Novel Uncomplexed and Complexed Structures of Plasmepsin II, an Aspartic Protease from Plasmodium falciparum

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · APRIL 2003

Impact Factor: 4.33 · DOI: 10.1016/S0022-2836(03)00036-6 · Source: PubMed

doi:10.1016/S0022-2836(03)00036-6 J. Mol. Biol. (2003) 327, 173–181

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Novel Uncomplexed and Complexed Structures of Plasmepsin II, an Aspartic Protease from *Plasmodium falciparum*

Oluwatoyin A. Asojo[1], Sergei V. Gulnik[1], Elena Afonina[1], Betty Yu[1] Jonathan A. Ellman[2], Tasir S. Haque[2] and Abelardo M. Silva[1]*

[1]*Structural Biochemistry Program, National Cancer Institute/SAIC, Frederick, MD 21702, USA*

[2]*B-84 Hildebrand Hall Department of Chemistry University of California Berkeley, Berkeley, CA 94720 USA*

**Corresponding author*

Malaria remains a human disease of global significance and a major cause of high infant mortality in endemic nations. Parasites of the genus Plasmodium cause the disease by degrading human hemoglobin as a source of amino acids for their growth and maturation. Hemoglobin degradation is initiated by aspartic proteases, termed plasmepsins, with a cleavage at the α-chain between residues Phe33 and Leu34. Plasmepsin II is one of the four catalytically active plasmepsins that has been identified in the food vacuole of *Plasmodium falciparum*. Novel crystal structures of uncomplexed plasmepsin II as well as the complex with a potent inhibitor have been refined with data extending to resolution limits of 1.9 Å and 2.7 Å, and to *R* factors of 17% and 18%, respectively. The inhibitor, *N*-(3-{(2-benzo[1,3]dioxol-5-yl-ethyl)[3-(1-methyl-3-oxo-1,3-dihydro-isoindol-2-yl)-propionyl]-amino}-1-benzyl-2-(hydroxypropyl)-4-benzyloxy-3,5-dimethoxy-benzamide, belongs to a family of potent non-peptidic inhibitors that have large P1′ groups. Such inhibitors could not be modeled into the binding cavity of the structure of plasmepsin II in complex with pepstatin A. Our structures reveal that the binding cavities of the new complex and uncomplexed plasmepsin II are considerably more open than that of the pepstatin A complex, allowing for larger heterocyclic groups in the P1′, P2′ and P2 positions. Both complexed and uncomplexed plasmepsin II crystallized in space group *P*2, with one monomer in the asymmetric unit. The structures show extensive interlocking of monomers around the crystallographic axis of symmetry, with areas in excess of 2300 Å^2 buried at the interface, and a loop of one monomer interacting with the binding cavity of the 2-fold related monomer. Electron density for this loop is only fully ordered in the complexed structure.

Keywords: crystal structure; aspartic protease; plasmepsin; *Plasmodium falciparum*; malaria

Introduction

Different species of Plasmodium are the etiologic agents of malaria, a devastating human disease afflicting several hundred million people a year and killing an estimated two million of them, primarily children.[1] The most severe and common forms of malaria are caused by *Plasmodium falciparum* and *Plasmodium vivax*. These parasites undergo a complex life cycle in both the human host and the mosquito vector. The parasites cause disease during the intra-erythrocytic phase and consume nearly all the hemoglobin of the human host to generate amino acids for growth and maturation.[2]

Present addresses: O. A. Asojo, Macromolecular Crystallography Laboratory, National Cancer Institute, Room 144 Building 539 Fort Detrick, Frederick, MD 21702, USA; B. Yu, S. V. Gulnik, E. Afonina and A. M. Silva, Sequoia Pharmaceuticals Inc., 401 Professional Drive, Gaithersburg, MD 20879, USA.

Abbreviations used: Plm I, II and IV, plasmepsin I, II and IV, respectively; HAP, histidine aspartic protease; proPlm, proplasmepsin; Plm, plasmepsin; PDB, Protein Data Bank; EH58, *N*-(3-{(2-benzo[1,3]dioxol-5-yl-ethyl)-[3-(1-methyl-3-oxo-1,3-dihydro-isoindol-2-yl)-propionyl]-amino}-1-benzyl-2-(hydroxypropyl)-4-benzyloxy-3,5-dimethoxy-benzamide; PEG, polyethylene glycol.

E-mail addresses of the corresponding authors: asojo@ncifcrf.gov; abelardo.silva@sequoiapharma.com

0022-2836/03/$ - see front matter

Hemoglobin is degraded by a series of proteases in an acidic digestive vacuole of the parasite. The aspartic proteases found in the vacuole, named plasmepsin, make an initial attack on the hemoglobin molecule, followed by proteolysis of the large fragments into small peptides by a cysteine protease named falcipain.[3,4] The initial cleavage by plasmepsin is critical to hemoglobin degradation because intact hemoglobin cannot be cleaved by falcipain unless it is first denatured.[5] The products of digestive vacuole proteolysis are small peptides, which are then exported for terminal degradation by cytoplasmic exopeptidases.

