2009

Structural determinants of autoproteolysis of the Haemophilus influenzae Hap autotransporter

Roma Kenjale
Duke University

Guoyu Meng
Duke University

Doran L. Fink
Washington University School of Medicine in St. Louis

Twyla Juehne
Washington University School of Medicine in St. Louis

Tomoo Ohashi
Duke University

See next page for additional authors

Infection and Immunity

Structural Determinants of Autoproteolysis of the *Haemophilus influenzae* Hap Autotransporter

Roma Kenjale, Guoyu Meng, Doran L. Fink, Twyla Juehne, Tomoo Ohashi, Harold P. Erickson, Gabriel Waksman and Joseph W. St. Geme III

Infect. Immun. **2009, 77(11):4704. DOI: 10.1128/IAI.00598-09. Published Ahead of Print 17 August 2009.**

INFECTION AND IMMUNITY, Nov. 2009, p. 4704–4713 Vol. 77, No. 11
0019-9567/09/$12.00 doi:10.1128/IAI.00598-09

Structural Determinants of Autoproteolysis of the *Haemophilus influenzae* Hap Autotransporter▽

Roma Kenjale,[1,2] Guoyu Meng,[4] Doran L. Fink,[5]† Twyla Juehne,[5]‡ Tomoo Ohashi,[3] Harold P. Erickson,[3] Gabriel Waksman,[4] and Joseph W. St. Geme III[1,2]*

Department of Pediatrics,[1] Department of Molecular Genetics and Microbiology,[2] and Department of Cell Biology,[3] Duke University, Durham, North Carolina; Institute of Structural and Molecular Biology at UCL/Birkbeck, Malet Street, London, United Kingdom[4]; and Department of Pediatrics, Washington University, St. Louis, Missouri[5]

Received 28 May 2009/Returned for modification 7 July 2009/Accepted 6 August 2009

Haemophilus influenzae **is a gram-negative bacterium that initiates infection by colonizing the upper respiratory tract. The *H. influenzae* Hap autotransporter protein mediates adherence, invasion, and microcolony formation in assays with respiratory epithelial cells and presumably facilitates colonization. The serine protease activity of Hap is associated with autoproteolytic cleavage and extracellular release of the Hap_S passenger domain, leaving the Hap_β C-terminal domain embedded in the outer membrane. Cleavage occurs most efficiently at the LN1036-37 peptide bond and to a lesser extent at three other sites. In this study, we utilized site-directed mutagenesis, homology modeling, and assays with a peptide library to characterize the structural determinants of Hap proteolytic activity and cleavage specificity. In addition, we used homology modeling to predict the S1, S2, and S4 subsite residues of the Hap substrate groove. Our results indicate that the P1 and P2 positions at the Hap cleavage sites are critical for cleavage, with leucine preferred over larger hydrophobic residues or other amino acids in these positions. The substrate groove is formed by L263 and N274 at the S1 subsite, R264 at the S2 subsite, and E265 at the S4 subsite. This information may facilitate design of approaches to block Hap activity and interfere with *H. influenzae* colonization.**

Haemophilus influenzae is a gram-negative coccobacillus that typically colonizes the nasopharynxes of children and adults. In addition, this organism is an important cause of localized respiratory tract and invasive disease. Nonencapsulated strains cause otitis media, sinusitis, conjunctivitis, and exacerbations of respiratory symptoms in individuals with underlying lung disease, bronchiectasis, and cystic fibrosis (21, 29). Encapsulated strains are an important cause of bacteremic diseases, including sepsis and meningitis (29). Colonization of the upper respiratory tract represents an early step in the pathogenesis of all *Haemophilus* disease and requires adherence to respiratory epithelium (19). Adherence is facilitated by a number of adhesins, including Hap, Hia, Hsf, HMW1/HMW2, P5, pili, and lipooligosaccharide (2, 18, 21, 26, 27).

The Hap adhesin is ubiquitous among isolates of *H. influenzae* and is a member of the autotransporter family of virulence factors that have been recognized among many gram-negative bacteria (10). Autotransporters are synthesized as precursor proteins with three functional regions, namely, an N-terminal signal sequence, an internal passenger domain, and a C-terminal β-barrel domain (11). The signal sequence targets the precursor protein to the inner membrane and is then cleaved. The C-terminal β-barrel domain inserts into the outer membrane and facilitates presentation of the passenger domain on the bacterial cell surface. Depending upon the protein, the passenger domain remains covalently attached to the β-barrel domain, is cleaved but remains loosely attached to the β-barrel domain, or is cleaved and released entirely from the cell surface (10–12). Although diverse autotransporters share a similar structural organization and a common secretion mechanism, they vary widely in function, possibly reflecting adaptations to particular bacterial pathogenic niches. Autotransporters may function as adhesins mediating tissue tropism, as proteases involved in tissue degradation, as toxins causing host tissue damage, or as mediators of serum resistance (11).

Hap is synthesized as a 155-kDa preprotein encompassing a 110-kDa passenger domain, Hap_S, and a 45-kDa β-barrel domain, Hap_β. The Hap_S passenger domain harbors adhesive activity that has been shown to promote interactions with human respiratory cells, as well as with extracellular matrix proteins such as fibronectin, laminin, and collagen IV (7). Hap_S is also responsible for bacterial aggregation via Hap-Hap interactions, contributing to microcolony formation (5). Adherence to epithelial cells and bacterial aggregation are mediated by the C-terminal 311 amino acids of Hap_S, whereas interaction with extracellular matrix proteins is mediated by the C-terminal 511 amino acids of Hap_S (7).

Beyond possessing adhesive activities, the Hap_S passenger domain functions as a protease that directs the autoproteolytic cleavage of Hap_S from Hap_β, resulting in release of Hap_S from the bacterial cell surface (6). Hap autoproteolysis has been determined to occur at least partly through intermolecular

* Corresponding author. Mailing address: Department of Pediatrics, Duke University Medical Center, Children's Health Center, Rm. T901, Durham, NC 27710. Phone: (919) 681-4080. Fax: (919) 681-2714. E-mail: j.stgeme@duke.edu.

† Present address: Department of Pathology, Washington University, St. Louis, Missouri.

‡ Present address: Department of Pediatrics, Johns Hopkins University, School of Medicine, Baltimore, Maryland.

▽ Published ahead of print on 17 August 2009.

TABLE 1. Bacterial strains and plasmids

Strain or plasmid[a]	Description[b]	Source or reference
Strains		
E. coli DH5α	ϕ80d*lacZ*ΔM15 Δ*lacU169 deoR rec*A *endA1*	Life Technologies
H. influenzae DB117	Laboratory strain, *rec-1*, capsule-deficient serotype d, with nonsense mutation in *hap*	23a
Plasmids		
pGJB103	*E. coli-H. influenzae* shuttle vector; Tc^r	27a
pJS106	pGJB103 with a 6.7-kb PstI fragment containing *hap*	26, 27
pMLD100	pUC19 with a 6.7-kb PstI fragment containing *hap*	13
pHapS243A	pJS106 with S243A mutation	13
pHapL1036S	pJS106 with L1036S mutation	5
pHapL1036F	pJS106 with L1036F mutation	This study
pHapL1035S	pJS106 with L1035S mutation	This study
pHapL1035T	pJS106 with L1035T mutation	This study
pHapL1035E	pJS106 with L1035E mutation	This study
pHapE1045L	pJS106 with E1045L mutation	This study
pHapE1045A	pJS106 with E1045A mutation	This study
pHapF1077L	pJS106 with F1077L mutation	This study
pHapF1077L/L1036S	pJS106 with F1077L/L1036S mutations	This study
pHapF1077L/L1036F	pJS106 with F1077L/L1036F mutations	This study
pHapF1067L	pJS106 with F1067L mutation	This study
pHapV1066L	pJS106 with V1066L mutation	This study
pHapQ1033A	pJS106 with Q1033A mutation	This study
pHapQ1033E	pJS106 with Q1033E mutation	This study
pHapS1034F	pJS106 with S1034F mutation	This study
pHapS1034E	pJS106 with S1034E mutation	This study
pHapN1037R	pJS106 with N1037R mutation	This study
pHapA1038S	pJS106 with A1038S mutation	This study
pHapL1039S	pJS106 with L1039S mutation	This study
pHapE1040S	pJS106 with E1040S mutation	This study
pHapY137A	pJS106 with Y137A mutation	This study
pHapK240A	pJS106 with K240A mutation	This study
pHapL263R	pJS106 with L263R mutation	This study
pHapN274R	pJS106 with N274R mutation	This study
pHapR264A	pJS106 with R264A mutation	This study
pHapE265A	pJS106 with E265A mutation	This study
pHapE265R	pJS106 with E265R mutation	This study
pHapE265W	pJS106 with E265W mutation	This study

