JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Crystal Structures of the Histo-Aspartic Protease (HAP) from *Plasmodium falciparum*

Prasenjit Bhaumik[1], Huogen Xiao[2], Charity L. Parr[2], Yoshiaki Kiso[3], Alla Gustchina[1]*, Rickey Y. Yada[2] and Alexander Wlodawer[1]

[1]*Protein Structure Section, Macromolecular Crystallography Laboratory, National Cancer Institute, Frederick, MD 21702, USA*

[2]*Department of Food Science, University of Guelph, Guelph, Ontario, Canada N1G 2W1*

[3]*Department of Medicinal Chemistry, Center for Frontier Research in Medicinal Science, Kyoto Pharmaceutical University, Yamashina-ku, Kyoto 607-8412, Japan*

Edited by G. Schulz

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 resolutions, respectively. In the apoenzyme crystals, HAP forms a tight dimer not seen previously in any aspartic protease. 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 (residues 238–245), as well as the orientation of 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 from that 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 with 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.

Published by Elsevier Ltd.

Introduction

Histo-aspartic protease (HAP) is one of the 10 plasmepsins (PMs) identified in the genome of *Plasmodium falciparum*, the parasite responsible for the most widespread form of malaria. It is one of the 4 PMs residing in the food vacuole of the parasite that are involved in degradation of human hemoglobin,[1] making them potential targets for novel antimalarial therapy.[2] Despite its high sequence identity (60%) with other PMs (PMI, PMII, and PMIV), which are typical pepsin-like aspartic proteases, the active site of HAP contains several significant deviations from the pepsin standard.[3] In particular, Asp32†, 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 and 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

---

*Corresponding author. E-mail address: alla@ncifcrf.gov.
Abbreviations used: HAP, histo-aspartic protease; PM, plasmepsin; NCS, non-crystallographic symmetry; mtHAP, mature tHAP; PEG, polyethylene glycol; TLS, translation–libration–screw.

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

have been subjects of considerable disagreement. In the model published by Andreeva *et al.*,[4] 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 Aqvist[5] 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 parasite[1] have not been sufficient for structural studies. Initial efforts to clone and express HAP yielded inactive enzyme only.[3] Although the specific activity of the enzyme with a fluorescent substrate optimized for PMI and PMII was comparatively low and the mode of inhibition and pH activity profile were not consistent between different purification protocols, we subsequently succeeded in producing active recombinant HAP.[6,7] In this article, we report the crystal structures of recombinant HAP in ligand-free form and in complex with two inhibitors, pepstatin A[8] and KNI-10006.[9] The results do not directly support either of the proposed mechanisms and raise a number of questions that will require further studies.

Results and Discussion

Overall fold and structure quality

Crystal structures of uncomplexed HAP and 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 resolutions, respectively. The fit of the final models to electron density maps is satisfactory, and the quality of the structures, as measured by such parameters as the *R*-factor or departure from stereochemical standards,[10] 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 predicted,[3] 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, and cathepsin D,[13] the fungal enzyme endothiapepsin,[14] and the *Plasmodium* enzymes PMII[15] and PMIV.[16] 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 interdomain β-sheet that serves to suture the domains together. The conserved sequence DT(S)G, present in one copy in

each domain and containing the catalytic aspartate residues, is the signature motif of aspartic proteases.[13] 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 is 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 the pepstatin A complex of HAP were compared using ALIGN[17] with the structures of unliganded pepsin [Protein Data Bank (PDB) ID 4PEP] and its pepstatin A complex (PDB ID 1PSO). The pepstatin A complex of HAP was also compared with the pepstatin A complexes of PMII (PDB ID 1XDH) and PMIV (PDB ID 1LS5). The four superpositions based on the C^α atoms gave rmsd values 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 2-fold axis and can be superimposed with an rmsd of 0.32 Å between the corresponding C^α 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 the first monomer. Asojo *et al.*[15] reported several dimeric forms of PMs created by crystallographic symmetry and non-crystallographic symmetry (NCS). However, the type of tight non-crystallographic dimer seen in the crystals of HAP apoenzyme is unique not only among PMs 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.
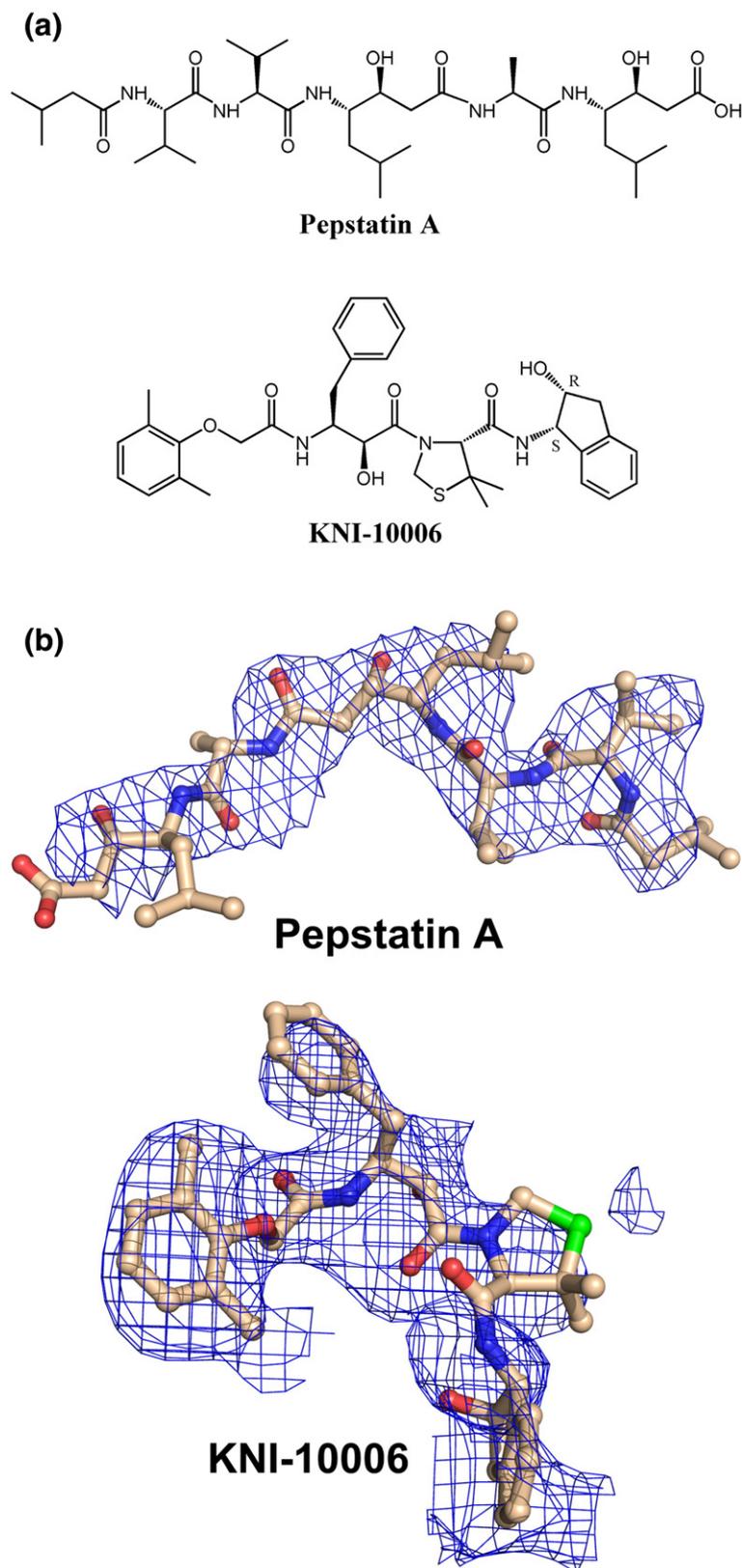
(a)

Pepstatin A

KNI-10006

(b)

Pepstatin A

KNI-10006

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

The surface area buried upon dimerization is 2284 Å^2 for each monomer, about twice the area buried in the majority of reported dimers of PMII (1270 Å^2; PDB ID 1XE6) and PMIV (1137 Å^2, PDB ID 1LS5). The only comparable buried surface area (2140 Å^2; Ref. 18) was reported for the crystal-lographic 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 PMs.[15] 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 discussed below, we are not certain that this is indeed the case.

