Published in final edited form as:

J Mol Biol. 2009 May 8; 388(3): 520–540. doi:10.1016/j.jmb.2009.03.011.

Crystal structure of the histo-aspartic protease (HAP) from *Plasmodium falciparum*

Prasenjit Bhaumik^a, Huogen Xiao^b, Charity L. Parr^b, Yoshiaki Kiso^c, Alla Gustchina^{a,*}, Rickey Y. Yada^b, and Alexander Wlodawer^a

^a Protein Structure Section, Macromolecular Crystallography Laboratory, National Cancer Institute, Frederick, MD 21702, USA

^b Department of Food Science, University of Guelph, Guelph, Ontario, Canada, N1G 2W1

^c Department of Medicinal Chemistry and Center for Frontier Research in Medicinal Science, Kyoto Pharmaceutical University, Yamashina-ku, Kyoto 607-8412, Japan

Abstract

The structures of recombinant histo-aspartic protease (HAP) from malaria-causing parasite *Plasmodium falciparum*, as apoenzyme and in complex with two inhibitors, pepstatin A and KNI-10006, were solved at 2.5, 3.3, and 3.05 Å resolution, respectively. In the apoenzyme crystals HAP forms a tight dimer, not seen previously in any aspartic proteases. The interactions between the monomers affect the conformation of two flexible loops, the functionally important “flap” (residues 70–83) and its structural equivalent in the C-terminal domain (238–245), as well as the orientation of the helix 225–235. The flap is found in an open conformation in the apoenzyme. Unexpectedly, the active site of the apoenzyme contains a zinc ion tightly bound to His32 and Asp215 from one monomer, and to Glu278A from the other monomer, with the coordination of Zn resembling that seen in metalloproteases. The flap is closed in the structure of the pepstatin A complex, whereas it is open in the complex with KNI-10006. Although the binding mode of pepstatin A is significantly different than in other pepsin-like aspartic proteases, its location in the active site makes unlikely the previously proposed hypothesis that HAP is a serine protease. The binding mode of KNI-10006 is unusual compared to the binding of other inhibitors from the KNI series to aspartic proteases. The novel features of the HAP active site could facilitate design of specific inhibitors used in the development of antimalarial drugs.

Introduction

Histo-aspartic protease (HAP) is one of the ten plasmepsins (PMs) identified in the genome of *Plasmodium falciparum*, the parasite responsible for the most widespread form of malaria. HAP is one of the four plasmepsins that reside in the food vacuole of the parasite which are involved in degradation of human hemoglobin¹, making them potential targets for novel antimalarial therapy². Despite high sequence identity (60%) with other plasmepsins (PMI,

*To whom reprint requests and correspondence should be addressed: Tel: 301-846-5338; FAX: 301-846-6322; E-mail: E-mail: alla@ncifcrf.gov.

¹For consistency, the numbering of residues in HAP follows the system used for porcine pepsin. Gaps in the numbering correspond to residues found in pepsin and not in HAP, and insertions in HAP are designated by letters A, B, C, etc. appended to residue numbers.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

PMII, and PMIV) which are typical pepsin-like aspartic proteases, the active site of HAP contains several significant deviations from the pepsin standard³. In particular, Asp32¹, which together with Asp215 creates the catalytic dyad in classic aspartic proteases, is replaced by a histidine, giving this enzyme its name. In addition, substitutions are found in the functionally important flexible loop called the “flap” (residues 70–83), which changes its conformation upon ligand binding and thus participates in catalysis. These substitutions include the strictly conserved Tyr75, as well as the highly conserved Val/Gly76, which are replaced by Ser and Lys, respectively.

Although the overall level of sequence similarity suggests that HAP should have a pepsin-like fold, the predicted details of the active-site architecture and, consequently, the mode of enzymatic activity have been subjects of considerable disagreement. In the model published by Andreeva and colleagues⁴, HAP was postulated to function as a trypsin-like serine protease, with the catalytic triad consisting of Ser35, His32, and Asp215. On the other hand, Bjelic and Åqvist⁵ postulated a reaction mechanism of HAP that would assign a direct catalytic role only to Asp215, whereas the role of His32 would be to provide critical ~10,000-fold stabilization along the reaction path. It has become clear that experimental structural evidence is necessary to validate or disprove these predictions.

Whereas some enzymatic properties of HAP isolated from *P. falciparum* could be determined, the amounts of the enzyme that could be purified from the parasite¹ have not been sufficient for structural studies. Initial efforts to clone and express HAP yielded only inactive enzyme³. We have subsequently succeeded in producing active recombinant HAP, although the specific activity of the enzyme with a fluorescent substrate optimized for PMI-II was comparatively low, and the mode of inhibition, as well as the pH activity profile, were not consistent between different purification protocols^{6,7}. In this work, we report the crystal structure of recombinant HAP in ligand-free form and in complex with two inhibitors, pepstatin A⁸ and KNI-10006⁹. The results do not directly support either of the proposed mechanisms and raise a number of new questions that will require further studies.

Results and Discussion

Overall fold and structure quality

Crystal structures of uncomplexed HAP, as well as of its complexes with pepstatin A and KNI-10006 (Fig. 1A), have been refined using data extending to 2.5, 3.3, and 3.05 Å resolution, respectively. The fit of the final models to electron density maps is satisfactory and the quality of the structures, as measured by parameters such as the *R*-factor or departure from stereochemical standards¹⁰, meets the acceptable criteria given the limited resolution of the diffraction data^{11,12}. The electron density maps for the inhibitors are of sufficient quality to define their configuration unambiguously (Fig. 1B).

As predicted³, the overall fold of HAP follows the canon of eukaryotic aspartic proteases defined by pepsin (Fig. 2A). Examples of typical pepsin-like proteases with known structures include mammalian enzymes such as chymosin, renin, or cathepsin D¹³, the fungal enzyme endothiasepsin¹⁴, and the *Plasmodium* enzymes PMII¹⁵ and PMIV¹⁶. These bilobal proteins are composed of two topologically similar N- and C-terminal domains, with a large substrate-binding cleft between them. The amino and carboxyl ends of the HAP chain are assembled into a characteristic six-stranded inter-domain β-sheet, which serves to suture the domains together. A conserved sequence DT(S)G, present in one copy in each domain and containing the catalytic aspartate residues, is the signature motif of aspartic proteases¹³. Although the two signatures are recognizable in the HAP sequence, they show unusual modifications. The catalytic aspartate of the N-terminal domain is substituted by His32, and both conserved glycines are replaced by alanines. The flap is open in the apoenzyme and closed in the complex

with pepstatin A, in a manner reminiscent of typical aspartic proteases. However, the conformation of the flap loop in the KNI-10006 complex resembles that of the apoenzyme due to an unusual binding mode of the inhibitor (see below).

The structures of the apo form and of the pepstatin A complex of HAP were compared using the program ALIGN¹⁷ to the structures of unliganded pepsin (4PEP) and its pepstatin A complex (1PSO). The pepstatin A complex of HAP was also compared to the pepstatin A complexes of PMII (1XDH) and PMIV (1LS5). The four superpositions based on the Ca atoms gave r.m.s. deviations of 1.72, 1.73, 1.07, and 1.27 Å, respectively. The largest deviations are observed in the flap area, as well as for the loops containing residues 238–245 and 276–283. The superposition of the pepstatin A complexes of HAP, pepsin, PMII and PMIV (Fig. 2B) has been used to create the structurally-based sequence alignment shown in Fig. 2C.

