Blood-feeding parasites, including schistosomes, hookworms, and malaria parasites, employ aspartic proteases to make initial or early cleavages in ingested host hemoglobin. To better understand the substrate affinity of these aspartic proteases, sequences were aligned with and/or three-dimensional, molecular models were constructed of the cathepsin D-like aspartic proteases of schistosomes and hookworms and of plasmepsins of Plasmodium falciparum and Plasmodium vivax, using the structure of human cathepsin D bound to the inhibitor pepstatin as the template. The catalytic subsites S5 through S4’ were determined for the modeled parasite proteases. Subsequently, the crystal structure of mouse renin complexed with the nonapeptidyl inhibitor t-butyl-CHO-His-Pro-Phe-His-Leu-[CH(OH)CH₂]Leu-Tyr-Tyr-Ser-NH₂ (CH-66) was used to build homology models of the hemoglobin-degrading peptides docked with a series of octapeptide substrates. The modeled octapeptides included representative sites in hemoglobin known to be cleaved by both Schistosoma japonicum cathepsin D and human cathepsin D, as well as sites cleaved by one but not the other of these enzymes. The peptidase-octapeptide substrate models revealed that differences in cleavage sites were generally attributable to the influence of a single amino acid change among the P5 to P4 residues that would either enhance or diminish the enzymatic affinity. The difference in cleavage sites appeared to be more profound than might be expected from sequence differences in the enzymes and hemoglobins. The findings support the notion that selective inhibitors of the hemoglobin-degrading peptidases of blood-feeding parasites at large could be developed as novel anti-parasitic agents.

Blood flukes, hookworms, and the malaria parasites are among the most important pathogens of humans in terms of both numbers of people infected and the consequent morbidity and mortality (1). Although phylogenetically unrelated, these parasites all share the same food source; they are obligate blood feeders, or hematophages. Hb from ingested or parasitized erythrocytes is their major source of exogenous amino acids for growth, development, and reproduction; the Hb, a ~64-kDa tetrameric polypeptide, is comprehensively catabolized by parasite enzymes to free amino acids or small peptides. Intriguingly, all these parasites appear to employ cathepsin D-like aspartic proteases to make initial or early cleavages in the Hb substrate (2–4).

The vertebrate endopeptidase, cathepsin D (EC 3.4.23.5), is a member of the aspartic protease category of hydrolases, which also includes renin, pepsin, chymosin, cathepsin E, HIV1 protease, and several other enzymes (5, 6). Cathepsin D is expressed in a diverse range of mammalian cells and tissues and is located predominantly in lysosomes (6). The molecule comprises two rather similar lobes, each incorporating a homologous Asp-Thr-Gly catalytic site motif, with the substrate binding groove located between these lobes. In aspartic proteases generally, the nucleophile that attacks the scissile bond of the substrate is an activated water molecule held in position by side chains of the two active-site aspartic acids (5). Well defined S4, S3, S2, S1, S1’, S2’, S3’, and S4’ subsite pockets for the amino acid side chains of the substrate (5) are additional hallmarks of these enzymes. A peptide analogue of microbial origin, pepstatin, is the definitive, general inhibitor of aspartic proteases (7). In general, human cathepsin D is specific for hydrophobic patches in proteins, with an ostensibly anomalous, additional preference for glutamate in the P2 position (8–11).

An improved understanding of the substrate affinity of cathepsin D-like proteases of blood-feeding parasites for Hb could facilitate the development of novel anti-parasite inhibitors. To this end, sequences or models of the cathepsin D-like aspartic proteases of schistosomes and hookworms and of plasmepsins of Plasmodium falciparum and Plasmodium vivax were aligned with and/or three-dimensional, molecular models were constructed, using the structure of human cathepsin D bound to pepstatin (1LYB) (12) as the template. Subsequently, the crystal structure of the peptidic inhibitor CH-66 complexed with mouse renin (ISMR) (13) was used to build homology models of octapeptide substrates. The molecular models were then used to examine the similarities and differences among known substrate cleavage sites in mammalian Hb reported previously for the cathepsin D of Schistosoma japonicum (4) and for human cathepsin D (11). The models revealed that the difference in cleavage sites was due, in general, to a
single amino acid alteration in the cleavage site (P4-P4') that either enhances or diminishes the enzymatic activity.