Four additional Zn ions are found on the surface of the HAP dimer. Two of them are located in equi-valent positions in each monomer and are coordi-nated 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 and by Glu57
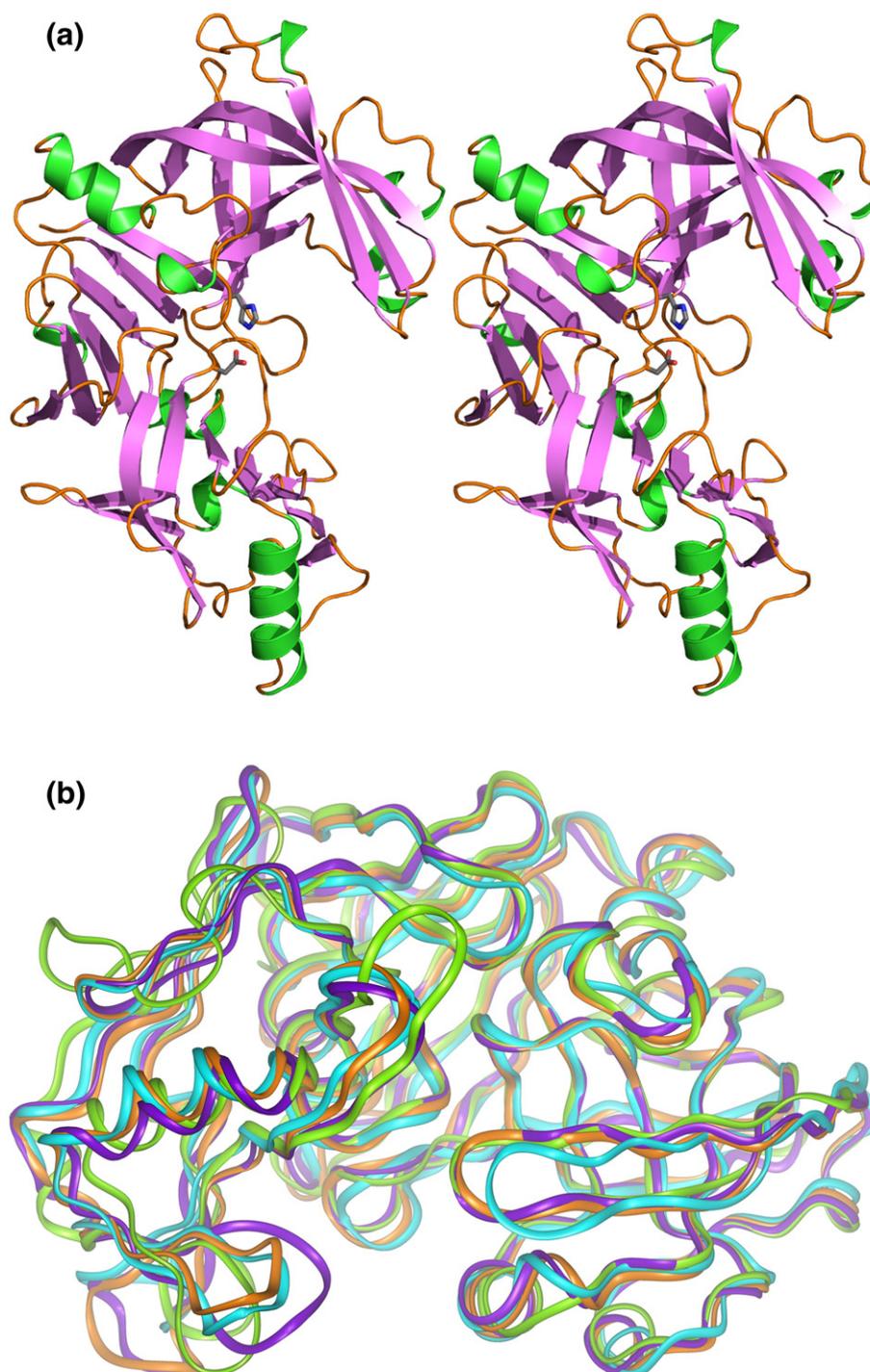

Fig. 2. Overall structure of HAP and a comparison with PMs. (a) Stereo ribbon diagram showing the monomer of HAP apoenzyme. Two active-site residues, His32 and Asp215, are shown in stick representation. The N-terminal domain is on top, and the C-terminal domain is at the bottom. (b) Ribbon diagram of 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 shown in white on red background, limited identies are in red, similarities are in blue, and differences are in black. The signature motifs of aspartic proteases are boxed.
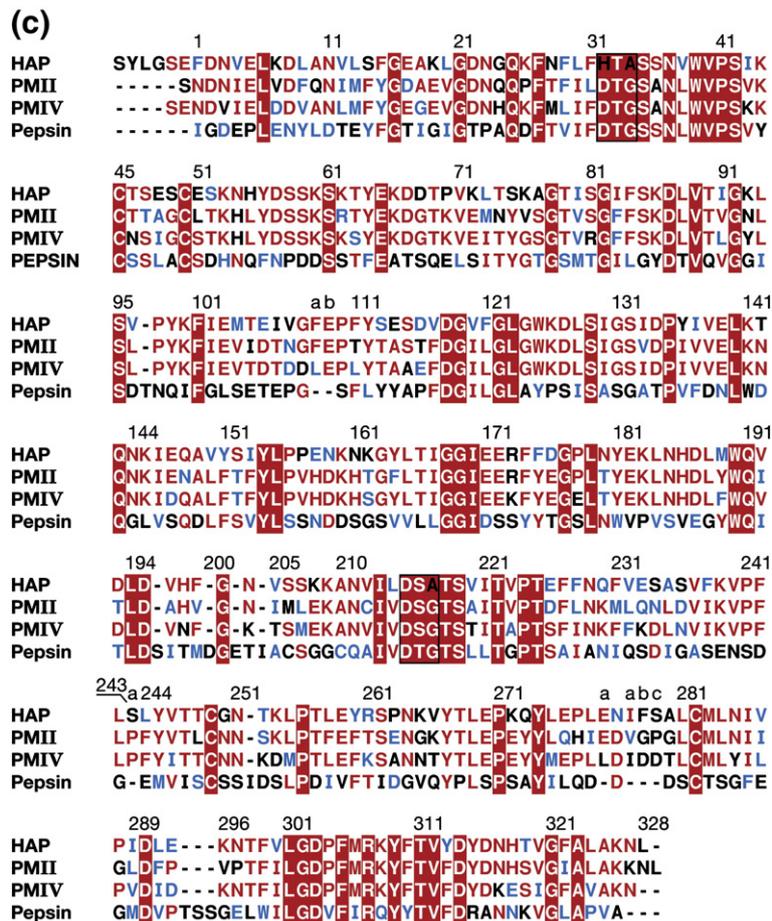
(c)

```
              1         11        21        31        41
HAP     SYLGSEFDNVELKDLANVLSFGEAKLGDNGQKFNLFHTASSNVWVPSIK
PMII    -----SNDNIELVDFQNIMFYGDAEVGDNQQPFTFILDTGSANLWVPSVK
PMIV    ----SENDVIELDDVANLMFYGEGEVGDNHQKFMLIFDTGSANLWVPSKK
Pepsin  ------IGDEPLENYLDTEYFGTIGIGTPAQDFTVIFDTGSSNLWVPSVY

        45        51        61        71        81        91
HAP     CTSESCESKNHYDSSKSKTYEKDDTPVKLTSKAGTISGIFSKDLVTIGKL
PMII    CTTAGCLTKHLYDSSKSRTYEKDGTKVEMNYVSGTVSGFFSKDLVTVGNL
PMIV    CNSIGCSTKHLYDSSKSKSYEKDGTKVEITYGSGTVRGFFSKDLVTLGYL
PEPSIN  CSSLACSDHNQFNPDDSSTFEATSQELSITYGTGSMTGILGYDTVQVGGI

        95        101     ab 111       121       131       141
HAP     SV-PYKFIEMTEIVGFEPFYSESDVDGVFGLGWKDLSIGSIDPYIVELKT
PMII    SL-PYKFIEVIDTNGFEPTYTASTFDGILGLGWKDLSIGSVDPIVVELKN
PMIV    SL-PYKFIEVTDTDDLEPLYTAAEFDGILGLGWKDLSIGSIDPIVVELKN
Pepsin  SDTNQIFGLSETEPG--SFLYYAPFDGILGLAYPSISASGATPVFDNLWD

        144       151       161       171       181       191
HAP     QNKIEQAVYSIYLPPENKNKGYLTIGGIEERFFDGPLNYEKLNHDLMWQV
PMII    QNKIENALFTFYLPVHDKHTGFLTIGGIEERFYEGPLTYEKLNHDLYWQI
PMIV    QNKIDQALFTFYLPVHDKHSGYLTIGGIEEKFYEGELTYEKLNHDLFWQV
Pepsin  QGLVSQDLFSVYLSSNDDSGSVVLLGGIDSSYYTGSLNWVPVSVEGYWQI

        194    200   205    210        221       231       241
HAP     DLD-VHF-G-N-VSSKKANVILDSATSVITVPTEFFNQFVESASVFKVPF
PMII    TLD-AHV-G-N-IMLEKANCIVDSGTSAITVPTDFLNKMLQNLDVIKVPF
PMIV    DLD-VNF-G-K-TSMEKANVIVDSGTSTITAPTSFINKFFKDLNVIKVPF
Pepsin  TLDSITMDGETIACSGGCQAIVDTGTSLLTGPTSAIANIQSDIGASENSD

        243 a244     251       261       271       a abc 281
HAP     LSLYVTTCGN-TKLPTLEYRSPNKVYTLEPKQYLEPLENIFSALCMLNIV
PMII    LPFYVTLCNN-SKLPTFEFTSENGKYTLEPEYYLQHIEDVGPGLCMLNII
PMIV    LPFYITTCNN-KDMPTLEFKSANNTYTLEPEYYMEPLLDIDDTLCMLYIL
Pepsin  G-EMVISCSSIDSLPDIVFTIDGVQYPLSPSAYILQD-D---DSCTSGFE

        289    296   301        311        321       328
HAP     PIDLE---KNTFVLGDPFMRKYFTVYDYDNHTVGFALAKNL-
PMII    GLDFP---VPTFILGDPFMRKYFTVFDYDNHSVGIALAKKNL
PMIV    PVDID---KNTFILGDPFMRKYFTVFDYDKESIGFAVAKN--
Pepsin  GMDVPTSSGELWILGDVFIRQYYTVFDRANNKVGLAPVA---
```

Fig. 2 (*legend on previous page*)

Comparison between apo HAP and its pepstatin A complex

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.

The complex of HAP with pepstatin A crystallized with one protein molecule in the asymmetric unit. Despite the overall similarity to the apoenzyme structure (rmsd values of 0.98 and 0.96 Å for the C^α atoms relative to monomers A and B of apo HAP, respectively), 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 C^α 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 re-