[a] Plasmids encoding wild-type *hap* or mutant *hap* derivatives were expressed in *H. influenzae* strain DB117.
[b] Tc^r, tetracycline resistance.

cleavage on the surface of the bacterium and involves a catalytic triad consisting of residues His98, Asp140, and Ser243. Ser243 is part of the GD*S*GS motif that is characteristic of many serine proteases (6, 13). In wild-type Hap, cleavage occurs most abundantly at the L1036-N1037 peptide bond, which is referred to as the primary cleavage site (13). Site-directed mutagenesis of this site and N-terminal sequencing of the resulting cleaved Hap fragments has identified three additional cleavage sites, including L1046-T1047, F1077-A1078, and F1067-S1068, termed the secondary, tertiary, and quaternary cleavage sites, respectively (see Table 2) (6). Alignment of the amino acid sequences at these cleavage sites has revealed a consensus target sequence motif that consists of (Q/R)(A/S)X(L/F) at the P4 through P1 positions (see Table 2) (6).

Hap protease activity can be blocked by selected serine protease inhibitors, including secretory leukocyte protease inhibitor, a component of human respiratory secretions (14). This inhibition results in accumulation of full-length Hap in the outer membrane, with Hap_S on the bacterial surface. Similar to the effect of serine protease inhibitors, mutation of the active site serine in the Hap protease domain causes retention of Hap_S on the bacterial surface, increased adherence to epithelial cells, increased adherence to extracellular matrix proteins, increased bacterial aggregation, and increased microcolony formation, suggesting that Hap protease activity might play a regulatory role in bacterial adherence. Indeed, autoproteolytic cleavage of Hap results in reduced adhesive activity (5).

In the present study, we sought to further elucidate the structural determinants of Hap proteolytic activity. We used site-directed mutagenesis and in vitro assays with a synthetic peptide library to characterize Hap cleavage site specificity, and we used homology modeling and site-directed mutagenesis to define the Hap substrate groove. We found that the two residues immediately N-terminal to Hap cleavage sites are critical for cleavage, with leucine preferred over larger hydrophobic residues or other amino acids in these positions. In addition, we defined the residues that likely form the Hap substrate groove.

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions. The bacterial strains and plasmids used in the present study are listed in Table 1. Plasmid constructs encoding *hap* were expressed in *H. influenzae* strain DB117, a *rec-1* derivative of

TABLE 2. Alignment of sequences at the Hap cleavage sites

Cleavage site	Sequence alignment[a]												
	P5	P4	P3	P2	P1	P1′	P2′	P3′	P4′	P5′			
Primary 1036-37	D	Q	S	L	L	N	A	L	E	A			
Secondary 1046-47	K	Q	A	E	L	T	A	E	T	Q			
Tertiary 1077-78	D	Q	S	L	F	A	L	E	A	A			
Quaternary 1067-68	K	R	A	V	F	S	D	P	L	L			
Peptide library[b]	M	A	X	X	X	X	X	N	A	L	E	A	K	K (biotin)

[a] The peptide bond where cleavage occurs is between the P1 and P1′ columns. According to nomenclature by Schechter and Berger (22), the residues N terminal to the cleavage site are labeled P1, P2, . . . P*n* in the N-terminal direction, and the residues C-terminal to the cleavage site are labeled P1′, P2′ . . . P*n*′.

[b] The "X" represents any amino acid except Cys. The P′ residues were fixed based on the P1′ to P5′ sequence at the Hap primary cleavage site. The M and A residues at the N terminus were added in order to validate the first two residues in the N-terminal sequencing reaction of each peptide.

strain Rd that contains a nonfunctional *hap* gene due to a spontaneous nonsense mutation in *hap*. *H. influenzae* strains were stored at −80°C in brain heart infusion (BHI) broth with 20% glycerol, grown overnight on supplemented BHI agar with appropriate antibiotics, and cultivated in BHI broth as described previously (1). *Escherichia coli* strain DH5α was grown on Luria-Bertani (LB) agar or in LB broth and was maintained at −80°C in LB broth with 20% glycerol. Tetracycline was used at concentrations of 5 μg/ml (*H. influenzae*) and 12.5 μg/ml (*E. coli*).

Purification of Hap_S protein. Hap_S was purified from culture supernatants of *H. influenzae* strain DB117/pJS106 as described previously (13).

Glycerol gradient sedimentation and electron microscopy. The sedimentation coefficient of Hap_S was estimated by glycerol gradient sedimentation. The purified protein was sedimented at 20°C through a 15 to 40% glycerol gradient in 0.2 M ammonium bicarbonate at 38,000 rpm for 16 h in a Beckman SW55.1 rotor (23). The glycerol gradients were calibrated with standard proteins of known S values (catalase, 11.3 S; aldolase, 7.3 S; bovine serum albumin, 4.6 S; and ovalbumin, 3.5 S). For rotary shadowing, samples from the glycerol gradient fractions were sprayed onto freshly cleaved mica, vacuum dried, and rotary shadowed with platinum (8, 23). For negative staining, purified Hap_S was applied to carbon-coated copper grids and stained with 2% uranyl acetate (8, 23).

Construction of mutant Hap derivatives. The plasmid pMLD100 is a pUC19 derivative that contains a 6.7-kb insert with the wild-type *hap* gene and was used as a template for PCR. Site-directed mutagenesis was performed using a QuikChange XL II kit from Stratagene (La Jolla, CA) according to the manufacturer's instructions. The mutations were confirmed by sequencing. After mutagenesis, the 6.7-kb PstI (Roche, Indianapolis, IN) fragment was excised from pMLD100 and ligated into PstI-digested pGJB103. The resulting plasmids were introduced into the *H. influenzae* strain DB117 by MII/MIV transformation (25).

Analysis of outer membrane proteins. *H. influenzae* strain DB117 derivatives expressing wild-type Hap or mutant Hap were grown in BHI broth with tetracycline to an optical density at 600 nm of 0.8. Sarkosyl insoluble outer membranes were purified as described previously (4). Outer membrane proteins were resuspended in a small volume of 10 mM HEPES (pH 7.4), resolved on a 10% sodium dodecyl sulfate-polyacrylamide gel, and electroblotted onto a nitrocellulose membrane. Immunoblots were probed with antiserum Rab290, which was raised against a C-terminal Hap fragment corresponding to residues 996 to 1395 and detects the membrane-bound Hap precursors, Hap_S and Hap_β. Anti-rabbit immunoglobulin G conjugated to alkaline phosphatase (Sigma Chemical Co., St. Louis, MO) was used as a secondary antibody.

Peptide library design and purification. A 12-mer C-terminally biotinylated peptide library with the sequence NH2-MAXXXXXNALEAKK(biotin)-COOH was synthesized at the Tufts University Core Facility. In this peptide library, the P′ residues were fixed, based on the Hap P1′-P5′ sequence at the primary cleavage site (Table 2). An "X" denotes random amino acids at the P1 to P5 positions that were synthesized using an isokinetic mixture of 19 amino acids (except cysteine). The peptide library was resuspended in phosphate-buffered saline (PBS; pH 7.4) at a concentration of 3.4 mg/ml. In order to ensure that the peptide library was homogenously biotinylated, 0.4 mg of the resuspended library was purified over monomeric avidin agarose (Pierce, Rockford, IL) as described previously (28). The resulting purified peptide library was lyophilized and resuspended in 100 μl of assay buffer composed of 50 mM sodium phosphate (pH 7.4).