Apoenzyme structure

The two molecules of HAP present in the tetragonal crystals of the apoenzyme form a tight dimer (Fig. 3A) involving very close contacts of their C-terminal domains, whereas the N-terminal domains point away from each other. The two monomers are related by a local two-fold axis and can be superimposed with an r.m.s. deviation of 0.32 Å between the corresponding Ca atoms. Upon dimerization, the fragment that contains helix 225–235 and the following loop 238–245 is displaced from its position commonly seen in aspartic proteases. The movement of this fragment is a consequence of the mutual insertion of loop 276–283 of the second monomer into the putative active site of monomer one.

Previously, Asojo et al.¹⁵ have reported several dimeric forms of plasmepsins created by crystallographic and non-crystallographic symmetry. However, the type of tight non-crystallographic dimer seen in the crystals of HAP apoenzyme is unique not only among plasmepsins, but also among all known pepsin-like aspartic proteases. An unusual feature found in the crystals of unliganded HAP is the presence of a zinc ion in the active site of each monomer, interacting with His32 and Asp215. The tetrahedral coordination of Zn in monomer A is completed by Glu278A, located in the intruding loop of monomer B (and vice versa), and by a water molecule (Fig. 3B). Two hydrophobic residues, Ile279A and Phe279B from the same loop are packed inside a hydrophobic pocket formed by Phe109A, Ile80, Met104, Ile107, and Val120 of the first monomer.

The surface area buried upon dimerization is 2284 Å² for each monomer, about twice the area buried in the majority of reported dimers of PMII (1270 Å², PDB ID 1XE6) or PMIV (1137 Å², PDB ID 1LS5). The only comparable buried surface area (2140 Å², Ref.18) was reported for the crystallographic dimer of PMII in complex with RS370 (PDB ID 1LF2). The presence of such extensive intermolecular interfaces was invoked as an indication that dimerization might play a role in the biological function of plasmepsins¹⁵. On the face of it, it would appear that the observed dimeric state of HAP should be even more indicative of its relevance to function. However, as will be discussed below, we are not certain that this is indeed the case.

Four additional Zn ions are found on the surface the HAP dimer. Two of them are located in equivalent positions in each monomer and are coordinated to His204 and Asp202. The third Zn ion interacts with Asp134 and His193 of monomer A and a water molecule, and the fourth one is coordinated by Glu54 and Glu57 of monomer A, as well as by Glu57 and a water molecule from the symmetry-related monomer B. It is unlikely that these additional Zn ions play any role in the mechanism of catalysis by HAP.

Comparison between apo HAP and its pepstatin A complex

The complex of HAP with pepstatin A crystallized with one protein molecule in the asymmetric unit. Despite the overall similarity to the apoenzyme structure (r.m.s. deviations of 0.98 and 0.96 Å for the Ca atoms relative to monomers A and B of apo HAP), pronounced differences are seen in the conformation of two fragments, namely the flap and the segment that contains helix 225–235 and loop 238–245 (Fig. 4). The differences between the Ca coordinates of the flap residues are in the range of 2.5–6.9 Å, with the largest shift at Lys76, a residue that is uniquely present in this location in HAP only. The whole fragment containing helix 225–235 and loop 238–245 moved in a concerted manner, with the largest deviation (19.7 Å at Val240) observed in the loop area. Several previous studies reported rigid-body movement of large fragments of the C-terminal domain of pepsin-like enzymes upon ligand binding^{19–21}, but the amplitude seen in the structure of HAP is unprecedented, revealing a remarkable flexibility of this part of the structure. In the crystal structure of HAP complexed with pepstatin A, the helix 225–235 is not involved in any crystal contacts, whereas the loop 238–245 is. The side chains of Phe238, Pro241, Phe242 and Leu244 are in contact with the hydrophobic part of the side chains of Glu278A, Asn11, Ala10 and Leu244 from a symmetry related molecule, respectively. It is also important to note that the side chain of Leu243 is also in contact with IVA1 of pepstatin A from a symmetry related molecule.

An important role of loop 238–245 in forming the extended active site area, and thus contributing to specificity and functional properties of aspartic proteases, has been the subject of debate for almost two decades. Structural studies of chymosin clearly indicated that specific cleavage of the bond between Phe105 and Met106 (P1-P1') in κ -casein, responsible for the high level of milk-clotting activity of this enzyme, was assisted by electrostatic interactions between a positively charged cluster His-Pro-His-Pro-His at positions P8-P4 (98–102) of the casein substrate and the negatively charged residues in loop 244–248 of chymosin (a structural equivalent of loop 238–245 in HAP)^{22–24}. N-terminal extension beyond the P6 position of the peptide substrates also increased the catalytic efficiency of cathepsin E and of several members of the plasmepsin family²⁵.

Loop 238–245 in plasmepsins has a notably different sequence pattern from its counterpart in other aspartic proteases. In other plasmepsins which are found in the digestive vacuole of *P. falciparum* and other *Plasmodium* species, the strictly hydrophobic nature of this loop is conserved (Fig. 5A,B). However, in plasmepsins V–X, which are not involved in hemoglobin degradation, the sequences of this loop are significantly different. It was proposed in several studies that, in plasmepsins, this loop is involved in distal interactions with the substrate, which enhance the enzymatic activity^{25,26}. Mutation of hydrophobic residues (Phe244) or deletion of Phe241 have resulted in impaired hemoglobin proteolysis, leading to the current hypothesis that recognition by this loop of hydrophobic patches on hemoglobin surface, distant from the cleavage site, weakens the substrate structure and thus initiates its degradation^{26,27}. The structure of the apoenzyme of HAP supports this idea, revealing remarkable flexibility of this loop that would be required to play such a functional role.

The overall mode of binding of pepstatin A to HAP resembles the mode of binding of this inhibitor that was previously reported in the structures of PMII (1XDH) and PMIV (1LS5). Pepstatin A is bound in extended conformation, with the statine hydroxyl positioned between Asp215 and His32. However, binding of the C-terminal half of the inhibitor is distinctly different from that found in complexes with plasmepsins and other pepsin-like proteases. Instead of wrapping around the flap, as observed with other enzymes, it is oriented towards loop 287–292, making extensive interactions with the residues comprising this fragment (Fig. 6A). The flap is closed in the structure of the complex and the Lys76 residue located at its tip, interacts with the inhibitor *via* hydrophobic contacts with the side chain of the Sta residue at P3' of pepstatin A (Fig. 6B). The ω -ammonium group of Lys76 is linked via a hydrogen bond

to the carbonyl oxygen of Ala at P2' position of pepstatin A. A charged residue with such a long side chain at this position in HAP (vs. glycine in PMIV or valine in PMII) is unique among plasmepsins, and most other pepsin-like proteases. It seems reasonable to propose that the shift of the C-terminal half of pepstatin A may be due to the presence of an unusually large residue at the tip of the flap in HAP.

Only two other hydrogen bonds between pepstatin A and HAP are clearly identified: one is between the carbonyl oxygen of Val at the P2 position and the amide of Ser219, whereas the second one is between the amide of Sta at P1 and O γ 1 of Thr218. Thus, the hydrogen bonding interactions of pepstatin A with HAP are not as extensive as those observed in the pepstatin A complex of PMII¹⁵, due to a different orientation of the C-terminal half of the inhibitor, which prevents formation of hydrogen bonds with flap residues. The side chains of pepstatin A at both termini of the molecule are also involved in extensive hydrophobic interactions with HAP residues. The isovaleryl group at the N terminus of the inhibitor and the side chain of Val at P3 interact with Phe111, Val12, and Leu13 in HAP. The P2 valine interacts with the methyl group of the side chain of Thr218, as well as with the side chains of Val287 and Ile289. The side chain of Sta P1 interacts with Phe109A, while Ala at P2' is involved in hydrophobic interactions with Met189 and Leu291. Finally, the side chain of Glu292 helps to stabilize the conformation of the C terminus of pepstatin A.