markable flexibility of this part of the structure. In the crystal structure of HAP complexed with pepstatin A, helix 225–235 is not involved in any crystal contact, whereas 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 in 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 that of several members of the PM family.[25]
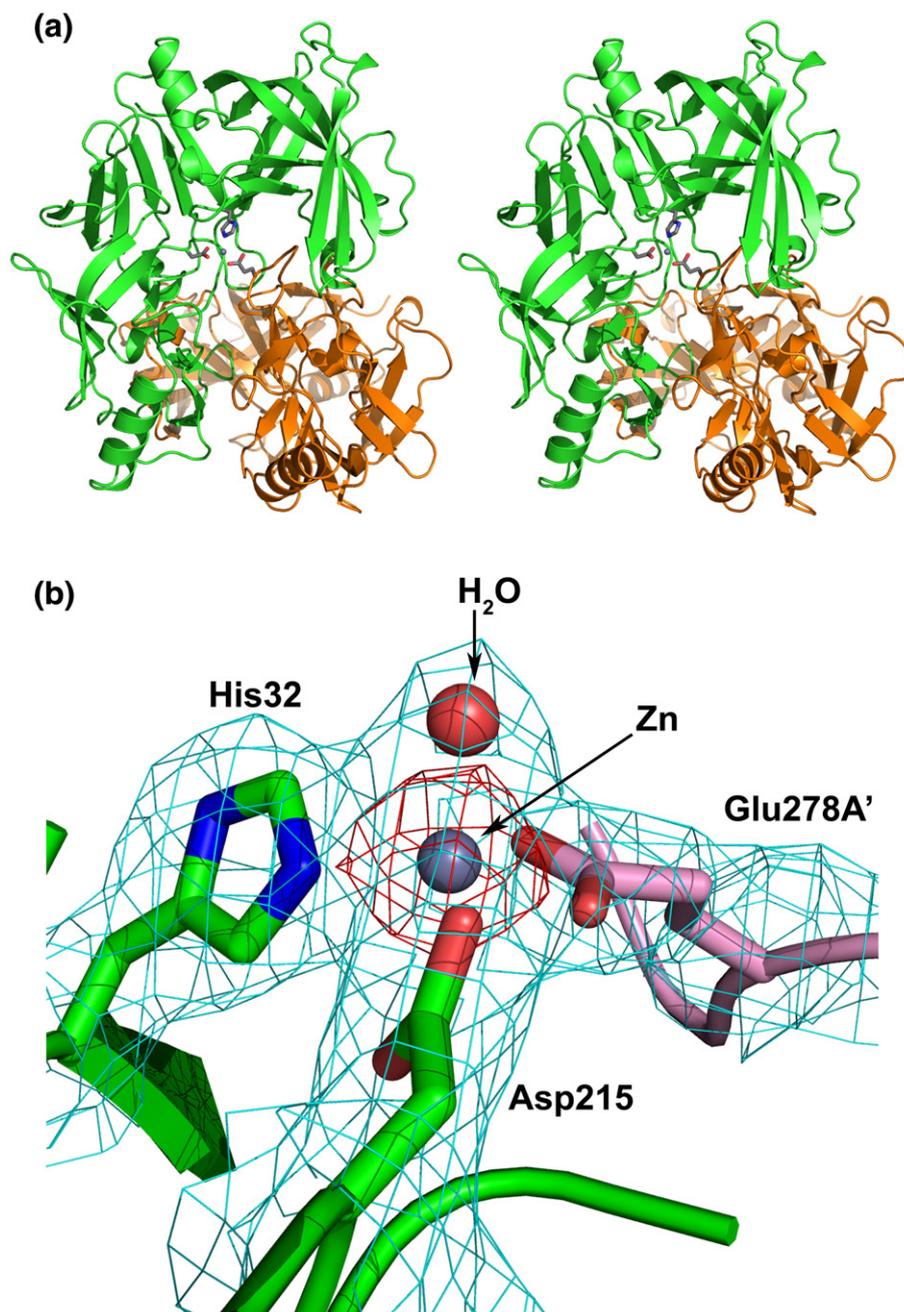
Fig. 3. Structure of the apoenzyme form of HAP. (a) A ribbon diagram of the HAP dimer, with the side chains in one of the active sites shown in stick representation. The Zn ion bound in the active site is shown as a sphere. (b) The residues coordinating the Zn ion covered by $2F_o-F_c$ electron density map (blue) contoured at the 1.0σ level and by F_o-F_c omit electron density map (red) contoured at the 6.0σ level, with 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.

Loop 238–245 in PMs has a sequence pattern notably different from its counterpart in other aspartic proteases. In other PMs that are found in the digestive vacuole of *P. falciparum* and other *Plasmodium* species, the strictly hydrophobic nature of this loop is conserved (Fig. 5a and b). However, in PMV to PMX, which are not involved in hemoglobin degradation, the sequences of this loop are significantly different. It was proposed in several studies that, in PMs, this loop is involved in distal interactions with the substrate that enhance the enzymatic activity.[25,26] Mutation of hydrophobic residues (Phe244) and 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 a remarkable flexi-
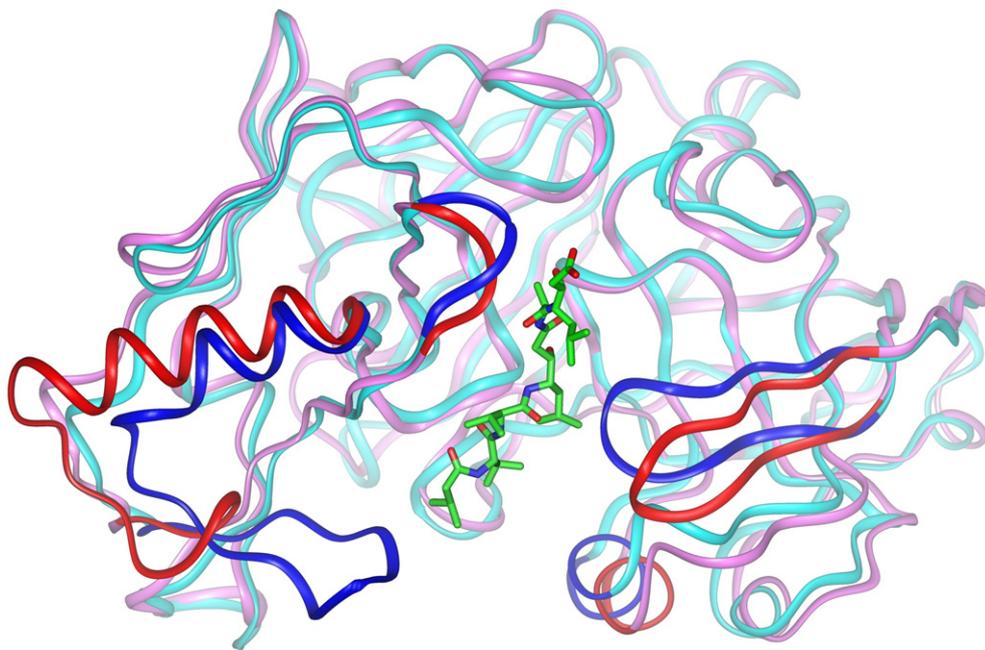
Fig. 4. Ribbon diagram showing a comparison of the HAP apoenzyme (pink) and its complex with pepstain A (cyan). The major structural differences are shown in red and blue for the apoenzyme and pepstatin A complex, respectively. The pepstatin A molecule bound in the active site is shown in stick representation.

bility 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 (PDB ID 1XDH) and PMIV (PDB ID 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 PMs and other pepsin-like proteases. Instead of wrapping around the flap, as observed with other enzymes, it is oriented toward 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 the P3′ position of pepstatin A (Fig. 6b). The ω-ammonium group of Lys76 is linked *via* a hydrogen bond to the carbonyl oxygen of Ala at the P2′ position of pepstatin A. A charged residue with such a long side chain at this position in HAP (*versus* glycine in PMIV or valine in PMII) is unique among PMs 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 $O^{\gamma 1}$ 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 PMII[15] due to a different orientation of the C-terminal half of the inhibitor that 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 the P2′ position is involved in hydrophobic interactions with Met189 and Leu291. Finally, the side chain of Glu292 helps 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, which contains a hydroxymethylcarbonyl isostere. It has been also reported that KNI compounds demonstrated effective inhibition of not only HIV-1 protease but also PMII and HTLV-1 protease.[31] Further chemical elaboration and optimization of the KNI inhibitors, supported by extensive
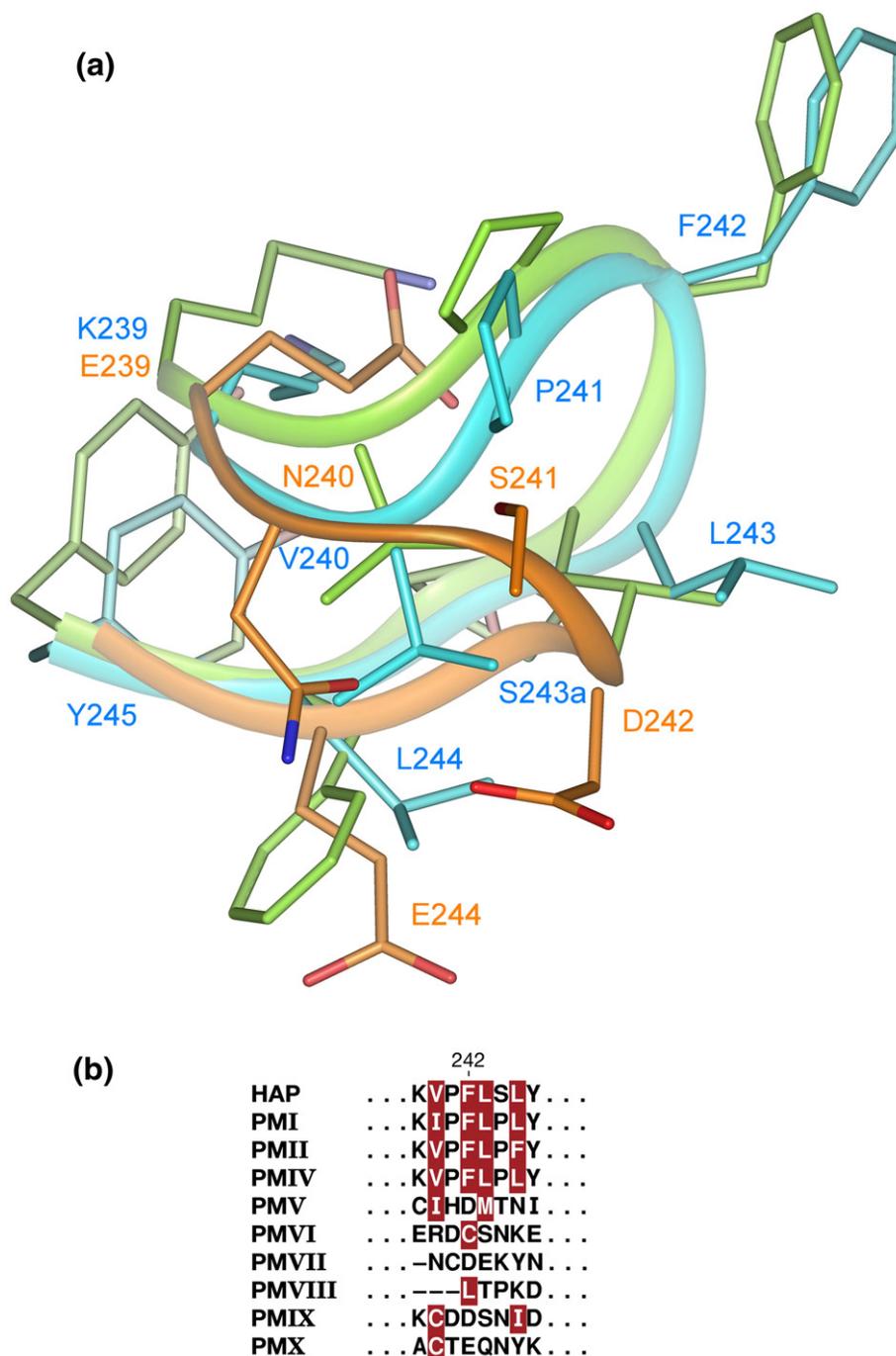
Fig. 5. Loop 238–245 in PMs and pepsin. (a) Superposition of loop 238–245 in HAP (cyan), PMII (green), and pepsin (orange), with the main chains shown as ribbons. The side chains of residues comprising these loops are shown in stick representation. (b) Sequence comparison of loop 238–245 in all known *P. falciparum* PMs.