Ten different plasmepsin molecules have been identified in the *P. falciparum* genome and four of them are active in the food vacuole during the intra-erythrocytic stage. The active enzymes are plasmepsin I (Plm I), plasmepsin II (Plm II), plasmepsin IV (Plm IV), and histidine aspartic protease (HAP), an aspartic protease with a histidine amino acid residue in the active site.[6] HAP, Plm II, Plm IV and Plm I are capable of making an initial cleavage in the hemoglobin α-chain between residues Phe33 and Leu34.[3,4,6] This bond is also the first cleaved when hemoglobin is incubated with an extract of purified digestive vacuoles [4]. It is believed that this cleavage results in the unraveling of the hemoglobin molecule, allowing further proteolysis to proceed efficiently. The plasmepsin molecules are also capable of several other cleavages after the initial event.[3]

Plm I and Plm II are 73% identical to each other while their mature coding regions share about 64% identity with Plm IV, 60% identity with HAP but only about 30–35% identity with mammalian aspartic proteases such as human cathepsin D.[7] The plasmepsin pro-pieces are substantially longer than their mammalian counterparts, having 124–125 amino acid residues instead of 40–50, and have a membrane-spanning domain.[7] The structures of Plm II from *P. falciparum* in complex with pepstatin A,[8] Plm from *P. vivax* in complex with pepstatin A (RCSB Protein Data Bank accession code 1QS8), pro-Plm II from *P. falciparum*,[9] Plm II in complex with non-peptidic Phe-Leu core inhibitors,[10] and Plm IV (Asojo *et al.*, unpublished results; PDB code 1LS5) have been determined by X-ray crystallographic techniques.

Inhibitors of Plm II with nanomolar K_i values have been identified, and some of them are lethal against cultured malarial parasites.[4,8,11–17] Using combinatory chemistry methods, several other non-peptidic compounds with low nM K_i values against Plm II were identified and some of them are more selective towards Plm II than human cathepsin D. One of these inhibitors is *N*-(3-{(2-benzo[1,3]dioxol-5-yl-ethyl)[3-(1-methyl-3-oxo-1,3-dihydro-isoindol-2-yl)-propionyl]-amino}-1-benzyl-2-(hydroxy-propyl)-4-benzyloxy-3,5-dimethoxy-benzamide (EH58), which has an inhibition constant of 100 nM against Plm II.[13] Attempts at modeling EH58 and like inhibitors into the binding cavity of the structure of Plm II–pepstatin A were largely unsuccessful, as the bulky inhibitors could not fit into the highly constrained binding cavity of the enzyme.[13] To understand the conformational flexibility of the binding cavity, it was thus important to determine the structure of Plm II in the presence of a representative member of the EH58 class of inhibitors as well as in the absence of any inhibitor. We therefore present here the structure of the uncomplexed form of plasmepsin II from *P. falciparum* as well as the structure of its complex with the non-peptidic inhibitor EH58.

Results and Discussion

Crystallization and model building

Complexed and uncomplexed forms of aspartic proteases often crystallize under similar conditions, and their structures do not show many differences. For instance, the uncomplexed form of human cathepsin D and its pepstatin A complex were crystallized with ammonium sulfate as the precipitant. Both structures are very similar except for the position of a β-hairpin structure, known as *flap*, which makes contact with the inhibitor on top of the active site.[18] Also, all published structures of the plasmepsins in complex with pepstatin A were grown using ammonium sulfate as the precipitant.[8,9] In spite of those results, we were unable to crystallize the uncomplexed form of Plm II in ammonium sulfate. We used polyethylene glycol (PEG) 4000 as precipitant and a lower pH. The crystals obtained diffract beyond a resolution limit of 1.9 Å and belong to the space group *P*2, with one plasmepsin molecule in the crystallographic asymmetric unit. *P*2 is a space group rarely found in protein crystals, in fact as of 13 May 2002 there were only seven structures belonging to space group *P*2 out of a total 14,758 X-ray diffraction structures deposited at the RCSB Protein Data Bank.[19]

Most of the residues of the uncomplexed enzyme were easily built into a $2F_o - F_c$ electron density map computed out of a molecular replacement solution. However, residues of a loop in close proximity to the crystallographic 2-fold axis of symmetry and extending from Val236 through Tyr245, known as the *flexible loop*, were disordered. The conformational flexibility of such a loop of the C-terminal domain was already noticed in the structure of the Plm II–pepstatin A complex[8] and in the cross-comparison of Plm II and human aspartic proteases.[11] In fact, this loop region displays the largest conformational variability in eukaryotic aspartic proteases.[11] Co-crystals of Plm II with the inhibitor EH58 also grew with PEG as precipitant. The Plm II–EH58 co-crystals also belong to the space group *P*2, and have similar cell constants to those of uncomplexed Plm II. The electron density for all residues of the protein, including those of the flexible loop Val236 through

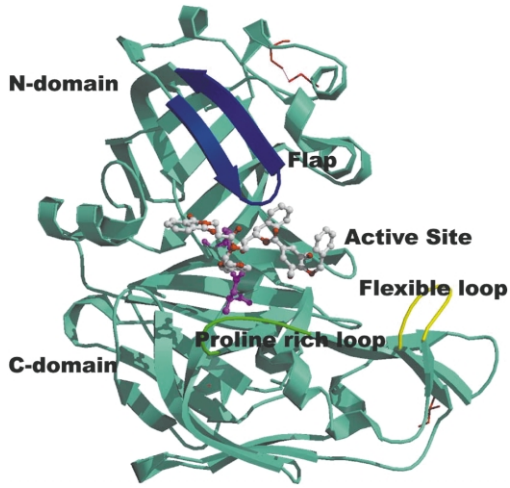


Figure 1. Ribbon diagram of the complex of plasmepsin II with EH58 showing disulfide bridges, catalytic dyad, inhibitor, flap, flexible loop and proline-rich loop in red, magenta, white, blue, yellow and green, respectively.

Tyr245, as well as the inhibitor was well defined in the $2F_o - F_c$ map computed out of a molecular replacement solution.