**EXPERIMENTAL PROCEDURES**

*Molecular Models of Cathepsin D-Like Peptidases—*Molecular modeling of the mature forms of the target aspartic proteases was carried out on a Silicon Graphics work station using the Insight II software package from Molecular Simulations Inc. (San Diego, CA), as described previously for several papain-like, cysteine proteases (14–16) and the Swiss Model server. Models were viewed using Swiss PDBViewer. The Insight II module, Homology, was used for the homology modeling. Homology operates by copying the backbone atoms and β-carbons from the template molecule to the model and adding the new side chains. Epitopes corresponding to α-helix or β-sheet in the template are assigned as structurally conserved regions, whereas other regions are assigned coordinates as designated loops. The Loop Search function was used to find suitable loops, from protein structures in the PDB data base, which have the correct number of residues and distance to bridge and in which the adjacent residues have the appropriate conformations. Molecular models of the cathepsin D-like enzymes of the blood flukes, *Schistosoma mansoni* (18), the hookworm, *Ancylostoma caninum* (19), and plasmepsins I and II of *Plasmodium falciparum* were built using the homology module. The crystal structure of human cathepsin D complexed with the inhibitor pepstatin, PDB number ILYB (12), was used as the template (12). In addition, the crystal structure of *P. vivax* plasmepsin complexed with pepstatin (1Q88) was compared and contrasted with these models. Construction of the homology models was based on previously determined sequence alignments (12, 13, 17–19) but refined according to known secondary structure. In the sequence alignments (shown below under “Results”), see Fig. 1), residues that have side chains participating in substrate binding pockets are indicated as the S5 to S4` subsites, respectively. In the case of the schistosome and hookworm proteases, the COOH-terminal domain extensions, which are ~40 amino acid residues in length (17–19), were omitted, because they could not be modeled in this way. These proteases have three loops that differ from human cathepsin D, including the hairpin hair, the long hair, and the loop between the α-helix and the β-sheet. The Loop Search function was used to assign possible coordinates for the loops of the *Schistosoma* cathepsin D proteases. The hookworm enzyme has an additional, unusual feature, two small loops, both of which are closed with disulfide bonds, Cys192-Cys97 and Cys210-Cys217. Although the function of the loops has not yet been determined, they do not appear to be structurally important for catalysis. A homology model was also built of human cathespin D, based on itself, with the residues in the hairpin loop deleted and replaced by a single glycine. This was done so that docking with substrate peptides could be carried out similarly with all the other proteases. In each case, the model was minimized to a root mean square deviation of 0.000001, employing conjugate gradients using the Discover module of Insight II. Subsequently, the accuracy and validity of the models was assessed with the Profiles 3D module of Insight II, which performs Eisenberg analysis (22).

*Homology Models of Peptidase-Octapeptide Substrate Complexes—*The crystal structure of mouse renin complexed with the peptide inhibitor CH-66 (1SMR) (13) was used to build homology models of octapeptide substrates within the catalytic clefts of *S. japonicum* cathepsin D, human cathepsin D, and several other proteases for which three-dimensional homology models had been constructed. CH-66 is z-butyl CO-His-Pro-Phes-His-Leu(CHOHCH2)Leu-Tyr-Tyr-Ser-NH2 and the complex of CH-66 within the active site groove of mouse renin is an informative example, with regard to the proteases and substrates targeted in this study, of an extended peptide (P5 to P4') analogue bound to an archetypal aspartic protease. It was anticipated that this complex would indicate substrate binding pockets in human cathepsin D and homologous aspartic proteases. Loop Search was used to replace the P1 and P1' isostere residues of the inhibitor with an extended dipeptide and to assign coordinates to the P1 and P1' residues.

The core and key substrate binding residues of mouse renin and human cathepsin D were superimposed using the α-carbons so that the peptidase and substrate models were in the same space coordinates to conserve the peptidase hydrogen bonds. Hydrogen bonds between the backbone peptide bonds of the substrate. For the mouse renin-CH-66 complex, the protease residues forming hydrogen bonds with the main chain of the inhibitor are Gly37, His79, Ser80, Gly81, Asp220, Gly222, Ser223, Ser225, and Thr227, whereas Asp35 and Asp220 are the catalytic dyad of aspartic acids (numbering according to Fig. 1) (13). Notably, these are consistent with the hydrogen bonding between human renin and the inhibitor CP-85339 (13). The backbone atoms of human cathepsin D and mouse renin are not completely identical; differences in the conformations have been noted for the S3, S2', and S3' pockets (12, 13). These differences produce renin substrate pockets that are less open, giving rise to the narrow specificity of this enzyme. Renin also has a higher degree of secondary structure, with some longer stretches of α-helix and β-sheet. Nonetheless, the appropriate pair of enzyme and substrate was associated as an assembly and minimized with only atoms in the side chains remaining unfixed. In some cases, the correct rotamer for the side chain of the substrate residue had to be selected empirically to avoid major clashes with the enzyme. As a result, the models were effectively docked with the peptidase models. 

### RESULTS

*Substrate Subsite Binding Pockets—*Fig. 1 presents the sequence alignments of the deduced mature enzyme amino acid sequence of *A. caninum* cathepsin D-like protease, human cathepsin D, *S. japonicum* cathepsin D, *S. mansoni* cathepsin D, *P. falciparum* plasmepsin I, plasmepsin II, and mouse renin. The three-dimensional structures of human cathepsin D, plasmepsin II, and mouse renin have been determined previously (12, 13, 24, 25). The sequence alignment was based firstly on structural alignment, with residues found in consensus regions of secondary structure (α-helix, β-sheet) in human cathepsin D, *P. falciparum* plasmepsin II, and mouse renin underlined. The remaining sequences were aligned taking into account the positions of conserved Cys, Gly, and Pro residues, as well as aromatic and hydrophobic residues in structurally conserved regions. The 30 residues that constitute each subsite of the various cathepsin D-like, Hb-degrading peptidases, including *P. vivax* plasmepsin (1Q88) are listed in Table I, illustrating not only the overall conservation of these residues but also subsite differences among these peptidases. The allocation of residues to the pockets differs somewhat from others reported previously, particularly with respect to the S3 and S4` subsites (8, 9, 12, 13), because it was assumed that the conformation of the inhibitor-complexed enzyme mimics that of its substrate-bound form. Profiles 3D analysis (22) of the molecular models indicated a high probability of correct folding, with very few residues defined as being misfolded (not shown).