Structure of HAP in complex with KNI-10006

KNI-10006 (Fig. 1A) is a peptidomimetic inhibitor from a series of the so-called “KNI compounds”. These inhibitors have been developed during the last 20 years in a program aimed at the creation of chemotherapeutic anti-HIV agents targeting the retroviral protease^{28,29}. The design of this series was based on the concept of “substrate transition-state mimicry”³⁰, with the central core made of an α -hydroxy- β -amino acid derivative, allophenylnorstatine (Apsns), which contains a hydroxymethylcarbonyl (HMC) isostere. It has been also reported that KNI compounds demonstrated effective inhibition of not only HIV-1 PR, but also PMII and HTLV-1 PR³¹. Further chemical elaboration and optimization of the KNI inhibitors, supported by extensive structural and biochemical data, resulted in the synthesis of many new compounds shown to be potent inhibitors of HIV-1 and HTLV-1 PRs^{9,32–36}. One of them, KNI-10006, was subsequently shown to be a potent inhibitor of HAP³⁷ with the IC₅₀ of 0.69 μ M⁹ and was thus chosen for the evaluation of its interaction with the enzyme.

The binding mode of KNI-10006 to HAP is drastically different from that of pepstatin A (Fig. 7), as well as from a number of other KNI inhibitors bound to various aspartic proteases. In the HAP complex, the hydroxyl group in the central part of the inhibitor points away from the catalytic residues, in contrast to its orientation in the structures of either HIV-1 PR³⁸ or PMIV¹⁵, where it is positioned between the active-site aspartates (Fig. 8A). The predominant interactions of KNI-10006 are with the flap and this inhibitor does not make any contacts with either the loop 287–292 or with several other hydrophobic residues, conserved in plasmepsins and in other pepsin-like enzymes, that were shown to anchor KNI-764 in a PMIV complex¹⁶ (Fig. 8B). However, there is a striking similarity in the binding mode of KNI-10006 to HAP and one of the molecules (designated 1) of an achiral inhibitor bound to PMII (2BJU)²⁶ (Fig. 9A). In the latter structure, two inhibitor molecules are bound to a single PMII molecule, with the second inhibitor molecule (2) oriented in a way that is reminiscent of the binding mode of pepstatin A to HAP. Both the *n*-pentyl chain of molecule 1 and the 2,6-dimethylphenyloxymethyl (DMP) moiety of KNI-10006 occupy the so-called “flap pocket”. The existence of such a pocket was noted before in human renin³⁹. In the HAP complex with KNI-10006, this pocket is open and the conformation of the flap is similar to its conformation in the apoenzyme (Fig. 7). Comparison of KNI-10006 bound to HAP with three crystal structures of aspartic proteases in complex with inhibitors featuring *n*-pentyl substituents

(2BJU, 2IGX, 2IGY)^{26,40,41} allows for the description of the residues comprising this binding pocket in the plasmepsin family and for the comparison to other aspartic proteases (Fig. 9B). The flap pocket is predominantly hydrophobic in both plasmepsins and other aspartic proteases. However, an insertion of Phe109A in HAP and PMII, or Leu109A in PMIV changes the architecture of this pocket and makes it even more hydrophobic in plasmepsins. The conformation of the inserted residue changes in the complexes with different ligands, in order to optimize the interactions with the moieties inserted in the pocket. It should be noted, however, that the side chain of Leu112 in pepsin is oriented in such a way that it occupies some of the space taken by the residue 109A in plasmepsins, thus contributing to the interactions with the ligand and partially compensating for the absence of the extra residue in the flap pocket (Fig. 9B). Another important residue, located at the entrance to the flap pocket, is Phe111 in HAP, substituted by threonine in PMII and by leucine in PMIV. These differences between plasmepsins may bear on their specific ligand preferences.

The active site

Similarly to its counterparts in all pepsin-like aspartic proteases, the active site of HAP is located in a large cleft formed by the N- and C-terminal domains of the protein. Whereas the overall architecture of the active site is preserved, a crucial difference in HAP is the replacement of the canonical aspartate from the N-terminal domain⁴² by His32. The other functionally important substitutions are found in the flap area, where the commonly conserved Tyr75 and Val/Gly76 residues are replaced in HAP by Ser and Lys, respectively. The conserved pattern of hydrogen bonds, known as the “fireman’s grip”⁴³, is found in the active sites of both the apoenzyme and the inhibitor-bound HAP, although the lengths of the individual hydrogen bonds are affected by the larger size of the side chain of His32 (vs. Asp32). As will be discussed below, some structural features of the active site of HAP also differ between the apoenzyme and the complexes with the inhibitors.

The two active sites in the dimeric apoenzyme are practically identical. Each of them contains a Zn ion bound to side chains that belong to both molecules of the dimer. The Zn ion is tetrahedrally coordinated by the side chains of His32 and Asp215 from one HAP monomer, Glu278A from the other monomer, and by a water molecule (Fig. 3B). Glu278A is located in the loop consisting of residues 274–285, which is inserted into the partner active site, leading to the observed opening of the flap. Previously, Asojo et al.¹⁵ have reported on a crystallographic dimer of PMII, in which the loop containing residues 237–247 is docked into the binding cavity of the two-fold-related monomer. However, the packing of the loop in the active site is not as tight as in the HAP dimer (see above). The presence of bound Zn in the active site of HAP has a profound effect on the hydrogen bond interactions. In the active sites of all pepsin-like aspartic proteases the catalytic residue Asp215 interacts with a neighboring Thr218^{15,44}, and this interaction is important in maintaining the proper protonation state of the catalytic carboxylate. However, no comparable interaction is present in the active site of apo HAP, where the distance between Oγ1 of Thr218 and Oδ2 of Asp215 is 4.04 Å. The latter atom also coordinates the Zn ion and is involved in hydrogen-bonded interactions with two neighboring water molecules. The other active site residue, His32, also interacts with the Zn ion *via* its Nε2 atom. The imidazole ring of His32 is fixed in suitable orientation by an Nδ1...Oγ hydrogen bond with the side chain of Ser35, the latter also being hydrogen-bonded to Nε1 of Trp39 *via* a water molecule (Fig. 10A). Since the flap is in an open position in the apoenzyme, Ser75 and Lys76 are far away from the active site. The side chain hydroxyl of Ser81 is hydrogen-bonded to the main chain NH groups of Lys76 and Ala77, whereas the side chain of Lys76 is solvent exposed. The presence of a metal ion in the active sites of aspartic proteases has not been observed before, although some heavy atom derivatives of retroviral proteases, prepared by soaking in very concentrated salts, included a uranyl ion bound to the two aspartates^{45,46}. Although the coordination of the Zn ion in the active site of HAP

resembles that observed in the active site of a metalloprotease DppA (D-aminopeptidase, PDB code 1HI9)⁴⁷ (Fig. 10B), the lack of inhibition by chelating agents such as EDTA indicates that HAP does not function as a metalloprotease¹, at least towards synthetic peptides.

