INFECTION AND IMMUNITY, Oct. 1989, p. 3097–3105
0019-9567/89/103097-09$02.00/0

Vol. 57, No. 10

Cloning and Sequencing of the Immunoglobulin A1 Protease Gene (*iga*) of *Haemophilus influenzae* Serotype b

K. POULSEN,[1] J. BRANDT,[2] J. P. HJORTH,[2] H. C. THØGERSEN,[3] AND M. KILIAN[1]*

Department of Oral Biology, Royal Dental College,[1] and Departments of Molecular Biology[2] and Chemistry,[3] University of Aarhus, DK-8000 Aarhus C, Denmark

Received 6 April 1989/Accepted 22 June 1989

Secretion of immunoglobulin A1 (IgA1) proteases is a characteristic of *Haemophilus influenzae* and several other bacterial pathogens causing infectious diseases, including meningitis. Indirect evidence suggests that the proteases are important virulence factors. In this study, we cloned the *iga* gene encoding immunoglobulin A1 (IgA1) protease from *H. influenzae* serotype b into *Escherichia coli*, in which the recombinant *H. influenzae iga* gene was expressed and the resulting protease was secreted. Sequencing a part of a 7.5-kilobase DNA fragment containing the *iga* gene revealed a large open reading frame with a strongly biased codon usage and having the potential of encoding a protein of 1,541 amino acids and a molecular mass of 169 kilodaltons. Putative promoter and terminator elements flanking the open reading frame were identified. Comparison of the deduced amino acid sequence of this *H. influenzae* IgA1 protease with that of a similar protease from *Neisseria gonorrhoeae* revealed several domains with a high degree of homology. Analogous to mechanisms known from the *N. gonorrhoeae* IgA protease secretion, we propose a scheme of posttranslational modifications of the *H. influenzae* IgA1 protease precursor, leading to a secreted protease with a molecular mass of 108 kilodaltons, which is close to the 100 kilodaltons reported for the mature IgA1 protease.

The immunoglobulin A1 (IgA1) proteases are extracellular bacterial enzymes that specifically cleave the heavy chain of the human IgA1 molecule in the hinge region. Production of IgA1 protease seems to be correlated with the ability of the bacteria to infect and invade mucosal membranes. Notably, the three leading etiological agents of bacterial meningitis (*Haemophilus influenzae*, *Neisseria meningitidis*, and *Streptococcus pneumoniae*) secrete IgA1 proteases (16, 27, 30), whereas closely related nonpathogenic species are devoid of similar enzyme activity. IgA1 protease production is also a property of certain bacterial species causing vaginal and urinary tract infections, such as *Neisseria gonorrhoeae* and *Ureaplasma urealyticum* (16, 27, 30). IgA1 proteases are likely to be important virulence factors, because they allow bacteria to interfere with the protective mechanisms of IgA1, which is the predominant class of immunoglobulin present on mucosal membranes (14).

Several enzymatically different types of IgA1 protease have been reported, each cleaving the human IgA1 molecule at a different site in the hinge region (Fig. 1) (16, 27, 30). In *H. influenzae* two such distinct enzymes are known (16, 27). In addition to this variation in substrate specificity, at least 15 antigenically different types of the IgA1 protease have been identified among *H. influenzae* isolates based on enzyme-neutralizing antibodies raised in rabbits (17). These "inhibition types" of the produced protease correlate strongly with the capsular serotype of the strains.

The *iga* genes encoding IgA1 protease produced by *N. gonorrhoeae* (type 1 [7] and type 2 [12, 19]), *H. influenzae* (type 1 of serotype d origin [5, 18] and type 2 [10]), *N. meningitidis* (18), and *Streptococcus sanguis* (8) have been cloned. *Escherichia coli* cells transformed with each of these *iga* genes express IgA1 protease activity; the *E. coli* hosts, which harbor the *H. influenzae* and *N. gonorrhoeae iga* genes, have been found to excrete the protease into the medium. At present, the type 2 IgA1 protease gene from *N. gonorrhoeae* is the only *iga* gene for which the nucleotide sequence is known (31). The deduced primary structure of the protein together with analyses of intermediate IgA1 protease precursors revealed that the protease is expressed as a preprotein, which is autoproteolytically processed during the secretion process (31).