*The cleavage of Hemoglobin—*The cleavage of mammalian Hb by human cathepsin D has been reported recently (11), as have the cleavage sites in human Hb for cathepsins D of *S. japonicum* and *S. mansoni* (4). Hydrolysis of Hb by *S. japonicum* cathepsin D involves a generally discrete set of cleavages for the α-chain, compared with cleavage sites for human cathepsin D, whereas these two enzymes share most of the same cleavage points for the β-chain (Fig. 2) (4, 11). The orthologous sequences at these known cathepsin D cleavage sites for bovine, human, and catfish Hb are also listed in Fig. 2. *Ancylostoma caninum* cathepsin D cleaves human Hb, but those sites indicated for other substrate species will be cleavage sites for that particular enzyme in every case. This scheme does not reflect the relative importance of each cleavage site or their order. Further, some sites in the Hb α-chain are adjacent, including Leu105 | Leu106 | Thr108 | Leu109, and Leu109 | Ala110, cleavage at any one of these would probably obviate cleavage at

---

Substrate Specificity of Parasite Hemoglobinases

38845

Downloaded from http://www.jbc.org/ at UQ Library on September 1, 2016
Substrate Specificity of Parasite Hemoglobinases

**Fig. 1.** Sequence alignment of amino acids of six aspartic proteases based on secondary structure and sequence homology. The mature enzyme domains only of the proteases enzymes are shown. **SJASP,** the cathepsin D-like protease of the Asian blood fluke *S. japonicum* (L41346); **SMASP,** the cathepsin D-like protease of the African blood fluke *S. mansoni* (U50995); **HCATHD,** human cathepsin D (P07339); **ACASP1,** the cathepsin D-like protease of the hookworm *A. caninum* (U34888); **PLAS1,** plasmins I of the malaria parasite *P. falciparum* (X75787); **PLAS2,** plasmins II of *P. falciparum* (L10740); and **MREN1,** renin from the mouse (P06281). Residues that have side chains participating in substrate binding pockets are indicated by their subsite numbers (S5–S4). The dyad of catalytic aspartic acid residues is indicated with asterisks. The arrow indicates the position of loop 3 in human cathepsin D, the site of processing to a two-chain form (see Refs. 6 and 12). Black boxes with white letters show identical residues, and gray boxes with black letters show chemically similar residues.

the vicinal sites. Perhaps the most striking of all these cathepsin D cleavage sites is FLSF ↓ PTTK (residues 33–40 of the human Hb α-chain), for schistosomal enzyme (4), because a substrate with a proline residue at P1 appears to be rare or even unique for a eukaryotic aspartic protease.

**Molecular Models**—Each of the protease and protease substrate complex models minimized to a plausible structure with no significant defects. However, for reasons of brevity and clarity, much of the analysis and discussion presented here focus on the cathepsin D-like protease from *S. japonicum.* Profiles 3D analysis (19) of the model of this enzyme gave a value of 151, compared with an overall self-compatibility score of 151 and a lowest possible score of 68. This is a high figure, indicating a high probability that the model is correct. The only value of 151, compared with an overall self-compatibility score of 151 and a lowest possible score of 68. This is a high figure, indicating a high probability that the model is correct. The only site of possible misfolding is the loop 46–54, which is well clear of the substrate binding cleft. Fig. 3 shows the general conservation of fold between the aminopeptides of human and *S. japonicum* cathepsins D; side chains of residues that constitute the subsite binding pockets are highlighted. To further investigate the magnitude of sequence variation in the substrate binding pockets and the associated implications for substrate affinities, molecular models of human cathepsin D complexed with KPIEFFRL, FLSF ↓ PTTK, and FLSF ↓ PTTK represents residues 107 to 114 of the α-chain of human Hb. It is a cleavage site for both human and schistosome cathepsins D (Fig. 2) (4, 11). Leu at P4 has considerable hydrophobic contact with Met122, Val at P3 has hydrophobic contacts with Ile114 and Gly14, whereas Thr2 P2 has hydrophobic contacts with Thr2 and Tyr198. Leu at P1 makes hydrophobic interactions with Phe104, Phe119, and Tyr78 of the protease, and Phe at P1 makes hydrophobic interactions with Ile113, Ile122, and Met122. Arg at P2 is mostly exposed to the solvent, though there are minor contacts with the β-methylene of His77 and with Ile130. Leu at P3 makes hydrophobic contact with Ile122 and Tyr198 (Fig. 4, panel A). The model explains why KPIEFFRL is a high affinity substrate for human cathepsin D; it makes numerous contacts with the active site of the peptidase.