Although the structure of HAP in complex with pepstatin A has been determined at a lower resolution, the electron density clearly indicates the presence of the inhibitor in the active site (Fig. 1B). The overall orientation of the pepstatin A molecule in the active site of HAP is similar to that in the active sites of mammalian aspartic proteases, as well as in plasmepsins II and IV¹⁵. The statine hydroxyl group of pepstatin A is positioned between the two catalytic residues and at hydrogen-bonding distances of the Nε2 and Oδ2 atoms of His32 and Asp215, respectively (Fig. 10C). The change, upon ligand binding, of the flap conformation from open to closed, is accompanied by a dramatic change in the conformation of Trp39 (Fig. 10C). A flip of the side chain of the corresponding tryptophan residue has also been observed in the liganded structures of PMII¹⁵ and PMIV¹⁶, as well as of pepsin⁴⁸. Closing of the flap is important for the catalytic mechanism in pepsin-like aspartic proteases⁴⁴. A motion of this segment brings an important residue, Tyr75, closer to the active site area, generating a network of hydrogen bonds leading to the catalytic aspartates. Since serine occupies an equivalent position in HAP, closing of the flap cannot generate the same effect. Thus, the exact role of the flap in substrate binding or in the catalytic mechanism is still not entirely clear.

A superposition of the structures of pepstatin A complexes of HAP and pepsin (1PSO), based on the Cα atoms of the proteins, shows that whereas the residues in the extended active site area are well superposed, the statine hydroxyls occupy different positions (Fig. 10D). For clarity of visualization, we subsequently compared the interactions of the statine hydroxyls with the catalytic residues in two structures by superposing the identical parts of the active sites of HAP and pepsin. These parts of the active sites are involved in the catalytic mechanism and include the loop carrying Asp215 and the central part of the inhibitor with its hydroxyl group that occupies the position of the nucleophilic water molecule. As a result of the latter superposition, Nε2 of His32 is close to the Oδ1 atom of Asp32 in pepsin (Fig. 10E), thus completing the superposition of all functional groups that are presumably involved in the catalytic mechanism in both enzymes. A comparison of these two superpositions (Fig. 10D,E) unambiguously indicates that the shift in the position of the statine hydroxyl group in HAP is most likely dictated by the increased dimensions of the histidine side chain at position 32 (vs, aspartate). This shift is determined by the distance between Oδ1 of Asp32 and Nε2 of His32 in the former superposition and roughly corresponds to the distance between Cδ2 and Nε2 atoms in the ring of a histidine side chain. Different location of the key functional group of HAP leads to a misalignment of the residues in the extended area of the active site in the N-terminal domain of HAP (Fig. 10E). These differences in the active sites of HAP and pepsin make the similarity in the pattern of hydrogen bonds between the statine OH group and the catalytic residues even more surprising (Fig. 10F).

In order to compare the active site structures of the apoenzyme, HAP-KNI-10006, and HAP-pepstatin A, the Cα atoms of these three structures have been superposed (Fig. 10G). The superposition shows that the positions of the catalytic His32 and Asp215 residues are very similar in all three structures. Also, the orientation of Trp39 in the apoenzyme and in the HAP-KNI-10006 complex is identical. The two water molecules located between Ser35 and Trp39 in both structures occupy the same sites. The core hydroxyl group of pepstatin A in the HAP-pepstatin A complex occupies the position corresponding to the water molecule bound between the side chains of His32 and Asp215 in the HAP-KNI-10006 complex. This latter water molecule is hydrogen bonded to Nε2 of His32 and Oδ2 of Asp215. By analogy with pepsin-like aspartic proteases¹³, it may be expected that this water molecule is directly involved in the catalytic mechanism of HAP.

As expected from the sequence similarity, the active-site architecture of HAP is similar to other pepsin-like aspartic proteases. However, the situation in the apoenzyme is quite unique because of the presence of the Zn ion. Although this ion is most likely originates from the crystallization mother liquor, its presence in the active site indicates that HAP could be sensitive to elevated zinc concentration. The binding mode of pepstatin A in the active site of HAP indicates that His32 and Asp215 are very likely involved in catalysis. As observed in the other pepsin-like aspartic proteases, Ser35, Trp39, and Thr218 may also play a role in the catalytic mechanism of HAP. High resolution crystal structures are, however, required to assign the correct role of all these residues in the reaction mechanism.

Conclusions

Whereas the structure of HAP from *P. falciparum* exhibits the expected overall similarity to the structure of pepsin-like aspartic proteases, and to plasmepsins in particular, some substantial differences exist as well. The structural data presented for the apoenzyme and two inhibitor complexes can be interpreted to argue against the hypothesis that HAP functions as an unusual serine protease with an aspartic protease fold. However, the available data do not provide an unambiguous answer about the exact nature of the catalytic mechanism of this unusual enzyme. Although the active site of the apoenzyme HAP can bind a Zn ion in a manner reminiscent of metalloproteases, such as D-aminopeptidase DppA, this similarity may be misleading. First, the metal ion is coordinated by two HAP molecules forming a tight dimer, and yet there is no proof, to date, that dimerization is required for the enzymatic activity. Second, the structures of the inhibitor complexes do not contain any metal ions and indeed, they cannot be accommodated between the inhibitor and the active-site residues. An additional complication is the fact that both inhibitor complexes (with pepstatin A and KNI-10006) were solved using crystals grown at pH well above the range in which the enzyme is active. Thus, the determination of the details of the enzymatic mechanism of HAP will require further studies, both biochemical and structural. However, an observation that parts of the inhibitor bind in a pocket that is unique to plasmepsins but not utilized for inhibitor binding by other aspartic proteases bodes well for the successful development of broad-specificity compounds capable of simultaneous inhibition of HAP and other plasmepsins. Such compounds could be developed into a new class of anti-malarial drugs.

Materials and Methods

Expression and purification of HAP

Expression of the recombinant Trx-tHAP fusion protein was conducted according to the method described previously⁶. Cell pellet of 1 l culture was suspended in 50 ml potassium phosphate buffer (pH=7.5) containing 1X BugBuster™ reagent (Novagen, Madison WI, USA) and 1μl of Benznase (Novagen), 250 U/μl. The suspension was incubated at room temperature for 40 min with gentle shaking. The sample was then centrifuged at 16,000g for 20 min at 4° C. The supernatant was applied on a HisSelect™ Cartridge (6.4 ml) (Sigma-Aldrich, Oakville ON, Canada); the column was first flushed with washing buffer (50 mM sodium phosphate, 0.3 M NaCl, 10 mM imidazole, pH 7.5), then with 10% elution buffer (50 mM sodium phosphate, 0.3 M NaCl, 250 mM imidazole, pH 7.5), and finally eluted with 50% elution buffer. The eluate was concentrated in 50 mM sodium phosphate buffer, pH7.5, containing 0.2% CHAPS in an Ultracel YM50 Centricron (Millipore, Billerica, USA) and further purified by gel filtration using a Superose™ 12 10/300 GL column (GE Healthcare, Uppsala, Sweden) equilibrated with 50 mM sodium phosphate buffer, pH 7.5, containing 150 mM NaCl and 0.2% CHAPS to obtain pure fusion Trx-tHAP protein. Two peaks were observed on the chromatogram. Fractions from peak 2 were collected and washed with 50 mM sodium phosphate buffer, pH 7.5, containing 0.2% CHAPS. Subsequently, the Trx-tHAP protein was

activated with enteropeptidase (Sigma, St. Louis, MO) (EK:Trx-tHAP = 1:25) in 50 mM MES buffer, pH 6.5, containing 0.2% CHAPS at 37 °C for 2 days. After checking the initial purity of the protein sample by SDS-PAGE, the activated mtHAP was purified by washing with 50 mM MES buffer (pH 6.5), in an Ultracel YM30 Centricon (Millipore, Billerica, USA) to remove the cleaved prosegment and thioredoxin, yielding mature tHAP (mtHAP). mtHAP was further purified by gel filtration with a SuperoseTM 12 10/300 GL column equilibrated with 50 mM MES buffer, pH 6.5, containing 150 mM NaCl and 0.2% CHAPS. The fractions from the second major peak were washed and concentrated in 50 mM MES buffer, pH 6.5, to about 10 mg/ml concentration, then frozen in liquid nitrogen and stored at -80 °C. As shown previously⁶ and again verified by mass spectrometry (data not shown), protein purified by this procedure consists of 332 amino acids. HAP used in this study contains four additional residues on its N terminus that originate from the propeptide and are not present in the native protein directly isolated from *P. falciparum*.