structural and biochemical data, resulted in the synthesis of many new compounds shown to be potent inhibitors of HIV-1 and HTLV-1 proteases.[9,32–36] One of them, KNI-10006, was subsequently shown to be a potent inhibitor of HAP[37] with an IC_{50} of 0.69 μM[9] 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 structure of either HIV-1 protease[38] or PMIV,[15] 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 contact either with loop 287–292 or with several other hydrophobic residues, conserved in PMs and in other pepsin-like enzymes, that were shown to anchor KNI-764 in a PMIV complex[16] (Fig. 8b).
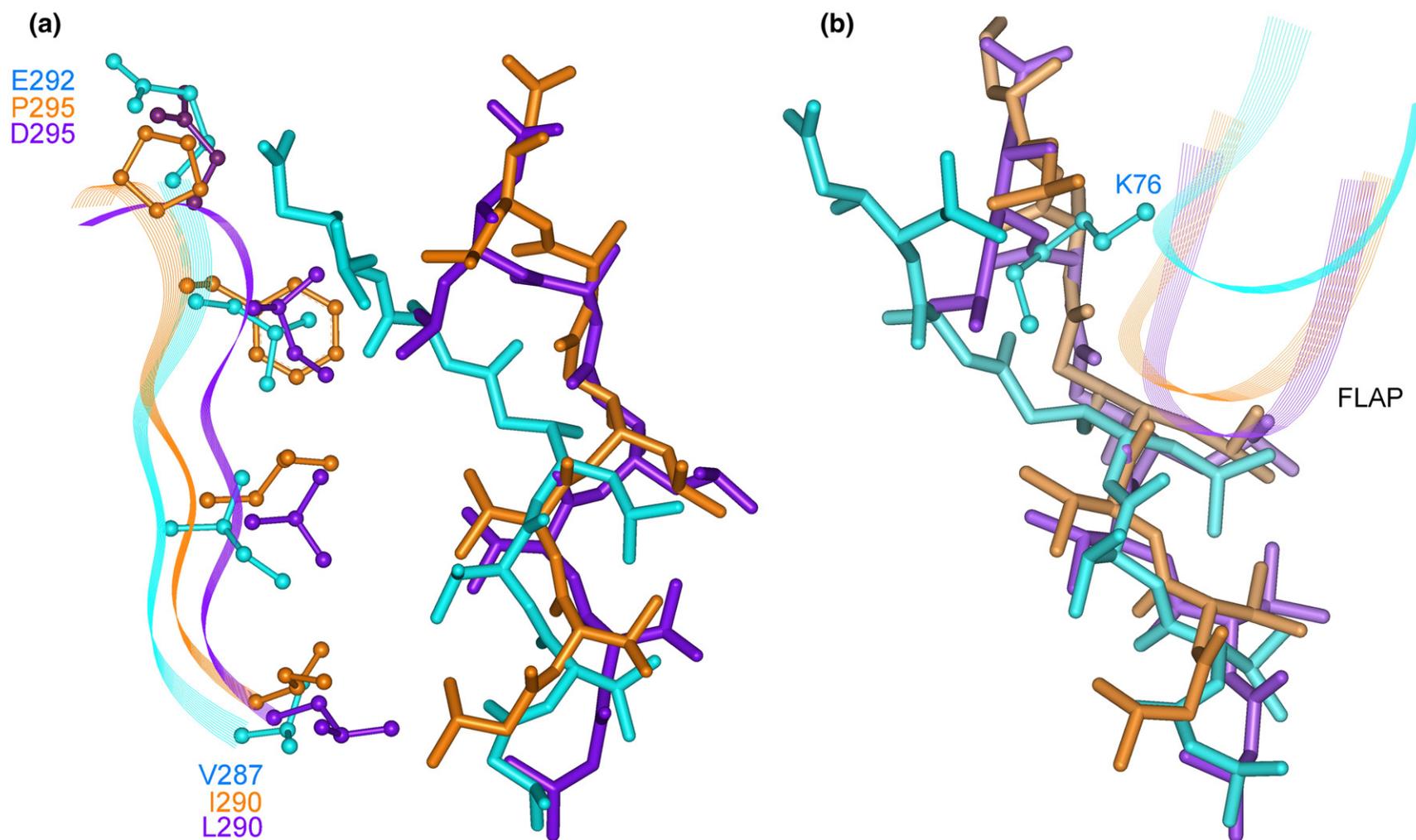
Fig. 6. Two views comparing the binding mode of pepstatin A with different PMs. (a) The vicinity of residues 287–292. (b) The area of the flap. The inhibitors are shown as sticks, protein main chains are shown as ribbons, and the selected side chains are shown in ball-and-stick representation. HAP is shown in cyan, PMII is in orange, and PMIV is in purple.
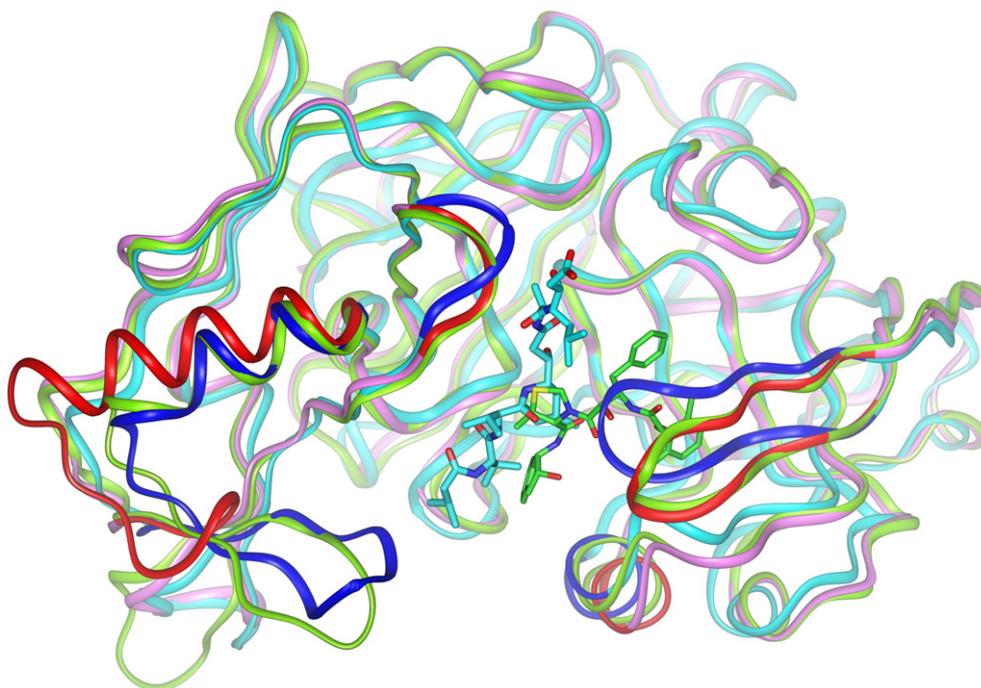
Fig. 7. Ribbon diagram showing the superposition of the structures of the HAP apoenzyme (pink) with the pepstatin A (cyan) and KNI-10006 (green) bound complexes. The inhibitor molecules are shown in stick representation.

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 (PDB ID 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 moiety of KNI-10006 occupy the so-called flap pocket. The existence of such a pocket was noted before in human renin.[39] 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 (PDB IDs 2BJU, 2IGX, and 2IGY)[26,40,41] allows for the description of the residues comprising this binding pocket in the PM family and for comparison with other aspartic proteases (Fig. 9b). The flap pocket is predominantly hydrophobic in PMs 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 PMs. 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 residue 109A in PMs, thus contributing to the interactions with the ligand and partially compen-

sating 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 PMs 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 domain[42] 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 (*versus* Asp32). As 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 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. Asojo *et al.*[15] reported on a crystallographic dimer of PMII in which the loop containing residues 237–247 is docked into the binding cavity of the 2-fold-related monomer. However, the packing of the loop in the active site is not as tight as that 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 the 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 $O^{\gamma 1}$ of Thr218 and $O^{\delta 2}$ 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 $N^{\varepsilon 2}$ atom. The imidazole ring of His32 is fixed in suitable orientation by a $N^{\delta 1}...O^{\gamma}$ hydrogen bond with the side chain of Ser35, the latter also being hydrogen bonded to $N^{\varepsilon 1}$ 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 ID 1HI9)[47] (Fig. 10b), the lack of inhibition by chelating agents such as ethylenediaminetetraacetic acid indicates that HAP does not function as a metalloprotease,[1] at least toward 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 PMII and PMIV.[15] The statine hydroxyl group of pepstatin A is positioned between the two catalytic residues and at hydrogen-bonding distances of the $N^{\varepsilon 2}$ and $O^{\delta 2}$ 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 PMII[15] and PMIV,[16] as well as of pepsin.[48] Closing of the flap is important for the catalytic mechanism in

pepsin-like aspartic proteases.[44] 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 (PDB ID 1PSO), based on the C^{α} 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, $N^{\varepsilon 2}$ of His32 is close to the $O^{\delta 1}$ 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 and 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 (*versus* aspartate). This shift is determined by the distance between $O^{\delta 1}$ of Asp32 and $N^{\varepsilon 2}$ of His32 in the former superposition and roughly corresponds to the distance between $C^{\delta 2}$ and $N^{\varepsilon 2}$ atoms in the ring of a histidine side chain. A 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, we superposed the C^{α} atoms of these three structures (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 that in the HAP–KNI-10006 complex are 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 $N^{\varepsilon 2}$ of His32 and $O^{\delta 2}$ of Asp215. By analogy with pepsin-like aspartic proteases,[13] 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 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