**Modeled Complex of S. japonicum Cathepsin D with LVTLAAHL at Its Active Site Cleft**—LVTLAAHL represents residues 107 to 114 of the α-chain of human Hb. It is a cleavage site for both human and schistosome cathepsins D (Fig. 2) (4, 11). Leu at P4 has considerable hydrophobic contact with Met122, Val at P3 has hydrophobic contacts with Ile114 and Gly14, whereas Thr2 P2 has hydrophobic contacts with Thr2 and Tyr198. Leu at P1 makes hydrophobic interactions with Phe104, Phe119, and Tyr78 of the protease, and Phe at P1 makes hydrophobic interactions with Ile113, Ile122, and Met122. Arg at P2 is mostly exposed to the solvent, though there are minor contacts with the β-methylene of His77 and with Ile130. Leu at P3 makes hydrophobic contact with Ile122 and Tyr198 (Fig. 4, panel A). The model explains why KPIEFFRL is a high affinity substrate for human cathepsin D; it makes numerous contacts with the active site of the peptidase.
Residues that constitute the substrate binding pockets, S5 to S4, of human cathepsin D (HuCathD), cathepsins D from Schistosoma japonicum and S. mansoni (abbreviated here as Sj and Sm CathD, respectively), and Ancylostoma caninum (AcASP-1), plasmepsins I and II of Plasmodium falciparum (Pf), and plasmepsin from P. vivax (Pv), as determined by three-dimensional homology modeling

### TABLE I

<table>
<thead>
<tr>
<th>Subsite</th>
<th>HuCathD</th>
<th>Sj/Sm CathD</th>
<th>AcASP-1</th>
<th>Pf plasmepsin I/II</th>
<th>Pf plasmepsin I/II</th>
</tr>
</thead>
<tbody>
<tr>
<td>S5</td>
<td>Ala15</td>
<td>Ala13/Ala13</td>
<td>Ala11</td>
<td>Ile14/Val14</td>
<td>Ile13</td>
</tr>
<tr>
<td>S3</td>
<td>Glu14</td>
<td>Gln14/Gln14</td>
<td>Gln13</td>
<td>Met15/Met15</td>
<td>Met15</td>
</tr>
<tr>
<td>S1</td>
<td>Val31</td>
<td>Val31/Val31</td>
<td>Val30</td>
<td>Ile32/Ile32</td>
<td>Ile32</td>
</tr>
<tr>
<td>Catalytic</td>
<td>Asp376</td>
<td>Asp374/Asp374</td>
<td>Asp39</td>
<td>Asp384/Asp384</td>
<td>Asp384</td>
</tr>
<tr>
<td>S1'</td>
<td>Gly35</td>
<td>Gly35/Gly35</td>
<td>Gly34</td>
<td>Gly36</td>
<td>Gly36</td>
</tr>
<tr>
<td>S2'</td>
<td>Ser26</td>
<td>Ser26/Ser26</td>
<td>Ser35</td>
<td>Ser37/Met15</td>
<td>Ser37</td>
</tr>
<tr>
<td>S2'</td>
<td>Ile43</td>
<td>Ile43/Val43</td>
<td>Ile5</td>
<td>Ile5</td>
<td>Ile5</td>
</tr>
<tr>
<td>S2'</td>
<td>His27</td>
<td>Arg7/His77</td>
<td>Glu76</td>
<td>Glu76</td>
<td>Thr76</td>
</tr>
<tr>
<td>S1</td>
<td>Tyr18</td>
<td>Tyr18/Tyr21</td>
<td>Tyr77</td>
<td>Tyr77/Met77</td>
<td>Tyr77</td>
</tr>
<tr>
<td>S4'</td>
<td>Ser10</td>
<td>Thr99/Thr100</td>
<td>Thr78</td>
<td>Thr78/Met78</td>
<td>Thr78</td>
</tr>
<tr>
<td>S3</td>
<td>Ile124</td>
<td>Leu113/Leu113</td>
<td>Leu112</td>
<td>Leu112</td>
<td>Leu112</td>
</tr>
<tr>
<td>S3/S1</td>
<td>Thr125</td>
<td>Val114/Val114</td>
<td>Thr113</td>
<td>Thr113</td>
<td>Thr113</td>
</tr>
<tr>
<td>S1</td>
<td>Phe126</td>
<td>Phe117/Asp117</td>
<td>Phe114</td>
<td>Phe114</td>
<td>Phe114</td>
</tr>
<tr>
<td>S4</td>
<td>Ala129</td>
<td>Ala117/Ala117</td>
<td>Ala117</td>
<td>Ala117</td>
<td>Ala117</td>
</tr>
<tr>
<td>S3/S1</td>
<td>Phe131</td>
<td>Phe119/Asp119</td>
<td>Phe119</td>
<td>Phe119</td>
<td>Phe119</td>
</tr>
<tr>
<td>S2'</td>
<td>Ile134</td>
<td>Ile122/Thr122</td>
<td>Ile122</td>
<td>Ile122</td>
<td>Ile122</td>
</tr>
<tr>
<td>S4'</td>
<td>Ile142</td>
<td>Val130/Val130</td>
<td>Val130</td>
<td>Val130</td>
<td>Val130</td>
</tr>
<tr>
<td>S2'</td>
<td>Val144</td>
<td>Val132/Val132</td>
<td>Thr131</td>
<td>Thr131</td>
<td>Thr131</td>
</tr>
<tr>
<td>S3</td>
<td>Tyr105</td>
<td>Tyr133/Tyr133</td>
<td>Tyr133</td>
<td>Tyr133</td>
<td>Tyr133</td>
</tr>
<tr>
<td>S1'/S3'</td>
<td>Ile225</td>
<td>Ile220/Ile220</td>
<td>Ile220</td>
<td>Ile220</td>
<td>Ile220</td>
</tr>
<tr>
<td>Catalytic</td>
<td>Asp231a</td>
<td>Asp231a/Asp231a</td>
<td>Asp231a</td>
<td>Asp231a</td>
<td>Asp231a</td>
</tr>
<tr>
<td>S1</td>
<td>Gly33</td>
<td>Gly34/Gly34</td>
<td>Gly33</td>
<td>Gly33</td>
<td>Gly33</td>
</tr>
<tr>
<td>S2</td>
<td>Thr134</td>
<td>Thr134/Thr134</td>
<td>Thr134</td>
<td>Thr134</td>
<td>Thr134</td>
</tr>
<tr>
<td>S4</td>
<td>Met243</td>
<td>Met244/Met244</td>
<td>Met244</td>
<td>Met244</td>
<td>Met244</td>
</tr>
<tr>
<td>S2</td>
<td>Val375</td>
<td>Ala228/Ala228</td>
<td>Ala228</td>
<td>Ala228</td>
<td>Ala228</td>
</tr>
<tr>
<td>S4</td>
<td>Met307</td>
<td>Met295/Met295</td>
<td>Met295</td>
<td>Met295</td>
<td>Met295</td>
</tr>
<tr>
<td>S1'/S2</td>
<td>Met309</td>
<td>Met297/Leu297</td>
<td>Met297</td>
<td>Met297</td>
<td>Met297</td>
</tr>
<tr>
<td>S1'/S3'</td>
<td>Ile311</td>
<td>Leu290/Leu299</td>
<td>Leu290</td>
<td>Leu290</td>
<td>Leu290</td>
</tr>
<tr>
<td>S3'</td>
<td>Pro212</td>
<td>Pro200/Pro200</td>
<td>Pro200</td>
<td>Pro200</td>
<td>Pro200</td>
</tr>
<tr>
<td>S1'</td>
<td>Ile320</td>
<td>Ile308/Ile308</td>
<td>Ile308</td>
<td>Ile308</td>
<td>Ile308</td>
</tr>
</tbody>
</table>