General features of the models

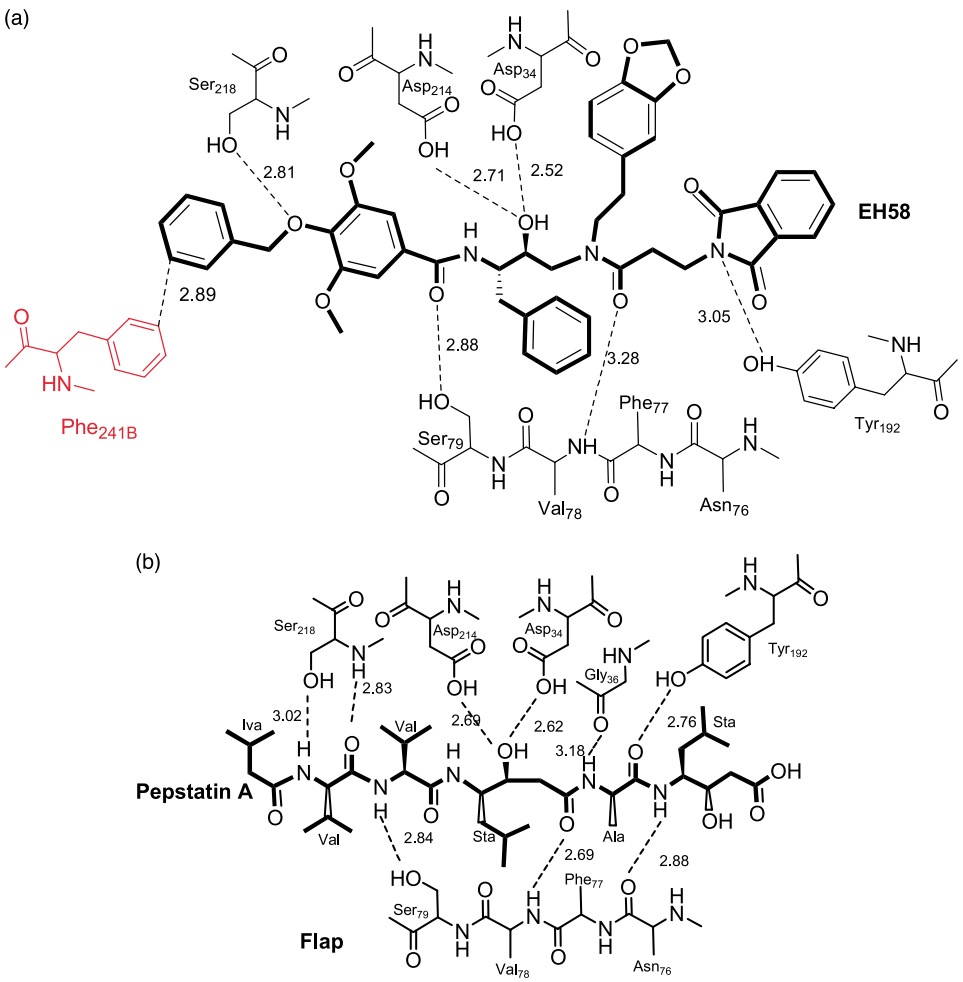
Plm II displays the typical fold of eukaryotic aspartic proteases that has been observed from mammalian enzymes such as human pepsin, renin, gastricsin and cathepsin D, to fungal enzymes like endothiapepsin and rhizopuspepsin.[20] The main secondary structure features of Plm II are retained regardless of which inhibitor is bound. The mature enzyme is formed by a single chain of 329 amino acid residues and is folded into two topologically similar N and C-terminal domains. The domains contact each other along the bottom of the binding cleft that contains the catalytic dyad, Asp34 and Asp214. The N-terminal domain includes a distinctive single β-hairpin structure, known as *flap*, which lies perpendicular over the binding cleft interacting with bound inhibitors and presumably substrates. The amino and carboxyl ends of the polypeptide chain are assembled into a characteristic six-stranded inter-domain β-sheet, which serves to anchor the domains together (Figure 1). The structures of uncomplexed Plm II and the Plm II–EH58 complex are very similar to each other. After superposition of the structures, the rms difference between corresponding C^{α} atoms is 0.45 Å. The major structural differences are in the flexible loop, which is disordered in the absence of an inhibitor, and in regions that are in contact with the binding cavity. Interestingly, these two new structures display a larger conformational difference with respect to the structure of the Plm II–pepstatin A complex.[8] The rms differences between C^{α} atoms of each independent Plm II–pepstatin A monomer and the uncomplexed structure are 1.00 Å and 1.17 Å. The comparison of each of the pepstatin A complex monomers with the EH38 complex yields rms differences that equal 1.32 Å and 1.49 Å. In these cases the main conformational differences are also at the binding cavity and the flexible loop, and the origin of those differences will become evident in the following sections.

Active site and inhibition

The P1–P1′ core of EH58 is a transition-state analog of the hydrated from of the peptide bond. The hydroxyl group of the core is at the characteristic hydrogen bond distances from carboxylic acid oxygen atoms of the catalytic dyad Asp214 and Asp34, of 2.71 Å and 2.52 Å, respectively (Scheme 1). Equivalent interaction in the Plm II–pepstatin A complex show distances of 2.69 Å and 2.62 Å, respectively. At the location of the core hydroxyl group, the structure of the uncomplexed enzyme shows the distinctive catalytic water molecule, which initiates bond cleavage by attack on the carbonyl carbon atom of the scissile bond. The catalytic water molecule is at 2.84 Å and 2.75 Å from the carboxylic acid oxygen atoms the catalytic dyad. The *flap* residues Val78 and Ser79 form hydrogen bonds with EH58 *via* the main-chain nitrogen atom of the former at 3.29 Å and the side-chain hydroxyl group of the latter at 2.88 Å bond lengths. Also, the hydroxyl groups of Ser218 and Tyr192 form hydrogen bonds with the inhibitor, at 2.91 Å and 3.05 Å, respectively. It is interesting to note that pepstatin A and EH58, both potent inhibitors of Plm II, have similar hydrogen bond interactions with the enzyme. In fact, all polar groups of the enzyme involved in hydrogen bonds with EH58 were also involved in hydrogen bonds with pepstatin A.