Peptide library cleavage assay. In order to determine the P residue specificity, 15 μl of the purified peptide library was incubated at 37°C overnight with 8 pmol of purified Hap_S in a 20-μl reaction mixture. Control reactions contained either substrate only (without Hap_S) or Hap_S only (without peptide). After incubation, the samples were heated to 100°C for 5 min to stop the cleavage reaction. The sample with Hap_S and peptide was incubated with 0.6 ml of avidin agarose (Sigma) and rotated at room temperature for 1 h to remove the C-terminal fragments of the cleaved products, as well as uncleaved peptides. The samples were subjected to repeated lyophilization, followed by resuspension in water, finally resuspended in 10 μl of water, and then subjected to N-terminal sequencing. The peptide fragments recovered after avidin treatment were sequenced for one cycle beyond the intended P1 position to test for the removal of unreacted peptides. To determine bias present in the peptide library itself, the enzyme-free reaction without avidin treatment was sequenced in the same manner. Raw data from the sequencing reactions were obtained as molar quantities of each amino acid present in each sequencing cycle. Corrected experimental data were obtained by dividing the molar percent abundance of each residue in any given sequencing cycle by the molar percent abundance of that residue in the starting peptide library. These data were normalized to an average value of 1, resulting in a selective value greater than 1 for residues that were positively selected and a value less than 1 for residues that were negatively selected.

RESULTS

Electron microscopy of purified Hap_S. In order to study the architecture of Hap_S, we began by purifying Hap_S from culture supernatants of *H. influenzae* strain DB117 expressing wild-type Hap. As shown in Fig. 1A, the sedimentation coefficient of purified Hap_S, as estimated by glycerol gradient sedimentation, was found to be 5.5 S. This S value of Hap_S, together with its known molecular weight, gives rise to an S_{max}/S value of 1.53, indicating that Hap_S has an elongated shape (Fig. 1A) (23). To further investigate the structure of Hap_S, we performed rotary shadowing electron microscopy. As shown in Fig. 1B, this technique revealed structures that were ~15 nm in length and resembled very short arrowheads (Fig. 1B, middle panel). Similar arrowhead-shaped structures were also visible with negative staining transmission electron microscopy (Fig. 1B, left panels). The arrowhead shape of Hap_S resembled the crystal structure of *E. coli* hemoglobin protease (Hbp, Fig. 1B, right panel), a Hap homolog with heme-binding and serine protease activities that is expressed by clinical isolates associated with intra-abdominal infections (20, 30). Alignment of the amino acid sequences of Hap_s and the protease domain of Hbp revealed an identity of 41% and a similarity of 54%, as highlighted in Fig. 2A.

Homology based modeling of the Hap_S structure. Sequence alignment using the Pfam server (http://pfam.sanger.ac.uk) suggested that Hap_S consists of an N-terminal domain (residues 70 to 550) sharing 41% sequence identity with Hbp and a C-terminal domain (residues 851 to 980) sharing 43% sequence identity with the *Bordetella pertussis* Pertactin protein. To gain further insight into the structure of Hap_S, we modeled residues 70 to 550 containing the protease domain based on the crystal structure of Hbp (PDB code 1WXR) and residues 851 to 980 based on the crystal structure of Pertactin (PBD code 1DAB) using Swiss-model (http://swissmodel.expasy.org) and 3Djigsaw (http://bmm.cancerresearchuk.org) homology modeling servers (Fig. 2B). Based on the Hbp-like shape of Hap_S observed by electron microscopy, we modeled the region between the protease and Pertactin-like domains on the crystal structure of Hbp, assuming that $Hap_{551\text{-}850}$ shares the same β-helix fold as observed in the Hbp crystal structure and other

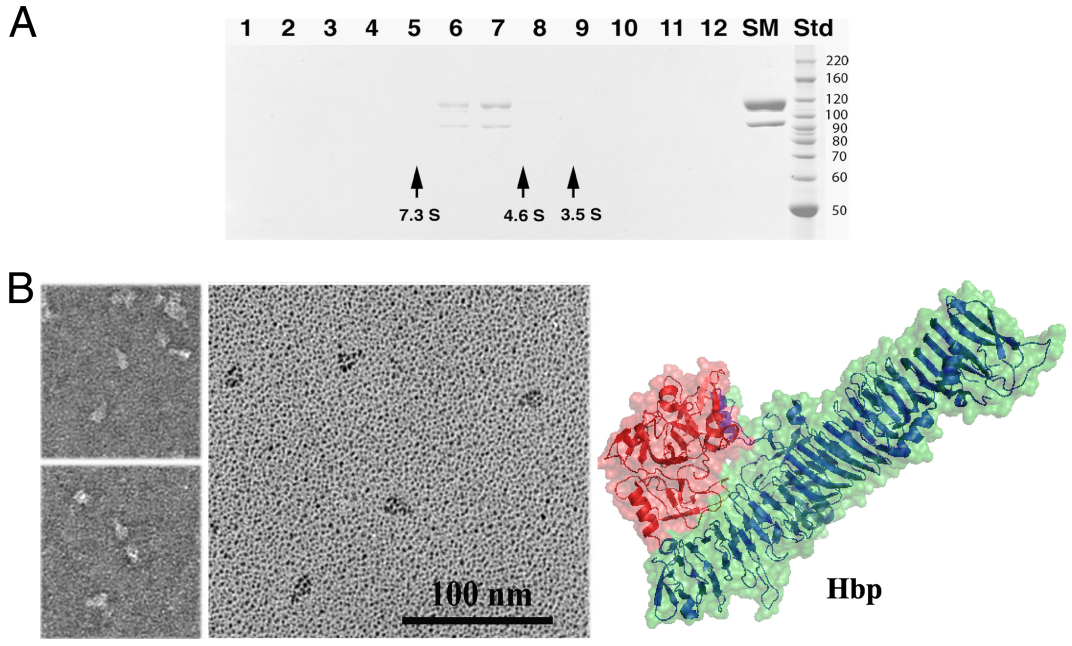

FIG. 1. (A) Glycerol gradient sedimentation of Hap_S. Hap_S sedimented at 5.5 S, with an S_{max}/S of 1.53. The positions of standard proteins are indicated by arrows from the left as follows: aldolase (7.3 S), bovine serum albumin (BSA) (4.6 S), and ovalbumin (3.5 S). SM, starting material. (B) Electron microscopy of purified Hap_S. Negative staining electron microscopy (upper and lower left panels) and rotary shadow electron microscopy images (middle panel) of purified Hap_S revealed an elongated arrowlike shape, which resembled the crystal structure of *E. coli* Hbp (right panel). Note that Hap in the rotary shadowed images is slightly larger than the negative stain images because of the platinum shell. The right panel shows the space-filling model and the overlaid ribbon model of the crystal structure of *E. coli* Hbp (PDB 1WXR), demonstrating an arrowlike shape. The serine protease domain is highlighted in red.

conventional autotransporters (31). $Hap_{551-850}$ was predicted by the PredictProtein server to be a domain consisting only of β strands (http://www.predictprotein.org/).

Using the homology model of Hap_S, we predicted the Hap substrate groove. First, we aligned Hap_{63-287} with the published crystal structures of alpha-chymotrypsin (PDB code 2CHA) and beta-trypsin (PDB code 5PTP). The structures of Hap_{63-287}, alpha-chymotrypsin, and beta-trypsin are very similar, with a root mean square deviation in Cα positions of 1.0 Å between Hap_{63-287} and alpha-chymotrypsin and 2.2 Å between Hap_{63-287} and beta-trypsin (Fig. 3A). Next, we inserted the inhibitor *p*-sulfinotoluene in these aligned structures to highlight the catalytic triad and to predict the S subsite residues of Hap_S (Fig. 3A). The coordinates of inhibitor *p*-sulfinotoluene were extracted from the superimposed alpha-chymotrypsin structure (2CHA) and merged into the homology model of Hap_{63-287}. By nomenclature, binding sites on a protease are called subsites (S residues), and each subsite residue interacts with a residue of the substrate (called P residues) (22). The predicted S1 subsite contains S243, N274, L263, and K240; the predicted S2 subsite contains H98, R264, and Y137; and the predicted S4 subsite contains E265 (Fig. 3B).

