* **9**: 1397-402.

Xie D, Gulnik S, Collins L, Gustchina E, Suvorov L, Erickson JW (1997) Dissection of the pH dependence of inhibitor binding energetics for an aspartic protease: direct measurement of the protonation states of the catalytic aspartic acid residues. *Biochemistry* **36**: 16166-72.

## Chapter 4: CRYSTAL STRUCTURE OF *P. vivax* PROPLASMEPSIN

### 4.1 Introduction

*P. vivax* is the most common malaria parasite that infects humans [Bruce-Chwatt, 1985]. It has not, however, been studied as extensively as its deadly relative, *P. falciparum*. The greater danger posed by *P. falciparum* is a powerful incentive for directing precious research resources at that deadly pathogen.

The hemoglobin degradation pathway has been studied in considerable detail in *P. falciparum*, as summarized in Chapter 1 (Figure 1.3). No analogous work has been described for the other malarial pathogens, in part because the lack of laboratory cultures makes investigation of these organisms very difficult. However, plasmepsin, a protein homologous to the hemoglobin-degrading aspartic proteinases of *P. falciparum*, has been identified in a number of other *Plasmodium* species, including the human pathogens and the rodent parasite, *P. berghei* (Table 2.1). Although the plasmepsins from *P. vivax* (PvPM), *P. ovale* (PoPM) and *P. malariae* (PmPM) are close in sequence to PfPMI and PfPMII, which have been shown to act as hemoglobinasases, they share even greater similarity with PfPMIV, an enzyme whose function has not yet been established (Table 2.2). PfPMIV does appear to be expressed during the blood stage of malaria (J. Dame, personal communication), but its role may be distinct from hemoglobin degradation.

PvPM is initially produced by the parasite as the zymogen proplasmepsin (pPvPM). The N-terminal prosegment of pPvPM is typical of the proplasmepsins, consisting of 123 amino acids and containing a 21-amino acid long hydrophobic segment 38 residues from its N-terminus. The corresponding hydrophobic segment in PfPMI and PfPMII has been shown to act as a signal anchor sequence [Francis *et al.*, 1997].

A recombinant form of pPvPM, with the prosegment shortened to 48 residues, has been cloned and expressed [Westling *et al.*, 1997]. This zymogen is inactive at neutral pH, and can undergo autocatalytic activation at acidic pH to yield the active enzyme. The activating cleavage, at Tyr 121p - Leu 122p, produces the mature enzyme extended at the N-terminus by two residues (Figure 3.3) [J. Dame, manuscript in preparation].

I have determined the crystal structure of the truncated pPvPM. A comparison of this structure to that of the mature enzyme, PvPM (presented in Chapter 2), makes it possible to analyze the mode of inactivation of this zymogen. Furthermore, comparison with the structure of pPfPMII (presented in Chapter 3) will help to define those features which are common to the proplasmepsins, and which are specific to the different members of this group.

## 4.2 Methods

### 4.2.1 Crystallization and data collection

Drs. Charles A. Yowell and John B. Dame (University of Florida, Gainesville) provided recombinant truncated *P. vivax* proplasmepsin, expressed in *E. coli*. This molecule contains the last 48 residues of the prosegment as well as an additional 13 amino acids (A S M T G G Q Q M G R G S) at the N-terminus, that were derived from the pET3a expression plasmid. The presence of the additional amino acids was confirmed by Edman degradation. Proplasmepsin was crystallized without any further purification by the vapor diffusion method by equilibrating sitting drops containing 2  $\mu$ L of protein solution (15 mg/mL in 10 mM Tris, pH8.0) and 3  $\mu$ L of reservoir solution (18% PEG 4000, 100 mM Tris, pH 8.0, 200 mM NH<sub>4</sub> acetate, 3% *t*-amyl alcohol and 15% glycerol) against 1 mL of the reservoir solution. Seeding was used to optimize the crystals for diffraction. The crystals belonged to the monoclinic space group C2 with  $a = 149.41 \text{ \AA}$ ,  $b = 92.74 \text{ \AA}$ ,  $c = 99.13 \text{ \AA}$ ,  $\beta = 130.10^\circ$  and contained two molecules in the asymmetric unit ( $V_m = 3.05 \text{ \AA}^3/\text{Da}$  for 2 molecules/asymmetric unit) [Matthews, 1968] (Table 4.1).

One data set (DS-1) was collected on a DIP 2030H image plate detector (Mac Science Co., Ltd.) using double-mirror focussing optics and CuK $\alpha$  radiation (1.5418  $\text{\AA}$ ) generated with a Rigaku rotating anode generator RU-200BH operating at 45 kV and 75 mA. For data collection the crystals were cryoprotected by quickly dipping them into a solution containing 20% PEG 4000, 100 mM Tris, pH 8.0, 200 mM NH<sub>4</sub> acetate and 30% glycerol, and flash-frozen in a cold N<sub>2</sub> gas stream at 100K. 182 images (1 $^\circ$  rotations) were collected at a crystal to detector distance of 120 mm. The data were processed and scaled with Denzo and Scalepack [Otwinowski & Minor, 1996]. This crystal had a mosaicity of 0.69 $^\circ$ , and the data set was 95% complete between 20 and 2.7  $\text{\AA}$ , with an

**Table 4.1** Summary of data and refinement

Space Group	C2
Unit cell	$a = 149.41 \text{ \AA}$ $b = 92.74 \text{ \AA}$ $c = 99.13 \text{ \AA}$ $\beta = 130.11^\circ$
Molecules/asymmetric unit	2
$V_m$ [Mathews, 1968]	$3.0 \text{ \AA}^3/\text{Da}$
Data completeness	99.3 [88.3] % (30 - 2.50 \AA)
$F/\sigma F > 1$ , [ $F/\sigma F > 3$ ]	98.9 [65.1] % (2.54 - 2.50 \AA)
$I/\sigma I$	15.52 [3.79 for 2.52 - 2.50 \AA]
Observed reflections	116818
Unique reflections	35657
Rmerge <sup>1</sup>	4.8 % [33.6 % for 2.59 - 2.50 \AA]
<b>Final refinement statistics</b>	
R-factor <sup>2</sup> (30 - 2.50 \AA, 33805 reflections)	23.7 %
R-free <sup>3</sup> (30 - 2.50 \AA, 1852 reflections)	29.4 %
Number of protein atoms	5985
Number of water molecules	67
R.m.s. deviations from ideal geometry	
bond lengths	0.007 \AA
bond angles	1.41 °
improper angles	0.78 °
Temperature factors	
mean B (overall)	60.3 \AA <sup>2</sup>
estimated from Wilson plot	62.1 \AA <sup>2</sup>
Estimated cross-validated coordinate error	
from luzzati plot	0.51 \AA
from sigmaa	0.57 \AA
Bulk solvent model	
Ksol	0.345 e/\AA <sup>3</sup>
Bsol	43.7 \AA <sup>2</sup>
Ramachandran plot statistics for 669 non-proline and non-glycine residues [Laskowski <i>et al.</i> , 1993]	
residues in most favored regions	548 (81.9 %)
residues in additional allowed regions	113 (16.9 %)
residues in generously allowed regions	7 (1.0 %)
residues in disallowed regions	1 (0.1 %)

$$^1 R_{\text{merge}} = \frac{\sum hklj |I_{hklj} - \langle I_{hklj} \rangle|}{\sum hklj \langle I_{hklj} \rangle}$$

<sup>2</sup>R-factor =  $\frac{\sum ||F_o| - |F_c||}{\sum |F_o|}$ , where  $|F_o|$  and  $|F_c|$  are the observed and calculated structure factor amplitudes, respectively.

<sup>3</sup>R-free was calculated using a random set containing 2.1% of observations which were omitted during refinement [Brunger, 1992].

R-merge of 0.07 and redundancy of 3.4.

Another data set (DS-2) was later collected on beamline X12B at the NSLS. 154 images (1° rotations) were collected at 90K on a MAR image plate in the 18 cm scanning mode at a wavelength of 0.978 Å and a crystal to detector distance of 200 mm. The data were processed and scaled with Denzo and Scalepack [Otwinowski & Minor, 1996]. Although some diffraction spots were observed at better than 2.3 Å resolution, the diffraction was highly anisotropic, and based on Scalepack statistics (binwise R-merge) the data was processed between 30 and 2.5 Å. The mosaicity of this crystal was 0.54° and the data set had a completeness of 99.3%, R-merge 0.048 and redundancy of 3.3 (Table 4.1).

#### 4.2.2 Structure determination and refinement

The structure of *P. vivax* proplasmepsin was solved by molecular replacement with AMoRe [Navaza, 1994] (Table 4.2). Initially the molecular replacement searches were carried out with the DS-1 data, using various forms of the current *P. falciparum* proplasmepsin II model. The models tested were proplasmepsin II with most of the prosegment (36 out of 48 residues, 79P - 115P) (corr. = 0.425, R = 49.6%), proplasmepsin II without the prosegment (corr. = 0.43, R = 49.6%), and proplasmepsin II divided into N- and C-terminal domains (7 - 134 and 135 - 327, respectively). Although all three searches yielded the same solutions, the last one, with separate domains, gave the best results (correlation = 0.443, R = 49.1% for 10 - 3 Å data). After introduction of the required sequence changes and some manual rebuilding in O [Jones & Kjeldgaard, 1995], the model was subjected to one round of refinement in X-PLOR [Brunger, 1993], including positional refinement, simulated annealing and temperature factor refinement. The R- and free R-factors of this model after the refinement were 0.372 and 0.421, respectively.

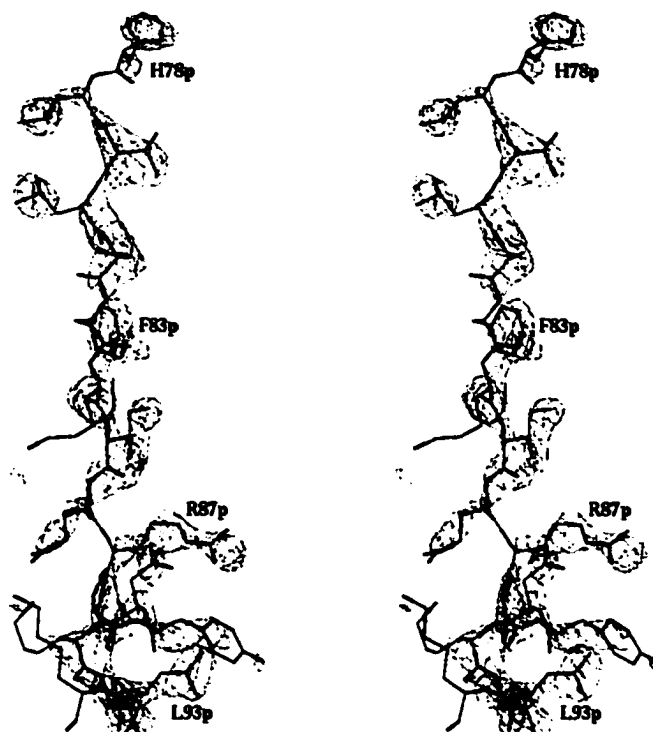
The above model was then used in molecular replacement with the DS-2 data set. This time the two molecular replacement solutions were much stronger than with any of the models used in the DS-1 MR searches: correlation = 0.683, R = 38.3% for 10 - 3 Å data. The resultant model was subjected to refinement in X-PLOR and CNS [Brunger, 1993; Brunger *et al.*, 1998] (Figures 4.1a, 4.1b). The refinement protocol included torsion angle simulated annealing, positional refinement (against the maximum likelihood