When the uncomplexed and EH58 complexed forms of Plm II are compared, it becomes evident that residues in the binding cavity shift away to accommodate the bulky groups of the inhibitor, while others close in to form hydrogen bonds and other contacts (Figure 2). Most noticeably side-chains of the *flap* and the proline-rich loop, Ile290-Pro297, and in particular residues Val78, Ile290, Leu292 and Phe294 undergo the bulk of the relative changes from the uncomplexed to complexed structures. There are not only relative changes in side change positions but also movements of the main-chain.

Overall, the binding cavity of the uncomplexed Plm II structure is more open than in the complex with EH58. The relative opening of the binding cavity can be measured by the distance between C^{α} Val78, at the tip of the *flap*, to C^{α} Leu292 lying opposite to Val78 at the hydrophobic rim of the binding cavity. The distance between these atoms is 12.6 Å in the uncomplexed structure and 12.0 Å in the EH58 complex. However, the conformational flexibility of the binding cavity becomes more evident when measuring that distance in the



Scheme 1. Interactions of Plm II with (a) EH58 and (b) pepstatin A, showing hydrogen bonds of distances less than 3.5 Å. A close contact of EH58 with Phe241 from the flexible loop of the 2-fold related monomer is also shown in (a).

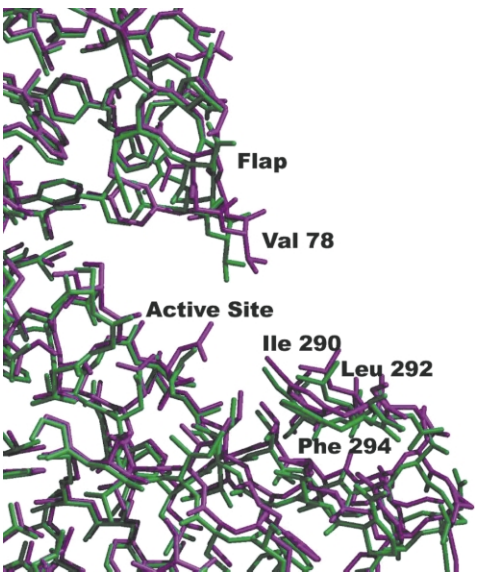


Figure 2. Overlay of the binding cavities of uncomplexed Plm II (aquamarine) and the complex with EH58 (magenta) revealing a shift in residues from the flap and the proline rich loop, Ile290–Pro297. The residues Val78, Ile290, Leu292 and Phe294 undergo the bulk of the relative changes from the uncomplexed to complexed structures.

Plm II–pepstatin A complex, which equals 9.9 Å. Such a large conformational change of the binding cavity illustrates how Plm II can accommodate the large heterocyclic group in the P1′ position of EH58. This could not be predicted merely on the basis of the pepstatin A complex structure (Figure 3).

Rather than referring to the "opening" of the binding cavity we should consider that the likely effect is the closing of the binding cavity to embrace the inhibitors. Indeed, the uncomplexed form of the enzyme displays the most open conformation, with subsites that can readily accept bulkier, P1′, P2′ and P2 groups than predicted from the Plm II–pepstatin A complex structure. Thus, we can envision that the binding cavity closes down for a tighter embrace of the less bulky and also more potent inhibitor pepstatin A. This tight embrace is evidenced by the collapse of the entire binding cavity, and most notably the S1′ subsite of the binding cavity in the structure of the Plm II–pepstatin A complex. The molecular surfaces shown in Figure 4 illustrate how tight the embrace is in the presence of each inhibitor. The binding cavity remains slightly open in the structure of the

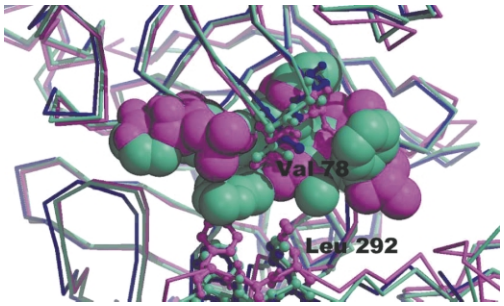


Figure 3. Overlay of the binding cavity of the monomers of uncomplexed Plm II (blue), and complexes of Plm II with EH58 (aquamarine) and pepstatin A (magenta), showing both inhibitors superimposed in CPK. The opening of the binding cavity is measured by the distance between C^{α} Val78 at the tip of the flap and C^{α} Leu292 lying opposite to Val78 at the hydrophobic rim of the binding cavity. This opening is over 4 Å wider in uncomplexed Plm II and the Plm II–EH58 complex than in the Plm II–pepstatin A complex.

complex with EH58, while in the pepstatin A complex structure the enzyme completely wraps around the inhibitor.

Dimerization

The monoclinic crystal form of Plm II reveals very well packed pairs of molecules with an extensive interlocking of monomers around the crystallographic 2-fold axis of symmetry (Figure 5(a)). According to Janin[21] molecular contacts induced by crystal packing bury, on average, an area of 500 Å^2 with an upper limit of 1300 Å^2. Therefore, crystal contacts burying an area in excess of 1300 Å^2 might indicate aggregation of monomers prior to crystallization. The area buried by 2-fold symmetry-related monomers of Plm II is 2351 Å^2 for the uncomplexed enzyme and 2580 Å^2 for the complex. In both cases the proportions of polar and hydrophobic buried areas are about 20% and 80%, respectively. It is thus possible that dimerization occurred prior to crystal formation, which in turn could help to explain crystallization in such an uncommon space group as *P*2. In addition, the flexible loop from each monomer docks into the binding cavity of the 2-fold related monomer, and is in close proximity to the inhibitor. Thus, the interaction with the inhibitor might help to stabilize the conformation of the flexible loop in the complex.