Mutagenesis analysis of the predicted substrate groove. To assess the accuracy of the predicted substrate groove, we performed site-directed mutagenesis, changing each of the predicted subsite residues individually and assessing the effect of each mutation on Hap autoproteolytic activity (Table 3). Wild-type residues were changed to amino acids that differ in terms of polarity, charge, or length of side chain. For the predicted S1 residues, we changed K240 from a charged amino acid to a nonpolar amino acid (K240A), L263 from a nonpolar amino acid to a charged polar amino acid (L263R), and N274 from an uncharged polar amino acid to a charged polar amino acid (N274R). For the predicted S2 subsite, we changed Y137 from an aromatic amino acid to a nonpolar amino acid (Y137A) and R264 from a charged polar amino acid to a nonpolar amino acid (R264A). For the predicted S4 subsite, we changed E265 from a charged polar amino acid to nonpolar amino acid (E265A) or an aromatic amino acid (E265W). In Western blot analysis of outer membrane preparations from strain DB117 expressing wild-type Hap, a 45-kDa band corresponding to Hap_β is apparent, representing cleavage at the primary cleavage site (Fig. 4A). By comparison, the N274R and L263R mutations at the predicted S1 subsite disrupted autoproteolysis, resulting in accumulation of the 155-kDa Hap precursor (Fig. 4A). Similarly, the R264A mutation at the predicted S2 subsite and the E265W mutation at the predicted S4 subsite also disrupted autoproteolysis. The E265A mutation had only a modest effect on autoproteolysis, as evidenced by partial accumulation of Hap precursor in the outer membrane. In contrast, the Y137A and K240A mutations had no effect on autoproteolysis, suggesting that Y137 and K240 may not be essential for mediating interactions between Hap_S and its substrate (Fig. 4A). Taken together, these mutagenesis data provide further support that residues N274, L263, R264, and E265 may indeed form the substrate binding groove in Hap_S predicted by our homology model (Fig. 3B).

Analysis of Hap cleavage specificity. In earlier studies we found that mutation of the P1 residue of the Hap cleavage site disrupted cleavage (13). To extend these studies, we studied Hap

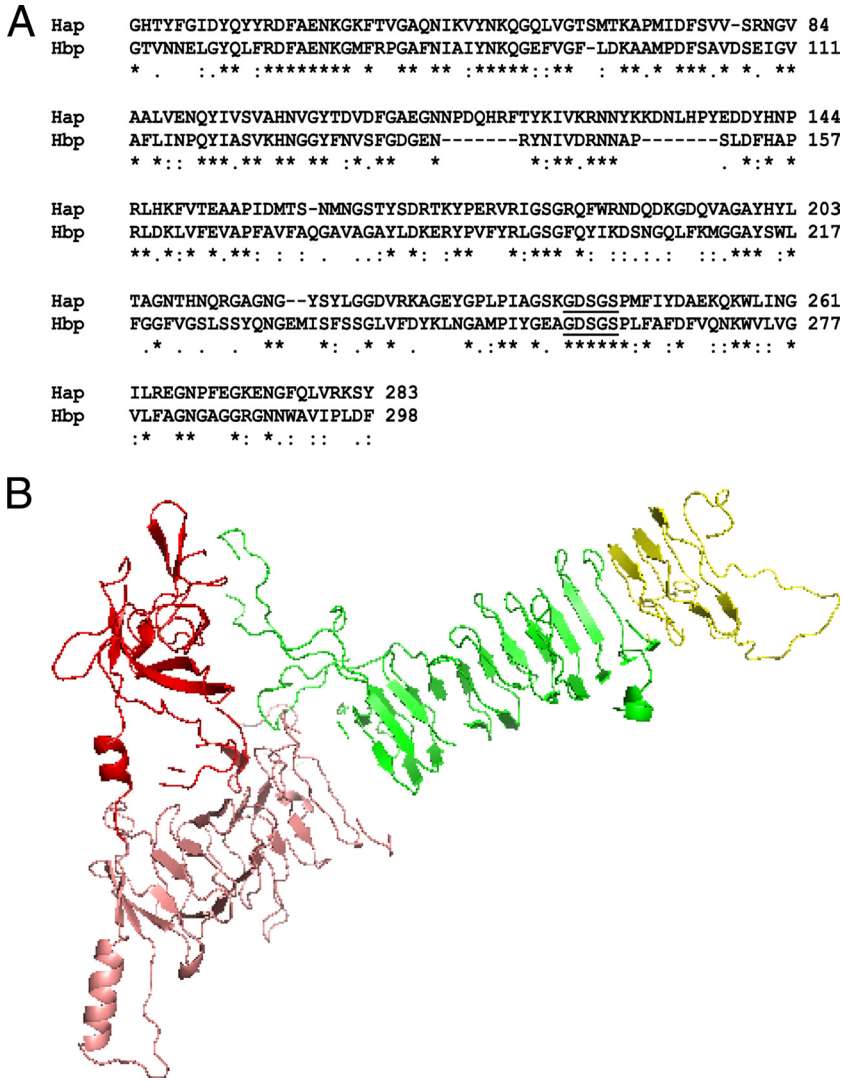

FIG. 2. (A) Sequence alignment between Hap and Hbp serine protease domains. Invariant residues are indicated by an asterisk. The conserved serine protease motif containing the catalytic serine is underlined. (B) Homology model of Hap_S. Hap_{70-550} (red and salmon) was modeled based on the Hbp structure (PDB code 1WXR). The region highlighted in red corresponds to the serine protease domain. $Hap_{851-980}$ (yellow) is modeled based on the Pertactin structure (PDB code 1DAB). $Hap_{551-850}$ (green) is predicted to be a β domain and is built on the crystal structure of Hbp based on the appearance of purified Hap_S by electron microscopy.

cleavage specificity by systematically mutating the P1 to P4 residues at cleavage sites and analyzing the effect of these mutations on autoproteolysis (Table 3). The P1 residue at the primary (L1036-N1037) and secondary (L1046-T1047) sites is a leucine. In contrast, the P1 residue at the tertiary (F1077-A1078) and quaternary (F1067-S1068) sites is a phenylalanine (Table 2). To address the possibility that leucine at the P1 position is preferred over the bulkier phenylalanine, leucine in the P1 position at the primary cleavage site was mutated to phenylalanine (L1036F). In addition, the bulky phenylalanine in the P1 position at both the tertiary and quaternary cleavage sites was mutated to leucine (F1077L and F1067L, respectively). As shown in Fig. 4B, the previously characterized L1036S mutation (6) that changes the P1 residue at the primary cleavage site to a serine resulted in loss of cleavage at the primary cleavage site and enhanced cleavage at the secondary, tertiary, and quaternary cleavage sites. The L1036F mutation in the present study resulted in slightly diminished cleavage at the primary cleavage site and enhanced cleavage at the secondary cleavage site (Fig. 4B). The F1077L mutation resulted in slightly increased cleavage at the tertiary cleavage site (Fig. 4B). When the L1036S and the F1077L mutations or the L1036F and the F1077L mutations were combined, cleavage at the tertiary site was markedly increased, whereas cleavage at the primary, secondary, and tertiary sites was markedly diminished (Fig. 4B). The F1067L mutation resulted in a marked increase in cleavage at the quaternary site (Fig. 4B). Taken together, these findings suggest that cleavage after leucine in the P1 position is more efficient than cleavage after the bulkier phenylalanine in that position.