We have previously used fragments of the cloned type 1 IgA1 protease gene of *H. influenzae* serotype d origin as probes in Southern blot experiments to study restriction site polymorphism of the *iga* gene among different strains of *H. influenzae* serotype b (32). Based on these *iga* gene restriction patterns, the strains could be divided into four groups. We found that 98% of the natural *H. influenzae* serotype b population could be assigned to *iga* gene types, I, II, or III. These three groups exclusively contained strains of the same unique IgA1 protease inhibition type. For the present study, we chose *H. influenzae* HK368 as a representative for the most common serotype b *iga* gene type (32), and we report here the cloning and sequence analysis of the *iga* gene from this strain. A comparison of the deduced amino acid sequence of this protease with that of *N. gonorrhoeae* reveals regions of homology.

MATERIALS AND METHODS

Bacterial strains. The wild-type *H. influenzae* serotype b strain HK368, isolated from cerebrospinal fluid from a meningitis patient, was the source of whole-cell DNA for molecular cloning. HK368 was grown in brain heart infusion broth (Difco Laboratories, Detroit, Mich.) supplemented with hemin and NAD (15). *E. coli* K802 (39) was used for propagation of lambda phages (25). The M13mp19 phage (29) and recombinant derivatives thereof were propagated in *E. coli* JM109 as described by Norrander et al. (29).

Enzymes and chemicals. Restriction endonucleases were purchased from Amersham International (Amersham, England) and Boehringer GmbH (Mannheim, Federal Republic of Germany). T4 DNA ligase, proteinase K, RNase A, and DNase I were obtained from Boehringer; DNA polymerase

* Corresponding author.

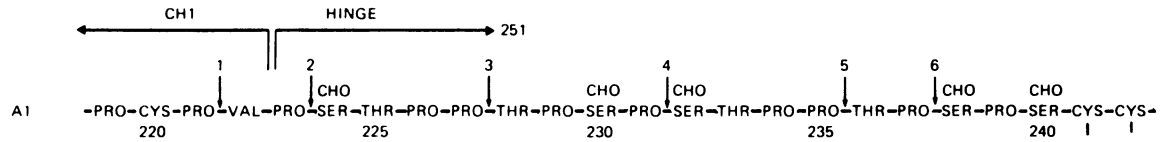

FIG. 1. Primary structure of the hinge region of human IgA1. The arrows indicate the peptide bonds cleaved by the individual IgA1 proteases as follows: 1, *Clostridium ramosum*; 2, *Bacteroides melaninogenicus* and *Capnocytophaga ochracea*; 3, *S. pneumoniae*, *Streptococcus oralis* ("*mitior*"), and *Streptococcus sanguis*; 4, *H. influenzae* protease type 1 (including serotype b) and *Haemophilus aegyptius*; 5, *H. influenzae* protease type 2, *N. gonorrhoeae* protease type 2, and *N. meningitidis* protease type 2; 6, *N. gonorrhoeae* protease type 1 and *N. meningitidis* protease type 1. CHO, carbohydrate side chains.

I, DNA polymerase I Klenow fragment, and radioactively labeled deoxynucleotides were from Amersham. Exonuclease III came from Pharmacia Fine Chemicals (Uppsala, Sweden), lysozyme was from Sigma Chemical Co. (St. Louis, Mo.), M13 pentadecamer primer was from New England BioLabs (Beverly, Mass.), and the Sequenase DNA sequencing kit was from United States Biochemical Corp. (Cleveland, Ohio).

Cloning of *H. influenzae* HK368 DNA in phage λL47.1. DNA from *H. influenzae* isolated as described previously (32) was partially digested with *Sau*3A and fractionated by agarose gel electrophoresis. Fragments ranging in size from 12 to 20 kilobases (kb) were extracted from the gel by electroelution and cloned by using λL47.1 (24) as the *Bam*HI substitution vector (26). Phages packaged in vitro as described by Maniatis et al. (25) were plated on *E. coli* K802, and positive plaques were identified by in situ hybridization. Positive plaques were purified by replating on *E. coli* K802, and DNA was isolated from 20-ml phage cultures (26).

Nucleic acid hybridizations and DNA manipulations. As a label in the hybridizations we used [α-^{32}P]dATP, which was incorporated into the DNA probes by nick translation (33). The in situ probing of phages fixed on nitrocellulose filters was carried out as described by Benton and Davis (1), and Southern blot experiments were performed as described previously (32). As probes we used fragments of plasmid pVD116, containing the *H. influenzae* type d *iga* gene (18), kindly provided by J. M. Koomey, The Rockefeller University, New York, N.Y. Subclonings into M13mp19 of restriction fragments isolated by electroelution and DNA purifications and manipulations were carried out as described by Maniatis et al. (25).