The Plm II–pepstatin A complex crystallized in space group $P3_121$ with two molecules in the crystallographic asymmetric unit.[8] A re-examination of that structure revealed, upon proper symmetry transformation, that the independent molecules form a dimer with approximate 2-fold non-crystallographic symmetry. The dimer is in fact tightly packed with a buried surface area of 2519 Å^2 and having 20% and 80% polar, and hydrophobic buried areas, respectively. While at first view this dimer form seems very different from that found in the monoclinic crystal form, it is remarkable that the flexible loop, Val236 through Tyr245, from one monomer also extends into the binding cavity of the other monomer. Moreover, we also observed a third dimeric form in the structure of Plm II in complex with a different class of non-peptidic inhibitors[10] which crystallized in space group *I*222. The 222 symmetry generates a tetramer and out of the three possible dimeric interfaces two of them have significant buried areas. The largest dimer interface, with a total buried surface area of 4427 Å^2, has the flexible loop of each monomer docking into the binding cavity of the 2-fold related monomer (Figure 5(c)).

In addition, an examination of the structures of the pepstatin A complex of plasmepsin from *P. vivax* (1QS8) (Figure 5(d)), Plm IV from *P. falciparum* (1LS5), (Figure 5(e)) also reveal dimers with extensive buried surface areas. The complexes with pepstatin A have similar tertiary and quaternary structures as Plm II with pepstatin A, and in both cases the flexible loop of one monomer docks into the binding cavity of the other one. Thus, in all known Plm

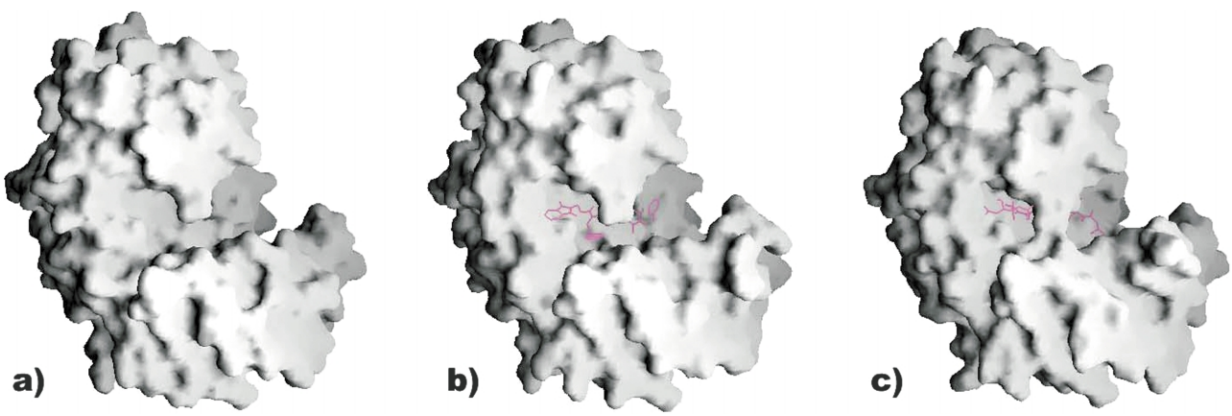


Figure 4. Surface representations of a monomer of (a) uncomplexed Plm II and Plm II in complex with (b) EH58 and (c) pepstatin A, with the binding cavity of the complex with pepstatin A revealing a much tighter embrace of the inhibitor.