Model of S. japonicum Cathepsin D with LDKFLASV at Its Active Site Cleft—LDKFLASV represents residues 126 to 133 of the α-chain of human Hb. S. japonicum cathepsin D cleaves at this site, whereas the human enzyme does not (Fig. 2) (4, 11). In the model (Fig. 4, panel C), Leu at P4 makes extensive hydrophobic contacts with Met224, Asp at P3 hydrogen bonds to Gln14, and Lys at P2 makes hydrophobic contacts with Thr222. Further, the epsilon amino group of the P2 Lys hydrogen bonds to main chain at residues Leu225 and Ala226. The latter is not possible with human cathepsin D (not shown), perhaps explaining why this site is not cleaved by human cathepsin D. The Phe at P1 makes hydrophobic interactions with Tyr77, Phe115, and Phe120 of the schistosome peptidase. Leu at P1 makes hydrophobic interactions with Ile217 and Tyr194 but would not contact Ile217 if the residue were larger. The Lys at P4 has extensive hydrophobic interactions with Ile290 but is mostly solvent-exposed. The Thr at P3 contacts Ile217 and Tyr194 but would not contact Ile217 if the P1 residue were larger. The Lys at P4 is mostly solvent-exposed but has some hydrophobic contact with Leu113/Val114 of the three plasmepsins presented here (Fig. 4, panels B, C, and D), yet none of the P2 residues apparently makes meaningful contact with Arg77.

Plasmepsins—Plasmepsins I and II are located in the digestive vacuole of intraerythrocytic stages of the malaria parasite, P. falciparum, where they function in the proteolysis of human Hb (20, 21). The crystal structures of plasmepsin II (24) and plasmepsin I (25) have been reported, and the crystal structure of a plasmepsin of P. vivax is available (1QS8). The plasmepsins are too dissimilar to human cathepsin D to be designated cathepsin D-like enzymes, as can be seen in Fig. 1. However, all of the parasite and mammalian enzymes examined here, including the plasmepsins, belong to the same family, family A1 of the pepsin clan (clan AA), according to the phylogeny and nomenclature of Barrett et al. (5). This structure-based sequence alignment (Fig. 1) showed 33% identity of

---

Downloaded from http://www.jbc.org/ on September 1, 2016
Human cathepsin D – Known cleavage sites in bovine hemoglobin are shown (Ref. 11), alongside orthologous sites in human and canine hemoglobin.