Although the structure of HAP from *P. falciparum* exhibits the expected overall similarity to the structure of pepsin-like aspartic proteases, and to

PMs 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 ion, 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 a pH level well above the range in which the enzyme is active. Thus, the
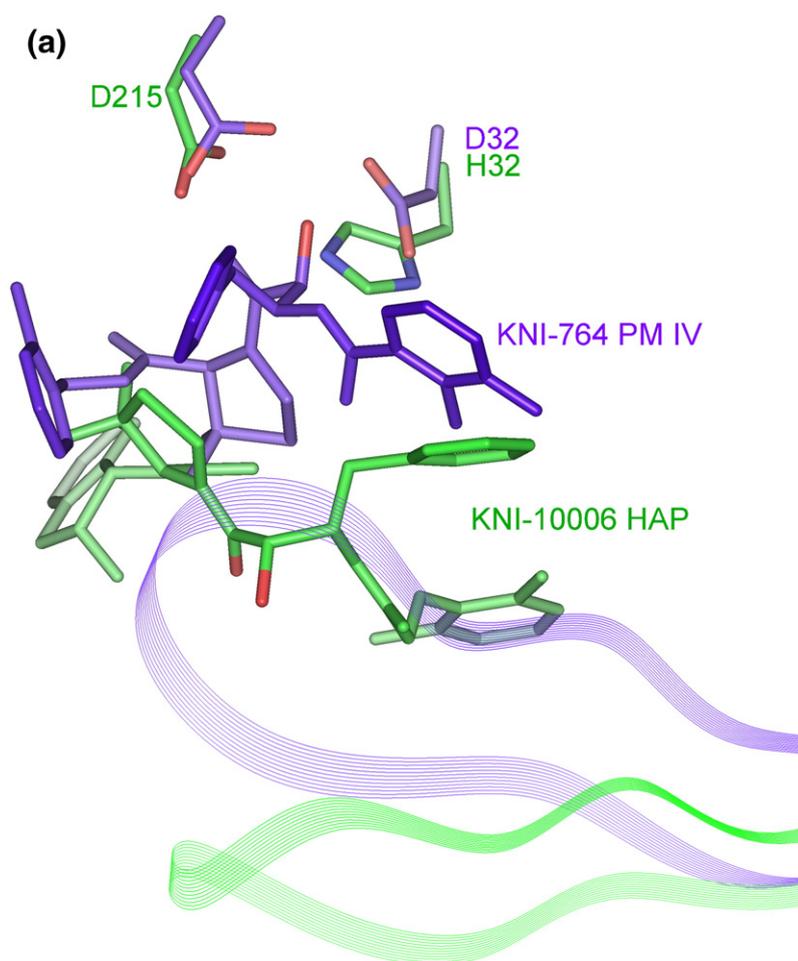

Fig. 8. Binding of KNI inhibitors in the active sites of HAP (green) and PMs (purple). (a) Different orientation of statine hydroxyls with respect to the catalytic residues in the active sites of HAP and PMIV. The flaps are shown in ribbon representation. (b) Differences in the binding modes of KNI inhibitors 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.
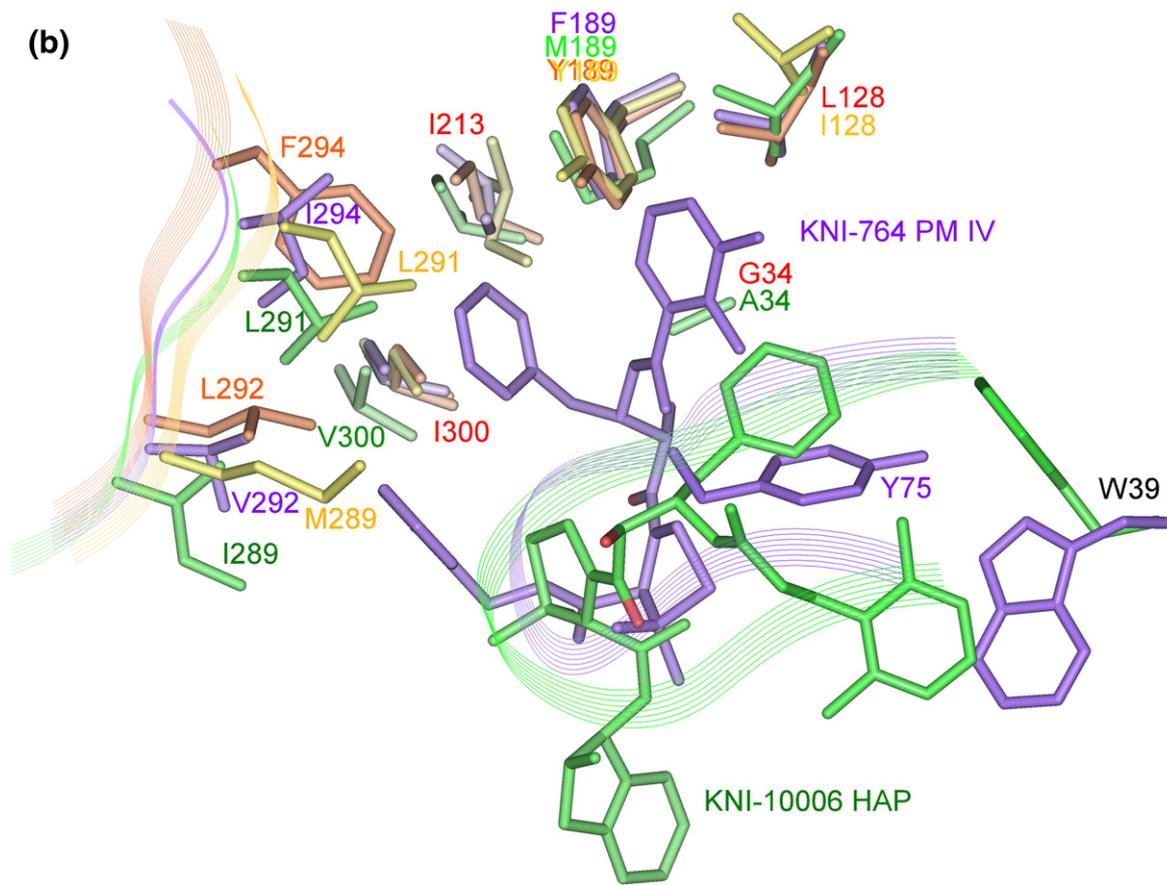
Fig. 8 (*legend on previous page*)

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 PMs 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 PMs. Such compounds could be developed into a new class of antimalarial drugs.

Materials and Methods

Expression and purification of HAP

Expression of the recombinant Trx-tHAP fusion protein was conducted according to a method described previously.[6] Cell pellet of 1-l culture was suspended in 50 ml of potassium phosphate buffer, pH 7.5, containing 1× BugBuster™ reagent (Novagen, Madison WI, USA) and 1 µl 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 16,000*g* for 20 min at 4 °C. The supernatant was applied on a HisSelect™ Cartridge (6.4 ml) (Sigma-Aldrich, Oakville, Ontario, Canada); the column was first flushed with washing buffer (50 mM sodium phosphate, 0.3 M NaCl, and 10 mM imidazole, pH 7.5) and then with 10% elution buffer (50 mM sodium phosphate, 0.3 M NaCl, and 250 mM imidazole, pH 7.5), and it was finally eluted with

50% elution buffer. The eluate was concentrated in 50 mM sodium phosphate buffer, pH 7.5, containing 0.2% Chaps in an Ultracel YM50 Centricon (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 Trx-tHAP fusion 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 Superose™ 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 previously[6] 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*.
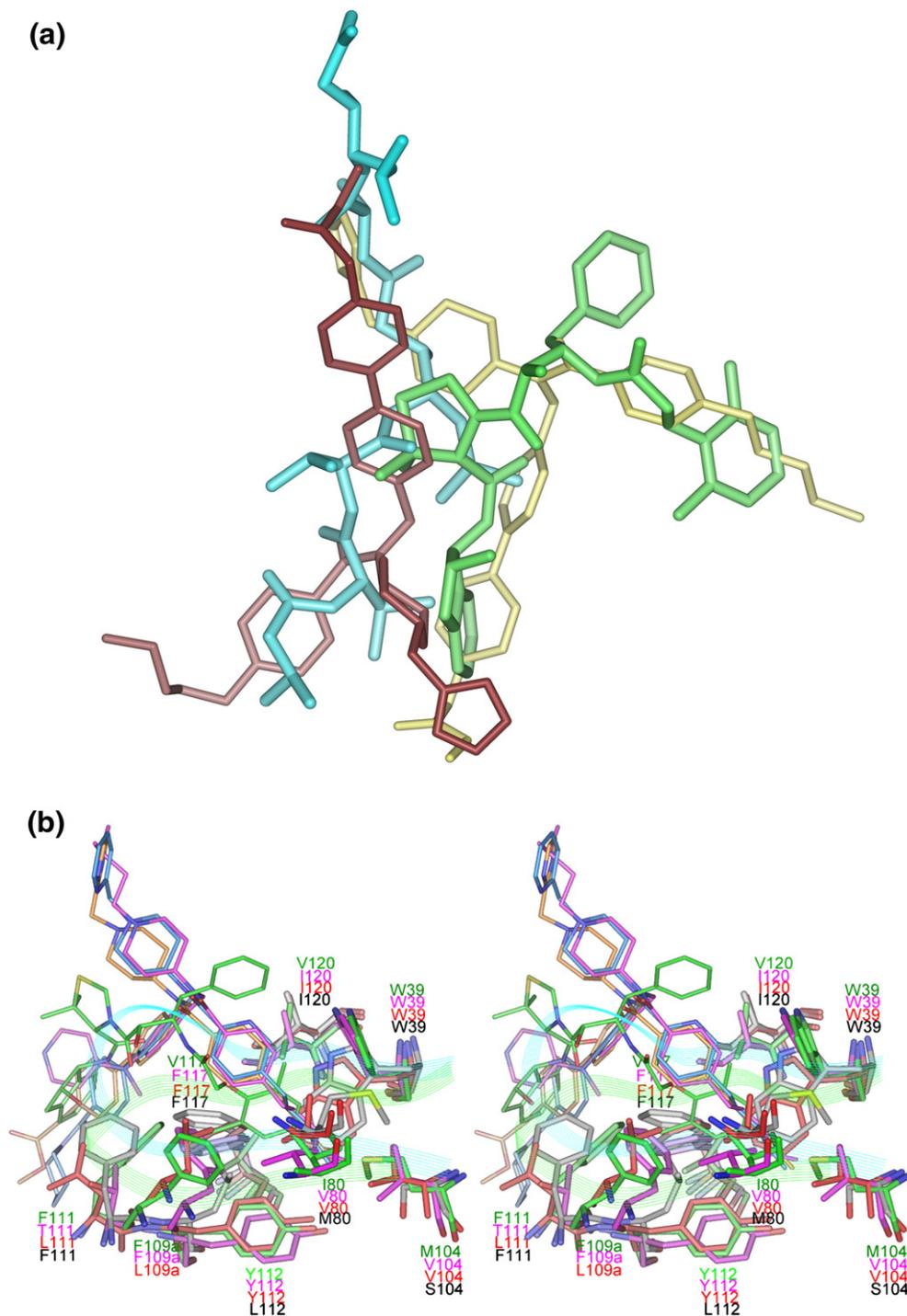
Fig. 9. The flap pocket in PMs. (a) Overlay of the inhibitor molecules based on the protein coordinates from their PM complexes: pepstatin A (cyan) from the HAP complex (this work), KNI-10006 (green) from the HAP complex (this work), and two molecules of an achiral inhibitor (violet and gold) bound to PMII (Ref. 26). (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 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).