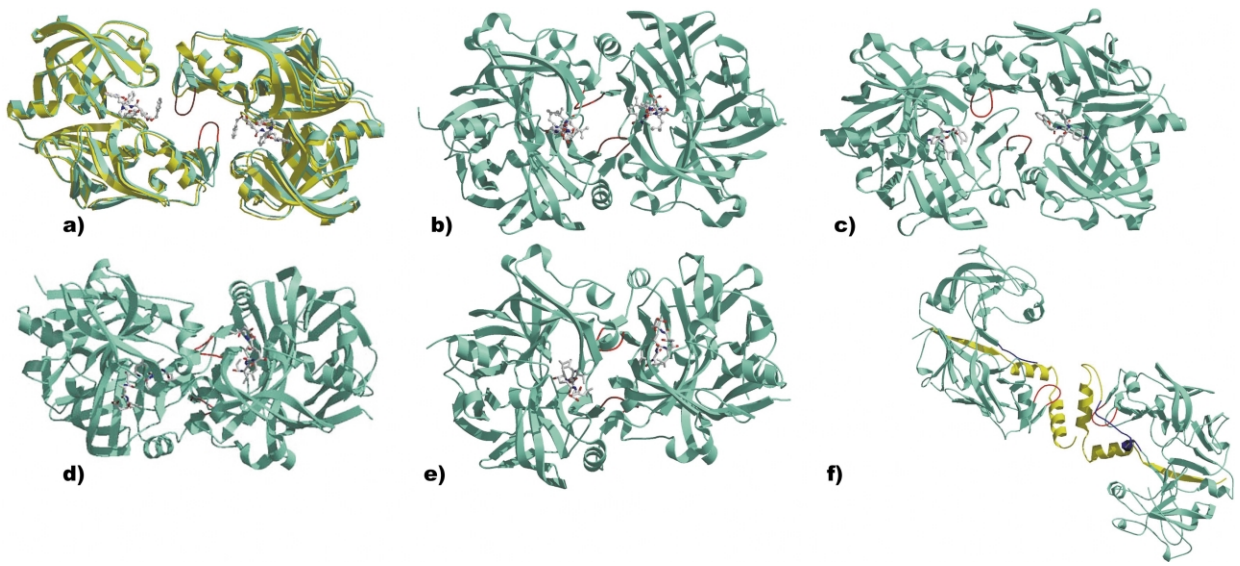


Figure 5. (a) Ribbon diagram of the alignment of dimers of uncomplexed Plm II (magenta) and the Plm II–EH58 complex (aquamarine), showing two very similar dimers. (b) Ribbon diagram of the dimers of the complex structures of Plm II with pepstatin A, and (c) with inhibitor rs367. The flexible loop of each monomer, shown in red, docks into the binding cavity of the other monomer and this interaction appears to mediate the formation of all three dimer forms. (d) A similar dimer is also observed in the complex structures with pepstatin A of both Plm from *P. vivax* and (e) Plm IV from *P. falciparum*. (f) The dimer of proPlm II of *P. falciparum* could be converted to the dimer of the active enzyme (a) after cleavage of the pro-parts (yellow) by a simple relative translation of the monomers. The translation will interlock the flexible loops (red) and re-orientation of the monomers is not required.

structures the same sides of the molecules come together to form similar, though not identical, 2-fold dimers that have in common the flexible loop acting as a hinge to mediate the quaternary structures.

We have also observed a similar dimer form mediated by the flexible loops in the crystal structure of proplasmepsin II (proPlm II). Proplm II is the inactive precursor of Plm II, including 124 additional residues forming the pro-segment, which is auto-catalytically cleaved off during activation of the enzyme. The structure of proplasmepsin II[9] shows that the pro-segment, disrupts the active conformation of the catalytic dyad, Asp34 and Asp214, through a major reorientation of the two domains of the mature protease. Residues of the pro-segment block the active site rendering the active site inaccessible to substrate. Despite this difference between the structures of the zymogen and the mature enzyme, the cleavage of pro-plasmepsin at Gly124p followed by a 12 Å translation of one monomer towards the other monomer, results in a dimer conformation similar to that of the monoclinic structures of mature Plm II (Figure 5(f)).

Concluding Remarks

The unexpected finding of a 2-fold dimeric association of monomers with a common structural hinge in all plasmepsin structures so far determined brings naturally the question of its biological role and significance. It has been proposed that plasmepsin molecules play an important role in hemoglobin catabolism inside the acidic vacuole of the parasite, an organelle with similarities to the lysosome. The characteristic features of such organelles are acidic pH, around 5, and very high proteolytic enzyme concentrations. For example, the concentration of cathepsin D inside the lysosome is as high as 0.7 mM. Some of the crystallization conditions used in this study may therefore resemble physiological conditions for Plm II. Moreover, the existence of dimeric structures for several different plasmepsins of *P. falciparum* and other parasite species may suggest their biological relevance.

Most of the differences between the structures described here and the structure of the Plm II–pepstatin A complex arise from the remarkable conformational flexibility of the binding cavity of the enzyme. An obvious model is that of an enzyme with a fully open binding site in the uncomplexed state, followed by closing of the cavity upon complex formation. Such adaptability to different inhibitors, though difficult to predict from our original structure of the pepstatin A complex, would likely be of benefit for the cleavage of hemoglobin at multiple sites. Whether the conformational flexibility makes this enzyme a better target for drug design or not is, no doubt, an arguable matter. However, our finding certainly poses a more complex problem than that of designing inhibitors against a rather rigid enzyme structure.

Materials and Methods

The expression and purification of plasmepsin II has been described.[8,22] Protein from both the wild-type and the Ser205 mutant were used for crystallographic experiments. The surface-exposed Met205 is an autolysis site, and was substituted with Ser, which is tolerated less in the hydrophobic S1 subsite of Plm II, thus reducing the chances of protein autolysis during purification. The mutant enzyme retained its catalytic activity.[22] All crystals were grown at 20 °C by vapor diffusion in hanging drops. Drops were prepared by mixing 2 μl of either, 5 mg/ml protein in 10 mM Tris–HCl (pH 7.0), or protein-inhibitor solution, with an equal volume of the appropriate reservoir solution. The protein-inhibitor solution contained 5 mg/ml protein in 10 mM Tris–HCl (pH 7.0) and 7 mM inhibitor in 10% (v/v) dimethyl sulfoxide. All data sets were collected at beam-line X9B at the Brookhaven National Synchrotron Light Source using a MAR 345 detector (Mar Research).

The reservoir solution for the crystallization of the uncomplexed form of Plm II contained 0.1 M sodium citrate (pH 5.6), 20% (w/v) PEG 4000 and 10% (v/v) butanol. Plate-like crystals which were 0.3 mm long on the largest dimension and less than 0.1 mm thickness were obtained after several days. The crystals belong to space group $P2$, with cell constants $a = 52.7$ Å, $b = 39.6$ Å, $c = 90.6$ Å, $\alpha = 90.0°$, $\beta = 105.6°$ and $\gamma = 90.0°$ and have one monomer in the crystallographic asymmetric unit. A crystal was transferred into a cryoprotecting mother liquor containing 0.1 M sodium citrate buffer (pH 5.6), 10% butanol, 20% PEG 400 and 10% PEG 4000 prior to flash freezing in a stream of N_2 gas at 90 K yielding a 99% complete X-ray diffraction data set extending to 1.9 Å.

The reservoir solution for the crystallization of the complex of Plm II with the inhibitor EH58 contained 0.1 M citric acid (pH 4.5), 0.3 M ammonium acetate and 25% PEG 3350. Plate crystals, of space group $P2$, dimensions 0.3 mm × 0.2 mm × 0.1 mm, with cell constants $a = 53.1$ Å, $b = 39.9$ Å, $c = 91.3$ Å, $\alpha = 90.0°$, $\beta = 104.9°$, $\gamma = 90.0°$, and $Z = 1$, grew within several days. All attempts to freeze these crystals led to diffraction patterns with very large mosaicity values. As a last resort a crystal was mounted in a capillary and a 90% complete monoclinic data set extending to 2.7 Å was collected at room temperature.