### Hemoglobin α-chain Cleaved Bovine Human Canine
1 V*LESA V*LESA V*LESA
24 KAVG*QSA KAVG*QSA KAVG*QSA
32 LEKH*LSFP LEKH*LSFP LEKH*LSFP
109 LLYT*LAHC LLYT*LAHC LLYT*LAHC
134 AVSV*TLYC AVSV*TLYC AVSV*TLYC
137 STL*TKSY STL*TKSY STL*TKSY

### Hemoglobin β-chain Cleaved Bovine Human Canine
6 TAF*EASA TAF*EASA TAF*EASA
14 VQSH*MNGD VQSH*MNGD VQSH*MNGD
30 LGRL*LQVY LGRL*LQVY LGRL*LQVY
31 GRLL*YVYP GRLL*YVYP GRLL*YVYP
40 TOQF*FEFS TOQF*FEFS TOQF*FEFS
44 PESF*GDSL PESF*GDSL PESF*GDSL
53 EDAV*MSDP EDAV*MSDP EDAV*MSDP

#### Schistosoma japonicum cathepsin D – Known cleavage sites in human hemoglobin are shown (Ref. 4), alongside equivalent sites in bovine and canine hemoglobin.

### Hemoglobin α-chain Cleaved Bovine Human Canine
29 DEAL*ERMF DEAL*ERMF DEAL*ERMF
33 ERMF*LSPF ERMF*LSPF ERMF*LSPF
48 VPPF*GDSL VPPF*GDSL VPPF*GDSL
109 LLYT*LAHC LLYT*LAHC LLYT*LAHC
110 LLYT*NLHL LLYT*NLHL LLYT*NLHL
129 LGRL*LAHY LGRL*LAHY LGRL*LAHY

### Hemoglobin β-chain Cleaved Bovine Human Canine
6 TAF*EASA TAF*EASA TAF*EASA
14 VQSH*MNGD VQSH*MNGD VQSH*MNGD
30 LGRL*LQVY LGRL*LQVY LGRL*LQVY
31 GRLL*YVYP GRLL*YVYP GRLL*YVYP
40 TOQF*FEFS TOQF*FEFS TOQF*FEFS
44 PESF*GDSL PESF*GDSL PESF*GDSL
53 EDAV*MSDP EDAV*MSDP EDAV*MSDP

**FIG. 2. Known and predicted cleavage sites for S. japonicum** and human cathepsins D in mammalian (human, bovine, and canine) hemoglobin. Peptide bond cleaved after residue indicated. Residues marked with *underlines* show the presence of substitutions in orthologous amino acids. The hemoglobin sequences were obtained from the public domain; the data base accession numbers for the bovine, canine, and human hemoglobin sequences are P01966 and P020270, for the canine hemoglobin α- and β-chains they are P01952 and P02056, and for the human hemoglobin α- and β-chains they are NP0000508 and NP0000509. Some of the information for this figure was obtained from Refs. 4 and 11.

plasmepsin II with human cathepsin D and 31% for plasmepsin I. Blastp analysis revealed that the mature form of the *P. vivax* plasmepsin (1Q86) is 31% identical to human cathepsin D and 70% identical to plasmepsin II of *P. falciparum*. In all three plasmepsins examined here, about half of the substrate binding residues of human cathepsin D are conserved (Table I). Plasmepsin I of *P. falciparum* cleaves the human Hb α-chain at residue 33 (ERMF | LSFP), residue 46 (FPHF | DLSH), and residue 98 (VPFK | LSHL) and the β-chain at residues 31 (LGRL | LQVY), 41 (TQRF | FEFS), and 129 (VQAA | YQKV) (2). *P. falciparum* plasmepsin II cleaves the human Hb α-chain at residues 33 (ERMF | LSFP), 108 (LLVT | LAAH), and 136 (STVL | TSKY) and the β-chain at residue 32 (GRLL | VYYP) (2). These differences in cleavage sites may be the reason for having (at least) two plasmepsins; the combination of cleavages by both plasmepsins may deliver Hb fragments small enough for further processing by other peptidases. The oligopeptide Ala-Leu-Glu-Arg-Thr-Phe | Phe-Ser-Phe-Pro-Thr has been proposed as an ideal plasmepsin II substrate, based on the ERMF | LSFP cleavage site (24).

As shown in Table I, the S4, S3, S2, S1, and S2′ subsites of the malarial plasmepsins differ considerably from those of the human, schistosome, and hookworm cathepsins D, possibly explaining the fewer Hb cleavage sites for the *P. falciparum* plasmepsins. The presence of Ala219 (instead of Leu or Met, present in the human, schistosome, and hookworm cathepsins D) ensures a more open S4 pocket, which would accommodate the binding by large hydrophobic residues such as Phe. It would also allow binding by hydrophilic and charged residues, such as Glu, although this is not evident from the known Hb cleavage sites (2). However, the *P. falciparum* plasmepsins I and II and the *P. vivax* plasmepsin all retain a Ser equivalent to Ser50 of human cathepsin D (Table I), which suggests that they would have a similar P2 specificity for Glu residues. Another similarity with the human enzyme is the presence of an equivalent residue to Thr125 in the S3/S1 pockets (Thr114 in plasmepsin I) (Table I), which would prevent plasmepsin I from cleaving FLSSFPKTTK (not shown), a site cleaved by the schistosome cathepsins D (see Fig. 2 and Fig. 4D) but not by the human cathepsin D or *P. falciparum* plasmepsins I and II (2, 11).