<b>MR results</b>		
	<b>Correlation</b>	<b>R-factor (%)</b>
<b>Data Set 1</b>		
Model 1	0.425	49.6
Model 2	0.430	49.6
Model 3	0.443	49.1
<b>Data Set 2</b>		
Model 4	0.683	38.3

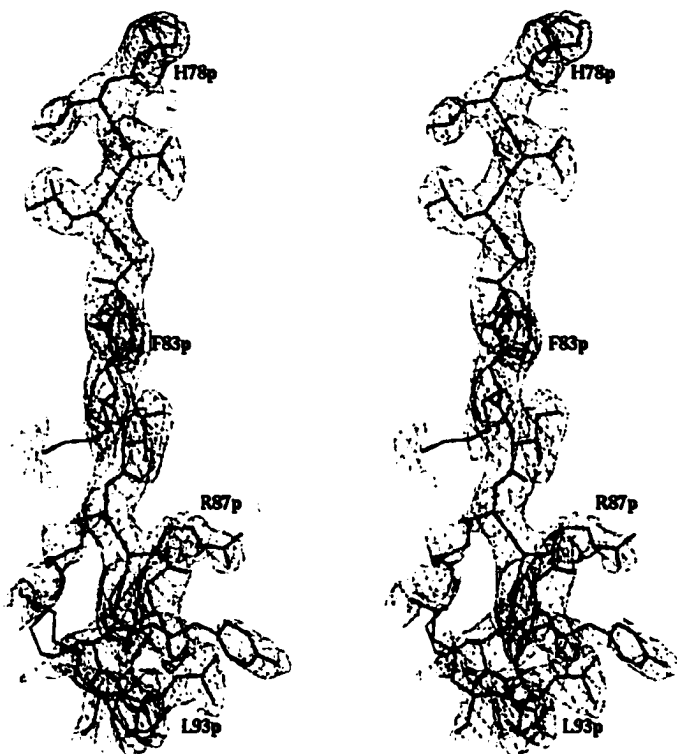
**Table 4.2** Molecular replacement results for pPvPM. Model 1 consisted of His 79p – Asn 115p, Glu 7 – Met 75 and Val 82 – Asn 328 of a partially refined pPfPMII structure. Model 2 was the same as model 1, but lacked the prosegment residues. Model 3 was the same as model 2, but searches were carried out with two separate domains (Glu 7 – Gly 134 and Ser 135 – Asn 328). Model 4 was the partial pPvPM structure (Glu 7 – Thr 71, Val 82 – Ile 277, Cys 285 – Asn 327) generated from Model 3, and subjected to one cycle of refinement in X-PLOR [Brunger, 1993].

(a)

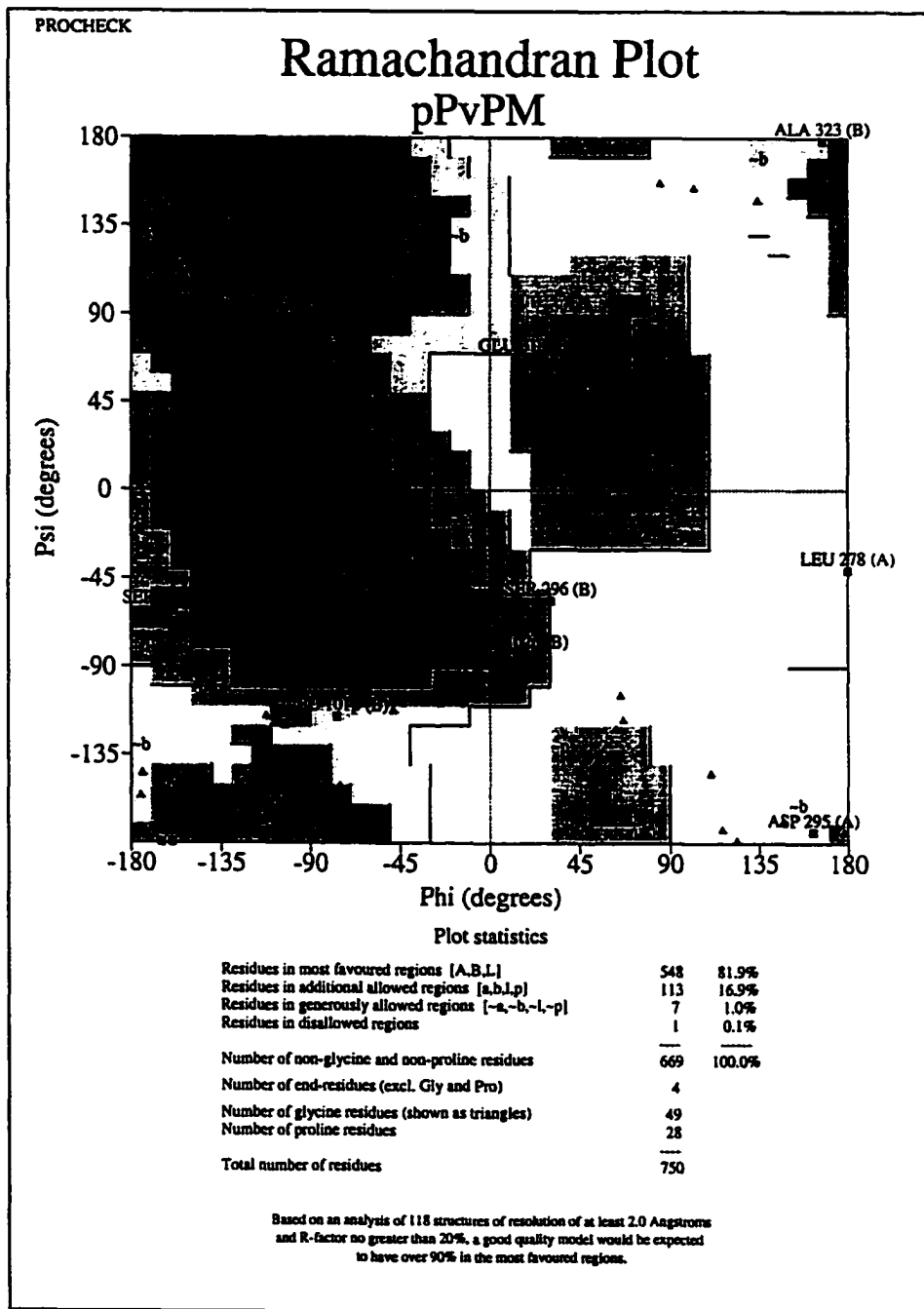


**Figure 4.1** Refinement progress for pPvPM. (a) Difference ( $|F_O| - |F_C|$ ) electron density for the prosegment in the molecular replacement map, phased with model 4 from Table 4.2, contoured at  $2\sigma$  about a portion of the prosegment. (b) Final refined ( $2|F_O| - |F_C|$ ) map contoured at  $1\sigma$  about the same portion of the prosegment. The final model is shown in (a) and (b). (c) Ramachandran plot for the pPvPM structure. Parts (a) and (b) of this figure, as well as figures 4.3 - 4.8 and 4.10 - 4.12, were produced with Bobscript [Esnouf, 1997; Merritt & Murphy, 1994].

(b)



(c)



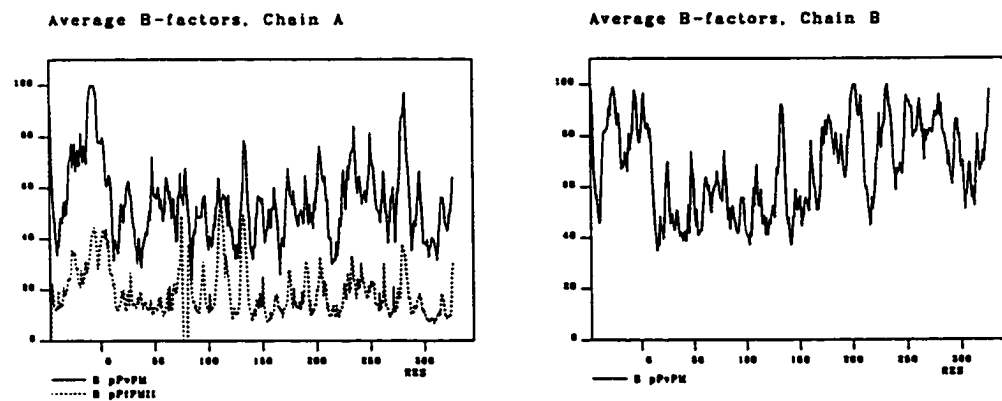
targets mli and mlf [Pannu & Read, 1996]) and restrained individual or grouped temperature factor refinement (against the least squares target). Use of the maximum likelihood targets in temperature factor refinement resulted in unreasonably low B-factors for several residues. Anisotropic B-factor scaling and a bulk solvent correction were applied to the data in CNS. 2-fold NCS restraints were imposed at the early stages of refinement, and were gradually removed. In all cases the improvement of the free R-factor was monitored. Cycles of refinement were alternated with manual model building in O [Jones & Kjeldgaard, 1995]. Water molecules were added conservatively, generally at positions that are either common in the two NCS-related molecules or at positions where water molecules are also found in the structure of *P. falciparum* proplasmepsin II. The final model contains 750 protein residues and 67 water molecules. Residues Thr A76p, Met A117p, Lys A118p, Thr B76p, Glu B77p, Phe B102p, Lys B203, Val A280 and Asp A281 have been included into the model as alanines and Gln A119p has been included as a glycine due to poor electron density for the side chains. The model is missing 13 residues at the N-termini of both molecules in the asymmetric unit. The final R- and free R-factors are 0.236 and 0.295, respectively, for all data between 30 and 2.5 Å (Table 4.1). Ser B296 and Gln A119p (which has been modelled as Gly) are in the disallowed regions of the Ramachandran plot. Seven residues are found in the generously allowed regions of the plot (Figure 4.1c). All the residues outside the favored regions of the Ramachandran plot are in areas of poor electron density.

## 4.3 Results and discussion

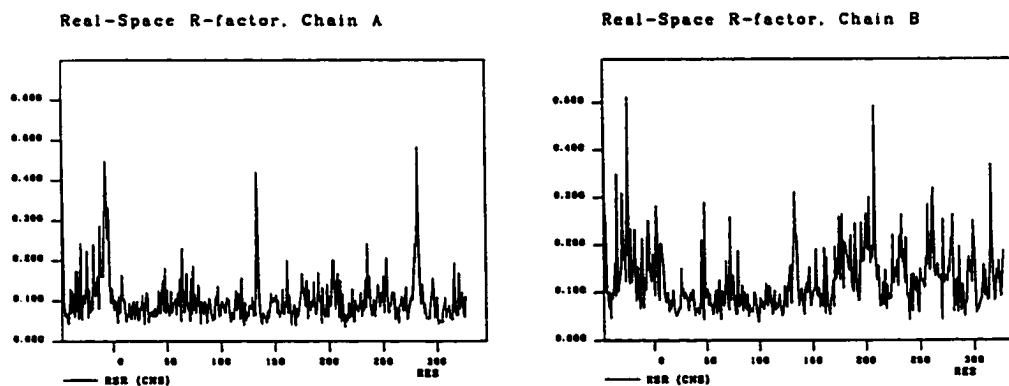
### 4.3.1 Quality of the model

Before discussing the *P. vivax* proplasmepsin structure, it is necessary to address the quality of this model. Two causes for concern with this structure are the high free R-factor (29.5%) and the high average temperature factor ( $60 \text{ \AA}^2$ ). Several lines of evidence attest to the general correctness of the model, however. Extremely high ( $> 80 \text{ \AA}^2$ ) temperature factors are not uniformly distributed, but are localized to portions of the structure that are expected to be flexible. These include surface loops, the C-subdomain of molecule B and parts of molecule A that contact this domain of molecule B (Figures 4.2a. 4.3a). As has been discussed earlier (in Chapters 2 and 3), the loops in question and