The complex with pepstatin A was crystallized from a reservoir solution containing 0.10 M sodium citrate buffer (pH 6.5) and 50% (w/v) ammonium sulfate. Rod-shaped crystals of the space group $P3_121$, with cell constants $a = b = 142.3$ Å and $c = 97.7$ Å, were obtained. A crystal was transferred to artificial mother liquor containing 25% (v/v) glycerol prior to flash freezing, to give an 80% complete X-ray diffraction data set extending to 2.4 Å.

All X-ray diffraction data sets were processed using the programs DENZO[23] and SCALEPACK.[24] The structures were solved by molecular replacement using the program AMoRe.[25] The search model was obtained by stripping a monomer from the 2.7 Å structure of plasmepsin II complexed with pepstatin A of its ligand and water moleculess.[8] Iterative cycles of model building with the program O[26] and structure refinement were carried out. The models were refined at low resolution with X-PLOR using positional refinement followed by group b-factor refinement of protein atoms and individual b-factor refinement for ligand and water molecules.[27–31] The model for both the uncomplexed structure and the structure of the complex with pepstatin A were further refined with CNS using a maximum-likelihood refinement procedure using Engh & Huber geometric parameters.[27–31] Data and model statistics are shown in Table 1.

Table 1. Statistics for data collection and model refinement

Data	Plm II	Plm II-EH58
Space group	$P2$	$P2$
Resolution (Å)	20–1.9 (2–1.9)	20–2.7 (2.9–2.7)
R_{merge} (%)[a]	4.2 (33.3)	6.69 (32.2)
Completeness (%)	99.0 (98.1)	93.0 (87.1)
Redundancy	7.0 (3.1)	3.0 (3.0)
$I/\sigma(I)$	13.0 (2.1)	12.1 (2.0)
# Unique reflections	28457	18608
Refinement [$I > 2\sigma(I)$]		
Resolution (Å)	20–1.9 (2–1.9)	20–2.7 (2.9–2.7)
R-factor (%)[b]	17.0 (27.0)	17.8 (27.6)
R-free[c]	22.8 (31.0)	22.4 (31.8)
Geometry		
Rms deviation		
Bond length (Å)	0.09	0.018
Bond angles (deg.)	2.1	2.5
Improper angles (deg.)	1.9	1.9
Dihedral angles (deg.)	32.2	28.0
Model information		
Mutant	Ser205	Wild
Number of monomers	1	1
Number of residues	331	331
Number of inhibitors	0	1
Number of waters	282	130

Numbers in parentheses apply to the highest resolution shell.
[a] $R_{merge} = (\sum|I - \langle I\rangle|)/\sum I$, where I is the observed intensity, and ⟨I⟩ is the average intensity obained from multiple observations of symmetry-related reflections after rejections.
[b] R-factor $= \sum||F_o| - |F_c||/\sum|F_o|$ where F_o are observed and F_c are calculated structure factors.
[c] R-free set uses 5% of randomly chosen reflections defined in Ref. 30.

The superposition of models and rms deviations were calculated using the program LSQKAB[32] within the CCP4 package (Collaborative Computational Project, Number 4, 1994). All Figures were generated using Raster3D,[33] MOLSCRIPT[34] and BOBSCRIPT[35] except surfaces that were generated using SPOCK.[36]

Accession numbers

The atomic coordinates have been deposited with the RCSB Protein Data Bank with accession codes 1LF4, and 1LF3, for the uncomplexed structure of plasmepsin II, and the complex structure of plasmepsin II with the inhibitors EH58.

Acknowledgements

We thank Dr J. W. Erickson for his support and encouragement. This research was supported in part by the National Cancer Institute under contract NOl CO-74102. The contents of this publication do not necessarily reflect the views or