**DISCUSSION**

Hemoglobin-degrading enzymes of hematophagous parasites are being targeted for the rational development of novel antiparasitic compounds (26, 27). Given that Hb is the natural substrate of these enzymes, comparison of their cleavage sites in the molecule with the cleavage site profile of homologous mammalian-host enzymes, such as human cathepsin D, is likely to provide specific leads that could be exploited in inhibitor design. However, information on the Hb-cleavage patterns for other schistosomes has not been available until recently. Accordingly, by focusing here on the specific substrate cleavage patterns known for schistosome cathepsin D (4) and human cathepsin D (11), we undertook a molecular modeling analysis of some representative shared and discrete cleavage sites. This involved the docking of octapeptides representing the P4-P4′ residues of Hb cleavage sites that were either common to both schistosome and human cathepsins D or cleaved by one but not the other.

Using the crystal structural models of human cathepsin D complexed with pepstatin and mouse renin complexed with CH-66 as guides for the models, ~30 residues could be identified as major contributors to substrate or inhibitor binding in the panel of target parasite and mammalian aspartic proteases (Table I). Eight of these catalytic subsite residues differed between human and *S. japonicum* cathepsins D. Such an ostensibly minor difference between these two enzymes raised the question as to why they do not cleave the α- and β-chains of Hb at exactly the same sites; of the 13 cleavage sites reported for both the mammalian (11) and *S. japonicum* cathepsin D (4), only six are shared (Fig. 2). As depicted in Fig. 4, homology models of the *S. japonicum* and/or human enzyme complexed with four informative octapeptides allowed examination of their subsite binding pockets. Whereas other, usually hydrophilic, residues can occur in any of the peptide subsites, their side chains do not interact with the enzyme. Glycine residues can also be found. Further, the exact nature of the Hb cleavage sites would depend on the order of cleavage, because an adjacent site may obscure its neighbor.

The S4 subsite is essentially a hydrophobic pocket. In human cathepsin D, it comprises Leu236 and Met307. These are replaced with Met or Ile in the *S. japonicum* and *S. mansoni* cathepsins D, changes that would have minimal influence on the specificity of this pocket. Residues preferred by human cathepsin D at P4 include Leu, Val, Thr, Pro, and Ala (8–11). The S3 pocket comprises Gln14, Ile224, Thr235, Ala129, and Phe131, and Met, Ile, Ser, and Thr are preferred P3 residues. S3 is both a hydrophilic and a hydrophobic pocket, with hydrophilic residues binding to the Gln14 side chain and hydrophobic residues binding to the remainder. Longer hydrophilic side chains can reach into the pocket as far as Ala129 and possibly
Phe$^{131}$, Ile$^{125}$ is probably too distant to be normally involved in binding, although it could be exploited in inhibitor design. Changing Thr$^{125}$ to a Val has little effect on the S3 pocket, because it is the methyl moiety of Thr$^{125}$, rather than the hydroxyl, that faces this pocket.

The S2 subsite of human cathepsin D comprises Ser$^{80}$, Thr$^{125}$, Val$^{238}$, and Met$^{309}$ (Table I) and exhibits a preference for Glu, Ile, Val, Ala, or Phe at P2 (8–11). Shorter residues, such as Val and Ala, bind to Thr$^{125}$, with longer residues such as Met, Ile, or Phe also binding to Val$^{238}$ and possibly Met$^{309}$. The major exception is a P2 Glu, which hydrogen bonds to Ser$^{80}$, in what is known in aspartic proteases as one of the “flaps” that close over the active site when a peptide is bound (28). As has been demonstrated with the equivalent mouse renin-CH-66 complex (13), the hydroxyl group of Ser$^{80}$ normally hydrogen bonds with the P2 amino group of the inhibitor. A putative hydrogen bond with the P2 Glu side chain would have to be a novel feature, unless the Ser hydroxyl can share two hydrogen bonds. This provides the likely reason why human cathepsin D cleaves AAELY↓GAEA (residues 21–28, α-chain of Hb) whereas loops and α-helices are shown as magenta carbon traces. Side chains of residues that constitute the subsite binding pockets are highlighted. The model was generated using the Swiss Model server and viewed in Swiss PDBViewer.

FIG. 3. Stereo view of molecular model of S. japonicum cathepsin D (panel A) based on the crystal structure of human cathepsin D (1LYA) (panel B). β-Sheets are shown as green ribbons, whereas loops and α-helices are shown as magenta carbon traces. Side chains of residues that constitute the subsite binding pockets are highlighted. The model was generated using the Swiss Model server and viewed in Swiss PDBViewer.