Crystallization

For initial crystallization experiments, a HAP sample was transferred to 0.1 M sodium acetate buffer, pH 4.0, and concentrated to 12.0 mg/ml. Concentrated DMSO solution of pepstatin A was mixed with the protein sample to yield the final inhibitor concentration of 0.3 mM in the mixture (1:1 protein:inhibitor molar ratio). The sample was subsequently incubated for two hours and centrifuged. Several crystallization screens were set up using sitting-drop vapor-diffusion method at 293 K, and a few conditions produced small needles. The conditions were optimized using hanging drop vapor-diffusion method at 293 K. Two types of crystals (type I and type II) were grown under two different crystallization conditions. Type I crystals appeared in drops containing 0.7 μ l protein solution and 0.7 μ l reservoir solution, equilibrated against 1 ml reservoir solution. The reservoir solution contained 10% PEG 3000, 0.2 M zinc acetate, and 0.1 M sodium acetate buffer at pH 4.5. These crystals were subsequently shown not to contain pepstatin A. Type II crystals were grown in drops containing 0.8 μ l protein solution and 0.4 μ l reservoir solution containing 15% PEG 20000 and 0.1 M Tris-HCl, pH 8.5. These crystals contained bound pepstatin A.

For crystallization of the HAP-KNI-10006 complex, HAP sample was first transferred to 0.1 M sodium acetate buffer, pH 5.0, and concentrated to 15.0 mg/ml. Concentrated DMSO solution of KNI-10006 was mixed with the protein sample to yield the final inhibitor concentration of 0.4 mM in the mixture (1:1 protein:inhibitor molar ratio), which was subsequently incubated for 24 h. Optimization of the crystallization condition for the complex between HAP and KNI-10006 was conducted in the same manner as for the complex with pepstatin A. The best quality crystals (designated type III) were grown using hanging-drop vapor-diffusion method at 293 K in drops mixed from 0.8 μ l protein solution and 0.8 μ l reservoir solution containing 0.2 M KH₂PO₄ and 20% PEG 3350.

Data collection

X-Ray diffraction data for type I and type III crystals were collected using a MAR300CCD detector and a wavelength of 0.9999 Å at the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory, Argonne, IL. Data for a type II crystal were collected using Cu K α radiation generated by a Rigaku H3R X-ray source and a MAR345dtb detector. All data were collected at 100 K, using, as cryoprotectant, 20% (v/v) MPD for type I crystals, and 20% ethylene glycol (v/v) for type II and type III crystals, added to reservoir solution. Type I crystals are tetragonal in space group *P*4₁2₁; type II crystals are trigonal, space group *P*3₂21, and type III crystals are tetragonal in space group *I*4₁22. All data sets were indexed and integrated using the program XDS⁴⁹. The data were converted to structure factors with modules F2MTZ and CAD of CCP4⁵⁰.

Structure solution and refinement

HAP has a high level of sequence identity with plasmepsins II (60.6%) and IV (61.8%). Several crystal structures of both plasmepsins are available in the Protein Data Bank (PDB), providing a wide choice of starting models for molecular replacement. The structures of the apoenzyme and of its complexes with pepstatin A and KNI-10006 were solved by molecular replacement using the programs MrBUMP⁵¹, MOLREP⁵², and PHASER⁵³.

The Matthews coefficient⁵⁴ indicated the likely presence of two molecules in the asymmetric unit of type I crystal. The automated search performed with MrBUMP identified the B-chain of PMIV (PDB code 2ANL) as the best search model. The final molecular replacement procedures utilized directly the programs MOLREP⁵² and PHASER⁵³. The preliminary model was refined with REFMAC5⁵⁵ and rebuilt with COOT⁵⁶. Very tight NCS restraints were used in the initial stages of refinement, but slowly the restraints were released as the model was becoming more complete; medium NCS restraints were applied for the final refinement cycles. Zn ions and water molecules were progressively introduced at peaks of electron density higher than 3 σ in ($F_o - F_c$) α_c σ_a -weighted maps while monitoring the decrease of R_{free} . Proper hydrogen bonding was required for placement of all solvent molecules. PROCHECK¹⁰ was used to monitor the stereochemistry of the protein model. Although the type I crystals were grown using HAP incubated in the presence of pepstatin A, careful analysis of the difference electron density map did not show the presence of the inhibitor in either of the active sites of the dimer. Seven residues (−5 to −1 and 327 to 328) are missing from each subunit in the final model of the apoenzyme; these residues could not be built because of lack of features in the electron density map. All other residues are well defined in the map.

The crystal structure of HAP in complex with pepstatin A was determined using data obtained from a type II crystal, which contained only one protein molecule in the asymmetric unit. Automated search by MrBUMP, using the A-chain of PMII (PDB code 1SME) as the probe, correctly placed the model in the asymmetric unit. After the first cycle of refinement in REFMAC5, the ($F_o - F_c$) α_c electron density map indicated the presence of pepstatin A in the active site (Fig. 1B). After modeling the inhibitor, iterative cycles of model refinement with REFMAC5 and model building in the electron density maps using COOT were carried out. The overall anisotropy was modeled with TLS parameters by dividing the protein molecule into two TLS groups, comprising residues 2–210 and 211–326. PROCHECK was used to monitor the stereochemistry of the model. Nine residues (−5 to 1 and 327–328) of the HAP molecule could not be modeled in the electron density map.

Solvent content analysis showed that the crystals of HAP in complex with KNI-10006 (type III) contain four protein molecules in the asymmetric unit. The structure of this complex was determined by molecular replacement using the coordinates of the protein from its pepstatin complex. Automated search by PHASER revealed the correct placement of the first two molecules in the asymmetric unit. These two molecules, which did not form a dimer similar to that observed in the apoenzyme structure, were subsequently used to find the correct orientation of the two remaining molecules in the asymmetric unit. Initially, a few cycles of refinement using REFMAC5 and rebuilding using COOT were performed for the protein only. Subsequent analysis of the ($F_o - F_c$) α_c electron density map (Fig. 1B) indicated the presence of the KNI-10006 inhibitors in the active site of each of the four monomers. After inhibitor modeling, iterative cycles of refinement in REFMAC5 and model building in the electron density maps using COOT, were carried out. Very tight NCS restraints were used during all refinement cycles. The overall anisotropy was modeled with TLS parameters by dividing each molecule into two TLS groups, comprising residues 0–194 and 195–327. In the final model, six residues (−5 to −1 and 328) are missing from all four molecules because of lack of features in the electron density map. The refinement statistics for all three structures are presented in Table I. The structures were analyzed using programs PROCHECK¹⁰ and COOT⁵⁶. Structural

superpositions were performed with programs SSM⁵⁷ and ALIGN¹⁷ and the figures were generated with the program PYMOL⁵⁸.