policies of the DHHS, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

References

1. WHO (1995). *Control of Tropical Diseases: Malaria Control*, WHO Office of Information, Geneva, Switzerland.
2. Francis, S. E., Sullivan, D. J., Jr & Goldberg, D. E. (1997). Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum*. *Annu. Rev. Microbiol.* **51**, 97–123.
3. Gluzman, I. Y., Francis, S. E., Oksman, A., Smith, C. E., Duffin, K. L. & Goldberg, D. E. (1994). Order and specificity of the *Plasmodium falciparum* hemoglobin degradation pathway. *J. Clin. Invest.* **93**, 1602–1608.
4. Glodberg, D. E., Slater, A. F., Beavis, R., Chait, B., Cerami, A. & Henderson, G. (1991). Hemoglobin degradation in the human malaria pathogen *Plasmodium falciparum*: a catabolic pathway initiated by a specific aspartic protease. *J. Exp. Med.* **173**, 961–969.
5. Francis, S. E., Gluzman, I. Y., Oksman, A., Banerjee, D. & Goldberg, D. E. (1996). Characterization of native falcipain, an enzyme involved in *Plasmodium falciparum* hemoglobin degradation. *Mol. Biochem. Parasitol.* **83**, 189–200.
6. Banerjee, R., Liu, J., Beatty, W., Pelosof, L., Klemba, M. & Goldberg, D. E. (2002). Four plasmepsins are active in the *Plasmodium falciparum* food vacuole, including a protease with an active-site histidine. *Proc. Natl Acad. Sci. USA*, **99**, 990–995.
7. Francis, S. E., Gluzman, I. Y., Oksman, A., Knickerbocker, A., Mueller, R., Bryant, M. L. *et al.* (1994). Molecular characterization and inhibition of a *Plasmodium falciparum* aspartic hemoglobinase. *EMBO J.* **13**, 306–317.
8. Silva, A. M., Lee, A. Y., Gulnik, S. V., Maier, P., Collins, J., Bhat, T. N. *et al.* (1996). Structure and inhibition of plasmepsin II, a hemoglobin degrading enzyme from *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA*, **93**, 10034–10039.
9. Bernstein, N. K., Cherney, M. M., Loetscher, H., Ridley, R. G. & James, M. N. (1990). Crystal structure of the novel aspartic proteinase zymogen proplasmepsin II from *Plasmodium falciparum*. *Nature Struct. Biol.* **6**, 32–37.
10. Asojo, O. A., Afonina, E., Gulnik, S. V., Yu, B., Erickson, J., Randad, R. *et al.* (2002). Crystal Structures of ser 205 mutant plasmepsin II from *Plasmodium falciparum* at 1.8 Å in complex with the inhibitors rs367 and rs370. *Acta Crystallog. sect. D*, **58**, 2001–2008.
11. Silva, A. M., Lee, A. Y., Erickson, J. W. & Goldberg, D. E. (1998). Structural analysis of plasmepsin II. A comparison with human aspartic proteases. *Advan. Exp. Med. Biol.* **436**, 363–373.
12. Westling, J., Cipullo, P., Hung, S. H., Saft, H., Dame, J. B. & Dunn, B. M. (1999). Active site specificity of plasmepsin II. *Protein Sci.* **8**, 2001–2009.
13. Haque, T. S., Skillman, A. G., Lee, C. E., Habashita, H., Gluzman, I. Y., Ewing, T. J. *et al.* (1999). Potent, low-molecular-weight non-peptide inhibitors of malarial aspartyl protease plasmepsin II. *J. Med. Chem.* **42**, 1428–1440.
14. Carroll, C. D., Johnson, T. O., Tao, S., Lauri, G., Orlowski, M., Gluzmanm, I. Y. *et al.* (1998). Evaluation of a structure-based statine cyclic diamino amide encoded combinatorial library against plasmepsin II and cathepsin D. *Bioorg. Med. Chem. Letters*, **8**, 3203–3206.
15. Carroll, C. D., Patel, H., Johnson, T. O., Guo, T., Orlowski, M., He, Z. M. *et al.* (1998). Identification of potent inhibitors of *Plasmodium falciparum* plasmepsin II from an encoded statine combinatorial library. *Bioorg. Med. Chem. Letters*, **8**, 2315–2320.
16. Carroll, C. D. & Orlowski, M. (1998). Screening aspartyl proteases with combinatorial libraries. *Advan. Exp. Med. Biol.* **436**, 375–380.
17. Goldberg, D. E. (1992). Plasmodial hemoglobin degradation: an ordered pathway in a specialized organelle. *Infect. Agents Dis.* **1**, 207–211.
18. Baldwin, E. T., Bhat, T. N., Gulnik, S., Hosur, M. V., Sowder, R. C., II, Cachau, R. E. *et al.* (1993). Crystal structures of native and inhibited forms of human cathepspin D: implications for lysosomal targeting and drug design. *Proc. Natl Acad. Sci. USA*, **90**, 6796–6800.
19. Sussman, J. L., Lin, D., Jiang, J., Manning, N. O., Prilusky, J., Ritter, O. & Abola, E. E. (1998). Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. *Acta Crystallog. sect. D*, **54**, 1078–1084.
20. Subramanian, E., Swan, I. D., Liu, M., Davies, D. R., Jenkins, J. A., Tickle, I. J. & Blundell, T. L. (1997). Homology among acid proteases: comparison of crystal structures at 3 Å resolution of acid proteases from *Rhizopus chinensis* and *Endothia Parasitica*. *Proc. Natl Acad. Sci. USA*, **74**, 556–559.
21. Janin, J. (1997). Surface and inside volumes in globular proteins. *Nature*, **277**, 491–492.
22. Gulnik, S. V., Afonina, E. I., Gustchina, E., Yu, B., Silva, A. M., Kim, Y. & Erickson, J. W. (2002). Utility of (his)(6) tag for purification and refolding of proplasmepsin-2 and mutants with altered activation properties. *Protein Expr. Purif.* **24**, 412–419.
23. Otwinowski, Z. (1993). *DENZO: An Oscillation Data Processing Program for Macromolecular Crystallography,* Yale University Press, New Haven, CT.
24. Otwinowski, Z. (1993). *SCALEPACK: Software for the Scaling Together of Integrated Intensities Measured on a Number of Separate Diffraction Images*, Yale University Press, New Haven.
25. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157–163.
26. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, G. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
27. Brunger, A. T., Krukowski, A. & Erickson, J. (1990). Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallog. sect. A*, **46**, 585–593.
28. Brunger, A. T., Adams, P. D. & Rice, L. M. (1997). New applications of simulated annealing in X-ray crystallography and solution NMR. *Structure*, **5**, 325–336.
29. Adams, P. D., Pannu, N. S., Read, R. J. & Brunger, A. T. (1997). Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc. Natl Acad. Sci. USA*, **94**, 5018–5023.

30. Brunger, A. T. (1992). The free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–474.
31. Brunger, A. T. (1992). *X-PLOR: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.
32. Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta. Crystallog. sect. A*, **32**, 922–923.
33. Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524.
34. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
35. Esnouf, R. M. (1997). An extensively modified version of MOLSCRIPT that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**, 132–134.
36. Christopher, J. A. (1998). *SPOCK: The Structural Properties Observations and Calculation Kit*, The Center for Macromolecular Design, Texas A&M University, College Station, TX.

Edited by R. Huber

(Received 7 August 2002; received in revised form 9 December 2002; accepted 16 December 2002)