The S1 subsite in human cathepsin D comprises Val$^{31}$, Asp$^{33}$, Tyr$^{28}$, Thr$^{125}$, Phe$^{126}$, Ile$^{229}$, Asp$^{231}$, and Gly$^{233}$ and exhibits a preference for Leu, Tyr, Phe, or Ile as P1 residues. Clearly, this is a very hydrophobic subsite that can accommodate the side chains of Leu and Phe. The two changes, Ser$^{80}$ to Thr and Thr$^{125}$ to Val, observed with both Schistosoma cathepsins D, render that side of the S1 pocket even more hydrophobic than that of human cathepsin D. This becomes important with the schistosomal cathepsin D cleavage of FLSF↓PTTK. The P1' proline residue readily fits into the S1' subsite. However, the change in the peptide backbone affects the location and conformation of the Phe at P1. With the S. japonicum cathepsin D, Phe at P1 can bind to the methyl groups of Thr$^{306}$ and Val$^{114}$. With human cathepsin D, by contrast, the hydrophilic hydroxyl of Thr$^{125}$ impedes the binding of a P1 Phe when proline is at P1'. These findings indicate that the octapeptide FLSFPTTK could be a useful lead for developing a specific inhibitor of schistosomal cathepsin D. Development of inhibitors of HIV protease followed the discovery of a similar preference by HIV, type I retropepsin for Pro at P1' (29, 30).

The S1' subsite of human cathepsin D comprises Gly$^{35}$, Asp$^{38}$, Met$^{309}$, Ile$^{231}$, and Ile$^{235}$, and its preferences include Leu, Val, Ala, and Phe at P1' (8–11). Phe side chains at P1' adopt one of two conformations. The Phe rotamer found with human and Schistosoma cathepsins D faces toward the S2 pocket, with the S-methyl of Met$^{309}$ (which arises in the S2 subsite) contacting the β-methylene of the Phe. With the hookworm cathepsin D, which has a S1' Phe instead of Ile$^{31}$, the P1' Phe adopts a different rotamer, it faces the S3' subsite, because it is sterically hindered from facing toward Met$^{309}$. This conformation is also adopted for a P1' Leu for any of the enzymes, because a P1' Leu cannot reach as far as Met$^{309}$. Consequently, the side chain of the P1' Phe makes more contacts with the enzyme.

The S2' subsite of human cathepsin D comprises Ser$^{86}$, Ile$^{76}$, His$^{77}$, Ile$^{134}$, Ile$^{142}$, and Val$^{144}$ and exhibits a preference for
Lys, Arg, Glu, or His at P2’. As with the S3 subsite, this pocket should accommodate hydrophobic and hydrophilic side chains, although P2’ hydrophobics are not common (8–11) . These models suggest that the functional ends of long basic P2’ residues (Lys and Arg) are mostly exposed to the solvent but that shorter hydrophilic side chains, such as those of Glu and Asp, would be able to hydrogen bond with Ser36 or possibly with His77. Replacement of His77 with either Arg or Gln will change the order of preferred residues at P2’ but not the overall mix. An Arg at this position may promote the possibility of a P2’ Glu by forming a salt bridge, whereas Gln at this position may promote the possibility of a Gln at P2’ through mutual hydrogen bonding. Hydrophobic residues such as valine can bind to Val144.

The S3’ subsite residues of human cathepsin D are Tyr205, Ile311, and Pro312, and preferences at P3’ include Leu, Lys, Val, Tyr, and Thr (8–11). This is another predominantly hydrophobic pocket. Ile311 is variable, affecting both the S1’ and the S3’ pockets, but this would have little influence on subsite specificity. The S4’ pocket is composed of Ile142 in human cathepsin D, and its P4’ preferences include Ala, Ser, Val, Lys, Phe, and Tyr. Most of these P4’ amino acids would be exposed to the solvent, but smaller hydrophobic residues such as Ala or Val can bind to Ile142. Mutating the Ile142 to Leu in S. japonicum cathepsin D (Ile311, see Table I) should have minimal effect on subsite specificity.

It is clear that the differences in Hb cleavage by the proteases examined here reflect subtle yet significant differences in their substrate binding pockets. Development of selective inhibitors of these parasite aspartic proteases as novel antifungal agents will depend on defining these differences and deriving compounds that exploit one or (preferably) several of them. From an evolutionary perspective, it is straightforward to conceptualize how mutations in Hb genes would lead to amino acid substitutions in their products and how the cumulative effect of such substitutions could then reduce the efficiency of the cognate hemoglobinases of the hematophagous parasite, delivering a selective advantage to the host by reducing the viability or virulence of the parasite. Moreover, as speculated previously (31), the compatibility between an Hb-degrading enzyme of the parasite and the Hb of its mammalian hosts (Fig. 2) is likely to be a critical factor in determining its host range, i.e., host specificity. In this regard, it is notable that plasmspin II has a lower affinity for fetal than for adult Hb, which contributes to the innate resistance of human neonates to malaria (32).
Hemoglobin-degrading, Aspartic Proteases of Blood-feeding Parasites: SUBSTRATE SPECIFICITY REVEALED BY HOMOLOGY MODELS
Ross I. Brinkworth, Paul Prociv, Alex Loukas and Paul J. Brindley

doi: 10.1074/jbc.M101934200 originally published online August 8, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M101934200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 25 references, 7 of which can be accessed free at
http://www.jbc.org/content/276/42/38844.full.html#ref-list-1