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
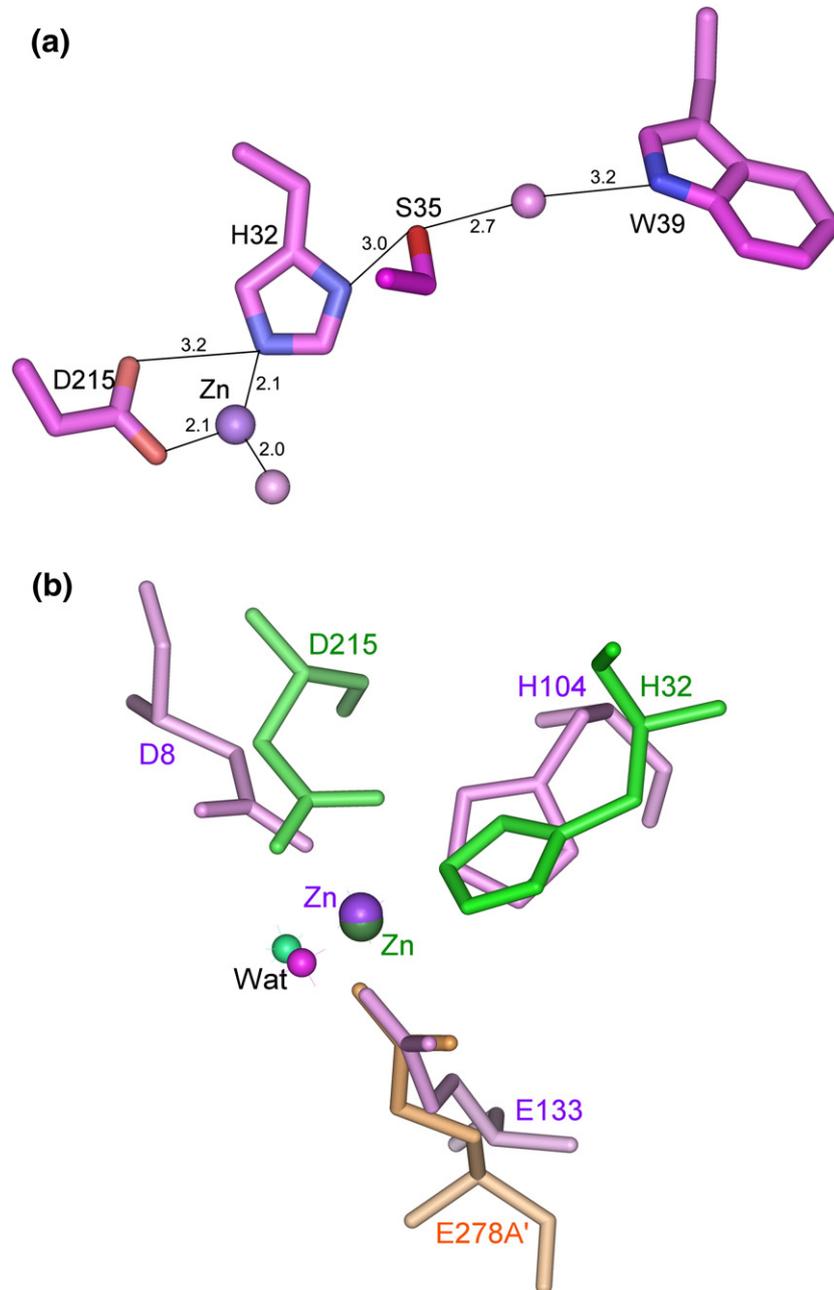
(1:1 protein/inhibitor molar ratio). The sample was subsequently incubated for 2 h and centrifuged. Several crystallization screens were set up using the sitting-drop vapor-diffusion method at 293 K, and a few conditions produced small needles. The conditions were optimized using the hanging-drop vapor-diffusion method at 293 K. Two types of crystals (type I and type II) were grown

under two crystallization conditions. Type I crystals appeared in drops containing 0.7 μl of protein solution and 0.7 μl of reservoir solution, equilibrated against 1 ml of reservoir solution. The reservoir solution contained 10% PEG (polyethylene glycol) 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 μl of protein solution and 0.4 μl of 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

Fig. 10. Active site of apo HAP and its complexes with pepstatin A and KNI-10006. 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-10006 (green). Only the central statine residue of pepstatin A is shown.
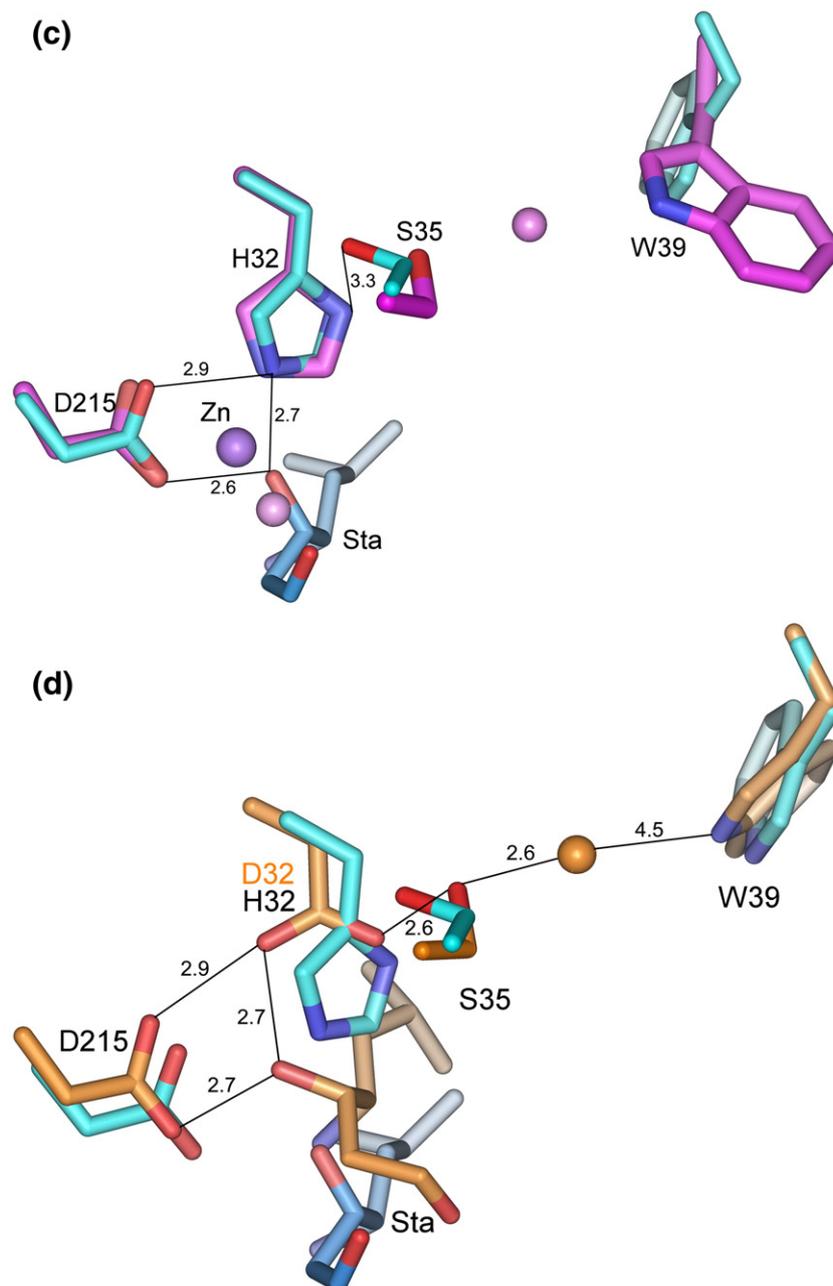
(c)

(d)

Fig. 10 (*legend on previous page*)

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 the hanging-drop vapor-diffusion method at 293 K in drops mixed from 0.8 μl of protein solution and 0.8 μl of reservoir solution containing 0.2 M KH_2PO_4 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 SER-CAT (Southeast Regional Collaborative Access Team) 22-ID beamline at the Advanced Photon Source (Argonne National Laboratory,

Argonne, IL). Data for a type II crystal were collected using Cu $K\alpha$ 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) methyl-2,4-pentanediol 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 $P4_12_12$, type II crystals are trigonal in space group $P3_221$, and type III crystals are tetragonal in space group $I4_122$. All data sets were indexed and integrated using XDS.[49] The data were converted to structure factors with modules F2MTZ and CAD of CCP4.[50]