PDB accession numbers

Atomic coordinates and structure factors have been deposited in the PDB with the codes 3FNS for the HAP apoenzyme, 3FNT for the complex with pepstatin A, and 3FNU for the complex with KNI-10006.

Acknowledgments

We would like to thank Dr. Mariusz Jaskolski for his invaluable comments on the manuscript. Diffraction data were collected at the Southeast Regional Collaborative Access Team (SER-CAT) beamline 22-ID, located at the Advanced Photon Source, Argonne National Laboratory. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38. This project was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. Financial support from the Natural Sciences and Engineering Research Council of Canada and the Canada Research Chairs Program is also gratefully acknowledged.

References

1. Banerjee R, Liu J, Beatty W, Pelosof L, Klemba M, Goldberg DE. Four plasmepsins are active in the *Plasmodium falciparum* food vacuole, including a protease with an active-site histidine. *Proc Natl Acad Sci USA* 2002;99:990–995. [PubMed: 11782538]
2. Ersmark K, Samuelsson B, Hallberg A. Plasmepsins as potential targets for new antimalarial therapy. *Med Res Rev* 2006;26:626–666. [PubMed: 16838300]
3. Berry C, Humphreys MJ, Matharu P, Granger R, Horrocks P, Moon RP, Certa U, Ridley RG, Bur D, Kay J. A distinct member of the aspartic proteinase gene family from the human malaria parasite *Plasmodium falciparum*. *FEBS Lett* 1999;447:149–154. [PubMed: 10214936]
4. Andreeva N, Bogdanovich P, Kashparov I, Popov M, Stengach M. Is histoaspartic protease a serine protease with a pepsin-like fold? *Proteins* 2004;55:705–710. [PubMed: 15103632]
5. Bjelic S, Aqvist J. Computational prediction of structure, substrate binding mode, mechanism, and rate for a malaria protease with a novel type of active site. *Biochemistry* 2004;43:14521–14528. [PubMed: 15544322]
6. Xiao H, Sinkovits AF, Bryksa BC, Ogawa M, Yada RY. Recombinant expression and partial characterization of an active soluble histo-aspartic protease from *Plasmodium falciparum*. *Protein Expr Purif* 2006;49:88–94. [PubMed: 16624575]
7. Parr CL, Tanaka T, Xiao H, Yada RY. The catalytic significance of the proposed active site residues in *Plasmodium falciparum* histoaspartic protease. *FEBS J* 2008;275:1698–1707. [PubMed: 18312598]
8. Umezawa H, Aoyagi T, Morishima H, Matsuzaki M, Hamada M. Pepstatin, a new pepsin inhibitor produced by *Actinomycetes*. *J Antibiot (Tokyo)* 1970;23:259–262. [PubMed: 4912600]
9. Nezami A, Kimura T, Hidaka K, Kiso A, Liu J, Kiso Y, Goldberg DE, Freire E. High-affinity inhibition of a family of *Plasmodium falciparum* proteases by a designed adaptive inhibitor. *Biochemistry* 2003;42:8459–8464. [PubMed: 12859191]
10. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 1993;26:283–291.
11. Jaskolski M, Gilski M, Dauter Z, Wlodawer A. Stereochemical restraints revisited: how accurate are refinement targets and how much should protein structures be allowed to deviate from them? *Acta Crystallogr* 2007;D63:611–620.
12. Wlodawer A, Minor W, Dauter Z, Jaskolski M. Protein crystallography for non-crystallographers or how to get the best (but not more) from the published macromolecular structures. *FEBS J* 2008;275:1–21. [PubMed: 18034855]
13. Dunn BM. Structure and mechanism of the pepsin-like family of aspartic peptidases. *Chem Rev* 2002;102:4431–4458. [PubMed: 12475196]

14. Erskine PT, Coates L, Mall S, Gill RS, Wood SP, Myles DA, Cooper JB. Atomic resolution analysis of the catalytic site of an aspartic proteinase and an unexpected mode of binding by short peptides. *Protein Sci* 2003;12:1741–1749. [PubMed: 12876323]
15. Asojo OA, Gulnik SV, Afonina E, Yu B, Ellman JA, Haque TS, Silva AM. Novel uncomplexed and complexed structures of plasmepsin II, an aspartic protease from *Plasmodium falciparum*. *J Mol Biol* 2003;327:173–181. [PubMed: 12614616]
16. Clemente JC, Govindasamy L, Madabushi A, Fisher SZ, Moose RE, Yowell CA, Hidaka K, Kimura T, Hayashi Y, Kiso Y, Agbandje-McKenna M, Dame JB, Dunn BM, McKenna R. Structure of the aspartic protease plasmepsin 4 from the malarial parasite *Plasmodium malariae* bound to an allophenylnorstatine-based inhibitor. *Acta Crystallogr* 2006;D62:246–252.
17. Cohen GE. ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J Appl Crystallogr* 1997;30:1160–1161.
18. Asojo OA, Afonina E, Gulnik SV, Yu B, Erickson JW, Randad R, Medjahed D, Silva AM. Structures of Ser205 mutant plasmepsin II from *Plasmodium falciparum* at 1.8 Å in complex with the inhibitors rs367 and rs370. *Acta Crystallogr* 2002;D58:2001–2008.
19. Sali A, Veerapandian B, Cooper JB, Moss DS, Hofmann T, Blundell TL. Domain flexibility in aspartic proteinases. *Proteins* 1992;12:158–170. [PubMed: 1603805]
20. Abad-Zapatero C, Rydel TJ, Erickson J. Revised 2.3 Å structure of porcine pepsin: evidence for a flexible subdomain. *Proteins* 1990;8:62–81. [PubMed: 2217165]
21. Andreeva NS, Pechik IV. [Comparison of three-dimensional structures of flexible protein molecules]. *Mol Biol (Mosk)* 1995;29:1102–1113. [PubMed: 8538605]
22. Gilliland GL, Winborne EL, Nachman J, Wlodawer A. The three-dimensional structure of recombinant bovine chymosin at 2.3 Å resolution. *Proteins* 1990;8:82–101. [PubMed: 2217166]
23. Safro MG, Andreeva NS. On the role of peripheral interactions in specificity of chymosin. *Biochem Int* 1990;20:555–561. [PubMed: 2111997]
24. Gustchina E, Rumsh L, Ginodman L, Majer P, Andreeva N. Post X-ray crystallographic studies of chymosin: the existence of two structural forms and the regulation of activity by the interaction with the histidine-proline cluster of kappa-casein. *FEBS Lett* 1996;379:60–62. [PubMed: 8566230]
25. Istvan ES, Goldberg DE. Distal substrate interactions enhance plasmepsin activity. *J Biol Chem* 2005;280:6890–6896. [PubMed: 15574427]
26. Prade L, Jones AF, Boss C, Richard-Bildstein S, Meyer S, Binkert C, Bur D. X-ray structure of plasmepsin II complexed with a potent achiral inhibitor. *J Biol Chem* 2005;280:23837–23843. [PubMed: 15840589]
27. Omara-Opyene AL, Moura PA, Sulsona CR, Bonilla JA, Yowell CA, Fujioka H, Fidock DA, Dame JB. Genetic disruption of the *Plasmodium falciparum* digestive vacuole plasmepsins demonstrates their functional redundancy. *J Biol Chem* 2004;279:54088–54096. [PubMed: 15491999]
28. Mimoto T, Imai J, Tanaka S, Hattori N, Kisanuki S, Akaji K, Kiso Y. KNI-102, a novel tripeptide HIV protease inhibitor containing allophenylnorstatine as a transition-state mimic. *Chem Pharm Bull (Tokyo)* 1991;39:3088–3090. [PubMed: 1799953]
29. Mimoto T, Imai J, Tanaka S, Hattori N, Takahashi O, Kisanuki S, Nagano Y, Shintani M, Hayashi H, Sakikawa H. Rational design and synthesis of a novel class of active site-targeted HIV protease inhibitors containing a hydroxymethylcarbonyl isostere. Use of phenylnorstatine or allophenylnorstatine as a transition-state mimic. *Chem Pharm Bull (Tokyo)* 1991;39:2465–2467. [PubMed: 1804562]
30. Kiso Y. Design and synthesis of substrate-based peptidomimetic human immunodeficiency virus protease inhibitors containing the hydroxymethylcarbonyl isostere. *Biopolymers* 1996;40:235–244. [PubMed: 8785365]
31. Maegawa H, Kimura T, Arai Y, Matsui Y, Kasai S, Hayashi Y, Kiso Y. Identification of peptidomimetic HTLV-I protease inhibitors containing hydroxymethylcarbonyl (HMC) isostere as the transition-state mimic. *Bioorg Med Chem Lett* 2004;14:5925–5929. [PubMed: 15501070]
32. Abdel-Rahman HM, Kimura T, Hidaka K, Kiso A, Nezami A, Freire E, Hayashi Y, Kiso Y. Design of inhibitors against HIV, HTLV-I, and *Plasmodium falciparum* aspartic proteases. *Biol Chem* 2004;385:1035–1039. [PubMed: 15576323]