The P2 residue at the primary and tertiary cleavage sites is leucine, while the P2 residue at the secondary cleavage site is glutamic acid and at the quaternary cleavage site is valine. To

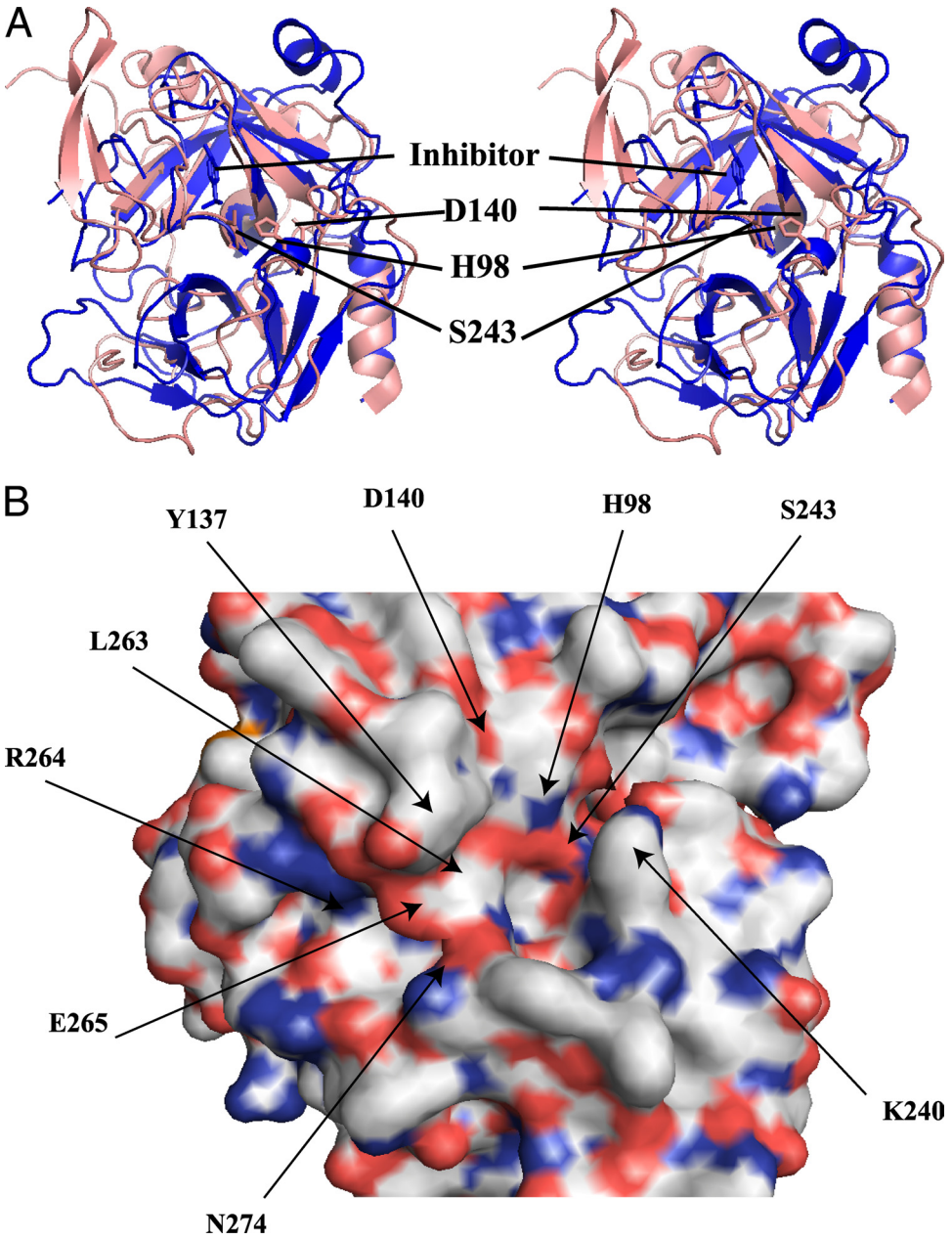

FIG. 3. (A) Stereo ribbon diagram of the superimposed Hap serine protease domain (salmon) and the *p*-sulfinotoluene bound catalytic alpha chymotrypsin (blue). The inhibitor, *p*-sulfinotoluene, and the adjacent catalytic triad of Hap H98, D140, and S243 are shown in stick. (B) Putative Hap substrate-binding groove. The catalytic triad and the residues that have been tested to be important for Hap autoproteolysis are labeled. L263 and N274 are predicted S1 residues, R264 is a predicted S2 residue, and E265 is a predicted S4 residue. Also labeled are residues K240 and Y137, which had no effect on autoproteolysis when mutated.

determine whether leucine is preferred in the P2 position, we initially changed the leucine at the primary cleavage site to polar residues serine (L1035S) and threonine (L1035T). The L1035S mutation resulted in diminished cleavage at the primary site and enhanced cleavage at the secondary, tertiary, and quaternary sites. The L1035T mutation virtually eliminated all cleavage, resulting in accumulation of the Hap precursor (Fig. 4B). These data indicate that the P2 position is important for substrate recognition. In order to further test the P2 position, we changed the P2 leucine at the primary cleavage site to a glutamic acid (L1035E), which is also present in the P2 position at the secondary cleavage site. The L1035E mutation showed wild-type cleavage (Fig. 4B). Next, we changed the glutamic acid in the P2 position at the secondary cleavage site to leucine (E1045L). The E1045L mutation resulted in prominent cleavage at the secondary cleavage site and almost no cleavage at the primary, tertiary, and quaternary cleavage sites (Fig. 4C). To further assess cleavage preference at the P2 position, we mutated the valine in the P2 position at the quaternary cleavage site to leucine (V1066L). As shown in Fig. 4C, the V1066L mutation resulted in enhanced

TABLE 3. Autoproteolysis phenotypes of Hap mutant derivatives

Wild type or mutation[a]	Cleavage site or subsite[b]	Description[c]
Hap wild type	NA	1°, 2°, and 3° cleavage
HapL1036S	P1 (1°)	No 1° cleavage; 2°, 3°, and 4° cleavage
HapL1036F	P1 (1°)	1° cleavage slightly diminished, enhanced secondary cleavage
HapF1077L	P1 (3°)	Slightly enhanced 3° cleavage
HapF1077L/L1036S	P1 (3°) + P1 (1°)	Enhanced 3° cleavage
HapF1077L/L1036F	P1 (3°) + P1 (1°)	Enhanced 3° cleavage
Hap F1067L	P1 (4°)	Enhanced 4° cleavage, some 1° and 3° cleavage
HapL1035S	P2 (1°)	Very little 1° cleavage, enhanced 2° cleavage
HapL1035T	P2 (1°)	Minimal autoproteolysis, accumulation of Hap precursor
HapL1035E	P2 (1°)	No effect on autoproteolysis
HapE1045L	P2 (2°)	Enhanced 2° cleavage
HapE1045A	P2 (2°)	Enhanced 2° cleavage
HapV1066L	P2 (3°)	Enhanced 4°, some 1° and 3° cleavage
HapS1034F	P3 (1°)	No effect on autoproteolysis
HapS1034E	P3 (1°)	No effect on autoproteolysis
HapQ1033A	P4 (1°)	No effect on autoproteolysis
HapQ1033E	P4 (1°)	No effect on autoproteolysis
HapN1037R	P1′ (1°)	No effect on autoproteolysis
HapA1038S	P2′ (1°)	No effect on autoproteolysis
HapL1039S	P3′ (1°)	No effect on autoproteolysis
HapE1040S	P4′ (1°)	No effect on autoproteolysis
HapK240A	S1	No effect on autoproteolysis
HapL263R	S1	Diminished autoproteolysis, accumulation of Hap precursor
HapN274R	S1	Diminished autoproteolysis, accumulation of Hap precursor
HapR264A	S2	Diminished autoproteolysis, accumulation of Hap precursor
HapY137A	S2	No effect on autoproteolysis
HapE265A	S4	Diminished autoproteolysis, accumulation of Hap precursor
HapE265W	S4	Diminished autoproteolysis, accumulation of Hap precursor

[a] All Hap derivatives were expressed in *H. influenzae* strain DB117.
[b] Numbers in parentheses indicate the primary (1°), secondary (2°), tertiary (3°), or quaternary (4°) cleavage sites. NA, not applicable.
[c] Descriptions are based on visual inspection of the blots in triplicate.

cleavage at the quaternary cleavage site. These observations indicate that the P2 residue is important in Hap cleavage specificity and suggest that leucine and glutamic acid are the preferred residues in this position for Hap proteolysis.