(a)

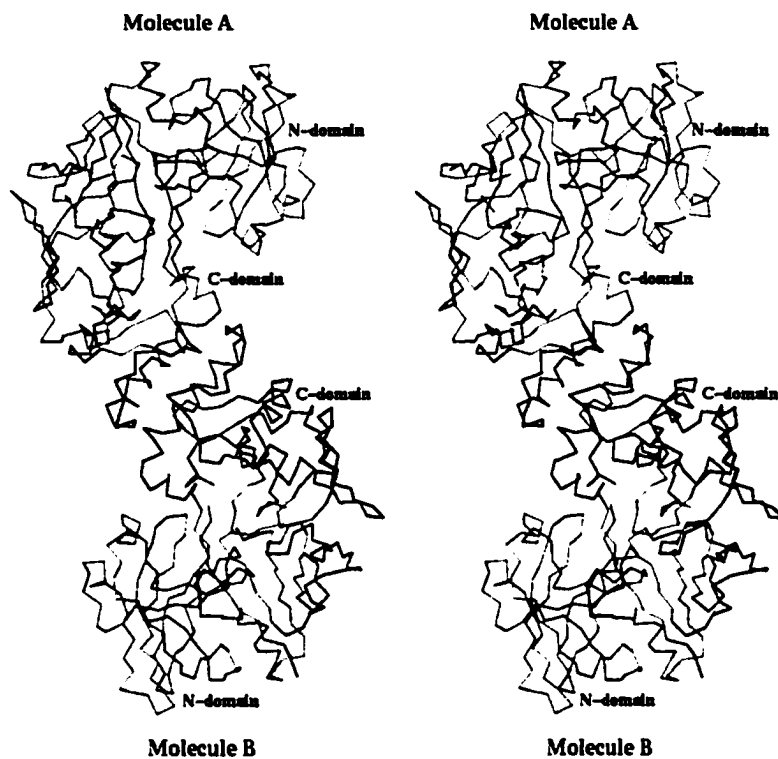


(b)



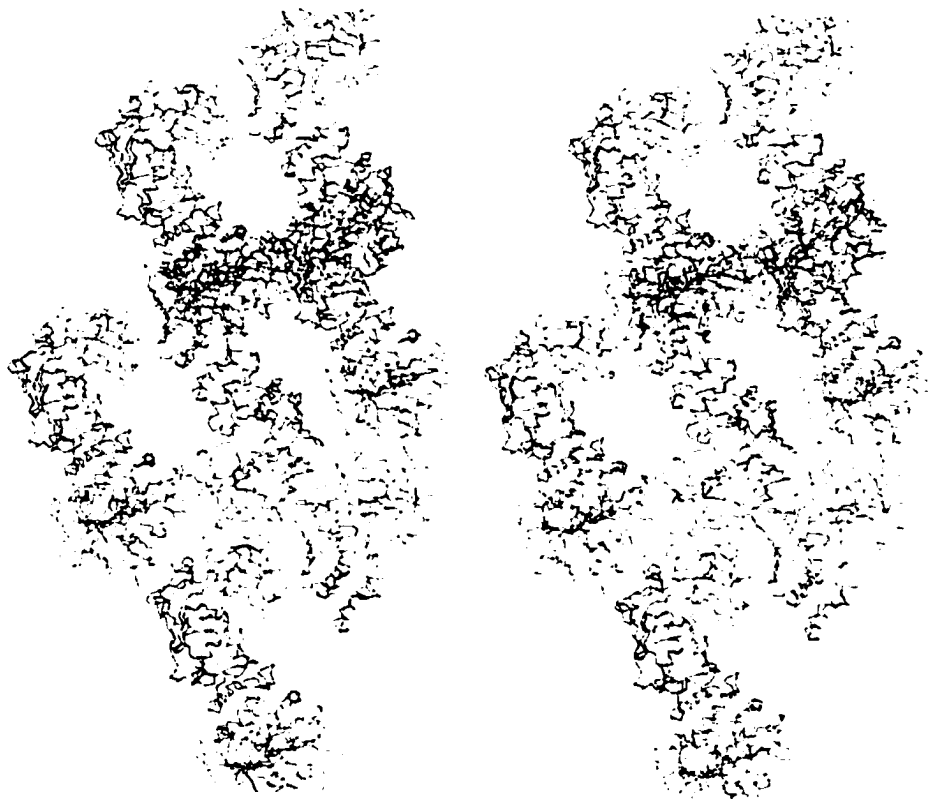
**Figure 4.2** Residue properties of pPvPM. (a) Temperature factor (B-factor) plot. (b) Real-space R-factor (rsr) plot.  $rsr = (\sum |\rho_{obs} - \rho_{calc}|) / (\sum |\rho_{obs} + \rho_{calc}|)$ .

(a)



**Figure 4.3** Distribution of temperature factors in pPvPM. (a) The dimer in the crystallographic asymmetric unit of pPvPM is shown, colored by B-factor. The color ramping is from green ( $B \leq 25 \text{ \AA}^2$ ) to red ( $B \geq 90 \text{ \AA}^2$ ). (b) Packing of the pPvPM dimers in the crystal.

**(b)**





the C-domain are generally found to be flexible in aspartic proteinases. The electron density associated with these regions of high B-factors was poor and very difficult to interpret (Figure 4.2b). In the C-domain, for instance, the use of two-fold NCS averaging improved the map, suggesting that the structure of this region is basically correct, and the electron density is poor due to domain flexibility. In several loops, however (Met 117p to Tyr 121p, Ile 133 to Gly 134 and Leu 278 to Asp 282), the two NCS-related molecules clearly differ in conformation, and therefore electron density averaging was not helpful. The average temperature factor of  $60 \text{ \AA}^2$  is in close agreement with the overall temperature factor estimated from the Wilson plot by the program CNS,  $62 \text{ \AA}^2$  between 4 and  $2.5 \text{ \AA}$  (Table 4.1). The rather high solvent content, 60%, may be allowing the protein a lot of opportunity for movement or disorder in the crystals. In addition, the packing arrangement in the crystal may be contributing to localized disorder, (see below). The regions of high B-factors with their nearly uninterpretable electron density are likely responsible for the refinement stalling at a free R-factor of 29.5%.

Other evidence for the overall correctness of the model includes: strong MR solutions, the appearance of prosegment difference electron density in the original map phased with the MR model which did not contain any of the prosegment residues (Figure 4.1a), and good agreement of most of the model with simulated annealing omit maps. Finally, the similarities of the structure of *P. vivax* proplasmepsin to the higher-resolution structure of *P. falciparum* proplasmepsin II further confirm that the former is, on the whole, accurate.

Due to the differences in crystal packing of the two molecules in the asymmetric unit, if a segment of polypeptide is disordered in one molecule, in most cases, the corresponding part is better defined in the other NCS-related molecule. Therefore in the following sections, the better ordered portion of the structure will be discussed.

#### 4.3.2 Description of structure

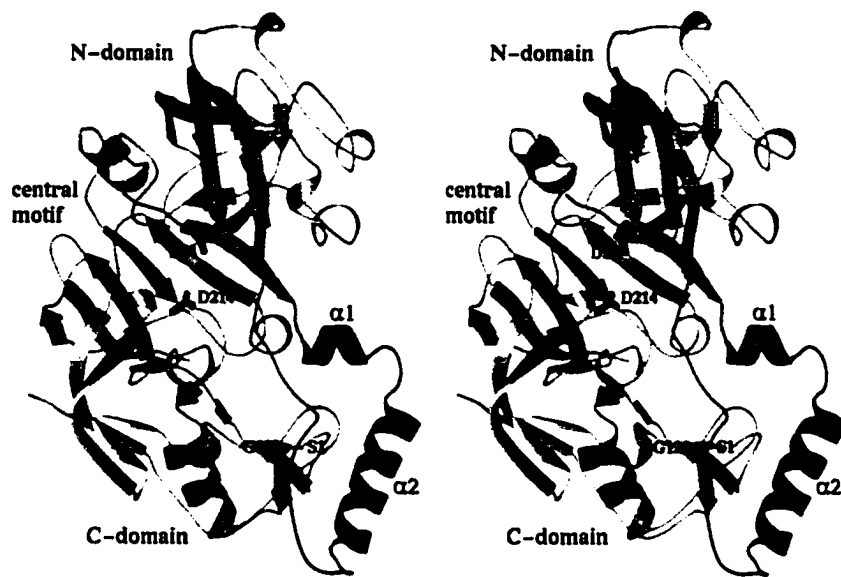
To facilitate expression, and based on the experience with proplasmepsin II from *P. falciparum*, *P. vivax* proplasmepsin was expressed in a truncated form that is missing the first 76 residues of the prosegment. The prosegment present in this construct of *P. vivax* proplasmepsin therefore contains 48 residues, of which the majority are defined in

the crystal structure. In this chapter, the term "proplasmepsin" and the abbreviation "pPvPM" will be used to refer to this recombinant, truncated proplasmepsin.

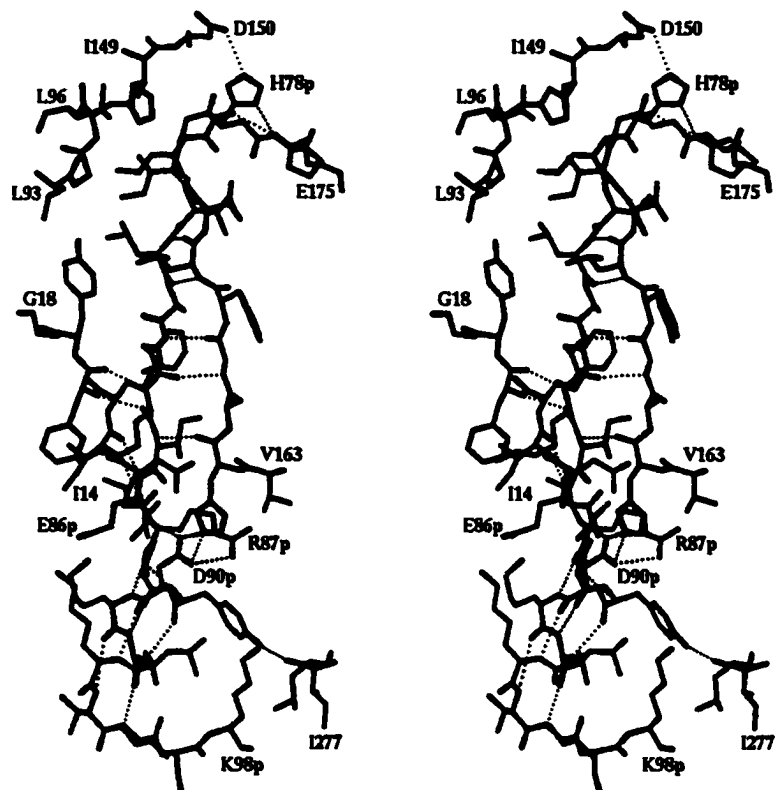
*P. vivax* proplasmepsin is folded into three domains: the N- and C-terminal domains anchored to the central motif, the six-stranded anti-parallel  $\beta$ -sheet that was described in the two preceding chapters (Figure 4.4). The domain composition is similar to that in PvPM (Figure 2.11), with several key differences. In pPvPM, the first strand of the central  $\beta$ -sheet consists of Thr 80p to Arg 87p. Pro 88p to Asn 13 belong to the C-domain in the zymogen, and the N-domain contains nearly the same residues as in the mature enzyme (Ile 14 - Asn 151).