33. Kimura T, Nguyen JT, Maegawa H, Nishiyama K, Arii Y, Matsui Y, Hayashi Y, Kiso Y. Chipping at large, potent human T-cell leukemia virus type 1 protease inhibitors to uncover smaller, equipotent inhibitors. *Bioorg Med Chem Lett* 2007;17:3276–3280. [PubMed: 17448657]
34. Nguyen JT, Zhang M, Kumada HO, Itami A, Nishiyama K, Kimura T, Cheng M, Hayashi Y, Kiso Y. Truncation and non-natural amino acid substitution studies on HTLV-I protease hexapeptidic inhibitors. *Bioorg Med Chem Lett* 2008;18:366–370. [PubMed: 18006315]
35. Zhang M, Nguyen JT, Kumada HO, Kimura T, Cheng M, Hayashi Y, Kiso Y. Locking the two ends of tetrapeptidic HTLV-I protease inhibitors inside the enzyme. *Bioorg Med Chem* 2008;16:6880–6890. [PubMed: 18558491]
36. Zhang M, Nguyen JT, Kumada HO, Kimura T, Cheng M, Hayashi Y, Kiso Y. Synthesis and activity of tetrapeptidic HTLV-I protease inhibitors possessing different P3-cap moieties. *Bioorg Med Chem* 2008;16:5795–5802. [PubMed: 18400502]
37. Hidaka K, Kimura T, Ruben AJ, Uemura T, Kamiya M, Kiso A, Okamoto T, Tsuchiya Y, Hayashi Y, Freire E, Kiso Y. Antimalarial activity enhancement in hydroxymethylcarbonyl (HMC) isostere-based dipeptidomimetics targeting malarial aspartic protease plasmepsin. *Bioorg Med Chem* 2008;16:10049–10060. [PubMed: 18952439]
38. Fitzgerald PMD, McKeever BM, VanMiddlesworth JF, Springer JP, Heimbach JC, Leu CT, Herber WK, Dixon RAF, Darke PL. Crystallographic analysis of a complex between human immunodeficiency virus type 1 protease and acetyl-pepstatin at 2.0 Å resolution. *J Biol Chem* 1990;265:14209–14219. [PubMed: 2201682]
39. 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, Iler M, Ridley RG, Stadler H, Vieira E, Wilhelm M, Winkler F, Wostl W. Renin inhibition by substituted piperidines: a novel paradigm for the inhibition of monomeric aspartic proteinases? *Chem Biol* 1999;6:127–131. [PubMed: 10074464]
40. Boss C, Corminboeuf O, Grisostomi C, Meyer S, Jones AF, Prade L, Binkert C, Fischli W, Weller T, Bur D. Achiral, cheap, and potent inhibitors of Plasmepsins I, II, and IV. *ChemMedChem* 2006;1:1341–1345. [PubMed: 17091526]
41. Hof F, Schutz A, Fah C, Meyer S, Bur D, Liu J, Goldberg DE, Diederich F. Starving the malaria parasite: inhibitors active against the aspartic proteases plasmepsins I, II, and IV. *Angew Chem Int Ed Engl* 2006;45:2138–2141. [PubMed: 16502446]
42. Barrett, AJ.; Rawlings, ND.; Woessner, JF. *Handbook of Proteolytic Enzymes*. Elsevier Academic Press; Amsterdam: 2004.
43. Pearl L, Blundell T. The active site of aspartic proteinases. *FEBS Lett* 1984;174:96–101. [PubMed: 6381096]
44. Andreeva NS, Rumsh LD. Analysis of crystal structures of aspartic proteinases: On the role of amino acid residues adjacent to the catalytic site of pepsin-like enzymes. *Protein Sci* 2001;10:2439–2450. [PubMed: 11714911]
45. Wlodawer A, Miller M, Jaskólski M, Sathyanarayana BK, Baldwin E, Weber IT, Selk LM, Clawson L, Schneider J, Kent SBH. Conserved folding in retroviral proteases: Crystal structure of a synthetic HIV-1 protease. *Science* 1989;245:616–621. [PubMed: 2548279]
46. Miller M, Jaskólski M, Rao JKM, Leis J, Wlodawer A. Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature* 1989;337:576–579. [PubMed: 2536902]
47. Remaut H, Bompard-Gilles C, Goffin C, Frere JM, Van Beeumen J. Structure of the *Bacillus subtilis* D-aminopeptidase DppA reveals a novel self-compartmentalizing protease. *Nat Struct Biol* 2001;8:674–678. [PubMed: 11473256]
48. Fujinaga M, Chernaia MM, Tarasova NI, Mosimann SC, James MN. Crystal structure of human pepsin and its complex with pepstatin. *Protein Sci* 1995;4:960–972. [PubMed: 7663352]
49. Kabsch W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* 1993;26:795–800.
50. CCP4. Collaborative Computational Project, Number 4, 1994. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr* 1994;D50:760–763.
51. Keegan RM, Winn MD. MrBUMP: an automated pipeline for molecular replacement. *Acta Crystallogr* 2008;D64:119–124.

52. Vagin A, Teplyakov A. MOLREP: an automated program for molecular replacement. *J Appl Cryst* 1997;30:1022–1025.
53. Read RJ. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr* 2001;D57:1373–1382.
54. Matthews BW. Solvent content of protein crystals. *J Mol Biol* 1968;33:491–497. [PubMed: 5700707]
55. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr* 1997;D53:240–255.
56. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr* 2004;D60:2126–2132.
57. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr* 2004;D60:2256–2268.
58. DeLano, WL. The PyMOL Molecular Graphics System. DeLano Scientific; San Carlos, CA: 2002.