Structure solution and refinement

HAP has a high level of sequence identity with PMII (60.6%) and PMIV (61.8%). Several crystal structures of
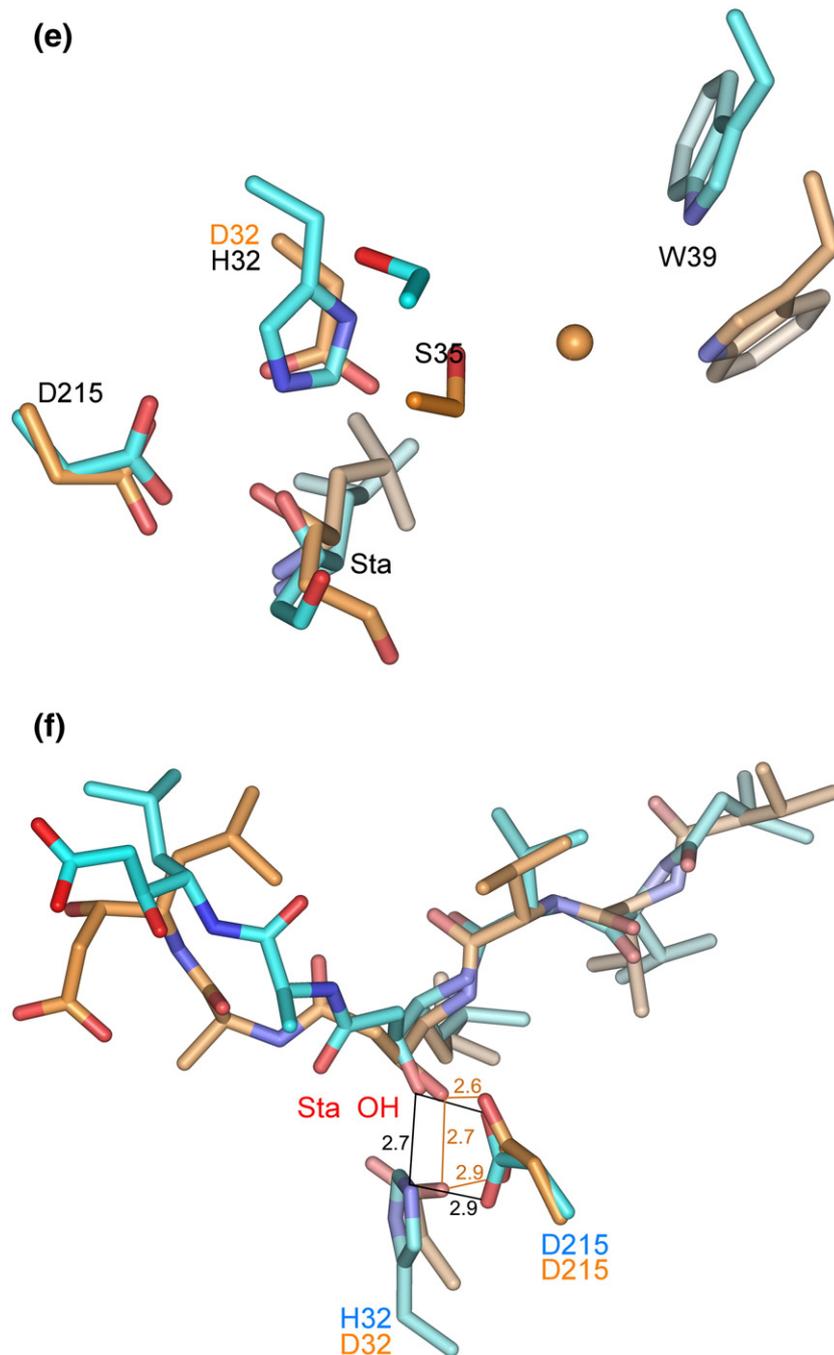
(e)

(f)

Fig. 10 (*legend on page 534*)

both PMs are available in the 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 MrBUMP,[51] MOLREP,[52] and PHASER.[53]

The Matthews coefficient[54] 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 ID 2ANL) as the best search model. The final molecular replacement procedures utilized directly the programs MOLREP[52] and PHASER.[53] The preliminary model was refined with REFMAC5[55] and rebuilt with COOT.[56] 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 3σ in $(F_o-F_c)\alpha_c$ σ_a-weighted maps while monitoring the decrease of R_{free}. Proper hydrogen bonding was required for placement of all solvent molecules. PROCHECK[10] 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 and 328) are missing from each subunit in the final model of the apoenzyme; these residues could not be built because of lack
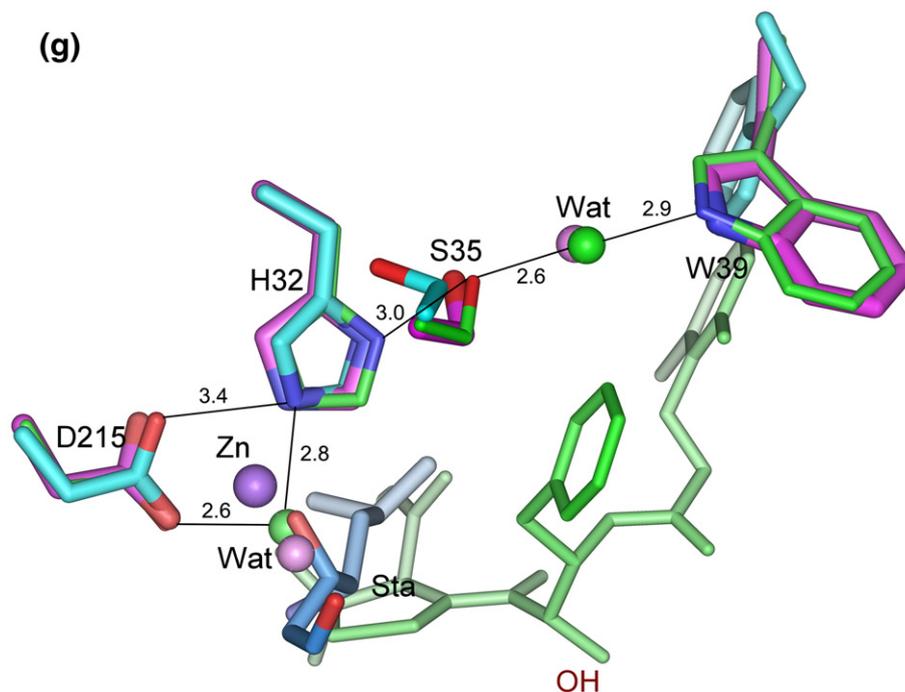
(g)

Fig. 10 (*legend on page 534*)

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 ID 1SME) as the probe, correctly placed the model in the asymmetric unit. After the first cycle of refinement in REFMAC5, the $(F_o - F_c)\alpha_c$ 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 translation–libration–screw (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 and 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 $(F_o - F_c)\alpha_c$ 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 1. The structures were analyzed using PROCHECK[10] and COOT.[56] Structural superpositions were performed with SSM[57] and ALIGN,[17] and the figures were generated with PYMOL.[58]

PDB accession numbers

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

Table 1. Data collection and refinement statistics

	Type I crystal (apo HAP)	Type II crystal (pepstatin A)	Type III crystal (KNI-10006)
PDB code	3FNS	3FNT	3FNU
Data collection[a]			
Space group	$P4_12_12$	$P3_221$	$I4_122$
Unit cell parameters *a, b, c* (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)
R_{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)
$I/\sigma(I)$	17.3 (3.1)	10.5 (2.2)	16.7 (2.0)
Unique reflections	28,895 (3120)	7350 (631)	37,179 (4908)
Redundancy	9.5 (9.7)	6.6 (6.7)	12.2 (12.5)
No. of molecules in the asymmetric unit	2	1	4
Refinement statistics			
Resolution (Å)	30.0–2.5	20.0–3.3	30.0–3.05
Working set: no. of reflections	28,018	6939	35,291
R-factor (%)	22.5	28.5	22.0
Test set: no. of reflections	866	366	1857
R_{free} (%)	27.4	35.3	25.2
Protein atoms	5182	2571	10,392
Zn ions	6	—	—
Water molecules	104	9	95
Geometry statistics			
rmsd			
Bond distance (Å)	0.013	0.02	0.01
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	—	—	—

[a] The values in parentheses are for the highest-resolution shell.
[b] $R_{merge} = \sum_h \sum_i |\langle I \rangle_h - I_{h,i}| / \sum_h \sum_i I_{h,i}$.
[c] As defined by PROCHECK.

References

1. 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.
2. Ersmark, K., Samuelsson, B. & Hallberg, A. (2006). Plasmepsins as potential targets for new antimalarial therapy. *Med. Res. Rev.* **26**, 626–666.
3. Berry, C., Humphreys, M. J., Matharu, P., Granger, R., Horrocks, P., Moon, R. P. *et al.* (1999). A distinct member of the aspartic proteinase gene family from the human malaria parasite *Plasmodium falciparum*. *FEBS Lett.* **447**, 149–154.
4. Andreeva, N., Bogdanovich, P., Kashparov, I., Popov, M. & Stengach, M. (2004). Is histoaspartic protease a serine protease with a pepsin-like fold? *Proteins*, **55**, 705–710.
5. Bjelic, S. & Aqvist, J. (2004). Computational prediction of structure, substrate binding mode, mechanism, and

rate for a malaria protease with a novel type of active site. *Biochemistry*, **43**, 14521–14528.
6. Xiao, H., Sinkovits, A. F., Bryksa, B. C., Ogawa, M. & Yada, R. Y. (2006). Recombinant expression and partial characterization of an active soluble histo-aspartic protease from *Plasmodium falciparum*. *Protein Expression Purif.* **49**, 88–94.
7. Parr, C. L., Tanaka, T., Xiao, H. & Yada, R. Y. (2008). The catalytic significance of the proposed active site residues in *Plasmodium falciparum* histoaspartic protease. *FEBS J.* **275**, 1698–1707.
8. Umezawa, H., Aoyagi, T., Morishima, H., Matsuzaki, M. & Hamada, M. (1970). Pepstatin, a new pepsin inhibitor produced by actinomycetes. *J. Antibiot. (Tokyo)*, **23**, 259–262.
9. Nezami, A., Kimura, T., Hidaka, K., Kiso, A., Liu, J., Kiso, Y. *et al.* (2003). High-affinity inhibition of a family of *Plasmodium falciparum* proteases by a designed adaptive inhibitor. *Biochemistry*, **42**, 8459–8464.
10. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
11. Jaskolski, M., Gilski, M., Dauter, Z. & Wlodawer, A. (2007). Stereochemical restraints revisited: how accurate are refinement targets and how much should protein structures be allowed to deviate from them? *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **63**, 611–620.
12. Wlodawer, A., Minor, W., Dauter, Z. & Jaskolski, M.