The P3 residue at all four cleavage sites is either a serine or an alanine. To test the importance of the P3 residue in Hap proteolysis, we mutated the serine at the primary cleavage site mutated to phenylalanine (S1034F), a much bulkier residue. We also changed the serine to a charged polar glutamic acid (S1034E). Both of these mutations had no effect on autoproteolysis (Fig. 4B), suggesting that this position does not significantly influence Hap cleavage specificity.

The P4 residue is glutamine at the primary, secondary, and tertiary cleavage sites and arginine at the quaternary site. To test the importance of the P4 residue, glutamine 1033 at the primary cleavage site was mutated to alanine, a smaller nonpolar residue (Q1033A), and to glutamic acid, a charged polar residue (Q1033E). These mutations did not have any effect on autoproteolysis (Fig. 4C), suggesting that the P4 position is not an important determinant of Hap proteolysis.

We also investigated the importance of the P′ residues in Hap cleavage specificity. However, the mutations N1037R, A1038S, L1039S, and E1040S at the P1′, P2′, P3′, and P4′ positions, respectively, had no effect on autoproteolysis (Fig. 4C).

Determination of the optimal Hap P motif using a peptide library. To complement our mutagenesis data and to determine the optimal Hap cleavage motif at the P4 through P1 positions, we incubated purified Hap_S with a C-terminally biotinylated peptide library containing fixed P′ motifs and randomized amino acids in the P positions. Figure 5 shows the selectivity ratios for amino acids selected at each position. Consistent with the mutagenesis data presented here, a leucine residue in both the P1 and the P2 positions strongly favored proteolytic cleavage. Arginine at P1 and glutamic acid at P2 seemed to favor cleavage to a lesser extent. Glutamic acid seems to be preferred at the P3 position. A number of amino acids had selectivity ratios greater than 1 at the P4 position, although no strong preference was evident for any particular amino acid.

DISCUSSION

In this study, we have investigated the Hap cleavage specificity and substrate groove in detail. Our results demonstrate that the P1 and P2 residues are important determinants of Hap cleavage specificity, that small hydrophobic amino acids are favored in the P1 and P2 positions, and that the substrate groove likely includes the L263, R264, E265, and N274 amino acids.

Our earlier work demonstrated that Hap is a member of the SA (chymotrypsin) clan of serine proteases (6). Interactions between the P1 residue and the S1 subsite pocket encompassing the residues adjacent to the active site serine are important determinants of cleavage specificity of chymotrypsinlike proteases (9). For chymotrypsin, cleavage specificity correlates to the hydrophobicity of the P1 residue and is characterized by preferential cleavage after tryptophan, tyrosine, or phenylala-

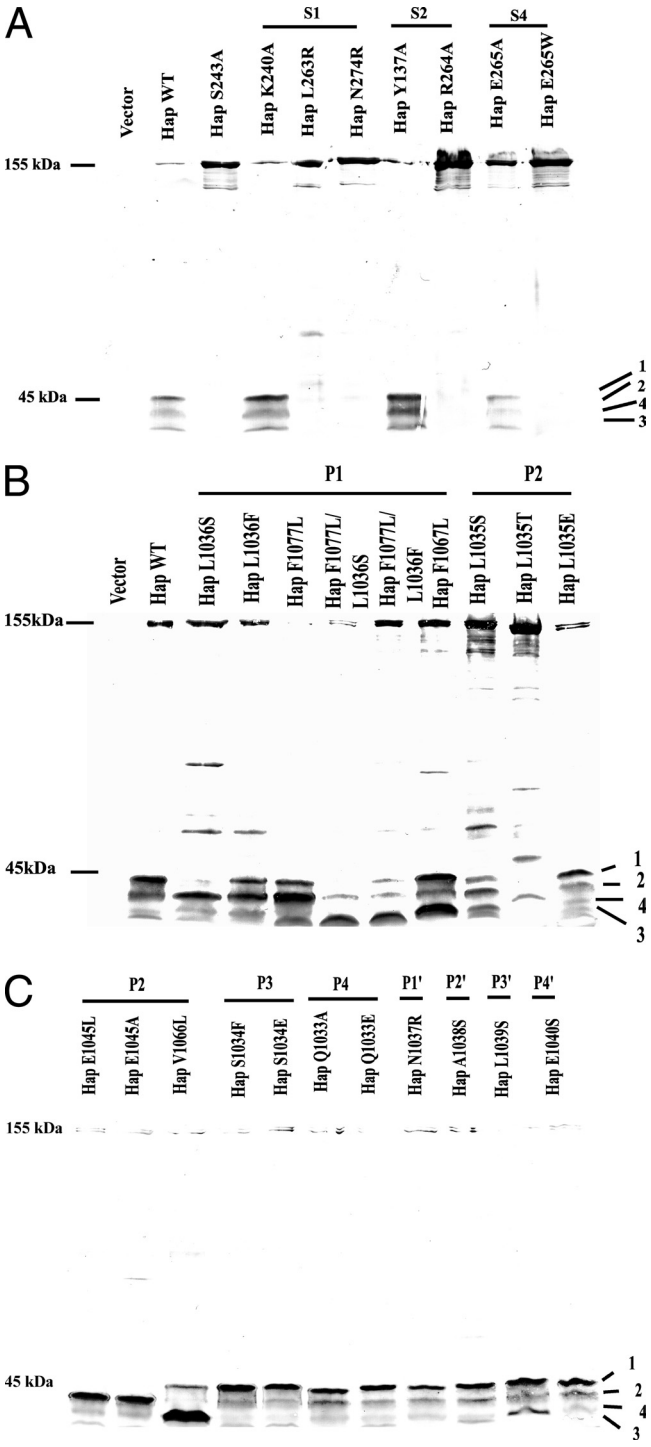

FIG. 4. Western blots of outer membrane proteins purified from Hap mutant derivatives. The presence of Hap in the outer membrane was detected using antibody Rab290 raised against the C-terminal Hap fragment encompassing residues 996 to 1395. (A) Mutagenesis analysis of predicted Hap substrate binding groove. (B and C) Mutagenesis analysis of Hap cleavage specificity. Full-length Hap is seen at 155 kDa, and Hap_{β} species are seen at 45 to 39 kDa. The different Hap_{β} derivatives result from cleavage at the different cleavage sites and are labeled "1" for primary cleavage (45 kDa), "2" for secondary cleavage (43 kDa), "3" for tertiary cleavage (39 kDa), and "4" for quaternary cleavage (41 kDa). Immunoblots shown here are representative of blots performed in triplicate. Assessment of cleavage at the various cleavage sites was performed by manual inspection.

nine, with all of these residues fitting within the deep hydrophobic pocket of the protein (16, 17). However, chymotrypsin can also cleave after leucine, methionine, and histidine in the P1 position, although with lower efficiency (15, 16). In contrast, elastase prefers small hydrophobic residues such as alanine in the P1 position, reflecting a smaller S1 subsite pocket (24). In the present study, we demonstrated that Hap cleaves with high preference for leucine over phenylalanine in the P1 position, suggesting that the S1 subsite pocket in Hap is relatively shallow and accommodates small hydrophobic residues better than bulky residues. This conclusion is consistent with the relative inefficiency of cleavage at the tertiary and quaternary sites in wild-type Hap, where the P1 residue is phenylalanine.