Although the electron density for certain portions of the prosegment is poor, it was still possible to define the structure of most of the prosegment residues, His 79p to Gly 124p (Figure 4.4). The prosegment folds into a  $\beta$ -strand (Thr 81p to Glu 87p), that serves as the first strand of the central  $\beta$ -sheet, two  $\alpha$ -helices (Arg 88p to Ile 97p and Ser 98p to Lys 114p), followed by a coil connection to the mature sequence. The prosegment forms numerous hydrogen bonds and van der Waals contacts with the mature portion of pPvPM (Figure 4.5). The side chain nitrogen atoms of His 79p form hydrogen bonds with the carboxylates of Asp 150 and Glu 174, that surround the imidazole ring. Residues Leu 80p to Ile 86p form anti-parallel  $\beta$ -sheet hydrogen bonds with residues Ile 170 to His 164. This set of interactions is part of the central  $\beta$ -sheet. The side chains of Leu 80p, Ile 82p, Phe 84p and Ile 86p fit into hydrophobic pockets lined by residues of the mature portion of pPvPM. Thr 81p O<sub>1</sub> forms a hydrogen bond with Thr 169 O<sub>1</sub>. Lys 85p and Glu 87p are engaged in 4 anti-parallel  $\beta$ -sheet hydrogen bonds with Ile 14 and Phe 16, as part of one of the N-domain  $\beta$ -sheets.

The portion of the prosegment following the  $\beta$ -strand interacts extensively with the C-domain (Figure 4.5). The first helix of the prosegment contains 2 turns (10 residues, Arg 88p to Ile 97p). The N-terminus of the helix is stabilized by hydrogen bonds from Arg 88p peptide NH and its side chain N $\zeta$  to O $\delta$ 2 of Asp 91p. The carboxylate of Asp 91p also participates in a salt bridge with His 164 N $\epsilon$ 2. The second residue of helix 1 is Pro 88p, whose constrained main chain dihedral angle  $\omega$  (-52°) favors helix formation. Tyr 89p donates a hydrogen bond from its OH to Ile 277 O. In addition, since the guanidinium group of Arg 87p is within 4 Å of the Tyr 89p aromatic ring, these two residues may be interacting *via* cation - aromatic ( ) stacking. The loop connecting helix



**Figure 4.4** The structure of pPvPM. The prosegment is colored magenta, residues 1 - 21 of the mature sequence are in cyan, the mature portion of the central motif is in green, and the remaining portion of the molecule is in yellow. The catalytic aspartates are shown, and the pro-mature junction is labeled (G123p - S1).



**Figure 4.5** Interactions of the prosegment with the mature portion of pPvPM. Residues H78p to K98p of the prosegment (carbon atoms in cyan) are shown. Residues from the mature segment have carbon atoms in grey.

1 to helix 2 is highly flexible. Helix 2 has rather poor electron density in both molecules, but slightly better in B. Helix 2 is longer than helix 1, consisting of 16 residues. Tyr 104p OH forms the only hydrogen bond between this helix and the mature portion, to Pro 240 O.

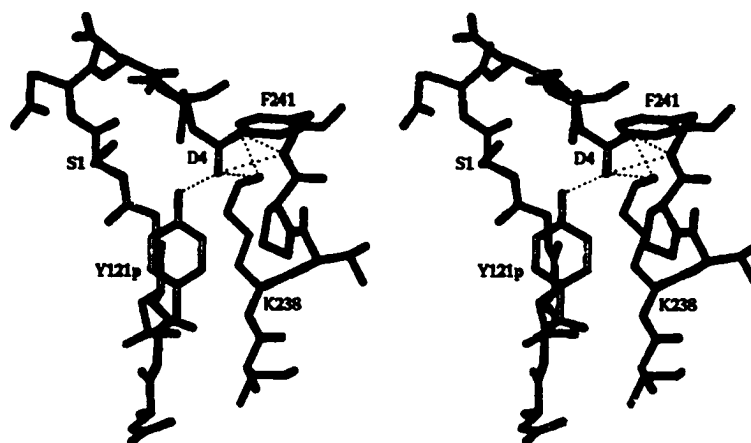
Otherwise, the two amphipathic helices pack with their hydrophobic faces against hydrophobic surfaces on the mature portion of the protein. The segment of polypeptide between helix 2 and the pro-mature junction is highly mobile in both molecules (Figures 4.2b, 4.3).

The connection to the mature segment of the molecule, the pro-mature junction (Gly 124p to Ser 1), is contained in a "Tyr-Asp" loop that is defined by the interaction between Tyr 121p and Asp 4 (Figure 4.6). The side chains of these residues form a hydrogen bond, which serves to keep this loop in its tight conformation. In addition, hydrogen bonds between Tyr 121p O and Lys 238 NH, between Asp 4 OD1 and Phe 241 NH, and between the Asp 4 carboxylate group and Lys 238 N $\zeta$  anchor the pro-mature junction to the C-domain of the protein.

The path of the first 14 residues of the mature sequence continues along the C-domain. This segment is predominantly in the random coil conformation, except for a single turn of  $3_{10}$  helix (Glu 7 to Val 11) (Figure 4.4). This helical turn resides at one end of the active site cleft in a position that corresponds approximately to the S4 substrate-binding site in the mature enzyme. The significance of this structural element is discussed in more detail below. Residues 13 to 17 form main chain hydrogen bonds to the active site Psi loops (Figure 4.7a). Asn 13 NH interacts with Gly 216 O, Met 15 O with Phe 33 NH, and Tyr 17 NH and O form hydrogen bonds with Leu 31 O and NH, respectively. After the type IV  $\beta$ -turn formed by Gly 23 to His 26, the first Psi loop begins. Subsequently, the secondary structure closely resembles that of PvPM.

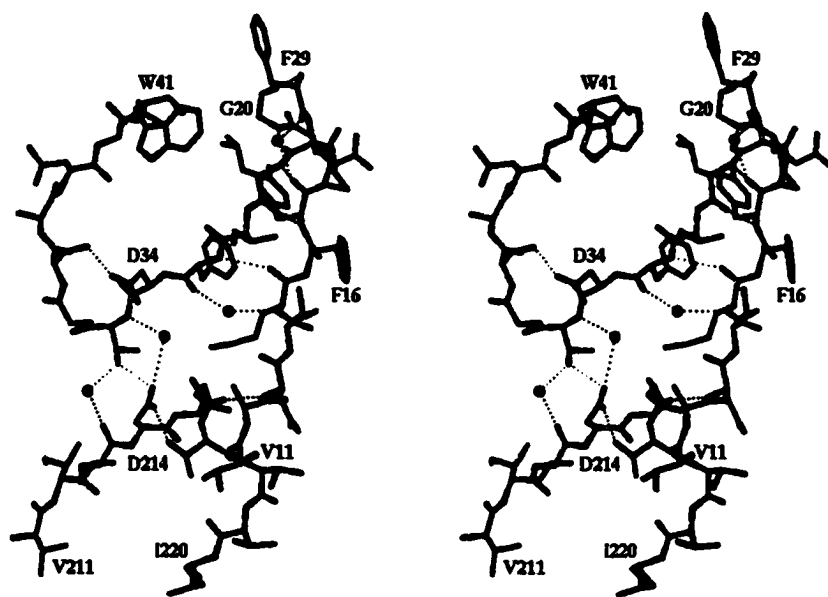
Residues Pro 88p to Val 14 wrap around the C-domain. An equivalent feature has been observed in *P. falciparum* proplasmepsin II, and has been called the "harness". As will be discussed in a later section, this "harness" plays a crucial role in keeping the zymogen inactive at neutral pH.

The flap is well defined in the electron density. This  $\beta$ -hairpin folds over the substrate-binding cleft, and Tyr 75 OH forms hydrogen bonds with Trp 41 N $\epsilon$ 1 and a water molecule, in a network that is structurally conserved among eukaryotic APs.



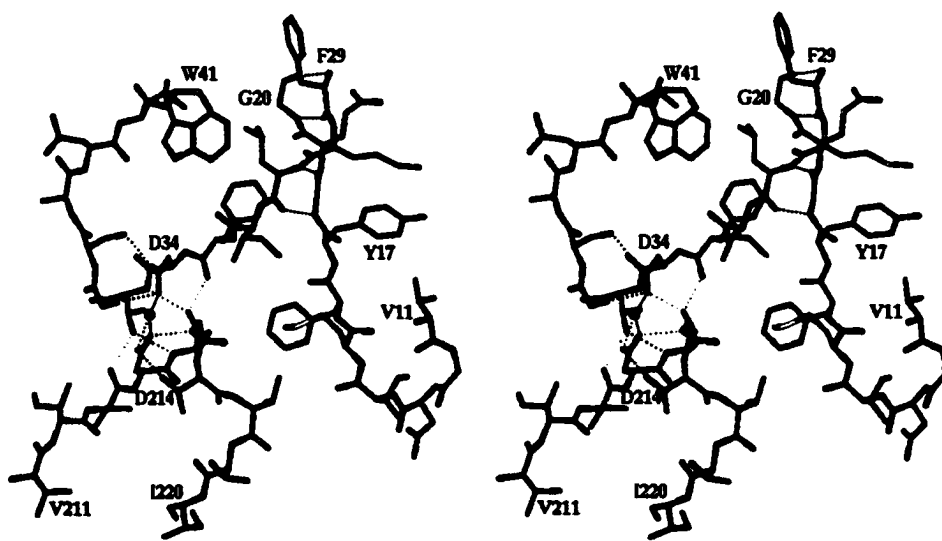
**Figure 4.6** The pro-mature junction of pPvPM. Residues I237 to F241 are shown with carbon atoms in cyan.

(a)



**Figure 4.7** The active site regions of pPvPM and PvPM. (a) The "immature" active site of pPvPM. The segment from V11 to F29, which forms hydrogen bonds with the Psi loops, is shown with carbons in cyan. (b) The active site of PvPM.

(b)



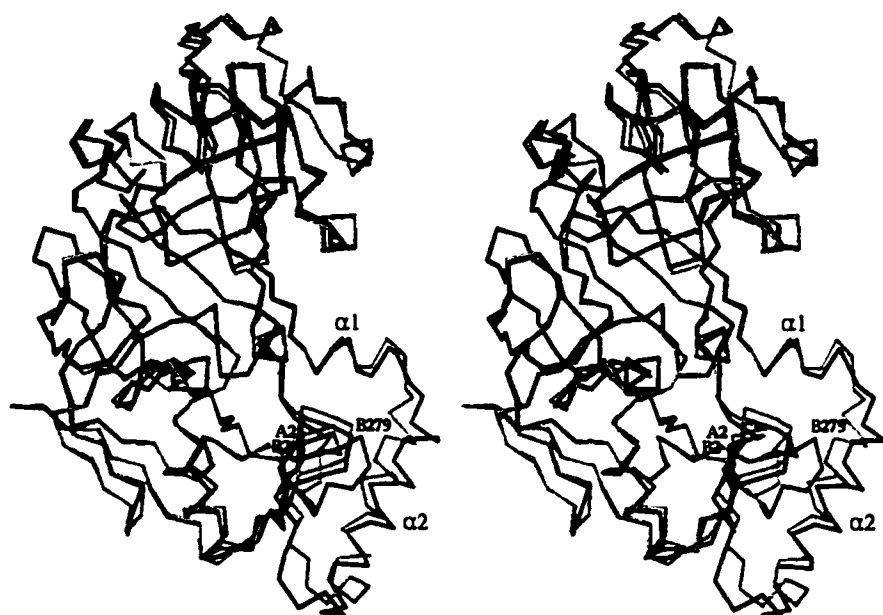


ProPvPM contains one *cis*-proline, residue 113. This residue begins an  $\alpha$ -helix, located near the flap in the N-domain and forms part of the S1 and S3 pockets in the substrate-binding cleft.