(2008). Protein crystallography for non-crystallographers or how to get the best (but not more) from the published macromolecular structures. *FEBS J.* **275**, 1–21.

13. Dunn, B. M. (2002). Structure and mechanism of the pepsin-like family of aspartic peptidases. *Chem. Rev.* **102**, 4431–4458.

14. Erskine, P. T., Coates, L., Mall, S., Gill, R. S., Wood, S. P., Myles, D. A. & Cooper, J. B. (2003). Atomic resolution analysis of the catalytic site of an aspartic proteinase and an unexpected mode of binding by short peptides. *Protein Sci.* **12**, 1741–1749.

15. Asojo, O. A., Gulnik, S. V., Afonina, E., Yu, B., Ellman, J. A., Haque, T. S. & Silva, A. M. (2003). Novel uncomplexed and complexed structures of plasmepsin II, an aspartic protease from *Plasmodium falciparum*. *J. Mol. Biol.* **327**, 173–181.

16. Clemente, J. C., Govindasamy, L., Madabushi, A., Fisher, S. Z., Moose, R. E., Yowell, C. A. *et al.* (2006). Structure of the aspartic protease plasmepsin 4 from the malarial parasite *Plasmodium malariae* bound to an allophenylnorstatine-based inhibitor. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **62**, 246–252.

17. Cohen, G. E. (1997). ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions.. *J. Appl. Crystallogr.* **30**, 1160–1161.

18. Asojo, O. A., Afonina, E., Gulnik, S. V., Yu, B., Erickson, J. W., Randad, R. *et al.* (2002). Structures of Ser205 mutant plasmepsin II from *Plasmodium falciparum* at 1.8 Å in complex with the inhibitors rs367 and rs370. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **58**, 2001–2008.

19. Sali, A., Veerapandian, B., Cooper, J. B., Moss, D. S., Hofmann, T. & Blundell, T. L. (1992). Domain flexibility in aspartic proteinases. *Proteins*, **12**, 158–170.

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

21. Andreeva, N. S. & Pechik, I. V. (1995). Comparison of three-dimensional structures of flexible protein molecules. *Mol. Biol. (Moscow)*, **29**, 1102–1113.

22. Gilliland, G. L., Winborne, E. L., Nachman, J. & Wlodawer, A. (1990). The three-dimensional structure of recombinant bovine chymosin at 2.3 Å resolution. *Proteins*, **8**, 82–101.

23. Safro, M. G. & Andreeva, N. S. (1990). On the role of peripheral interactions in specificity of chymosin. *Biochem. Int.* **20**, 555–561.

24. Gustchina, E., Rumsh, L., Ginodman, L., Majer, P. & Andreeva, N. (1996). 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.* **379**, 60–62.

25. Istvan, E. S. & Goldberg, D. E. (2005). Distal substrate interactions enhance plasmepsin activity. *J. Biol. Chem.* **280**, 6890–6896.

26. Prade, L., Jones, A. F., Boss, C., Richard-Bildstein, S., Meyer, S., Binkert, C. & Bur, D. (2005). X-ray structure of plasmepsin II complexed with a potent achiral inhibitor. *J. Biol. Chem.* **280**, 23837–23843.

27. Omara-Opyene, A. L., Moura, P. A., Sulsona, C. R., Bonilla, J. A., Yowell, C. A., Fujioka, H. *et al.* (2004). Genetic disruption of the *Plasmodium falciparum* digestive vacuole plasmepsins demonstrates their functional redundancy. *J. Biol. Chem.* **279**, 54088–54096.

28. Mimoto, T., Imai, J., Tanaka, S., Hattori, N., Kisanuki, S., Akaji, K. & Kiso, Y. (1991). KNI-102, a novel tripeptide HIV protease inhibitor containing allophenylnorstatine as a transition-state mimic. *Chem. Pharm. Bull. (Tokyo)*, **39**, 3088–3090.

29. Mimoto, T., Imai, J., Tanaka, S., Hattori, N., Takahashi, O., Kisanuki, S. *et al.* (1991). 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)*, **39**, 2465–2467.

30. Kiso, Y. (1996). Design and synthesis of substrate-based peptidomimetic human immunodeficiency virus protease inhibitors containing the hydroxymethylcarbonyl isostere. *Biopolymers*, **40**, 235–244.

31. Maegawa, H., Kimura, T., Arii, Y., Matsui, Y., Kasai, S., Hayashi, Y. & Kiso, Y. (2004). Identification of peptidomimetic HTLV-I protease inhibitors containing hydroxymethylcarbonyl (HMC) isostere as the transition-state mimic. *Bioorg. Med. Chem. Lett.* **14**, 5925–5929.

32. Abdel-Rahman, H. M., Kimura, T., Hidaka, K., Kiso, A., Nezami, A., Freire, E. *et al.* (2004). Design of inhibitors against HIV, HTLV-I, and *Plasmodium falciparum* aspartic proteases. *Biol. Chem.* **385**, 1035–1039.

33. Kimura, T., Nguyen, J. T., Maegawa, H., Nishiyama, K., Arii, Y., Matsui, Y. *et al.* (2007). Chipping at large, potent human T-cell leukemia virus type 1 protease inhibitors to uncover smaller, equipotent inhibitors. *Bioorg. Med. Chem. Lett.* **17**, 3276–3280.

34. Nguyen, J. T., Zhang, M., Kumada, H. O., Itami, A., Nishiyama, K., Kimura, T. *et al.* (2008). Truncation and non-natural amino acid substitution studies on HTLV-I protease hexapeptidic inhibitors. *Bioorg. Med. Chem. Lett.* **18**, 366–370.

35. Zhang, M., Nguyen, J. T., Kumada, H. O., Kimura, T., Cheng, M., Hayashi, Y. & Kiso, Y. (2008). Locking the two ends of tetrapeptidic HTLV-I protease inhibitors inside the enzyme. *Bioorg. Med. Chem.* **16**, 6880–6890.

36. Zhang, M., Nguyen, J. T., Kumada, H. O., Kimura, T., Cheng, M., Hayashi, Y. & Kiso, Y. (2008). Synthesis and activity of tetrapeptidic HTLV-I protease inhibitors possessing different P3-cap moieties. *Bioorg. Med. Chem.* **16**, 5795–5802.

37. Hidaka, K., Kimura, T., Ruben, A. J., Uemura, T., Kamiya, M., Kiso, A. *et al.* (2008). Antimalarial activity enhancement in hydroxymethylcarbonyl (HMC) isostere-based dipeptidomimetics targeting malarial aspartic protease plasmepsin. *Bioorg. Med. Chem.* **16**, 10049–10060.

38. Fitzgerald, P. M. D., McKeever, B. M., VanMiddlesworth, J. F., Springer, J. P., Heimbach, J. C., Leu, C. T. *et al.* (1990). Crystallographic analysis of a complex between human immunodeficiency virus type 1 protease and acetyl-pepstatin at 2.0 Å resolution. *J. Biol. Chem.* **265**, 14209–14219.

39. Oefner, C., Binggeli, A., Breu, V., Bur, D., Clozel, J. P., D'Arcy, A. *et al.* (1999). Renin inhibition by substituted piperidines: a novel paradigm for the inhibition of monomeric aspartic proteinases? *Chem. Biol.* **6**, 127–131.

40. Boss, C., Corminboeuf, O., Grisostomi, C., Meyer, S., Jones, A. F., Prade, L. *et al.* (2006). Achiral, cheap, and potent inhibitors of plasmepsins I, II, and IV. *ChemMedChem*, **1**, 1341–1345.

41. Hof, F., Schutz, A., Fah, C., Meyer, S., Bur, D., Liu, J. *et al.* (2006). Starving the malaria parasite: inhibitors active against the aspartic proteases plasmepsins I, II, and IV. *Angew. Chem., Int. Ed. Engl.* **45**, 2138–2141.

42. Barrett, A. J., Rawlings, N. D. & Woessner, J. F. (2004). Handbook of Proteolytic Enzymes Elsevier Academic Press, Amsterdam, The Netherlands.

43. Pearl, L. & Blundell, T. (1984). The active site of aspartic proteinases. *FEBS Lett.* **174**, 96–101.
44. Andreeva, N. S. & Rumsh, L. D. (2001). 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.* **10**, 2439–2450.
45. Wlodawer, A., Miller, M., Jaskólski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T. *et al.* (1989). Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science*, **245**, 616–621.
46. Miller, M., Jaskólski, M., Rao, J. K. M., Leis, J. & Wlodawer, A. (1989). Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature*, **337**, 576–579.
47. Remaut, H., Bompard-Gilles, C., Goffin, C., Frere, J. M. & Van Beeumen, J. (2001). Structure of the *Bacillus subtilis* D-aminopeptidase DppA reveals a novel self-compartmentalizing protease. *Nat. Struct. Biol.* **8**, 674–678.
48. Fujinaga, M., Chernaia, M. M., Tarasova, N. I., Mosimann, S. C. & James, M. N. (1995). Crystal structure of human pepsin and its complex with pepstatin. *Protein Sci.* **4**, 960–972.
49. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800.
50. Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 760–763.
51. Keegan, R. M. & Winn, M. D. (2008). MrBUMP: an automated pipeline for molecular replacement. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **64**, 119–124.
52. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025.
53. Read, R. J. (2001). Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **57**, 1373–1382.
54. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
55. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53**, 240–255.
56. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
57. Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2256–2268.
58. DeLano, W. L. (2002). The PyMOL Molecular Graphics System DeLano Scientific, San Carlos, CA.