Fig. 1. Inhibitors of HAP used in this study. (A) Chemical diagrams of pepstatin A and KNI-10006. (B) Initial $F_o - F_c$ omitmaps for pepstatin A and KNI-10006 contoured at 1.9σ level, with the final models superimposed.

Fig. 2.

Overall structure of HAP and a comparison with plasmepsins. (A) Stereo ribbon diagram showing the monomer of HAP apoenzyme. Two active site residues, His32 and Asp215, are shown in stick representation. The N-terminal domain is on top, and the C-terminal domain is on the bottom. (B) Ribbon diagram of the superposed structures of pepstatin A-bound complexes of HAP (cyan), PMII (orange), PMIV (purple), and pepsin (green). (C) Structure-based sequence alignment of HAP, PMII, PMIV, and pepsin. Strictly conserved residues are white on red background, limited identities are red, similarities blue, and differences black. The signature motifs of aspartic proteases are boxed.

Fig. 3.

Structure of the apoenzyme form of HAP. (A) A ribbon diagram of the HAP dimer, with the side chains in one of the active sites shown in stick representation. The Zn ion bound in the active site is shown as a sphere. (B) The residues coordinating the Zn ion covered by $2F_o-F_c$ electron density map (blue) contoured at the 1.0σ level, as well as by F_o-F_c omit electron density map (red) contoured at the 6.0σ level, the latter calculated after refinement of the final model without the Zn ions. The prime mark on Glu278A indicates that this residue originates from a different monomer than His32 and Asp215.

Fig. 4. Ribbon diagram showing a comparison of the HAP apoenzyme (pink) and its complex with pepstatin A (cyan). The major structural differences are shown in red and blue for the apoenzyme and pepstatin A complex, respectively. The pepstatin A molecule bound in the active site is shown in stick representation.

Fig. 5.

Loop 238–245 in plasmepsins and pepsin. (A) Superposition of loops 238–245 in HAP (cyan), PMII (green), and pepsin (orange), with the main chains shown as ribbons. The side chains of residues comprising these loops are shown in stick representation. (B) Sequence comparison of loop 238–245 in all known *P. falciparum* plasmepsins.

Fig. 6.

Two views comparing the binding mode of pepstatin A to different plasmepsins. (A) The vicinity of residues 287–292. (B) The area of the flap. The inhibitors are shown as sticks, protein main chains as ribbons, and the selected side chains in ball-and-stick representation. HAP is colored cyan, PMII is orange, and PMIV is purple.

Fig. 7. Ribbon diagram showing the superposition of the structures of the HAP apoenzyme (pink) with the pepstatin A (cyan) and KNI-10006 (green) bound complexes. The inhibitor molecules are shown in stick representation.

Fig. 8. Binding of KNI inhibitors in the active sites of HAP (green) and plasmepsins (purple). (A) Different orientation of statine hydroxyls with respect to the catalytic residues in the active sites of HAP and PMIV. The flaps are shown in ribbon representation. (B) Differences in the binding modes of KNI inhibitors in the vicinity of loop 287–292 in the complexes with HAP (green) and PMIV (purple). Residues, occupying equivalent positions in PMII and pepsin are shown in orange and yellow. Residues of identical type are labeled in red.

Fig. 9.

A flap pocket in plasmepsins. (A) Overlay of the inhibitor molecules based on the protein coordinates from their plasmepsin complexes: pepstatin A (cyan) from the HAP complex (this work), KNI-10006 (green) from the HAP complex (this work), and two molecules of an achiral inhibitor (violet and gold) bound to PMII²⁶. (B) Stereo view of the flap pocket in HAP complexed to KNI-10006 (green), superimposed on the structures of PMII in the complexes with three different achiral inhibitors (shown in magenta, blue, and orange, respectively). The residues occupying structurally equivalent positions in the flap pocket in pepsin are shown for comparison (in black).

Fig. 10.

Active site of apo HAP and its complexes with pepstatin A and KNI-100006. The active site residues are shown in stick representation. (A) A close-up view of the active site of apo HAP. The Zn ion (purple) and the important water molecules (pink) are shown as spheres. (B) A comparison of the active sites of apo HAP and the metalloprotease DppA (D-aminopeptidase). (C) A comparison of the active sites of apo HAP (purple) and its complex with pepstatin A (cyan), based on a superposition of protein C α atoms. (D) A comparison of the active sites of pepstatin A complexes of HAP (cyan) and pepsin (orange). This superposition is based on protein C α atoms. (E) Overlay based on a superposition of the identical functional groups in the active sites of HAP (cyan) and pepsin (orange) complexed to pepstatin A. (F) Hydrogen bonding of the statine hydroxyls with the catalytic residues in HAP (cyan) and pepsin (orange). (G) Comparison of the active sites of HAP apoenzyme (pink) and its complexes with pepstatin A (cyan) and KNI-100006 (green). Only the central statine residue of pepstatin A is shown.

Table I
Data collection and refinement statistics

	Type I crystal(apo HAP)	Type II crystal(pepstatin A)	Type III crystal (KNI-10006)
A. Data collection statistics^a			
Space group	<i>P</i> 4 ₁ 2 ₁ 2	<i>P</i> 3 ₂ 21	<i>I</i> 4 ₁ 122
Unit cell parameters <i>a</i> , <i>b</i> , <i>c</i> (Å)	89.8, 89.8, 198.7	70.7, 70.7, 158.7	166.1, 166.1, 276.9
Temperature (K)	100	100	100
Wavelength (Å)	0.99999	1.5418	0.99999
Resolution (Å)	40.0–2.5 (2.6–2.5)	40.0–3.3 (3.4–3.3)	40.0–3.05 (3.2–3.05)
<i>R</i> _{merge} (%) ^b	9.3 (82.2)	20.4 (88.5)	12.3 (139.5)
Completeness (%)	99.7 (99.6)	99.7 (99.5)	99.9 (100)
<i>I</i> / <i>σ</i> (<i>I</i>)	17.3 (3.1)	10.5 (2.2)	16.7 (2.0)
Unique reflections	28895 (3120)	7350 (631)	37179 (4908)
Redundancy	9.5 (9.7)	6.6 (6.7)	12.2 (12.5)
No. of molecules in the asymmetric unit	2	1	4
B. Refinement statistics			
Resolution (Å)	30.0–2.5	20.0–3.3	30.0–3.05
Working set: number of reflections	28018	6939	35291
<i>R</i> _{factor} (%)	22.5	28.5	22.0
Test set: number of reflections	866	366	1857
<i>R</i> _{free} (%)	27.4	35.3	25.2
Protein atoms	5182	2571	10392
Zn ions	6	-	-
Water molecules	104	9	95
C. Geometry statistics			
rmsd (bond distance) (Å)	0.013	0.02	0.01
rmsd (bond angle) (deg.)	1.55	2.25	1.4
Ramachandran plot ^c			
Most favored region (%)	85.4	78.0	84.3
Additionally allowed regions (%)	14.4	21.3	14.1
Generously allowed regions (%)	0.2	0.7	1.4
Disallowed regions (%)	-	-	-
D. PDB code			
	3FNS	3FNT	3FNU

^aThe values in parentheses are for the highest resolution shell

^b $R_{\text{merge}} = \sum_h \sum_i |<I>_h - I_{h,i}| / \sum_h \sum_i I_{h,i}$

^cAs defined by PROCHECK.