#### 4.3.3 Comparison of NCS-related molecules

The two molecules in the asymmetric unit are related by a rotation of approximately  $170^\circ$  and a negligible ( $0.3 \text{ \AA}$ ) translation along the rotation axis. The pPvPM dimer looks like a bent letter "S", where the monomers interact chiefly through the second helix of their prosegments (Figure 4.3a).  $1558 \text{ \AA}^2$  of accessible surface area is buried at this dimerization interface, including 40% ( $621 \text{ \AA}^2$ ) of hydrophobic surface. It is interesting to note that the most poorly ordered regions, the second helices of the prosegments and the C-domain of molecule B, are located near the center of this dimer. The disorder in these regions may be explained by the fact that the ends of the dimer participate in crystal contacts, but the central part of the dimer does not (Figure 4.3b). It is possible that slight differences in crystal packing at the ends of the dimer become projected to the center, where their effect is amplified.

Overall, the two molecules are quite similar in structure. The rmsd for the superposition of 373 atoms is  $1.26 \text{ \AA}$ . A superposition of only 270  $C\alpha$  atoms of residues that are within  $3\sigma$  of each other, results in an rmsd of  $0.39 \text{ \AA}$ . Most of the differences between the two molecules related by non-crystallographic are localized to the dimer interface, namely, prosegment helix 2 and the surrounding structure (Figure 4.8). Helix 2 of the prosegment has a slightly different orientation in the two molecules. The conformation of the coil segment that links this helix to the pro-mature junction is clearly different in the two molecules. Although this region is particularly poorly defined in molecule A, the electron density that is present is definitely inconsistent with the same conformation as in molecule B. The loop from Leu 278 to Asp 282 is located near the N-terminus of helix 2. The electron density associated with this loop is weak in both molecules, but once again, it is consistent with two different loop conformations, that are correlated with the displacement of helix 2. The segment from Asn 228 to Glu 253 shows a slight change in orientation between the NCS-related molecules. This is also likely



**Figure 4.8** Superposition of the two molecules in the pPvPM crystallographic asymmetric unit. Molecule A is in cyan, and molecule B is in dark blue. Prosegment  $\alpha$ -helices are labeled.

associated with the motion of helix 2, since this helix contacts the loop from Ile 237 to Pro 243.

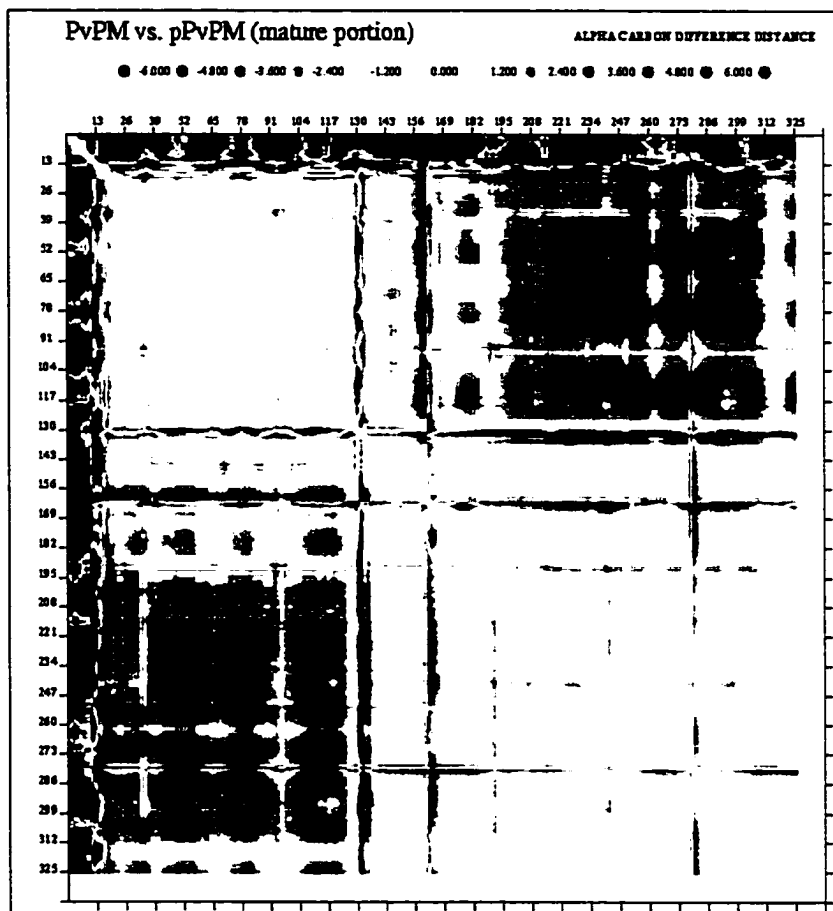
The loop from Ser 132 to Ser 135 changes conformation between the two molecules. This loop is distant from the dimer interface, but it is involved in different crystal contacts. In molecule A, Asp 250 of a symmetry-related molecule A is within 4 Å of Ile 133. Molecule B has no symmetry mates anywhere near this loop.

#### 4.3.4 Comparison to *P. vivax* plasmepsin

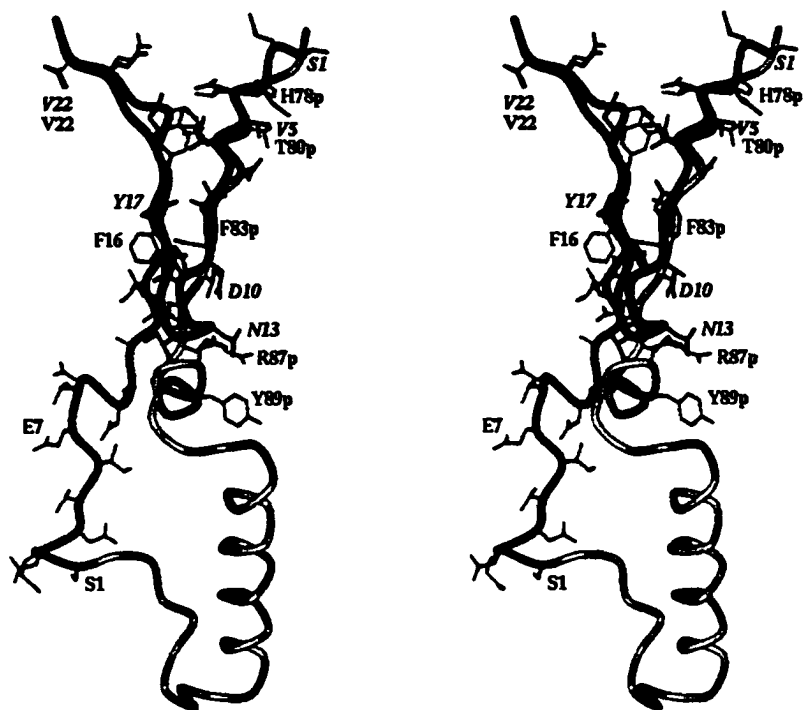
As in the case of pPvPMII, it is informative to compare the pPvPM crystal structure to the mature PvPM structure, which was presented in Chapter 2. From the difference distance matrix (Figure 4.9), two major conformational changes are apparent. The first 14 residues of the mature sequence undergo a substantial rearrangement, and the N-domain changes orientation between the zymogen and the mature enzyme. This is very similar to what was observed in the *P. falciparum* plasmepsin system, as will be discussed in more detail in the next section. In addition, three loops experience distinct conformational changes.

The first 14 residues of the mature plasmepsin sequence undergo a dramatic conformational change in the conversion of the zymogen to the active enzyme (Figure 4.10). Residues Met 15 to Glu 21 also experience a conformational change upon activation, albeit a less dramatic one than the preceding 14. In the zymogen, the mature N-terminus (Ser 1 to Ile 14) forms part of the "harness" that wraps around the C-domain. In this capacity, the N-terminus is packed along the C-domain, with the  $3_{10}$  helical turn formed by residues Glu 7 to Val 11 located at one end of the active site cleft, in the vicinity of the S4 substrate-binding site. In the mature enzyme, the N-terminus (Asp 4 to Leu 8) forms the first strand of the central six-stranded  $\beta$ -sheet. In the zymogen, the first strand of this  $\beta$ -sheet is formed by the N-terminus of the prosegment (residues Leu 81p to Arg 87p). Thus, upon activation, the mature N-terminus must refold from its random coil and  $3_{10}$  helical conformation into a  $\beta$ -strand. This segment must also relocate across the molecule, from the surface of the C-domain into the central  $\beta$ -sheet at the back of the active site cleft, with a maximum displacement of 43 Å.

In pPvPM, Ala 12 to Ile 14 are in an extended conformation, connecting the  $3_{10}$  helix (Ala 7 - Val 11) to the  $\beta$ -strand formed by Met 15 to Ala 22. In PvPM, residues



**Figure 4.9** DDM comparing the mature portion of pPvPM with PvPM. Blue indicates residues that are closer in PvPM, and red denotes residues that are farther apart in PvPM than in pPvPM.

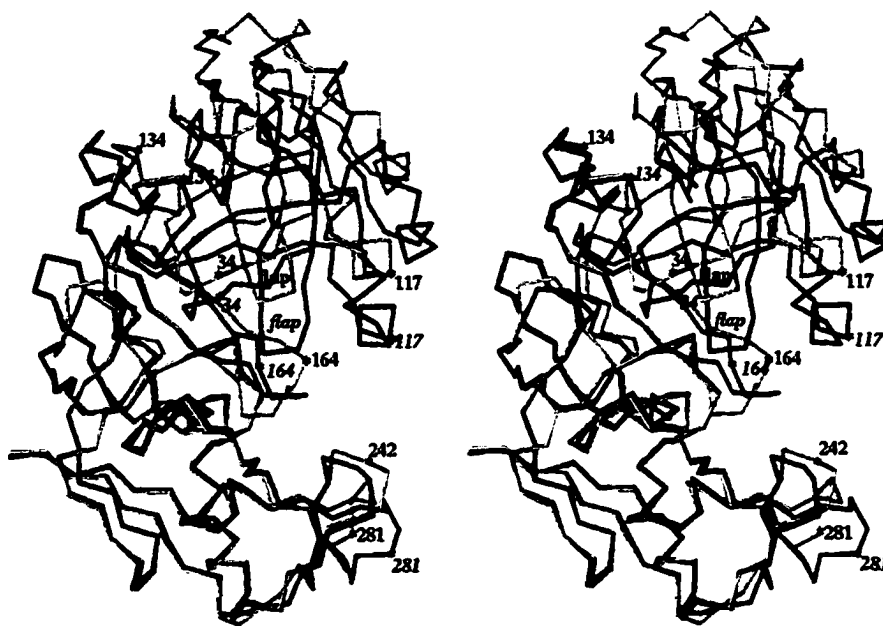


**Figure 4.10** The N-terminal rearrangement between pPvPM and PvPM. The prosegment (yellow) and residues 1 - 22 (green) of pPvPM are superimposed on the N-terminus, residues 1 - 22 (red), of PvPM. The side chains of residues 90p to 122p have been omitted for clarity. Labels in italics refer to PvPM.