In chymotrypsin, proteolysis is also influenced by interaction between the P2 residue and the S2 subsite pocket (3). In this context, it is noteworthy that our work supports an important role for the P2 residue in Hap cleavage specificity, with leucine or glutamic acid favored over other amino acids. Interestingly, insertion of leucine in the P2 position at the secondary cleavage site (HapE1045L) results in a preference for this cleavage site to the exclusion of other cleavage sites. It is possible that these mutations result in local steric hindrance that causes increased binding of the mutant substrate and more efficient cleavage at the secondary cleavage site. However, without a high-resolution crystal structure, we cannot explain the apparent preference of the secondary cleavage site over the primary cleavage site in this mutant. We also found that insertion of glutamic acid at the P2 position (L1035E) of the primary cleavage site had no significant effect on autoproteolysis. Consistent with this finding, glutamic acid is the P2 residue at the native secondary cleavage site (E1045).

Hap cleavage of a random peptide library revealed high preference for leucine at the P1 and P2 positions, agreeing with our mutagenesis results. Interestingly, phenylalanine was not selected at either of these positions, despite the fact that cleavage occurs naturally at F1067 and F1077 in wild-type Hap (albeit to a lesser extent than with leucine in the P1 position). In considering this information, it is possible that the S1 and S2 subsites in the Hap proteolytic pocket have reduced affinity for phenylalanine in the context of linear peptides compared to native folded protein. Based on selectivity ratios, arginine at the P1 position and glutamic acid at the P2 position seem to be the favored amino acids after leucine. Serine and threonine have the lowest selectivity ratio at the P1 and P2 positions, indicating that they have an inhibitory effect on proteolytic cleavage, in agreement with our mutagenesis data (L1036S, L1035S, and L1035T). Further analysis of the peptide library cleavage assay revealed that glutamic acid was selected at the P3 position and that a number of amino acids were selected at the P4 position.

Considered together, our mutational analysis and our peptide library cleavage results indicate that the P1 and P2 residues are important determinants of Hap cleavage specificity. Previously, we reported a consensus cleavage site sequence based only on alignment of the primary, secondary, tertiary, and quaternary cleavage sites (6). In the present study, mutational analysis reveals a target motif that consists of XX(L/E/V)(L/F) and the peptide library cleavage assay reveals a target motif of XE(L/E)(L/R) at the P4 through P1 positions for the Hap_S protease. The difference in results between mutational

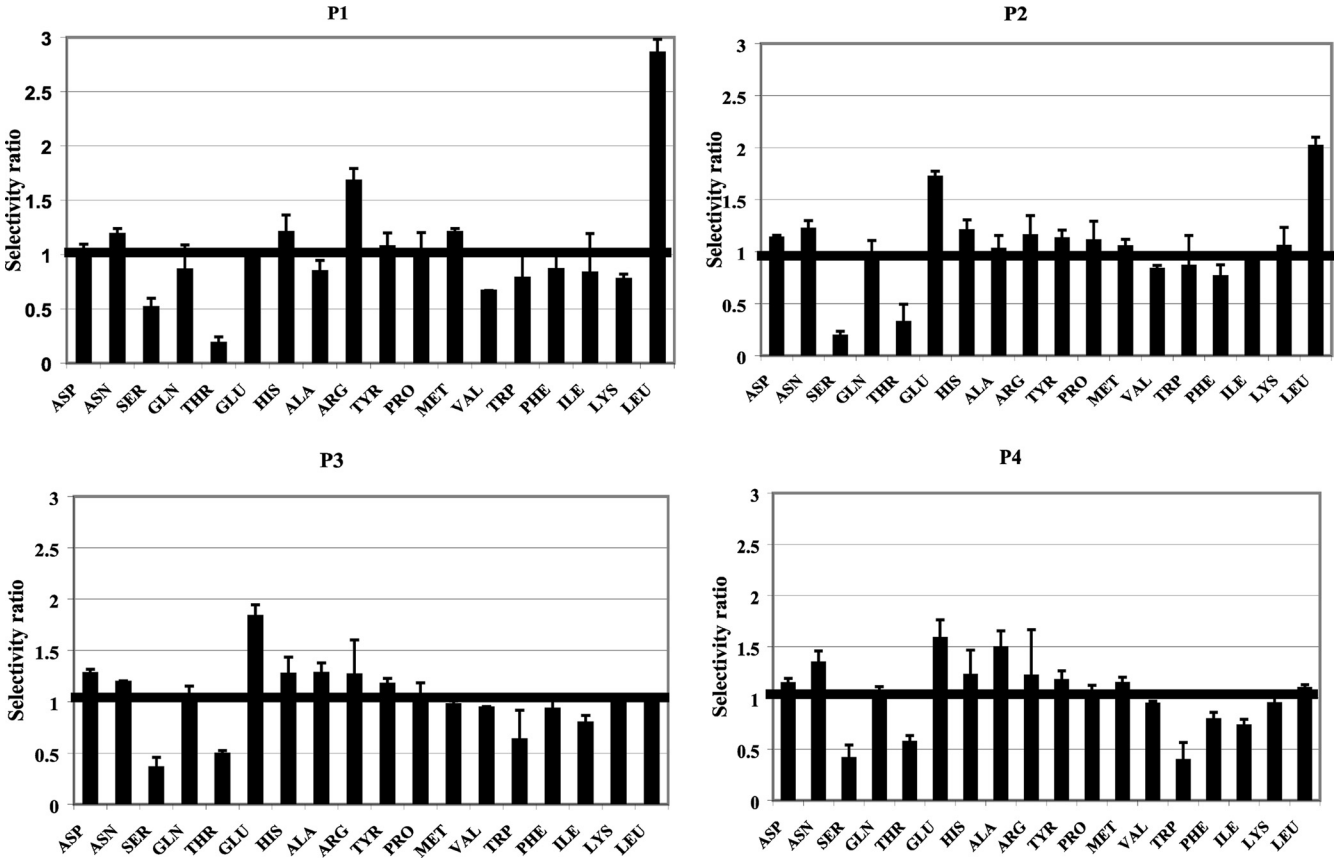

FIG. 5. Selectivity ratios obtained for each amino acid at the P1 to P4 positions in the peptide library cleavage assay using purified Hap_S. Selectivity ratios represent an average of three different data sets. Ratios are normalized to 1, so that residues with values greater than 1 are selected and residues with values at or less than 1 are not selected. The data for glycine were not included, since sequencing resulted in background peaks for this residue in the experimental as well as control samples.

analysis and peptide library cleavage may be due to the intrinsic difference between cleavage of full-length folded protein and linear peptides. Mutations introduced at the P′ residues of the primary cleavage site had no effect on autoproteolysis, indicating that these residues do not significantly affect proteolytic cleavage.

In an effort to define the Hap substrate binding groove, we developed a homology model based on the crystal structures of the *E. coli* Hbp and the *B. pertussis* Pertactin autotransporters. Guided by this model, we performed site-directed mutagenesis, changing residues at the predicted S1, S2, and S4 subsites. Our results demonstrated that L263, N274, R264, and E265 are critical for efficient autoproteolysis, supporting a role for these residues in forming the Hap substrate groove and interacting intimately with substrate residues. Mutations at Y137 and K240 that are located at the rim of S1 and S2 subsites in our model did not affect autoproteolysis.

The implication of the Hap serine protease domain in the pathogenesis of *H. influenzae* disease could be twofold. First, autoproteolytic release of Hap_S from the bacterial surface may allow the organism to modulate interactions with host tissues during various stages of colonization and infection. Second, the secreted Hap protease domain may have activity against host substrates such as extracellular matrix proteins or immune factors. We have previously determined that Hap can bind to various extracellular matrix proteins such as fibronectin, collagen IV, and laminin (7). Further investigation is needed to determine whether Hap protease activity might facilitate spread to deeper tissues or promote persistence in the face of the host immune response.

In conclusion, in this study we have provided important insights into the structural determinants of *H. influenzae* Hap proteolytic activity. Complete knowledge of the Hap substrate groove and cleavage specificity may facilitate the design of therapeutic agents that block Hap proteolytic activity and thereby attenuate *H. influenzae* disease.