11 - 14 form a type IV  $\beta$ -turn, connecting two strands (Asp 4 - Val 11 and Met 15 - Ala 22). The 14 - 22  $\beta$ -strand in the zymogen is not entirely equivalent to the 15 - 22 strand in the mature enzyme. In pPvPM, this strand contains a bulge at Gly 18 and Glu 19. Before the bulge, the two strands are out of register by one residue. The bulge brings the two strands into register at Gly 20. In pPvPM, seven main chain hydrogen bonds are made between Val 11 and Thr 218, Asn 13 and Gly 216, Met 15 and Phe 33, Tyr 17 and Leu 31, Gly 20 and Phe 29, and Val 22 and Gln 27. That is, the N-terminal segment forms hydrogen bonds to both active site Psi loops. In PvPM, there are six main chain hydrogen bonds in this region (between Gly 18 and Leu 31, Gly 21 and Phe 29, and Val 22 and Gln 27), but they all involve a single Psi loop. The change in association of the N-terminal residues with the Psi loops in the enzyme and zymogen is likely to be a contributing factor to the other major structural difference between these molecules, the domain shift.

Similar to the observation made with the *P. falciparum* plasmepsin II system, *P. vivax* plasmepsin and proplasmepsin exhibit a large ( $18^\circ$ ) rotation of the N-domain (Val 22 - Gly 125) relative to the central motif and C-domain (Asp 137 - Asn 327) (Figure 4.11). In pPvPM, the domains are farther apart, and the cleft is consequently wider than in PvPM. Focussing on the active site region, we see that the Psi loops are significantly farther apart in the zymogen than in the mature enzyme (Figures 4.7a, b). In pPvPM, the inner oxygens of Asps 34 and 214 are 6.3 Å apart, while in PvPM, they are separated by only 2.7 Å. This "immature active site" was also observed in pPfPMII (Figure 3.15b). Although the electron density in the active site cleft was not completely clear, it was possible to model several water molecules at the immature active site. The "fireman's grip" interactions surrounding the catalytic residues at an AP active site are absent, but the Psi loops are engaged in a different hydrogen bonding network. The interactions of the carboxylates of the catalytic Asps with the surrounding residues, are the same as in PvPM. In fact, the Psi loops themselves are essentially preformed in the zymogen, and are poised for collapse to the mature AP active site upon activation.

Gly 36 and Ser 37, at the tip of the N-domain Psi loop, are slightly (0.8 - 0.9 Å) displaced in the zymogen relative to the enzyme. The difference is due to the refolding of the 130 loop, since in plasmepsin Leu 131 points toward the Psi loop, while in pPvPM this loop points away.



**Figure 4.11** Superposition of residues 22 - 327 of pPvPM and PvPM. pPvPM (yellow) and PvPM (red) have been superimposed with their C-domains and central motifs. Labels in italics refer to PvPM.

On a local scale, three loops exhibit conformational changes between pPvPM and PvPM: Leu 126 to Asp 137, Leu 158 to Gly 166 and Asn 275 to Thr 283 (Figures 4.9, 4.11).

Loop 126 - 137 is at the junction of the rigid bodies in the domain shift, the N-domain and the central motif and C-domain. This loop is located at the edge of the active site cleft, and forms part of the S2' binding pocket. Refolding of this loop changes the shape of this pocket between the enzyme and zymogen. In the zymogen, Leu 131 points into the S2' pocket, and forming van der Waals contacts with Tyr 77 of the flap and with Ser 37. In PvPM, this loop points out of the S2' pocket. Due to the domain shift and refolding of this loop, the S2' pocket is wider and shallower in pPvPM than in PvPM. Also, the ends of this loop are in different positions in the zymogen and enzyme, due to the domain reorientation.

The rearrangement of loop 158 - 166 is related to the replacement of the first strand of the central  $\beta$ -sheet. As discussed above, in proplasmepsin, this strand is the prosegment  $\beta$ -strand, and in plasmepsin, this strand becomes replaced by the mature N-terminus. Residues 158 - 166 comprise the loop that connects the second and third strands of the central  $\beta$ -sheet, i.e., neighboring the replaced strand. A superposition of proplasmepsin and plasmepsin shows that, while most of the central  $\beta$ -sheet superimposes closely in the two structures, the first strands do not (Figure 4.11). The orientation of this strand in the zymogen leaves more room for the loop 158 - 166 in the plane of the sheet. The central  $\beta$ -sheet has a slightly greater twist in the zymogen.

Finally, in proplasmepsin the conformation of loop 275 - 283 is strongly influenced by its interaction with helix 2 of the prosegment (Figure 4.4). The removal of the prosegment during activation, allows this loop to adopt a different conformation in the mature enzyme.

#### **4.3.5 Comparison to *P. falciparum* proplasmepsin II**

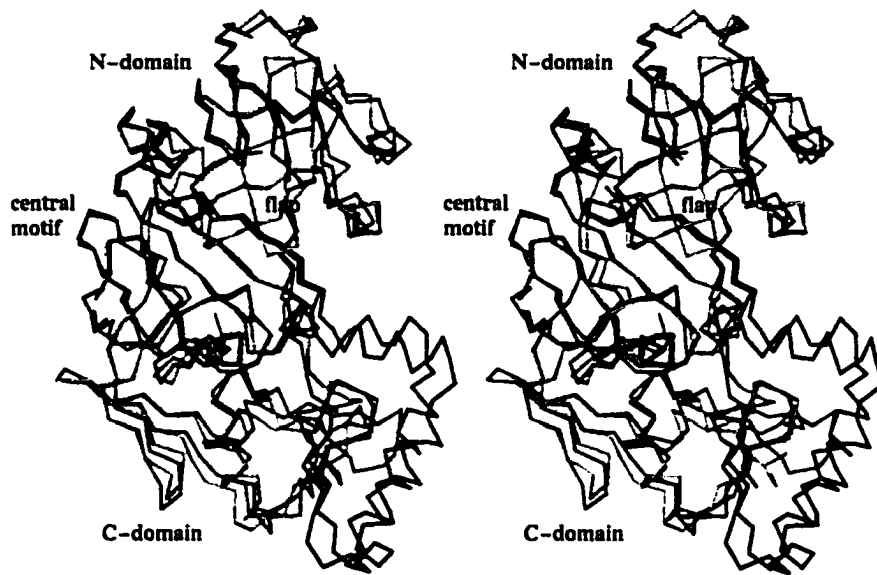
pPvPM is the second proplasmepsin to have its three-dimensional structure determined by X-ray crystallography. It is very interesting to compare this structure with the first, pPfPMII, to see which features can be generalized for the proplasmepsins.



Both proplasmepsins exhibit N-terminal refolding and a large N-domain rotation upon conversion to their corresponding mature enzymes (Figure 4.12). While dramatic N-terminal rearrangements have been seen in other AP zymogens, the gastric zymogens and prophytpsin [Sielecki *et al.*, 1991; Hartsuck *et al.*, 1992; Moore *et al.*, 1995; Bateman *et al.*, 1998; Kervinen *et al.*, 1999], the domain rotation appears to be limited to the plasmepsins. The observation of a domain shift comparable in magnitude and direction in both proplasmepsins confirms this to be a true structural feature that is common and, so far, unique to the plasmepsins. In both proplasmepsins, catalytic activity is prevented by the same means, the lack of a fully formed active site. At the "immature active site" of proplasmepsin, the catalytic machinery is approximately preformed on the Psi loops, but the loops are kept apart by the "harness" that consists of the prosegment and the mature N-terminus and impedes the formation of a functional active site.

The fold of the prosegment is similar, although not exactly the same in the two proplasmepsins (Figure 4.12). In both zymogens the prosegment consists of a  $\beta$ -strand, followed by two  $\alpha$ -helices and some random coil. In pPfPMII, both helices are similar in length, while in pPvPM, helix 1 is significantly shorter than helix 2. While the orientation of the helix-turn-helix unit changes slightly between the two structures, particularly at helix 2, their positions are still comparable. That is, helix 2 lies in a groove between the loops 237 - 245 and 275 - 283 in both structures. Tyr 104p, which is conserved in all the proplasmepsins, makes the only hydrogen bond from the second helix of the prosegment to the mature portion, to Pro 240 O. The pro-mature junction is contained in the "Tyr-Asp" loop in both proplasmepsins (Figures 3.9, 4.6). The disruption of this loop by low pH has been proposed to be an important step in the activation of plasmepsins.

The "harness" acts in the same way in both proplasmepsins to enforce the domain opening. Participation of the prosegment in the central  $\beta$ -sheet ensures a secure attachment to the rest of the protein. A conformationally constrained structure between the  $\beta$ -strand and helix 1 serves to orient the helices along the C-domain. The N-terminal 310 helix sits approximately at the S4 pocket of the substrate-binding cleft, potentially bumping into the flap in the domain-closed configuration. Finally, the hydrogen bonds shared by the N-terminal residues 13 - 22 with both Psi loops, may assist in keeping the domains apart.



**Figure 4.12** Superposition of pPvPM and pPfPMII. The mature sequences of pPvPM and pPfPMII are in yellow and green, respectively. The prosegments of pPvPM and pPfPMII are in magenta and dark blue, respectively.

Although the structures are, on the whole, very similar, there are some interesting contrasts between them. In the N-domain, most of the differences are centered around the flap, which behaves differently in the two proplasmepsins (Figure 4.12). In pPfPMII, the tip of the flap (Asn 76 to Thr 81) is disordered, and is missing from the model in all four NCS-related molecules. In pPvPM, on the other hand, the flap is perfectly well ordered, and clearly defined in the electron density map. The conformations of the flap are not influenced by crystal contacts in either structure. In Chapter 2 it was proposed that the flap is intrinsically more flexible in *P. falciparum* plasmepsin II than in *P. vivax* plasmepsin. The zymogen structures support this idea. The disorder in the flap of pPfPMII affects the surrounding structural elements. Loop 129 - 137 borders the flap on the Sn' side. In pPvPM, this loop folds toward the cleft, pointing Leu 131 into the S2' pocket. In pPfPMII, this loop points away from the cleft, leaving a deeper pocket. The segment from 108 to 113 borders the flap on the Sn side, and the loop 48 - 52 flanks the former. All these loops differ in conformation between the two zymogens.

The disordered flap in pPfPMII exerts an effect on the "immature active site". Since Tyr 77 is disordered, it does not form a hydrogen bond with Trp 41. Trp 41 therefore adopts a different conformation, where it engages in a hydrogen bond with Ser 37. In this orientation, Ser 37 cannot form a hydrogen bond with Asp 34 O $\delta$ 1. The normal hydrogen bonding pattern must be generated during activation. In pPvPM, the flap is ordered, and the mature hydrogen bonding network is in place prior to activation.

In the C-domain, there is a slight rigid body movement of the AP flexible subdomain (226 - 258 and 268 - 298). In addition, the loop 278 - 284 has a distinct conformational change due to interactions with the N-terminus of prosegment helix 2, whose orientation varies in the two zymogens. The stretch from 294 to 296 is generally known in APs as the "poly-proline loop" [Dhanaraj *et al.*, 1992]. In pPfPMII this loop contains two prolines, but pPvPM has no prolines in this loop. The Pro residues likely enforce a different conformation in pPfPMII than in pPvPM.

#### 4.4 Conclusion

The crystal structure of *P. vivax* proplasmepsin has revealed the same mode of inactivation as was observed previously in *P. falciparum* proplasmepsin II. The

observation of many of the same structural features in pPvPM as in pPfPMII. specifically, refolding of the mature N-terminus and a major reorientation of the N-domain relative to the mature enzyme, confirms that these are real and not an artifact of crystallization. In the proplasmepsins, proteolytic activity is prevented by a novel mechanism, one that has not been seen in any other AP zymogen.

In the other AP zymogens of known structure, pepsinogen, progastricsin and prophytepsin, activity is inhibited by a blockage of a pre-formed active site. Thus, in these molecules most of the mature proteinase is fully assembled, and the active site must be somehow blocked in order to maintain inactivity. Removing the occluding moiety from the active site releases the proteolytic activity.

Proplasmepsins are one step behind the gastric zymogens in assembly. That is, the two halves of the active site, the Psi loops, are pre-formed, but their relative orientation is incompatible with catalytic activity. The "immature" active sites of the proplasmepsins do not need to be occluded, since they do not have any inherent activity. Activity is generated by bringing the Psi loops together.

The principle of blocking a pre-formed active site is likely to apply to the inactivation of many AP zymogens, since a large number of them possess the conserved Lys-Tyr pair in their prosegments (gastric zymogens, cathepsins D), or the conserved Lys-Asn-Tyr sequence at the mature N-terminus (plant zymogens, cathepsins D). Many other AP zymogens, however, do not contain either of these motifs, and may well use completely different modes of inactivation. The proplasmepsin structures have revealed one new way of preventing proteolysis. Further biochemical and structural characterization of AP zymogens will show if the variation in sequence translates into different mechanisms of zymogen inactivation.

The crystal structures of the proplasmepsins and of prophytepsin have also provided new templates for modelling of AP zymogens. However, the novel and unexpected aspects of these structures suggest that there may still be many surprises waiting to be discovered regarding AP zymogens, and that structure prediction for these proteins still promises to be a challenging endeavor.

## 4.5 References

- Bateman KS, Chernaiia MM, Tarasova NI, James MNG. (1998) Crystal structure of human pepsinogen A. *Adv Exp Med Biol* **436**: 259-63.
- Bruce-Chwatt LJ (1985) *Essential malariology*. John Wiley & Sons, New York.
- Brunger AT (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**: 472-475.
- Brunger AT (1993) *XPLOR: a system for X-ray crystallography and NMR* (Yale University Press, New Haven, Connecticut).
- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**: 905-21.
- Esnouf RM (1997) An extensively modified version of MolScript that includes greatly enhanced colouring capabilities. *J Mol Graphics* **15**: 133-138.
- Hartsuck JA, Koelsch G, Remington SJ (1992) The high-resolution crystal structure of porcine pepsinogen. *Proteins* **13**: 1-25.
- Jones TA, Kjeldgaard M (1995) *O the Manual, Version 5.11* (Uppsala, Sweden).
- Kervinen J, Tobin GJ, Costa J, Waugh DS, Wlodawer A, Zdanov A (1999) Crystal structure of plant aspartic proteinase prophytepsin: inactivation and vacuolar targeting. *EMBO J* **18**: 3947-55.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**: 283-291.
- Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* **33**: 491-497.
- Merritt EA, Murphy MEP (1994) Raster3D version 2.0: a program for photorealistic molecular graphics. *Acta Cryst D* **50**: 869-873.

Moore SA, Sielecki AR, Chernaia MM, Tarasova NI, James MNG (1995) Crystal and molecular structures of human progastricsin at 1.62 Å resolution. *J Mol Biol* **247**: 466-85.

Navaza J (1994) AMoRe: an automated package for molecular replacement. *Acta Cryst*, **A50**: 157-163.

Otwinowski Z, Minor W (1996) Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology*, 276; (Carter Jr CW, Sweet R, eds., Academic Press, New York and London) pp. 307-326.

Pannu NS, Read RJ (1996) Improved structure refinement through maximum likelihood. *Acta Cryst* **A52**: 659-668.

Sielecki AR, Fujinaga M, Read RJ, James MNG (1991) Refined structure of porcine pepsinogen at 1.8 Å resolution. *J Mol Biol* **219**: 671-92.

## Chapter 5: LOOKING BEYOND THE STRUCTURES

The two preceding chapters have described the three-dimensional structures of two novel AP zymogens, proplasmepsin II from *P. falciparum* and proplasmepsin from *P. vivax*. The unusual features of these molecules, such as their global contortion and their incompletely formed active sites were discussed in detail. This chapter explores the potential usefulness of proplasmepsins as targets for novel anti-malarial drugs and presents some ideas about the parts of these zymogens that are absent from the structures, the signal anchor sequence and the cytoplasmic portion of proplasmepsin.

### 5.1 Proplasmepsins as drug design targets

In certain APs, the isolated prosegment, or portions of the prosegment, can act as inhibitors [Dunn *et al.*, 1983; Fusek *et al.*, 1991; Richards *et al.*, 1992]. This type of inhibition had not been observed with the plasmepsins, and the crystal structures explain why. The prosegment in proplasmepsins does not block a functional, fully formed active site, but rather it acts together with the mature N-terminus as a harness that distorts the molecule, preventing the active site from forming. Clearly, the prosegment alone or portions thereof could not achieve this type of inactivation of the enzyme. Thus, unfortunately, the prosegment itself is not a lead for inhibitor design. The structures are encouraging, however. The distortion of the mature portion of the molecule alters the shape of the active site cleft, and effectively presents a new site for potential inhibitor binding. If one could design a molecule that will bind to the wide cleft of proplasmepsin, it could act as a wedge that would prevent the domains from coming together, thereby inhibiting activation. One advantage of this strategy is that this type of inhibitor should be automatically selective for proplasmepsin, and should not cross-react with any of the human aspartic proteinases or zymogens, if they do not have such a distorted active site cleft. The gastric zymogens do not have a distorted cleft, and, judging from the sequence of its prosegment, cathepsin D probably does not either. Whether this is also the case for the zymogens of renin, cathepsin E, the napsins [Tatnell *et al.*, 1998] or the memapsins [Lin *et al.*, 2000], remains to be determined.

Two novel pockets have been identified in the active site cleft of pPfPMII. One is a hydrophobic pocket that extends the S1/S3 substrate-binding site. This pocket results from the unusual orientation of Trp 41 in pPfPMII. Normally in the APs and their zymogens, this pocket is filled by the indole moiety of Trp 41. This conformation of the tryptophan is stabilized by a hydrogen bond between its Nε1 and OH of Tyr 77 (Tyr 75 in pepsin numbering) on the tip of the flap. In pPfPMII, the tip of the flap, including Tyr 77, is disordered. The bond between Tyr 77 and Trp 41 is not formed, and consequently, the Trp 41 side chain rotates, leaving an empty hydrophobic pocket. This pocket appears promising for drug design, because an analogous pocket has been used in the design of a renin inhibitor [Oefner *et al.*, 1999; Vieira *et al.*, 1999]. The new hydrophobic pocket was filled by an *o*-chlorobenzoate group of the inhibitor. In that case, the re-orientation of Trp 41 (plasmepsin numbering) and a compensating displacement of the tip of the flap were induced by inhibitor binding. In pPfPMII, this pocket is already formed, so binding an inhibitor to it should be entropically more favorable than in renin.

The second new pocket is located at the bottom of the S2' site, and results from a different conformation of the 131 - 135 loop than in the mature enzyme. This pocket is quite narrow, but could potentially accommodate a small group from the inhibitor.

Industrial drug design programs usually begin with high throughput screening of vast libraries of chemical compounds. Molecules that give promising results are then chosen as lead compounds and subjected to optimization. In an academic setting, we do not have the capability to carry out such a massive experiment.

Another potential concept to approach this problem, may be to design a molecule *de novo* by screening a database of small molecules with a molecular docking computer program that should find molecules complementary to the proposed binding site on the protein [Selzer *et al.*, 1997; Tondi *et al.*, 1999]. For instance, the new ligand can exploit two of the novel binding sites: the "immature" active site and the Trp 41 hydrophobic pocket. The "immature" active site, with its unusually disposed Psi loops, should provide selectivity for the ligand. Since this is a novel structure, however, we do not know what type of molecules might bind to it. A database docking experiment was performed, using the program DOCKVISION [T. Hart & S. Ness, University of Alberta] to dock the Available Chemicals Directory to the "immature" active site. The predominant molecules among the hits were triazoles and tetrazoles, five-membered aromatic rings containing



three or four nitrogen atoms. For the hydrophobic pocket, *o*-chlorobenzoate, the same moiety that was found to be optimal in the renin inhibitor, was used. *o*-chlorobenzoate was docked as a methyl ester to the hydrophobic pocket using DOCKVISION, running in the RESEARCH mode [Hart *et al.*, 1997]. The best docking result for both groups is shown in figure 5.1.

The next step is to test if the selected molecules actually bind to the protein. If binding assays give a positive result, crystal structures of the zymogen in complex with the small molecules will be needed to elucidate the precise orientation of the ligands on the protein. Based on this information, the two ligands can be linked, preferably with a conformationally constrained linker [Khan *et al.*, 1998], with the aim of producing a much tighter binding inhibitor.

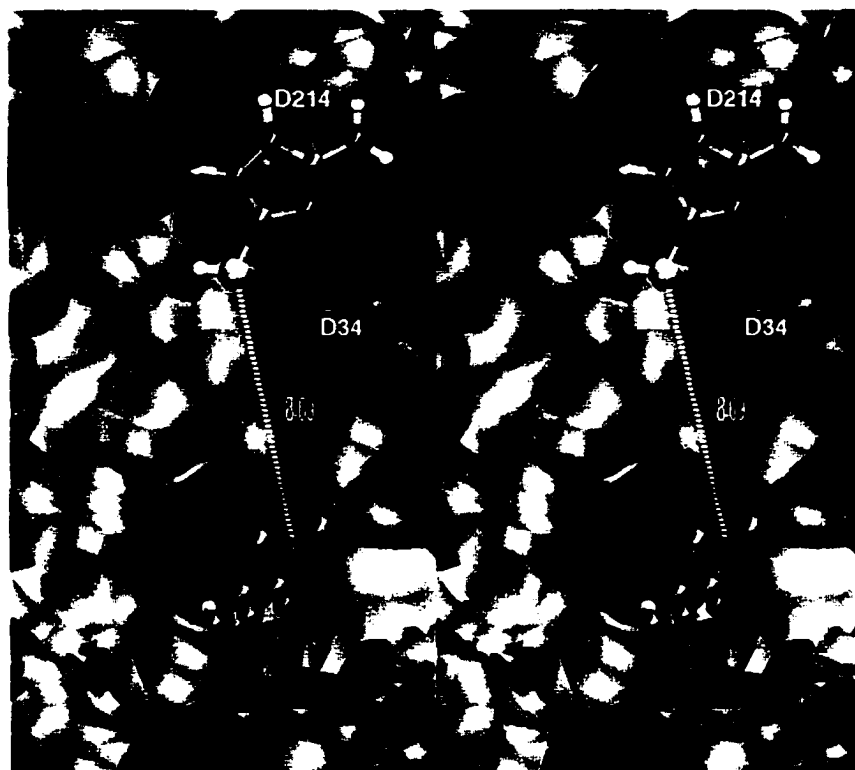
Due to its unusual active site configuration and two new pockets, pPfPMII appears to be a promising candidate for the design of selective inhibitors that could interfere with its conversion to the mature enzyme. pPvPM lacks the hydrophobic Trp 41 pocket and may prove a more elusive target for inhibitor design. However, an ortholog of PvPM has been found in *P. falciparum* (PfPMIV), and it is quite possible that an ortholog of pPfPMII exists in *P. vivax*. If this proves to be the true, then inhibitors of pPfPMII activation have the potential to be developed into drugs that will be effective against all species of the human malaria parasite.

## 5.2 Membrane attachment

The archetypal APs and zymogens are soluble proteins. However, several more recently discovered zymogens and mature enzymes utilize various methods of membrane attachment. The membrane-binding portion of the protein may be at the N-terminus, in the middle or at the C-terminus (Figure 5.2).

The unusually long prosegments of the proplasmepsins harbor a signal anchor sequence. The full-length zymogens are type II membrane proteins, but the mature plasmepsins become soluble when the prosegment is removed during activation.