ACKNOWLEDGMENTS

We thank Ben Turk for advice on the peptide library cleavage assays.

This study was supported by Public Health Service grant AI44322 to J.S.G. and Wellcome Trust grant 08087 to G.W. and G.M.

REFERENCES

1. **Anderson, P., R. B. Johnston, Jr., and D. H. Smith.** 1972. Human serum activities against *Haemophilus influenzae*, type b. J. Clin. Investig. **51:**31–38.
2. **Barenkamp, S. J., and J. W. St. Geme III.** 1994. Genes encoding high-molecular-weight adhesion proteins of nontypeable *Haemophilus influenzae* are part of gene clusters. Infect. Immun. **62:**3320–3328.
3. **Brady, K., and R. H. Abeles.** 1990. Inhibition of chymotrypsin by peptidyl trifluoromethyl ketones: determinants of slow-binding kinetics. Biochemistry **29:**7608–7617.
4. **Carlone, G. M., M. L. Thomas, H. S. Rumschlag, and F. O. Sottnek.** 1986.

Rapid microprocedure for isolating detergent-insoluble outer membrane proteins from *Haemophilus* species. J. Clin. Microbiol. **24:**330–332.
5. **Fink, D. L., A. Z. Buscher, B. Green, P. Fernsten, and J. W. St. Geme III.** 2003. The *Haemophilus influenzae* Hap autotransporter mediates microcolony formation and adherence to epithelial cells and extracellular matrix via binding regions in the C-terminal end of the passenger domain. Cell Microbiol. **5:**175–186.
6. **Fink, D. L., L. D. Cope, E. J. Hansen, and J. W. St. Geme III.** 2001. The *Haemophilus influenzae* Hap autotransporter is a chymotrypsin clan serine protease and undergoes autoproteolysis via an intermolecular mechanism. J. Biol. Chem. **276:**39492–39500.
7. **Fink, D. L., B. A. Green, and J. W. St. Geme III.** 2002. The *Haemophilus influenzae* Hap autotransporter binds to fibronectin, laminin, and collagen IV. Infect. Immun. **70:**4902–4907.
8. **Fowler, W. E., and H. P. Erickson.** 1979. Trinodular structure of fibrinogen: confirmation by both shadowing and negative stain electron microscopy. J. Mol. Biol. **134:**241–249.
9. **Hedstrom, L.** 2002. Serine protease mechanism and specificity. Chem. Rev. **102:**4501–4524.
10. **Henderson, I. R., and J. P. Nataro.** 2001. Virulence functions of autotransporter proteins. Infect. Immun. **69:**1231–1243.
11. **Henderson, I. R., F. Navarro-Garcia, M. Desvaux, R. C. Fernandez, and D. Ala'Aldeen.** 2004. Type V protein secretion pathway: the autotransporter story. Microbiol. Mol. Biol. Rev. **68:**692–744.
12. **Henderson, I. R., F. Navarro-Garcia, and J. P. Nataro.** 1998. The great escape: structure and function of the autotransporter proteins. Trends Microbiol. **6:**370–378.
13. **Hendrixson, D. R., M. L. de la Morena, C. Stathopoulos, and J. W. St. Geme III.** 1997. Structural determinants of processing and secretion of the *Haemophilus influenzae* Hap protein. Mol. Microbiol. **26:**505–518.
14. **Hendrixson, D. R., and J. W. St. Geme III.** 1998. The *Haemophilus influenzae* Hap serine protease promotes adherence and microcolony formation, potentiated by a soluble host protein. Mol. Cell **2:**841–850.
15. **Keil, B.** 1987. Proteolysis Data Bank: specificity of alpha-chymotrypsin from computation of protein cleavages. Protein Seq. Data Anal. **1:**13–20.
16. **Keil, B.** 1992. Specificity of proteolysis, p. 335. Springer-Verlag, Berlin, Germany.
17. **Knowles, J. R.** 1965. Enzyme specificity: alpha-chymotrypsin. J. Theor. Biol. **9:**213–228.
18. **Kubiet, M., R. Ramphal, A. Weber, and A. Smith.** 2000. Pilus-mediated adherence of *Haemophilus influenzae* to human respiratory mucins. Infect. Immun. **68:**3362–3367.
19. **Murphy, T. F., J. M. Bernstein, D. M. Dryja, A. A. Campagnari, and M. A. Apicella.** 1987. Outer membrane protein and lipooligosaccharide analysis of paired nasopharyngeal and middle ear isolates in otitis media due to nontypeable *Haemophilus influenzae*: pathogenetic and epidemiological observations. J. Infect. Dis. **156:**723–731.
20. **Otto, B. R., R. Sijbrandi, J. Luirink, B. Oudega, J. G. Heddle, K. Mizutani, S. Y. Park, and J. R. Tame.** 2005. Crystal structure of hemoglobin protease, a heme binding autotransporter protein from pathogenic *Escherichia coli*. J. Biol. Chem. **280:**17339–17345.
21. **Rao, V. K., G. P. Krasan, D. R. Hendrixson, S. Dawid, and J. W. St. Geme III.** 1999. Molecular determinants of the pathogenesis of disease due to non-typeable *Haemophilus influenzae*. FEMS Microbiol. Rev. **23:**99–129.
22. **Schechter, I., and A. Berger.** 1967. On the size of the active site in proteases. I. Papain. Biochem. Biophys. Res. Commun. **27:**157–162.
23. **Schurmann, G., J. Haspel, M. Grumet, and H. P. Erickson.** 2001. Cell adhesion molecule L1 in folded (horseshoe) and extended conformations. Mol. Biol. Cell **12:**1765–1773.
23a. **Settow, J. K., D. C. Brown, M. E. Boling, A. Mattingly, and M. P. Gordon.** 1968. Repair of deoxyribonucleic acid in *Haemophilus influenzae*. I. X-ray sensitivity of ultraviolet-sensitive mutants and their behavior as hosts to ultraviolet-irradiated becteriophage and transforming deoxyribonucleic acid. J. Bacteriol. **95:**546–558.
24. **Shotton, D. M., and H. C. Watson.** 1970. Three-dimensional structure of tosyl-elastase. Nature **225:**811–816.
25. **Steinhart, W. L., and R. M. Herriott.** 1968. Genetic integration in the heterospecific transformation of *Haemophilus influenzae* cells by *Haemophilus parainfluenzae* deoxyribonucleic acid. J. Bacteriol. **96:**1725–1731.
26. **St. Geme, J. W., III, M. L. de la Morena, and S. Falkow.** 1994. A *Haemophilus influenzae* IgA protease-like protein promotes intimate interaction with human epithelial cells. Mol. Microbiol. **14:**217–233.
27. **St. Geme, J. W., III, S. Falkow, and S. J. Barenkamp.** 1993. High-molecular-weight proteins of nontypeable *Haemophilus influenzae* mediate attachment to human epithelial cells. Proc. Natl. Acad. Sci. USA **90:**2875–2879.
27a. **Tomb, J. F., G. J. Barcak, M. S. Chandler, R. J. Redfield, and H. D. Smith.** 1989. Transposon mutagenesis, characterization, and cloning of transformation genes of *Haemophilus influenzae*. J. Bacteriol. **171:**3796–3802.
28. **Turk, B. E., and L. C. Cantley.** 2004. Using peptide libraries to identify optimal cleavage motifs for proteolytic enzymes. Methods **32:**398–405.
29. **Turk, D. C.** 1984. The pathogenicity of *Haemophilus influenzae*. J. Med. Microbiol. **18:**1–16.
30. **Unzai, S., S. Y. Park, and J. R. Tame.** 2005. Crystal structure of heme binding protein, an autotransporter hemoglobin protease from pathogenic *Escherichia coli*. Tanpakushitsu Kakusan Koso **50:**1322–1327. (In Japanese.)
31. **Yen, Y. T., M. Kostakioti, I. R. Henderson, and C. Stathopoulos.** 2008. Common themes and variations in serine protease autotransporters. Trends Microbiol. **16:**370–379.

Editor: A. J. Bäumler