Plant AP zymogens contain the plant-specific insert (PSI), a conserved domain that folds into a five-helix bundle that mediates membrane attachment of the zymogen. The PSI is removed during conversion to active enzyme, producing a soluble AP [Ramalho-Santos *et al.*, 1998; White *et al.*, 1999]. The three-dimensional structure of the PSI of



**Figure 5.1** Docking results for pPfPMII. The molecular surface of pPfPMII is shown, colored by atom type (carbon = cyan, oxygen = red, nitrogen = blue, sulfur = yellow). The two ligands (methyl *o*-chlorobenzoate and 3-amino-5-mercapto-1,2,4-triazole, protonated) are shown in ball-and-stick representation. (Hydrogen atoms are colored white and chlorine is purple.) This figure was produced with the program VMD [Humphrey *et al.*, 1996].

archetypal gastric AP zymogen



proplasmepsin



*Eimeria acervulina* AP zymogen



memapsin precursor



plant AP zymogen



**Figure 5.2** Schematic representation of membrane attachment in various AP zymogens (tm = trans-membrane helix; PSI = plant-specific insert; GPI = glycosylphosphatidylinositol).

prophytepsin revealed a structural similarity with NK-lysin, an anti-bacterial protein that acts by forming pores in the bacterial membrane [Kervinen *et al.*, 1999; Ruyschaert *et al.*, 1998]. PSI and NK-lysin also share significant sequence similarity with saposins, proteins that promote lysosomal degradation of sphingolipids [Vaccaro *et al.*, 1999]. The three-dimensional structure of PSI suggests that the mode of action of this domain may be similar to that of the exchangeable apolipoproteins apolipoprotein III and apolipoprotein E, which reversibly associate with lipoprotein particles and play important roles in lipoprotein metabolism. These proteins are helix bundles in solution, that open to reveal a hydrophobic surface in the presence of lipid [Narayanaswami & Ryan, 2000].

Yapsins 1, 2 and 3 and the *Eimeria* spp. aspartic proteinase possess C-terminal extensions that anchor these enzymes to the membrane *via* a glycosylphosphatidylinositol (GPI) moiety [Ash *et al.*, 1995; Cawley *et al.*, 1995; Komano & Fuller, 1995; Laurent *et al.*, 1993]. The newly-identified memapsins (membrane-anchored aspartic proteinases of the pepsin family) have C-terminal extensions of more than 75 residues, which contain a putative trans-membrane helix [Lin *et al.*, 2000].

A common theme in membrane attachment appears to be the trafficking of the APs, as membrane-bound zymogens, to the proper location for activity. The potentially harmful activity of aspartic proteinases is controlled by numerous means, such as production of inactive zymogens and the requirement of low pH for catalysis. Binding to the membrane provides yet another level of regulation of AP activity.

### **5.3 Do the proplasmepsins assist in their own trafficking?**

The crystal structures of the proplasmepsins reveal a great deal of information, but only about the C-terminal 48 amino acids of the prosegment and the mature portion. The segment connecting this portion of the zymogen to the signal anchor sequence has been predicted to form an  $\alpha$ -helix (see section 3.3.8). However, no structural or functional information is available on the portion of the prosegment that precedes the signal anchor sequence.

A search of the non-redundant protein sequence databases using the BLAST algorithm [Altschul *et al.*, 1990] has identified sequence similarity (12/24 identity and 17/24 similarity) between residues Asp 2p and Asn 25p of the prosegment and residues 167 - 190 of the yeast coatomer  $\delta$  subunit,  $\delta$ -COP (Swiss-Prot accession number P43621)

(Figure 5.3). Coatomer (COPI) is a protein complex that controls the transport of proteins between cellular compartments in eukaryotic organisms. COPI-coated vesicles carry proteins between the ER and the Golgi apparatus, between the Golgi cisternae and may play a role in endocytosis [Lowe & Kreis, 1998].  $\delta$ -COP is an essential protein and is one of the seven components of COPI [Cosson *et al.*, 1996; Lowe & Kreis, 1996; Faulstich *et al.*, 1996]. It is tempting to speculate that the 38 N-terminal residues of the proplasmepsin prosegment fold into a domain that protrudes into the parasite's cytoplasm and somehow controls the trafficking of proplasmepsin through the secretory pathway. Since no structure-function studies of  $\delta$ -COP have been reported, the role for residues 167 - 190 of  $\delta$ -COP is currently unknown. A BLAST search of the non-redundant protein sequence database showed that residues 1 - 150 and 300 - 550 are well conserved among  $\delta$ -COPs from different organisms. The segment between residues 150 and 300, appears to be specific for the yeast  $\delta$ -COP. This suggests that residues 167 - 190 may not form a functionally important structure. Further analysis of the coatomer proteins will be needed to resolve this issue.

Conversely, investigation of the structure and function of the N-terminal (cytoplasmic) part of the proplasmepsin prosegment will be interesting. One possible approach is a structural characterization of the cytoplasmic segment, which can be either chemically synthesized or expressed. The small size of the cytoplasmic region (38 amino acids) should make this peptide amenable to structure by NMR. This strategy should be complemented by the continuing progress on the *P. falciparum* genome project [Dame *et al.*, 1996], which will show if more coatomer-related proteins are present in the parasite.

pPfPMII: 2 DITVREHDFKHGFIKSNSTFDGLN 25  
+I +EH+ KHGF+ SN +DG N  
 $\delta$ -COP: 167 E I A R K E H E R K H G F M S S N G D Y D G A N 190

**Figure 5.3** Sequence similarity of the cytoplasmic portion of pPfPMII and yeast  $\delta$ -COP.

## 5.4 References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-10.
- Ash J, Dominguez M, Bergeron JJ, Thomas DY, Bourbonnais Y (1995) The yeast proprotein convertase encoded by YAP3 is a glycosylphosphatidylinositol-anchored protein that localizes to the plasma membrane. *J Biol Chem* **270**: 20847-54.
- Cawley NX, Wong M, Pu LP, Tam W, Loh YP (1995) Secretion of yeast aspartic protease 3 is regulated by its carboxy-terminal tail: characterization of secreted YAP3p. *Biochemistry* **34**: 7430-7.
- Cosson P, Demolliere C, Hennecke S, Duden R, Letourneur F (1996) Delta- and zeta-COP, two coatamer subunits homologous to clathrin-associated proteins, are involved in ER retrieval. *EMBO J* **15**: 1792-8.
- Dame JB, Arnot DE, Bourke PF, Chakrabarti D, Christodoulou Z, Coppel RL *et al* (1996) Current status of the Plasmodium falciparum genome project. *Mol Biochem Parasitol* **79**: 1-12.
- Dunn BM, Lewitt M, Pham C (1983) Inhibition of pepsin by analogues of pepsinogen-(1-12)-peptide with substitutions in the 4-7 sequence region. *Biochem J* **209**: 355-62.
- Faulstich D, Auerbach S, Orci L, Ravazzola M, Wegchinger S, Lottspeich F, Stenbeck G, Harter C, Wieland FT, Tschochner H (1996) Architecture of coatamer: molecular characterization of delta-COP and protein interactions within the complex. *J Cell Biol* **135**: 53-61 *FEBS Lett* **454**: 210-4.
- Fusek M, Vetvicka V (1995) *Aspartic proteinases: physiology and pathology*. (CRC Press Inc., Boca Raton).
- Hart TN, Ness SR, Read RJ (1997) Critical evaluation of the research docking program for the CASP2 challenge. *Proteins Suppl* **1**:205-9.
- Humphrey W, Dalke A, Schulten K (1996) VMD – Visual Molecular Dynamics. *J Molec Graphics* **14**: 33-38.

- Khan AR, Parrish JC, Fraser ME, Smith WW, Bartlett PA, James MNG (1998) Lowering the entropic barrier for binding conformationally flexible inhibitors to enzymes. *Biochemistry* **37**: 16839-45.
- Komano H, Fuller RS (1995) Shared functions in vivo of a glycosyl-phosphatidylinositol-linked aspartyl protease, Mkc7, and the proprotein processing protease Kex2 in yeast. *Proc Natl Acad Sci U S A* **92**: 10752-6.
- Laurent F, Bourdieu C, Kaga M, Chilmonczyk S, Zgrzebski G, Yvone P, Pery P (1993) Cloning and characterization of an *Eimeria acervulina* sporozoite gene homologous to aspartyl proteinases. *Mol Biochem Parasitol* **62**: 303-12.
- Lin X, Koelsch G, Wu S, Downs D, Dashti A, Tang J (2000) Human aspartic protease memapsin 2 cleaves the beta -secretase site of beta -amyloid precursor protein. *Proc Natl Acad Sci U S A* **97**: 1456-1460.
- Lowe M, Kreis TE (1996) In vivo assembly of coatamer, the COP-I coat precursor. *J Biol Chem* **271**: 30725-30.
- Lowe M, Kreis TE (1998) Regulation of membrane traffic in animal cells by COPI. *Biochim Biophys Acta* **1404**: 53-66.
- Narayanaswami V, Ryan RO (2000) Molecular basis of exchangeable apolipoprotein function. *Biochim Biophys Acta* **1483**: 15-36.
- Oefner C, Binggeli A, Breu V, Bur D, Clozel JP, D'Arcy A, Dorn A, Fischli W, Gruninger F, Guller R, Hirth G, Marki H, Mathews S, M ller M, Ridley RG, Stadler H, Vieira E, Wilhelm M, Winkler F, Wostl W (1999) Renin inhibition by substituted piperidines: a novel paradigm for the inhibition of monomeric aspartic proteinases? *Chem Biol* **6**: 127-31.
- Ramalho-Santos M, Verissimo P, Cortes L, Samyn B, Van Beeumen J, Pires E, Faro C (1998) Identification and proteolytic processing of procardosin A. *Eur J Biochem* **255**: 133-8.
- Richards AD, Kay J, Dunn BM, Bessant CM, Charlton PA (1992) Inhibition of aspartic proteinases by synthetic peptides derived from the propart region of human prorenin. *Int J Biochem* **24**: 297-301.



Ruyschaert JM, Goormaghtigh E, Homble F, Andersson M, Liepinsh E, Otting G (1998) Lipid membrane binding of NK-lysin. *FEBS Lett* **425**: 341-4.

Selzer PM, Chen X, Chan VJ, Cheng M, Kenyon GL, Kuntz ID, Sakanari JA, Cohen FE, McKerrow JH (1997) *Leishmania major*: molecular modeling of cysteine proteases and prediction of new nonpeptide inhibitors. *Exp Parasitol* **87**: 212-21.

Tatnell PJ, Powell DJ, Hill J, Smith TS, Tew DG, Kay J (1998) Napsins: new human aspartic proteinases. Distinction between two closely related genes. *FEBS Lett* **441**: 43-8.

Tondi D, Slomczynska U, Costi MP, Watterson DM, Ghelli S, Shoichet BK (1999) Structure-based discovery and in-parallel optimization of novel competitive inhibitors of thymidylate synthase. *Chem Biol* **6**: 319-31.

Vaccaro AM, Salvioli R, Tatti M, Ciaffoni F (1999) Saposins and their interaction with lipids. *Neurochem Res* **24**: 307-14.

Vieira E, Binggeli A, Breu V, Bur D, Fischli W, Guller R, Hirth G, Marki HP, Muller M, Oefner C, Scalone M, Stadler H, Wilhelm M, Wostl W (1999) Substituted piperidines--highly potent renin inhibitors due to induced fit adaptation of the active site. *Bioorg Med Chem Lett* **9**: 1397-402.

White PC, Cordeiro MC, Arnold D, Brodelius PE, Kay J (1999) Processing, activity, and inhibition of recombinant cyprosin, an aspartic proteinase from cardoon (*Cynara cardunculus*). *J Biol Chem* **274**: 16685-93.