Progressive deletions of recombinant M13mp19 phages for the sequencing procedure were produced by varying the time of exonuclease III digestion of *Bam*HI-*Sac*I-opened replicative-form DNA (40). For removal of the resulting single-stranded ends, we used S1 nuclease instead of exonuclease VII.

DNA sequencing. The sequences of individual clones were determined by the dideoxynucleotide-chain termination method (3, 35) with the 15-mer universal primer and [α-^{35}S] dATP. For some of the sequences the Sequenase DNA sequencing kit was used according to the manufacturer's recommendations. The program of Larson and Messing (21) was used to assemble the sequence data with an Apple IIe computer.

Sequence alignment. Computer-assisted analysis of amino acid sequence data was performed by using the MULTALI program for alignments (23).

Assay for IgA1 protease activity. IgA1 protease activity in phage lysates was detected by mixing 1 volume of the lysate supernatant with 1 volume of a 2-mg/ml solution of human myeloma IgA1. After overnight incubation, cleavage of the substrate IgA1 was demonstrated by immunoelectrophoresis as described previously (17).

Immunization and inhibition assay. The supernatant of an *E. coli* lysate of a recombinant λ phage containing the *iga* gene was used for immunization of a rabbit. The immunization procedure and protease inhibition assay were as described previously (17).

RESULTS

Cloning of the *H. influenzae* serotype b *iga* gene in *E. coli*. The *E. coli* plasmid pVD116 contains the *iga* gene from *H. influenzae* Rd$^-$/b$^+$ (19). A 2.0-kb *Cla*I-*Pst*I fragment of this plasmid containing the 5′ end of the *iga* gene of serotype d origin was used as a radioactively labeled probe to isolate the *iga* gene from a lambda-phage library of *H. influenzae* serotype b strain HK368 DNA with the vector L47.1. Replica filters of approximately 2×10^3 recombinant phages from the nonamplified library were screened with the probe. Sixteen positive clones labeled λ368*iga*-1 to λ368*iga*-16 were found. Assuming that *iga* is a single-copy gene, this frequency is in acceptable agreement with the estimated *H. influenzae* genome size of 1.8×10^3 kb (22) and an average insert of 15 kb in the recombinant λ phages. Eleven of these positive clones were purified, and recombinant phage DNA was isolated. The localization of recognition sites for the restriction endonucleases *Eco*RI and *Hin*dIII in the insert DNA were determined by complete and partial digestions of DNA from the individual 368*iga* phages as previously described (26). In addition, DNA from λ368*iga*-8 and λ368*iga*-16 was mapped with respect to the restriction enzymes *Bam*HI, *Cla*I, and *Pst*I. The restriction maps of the inserts overlap as shown in Fig. 2. This strongly suggests that the *iga* gene in *H. influenzae* HK368 is a single-copy gene.

Localization of the *iga* gene. The previously described 5′-*iga*-specific probe together with a 3′-*iga*-specific 2.8-kb *Pst*I-*Eco*RI fragment of pVD116 were hybridized to replica Southern blots of *Eco*RI-, *Hin*dIII-, and *Sau*3A-restricted DNA from each of the 11 λ368*iga* clones analyzed. Thereby, the *iga* gene was deduced to be localized within a 7.5-kb *Eco*RI-*Hin*dIII fragment and oriented as shown in Fig. 2.

The restriction enzyme *Sau*3A sites at the insert-vector junctions were preserved during the cloning. For 7 of the 11 recombinant clones, the sizes and numbers of DNA fragments generated by *Sau*3A and hybridizing to the two *iga* probes were identical. The same pattern of bands was observed in Southern blots of *Sau*3A-digested whole-cell DNA from *H. influenzae* HK368 hybridized with the same two probes (data not shown). This indicates that no rearrangements of the hybridizing region in these lambda phages had occurred during the cloning procedure. In addition, this result confirms that the same *iga*-specific sequence had been cloned in these phages.

Liquid lysates from these seven recombinant phage clones were found to possess IgA1 protease activity. This confirms the cloning of the *iga* gene. In the lysates from 3 of the 11 clones, i.e., λ368*iga*-2, λ368*iga*-4, and λ368*iga*-5, we could

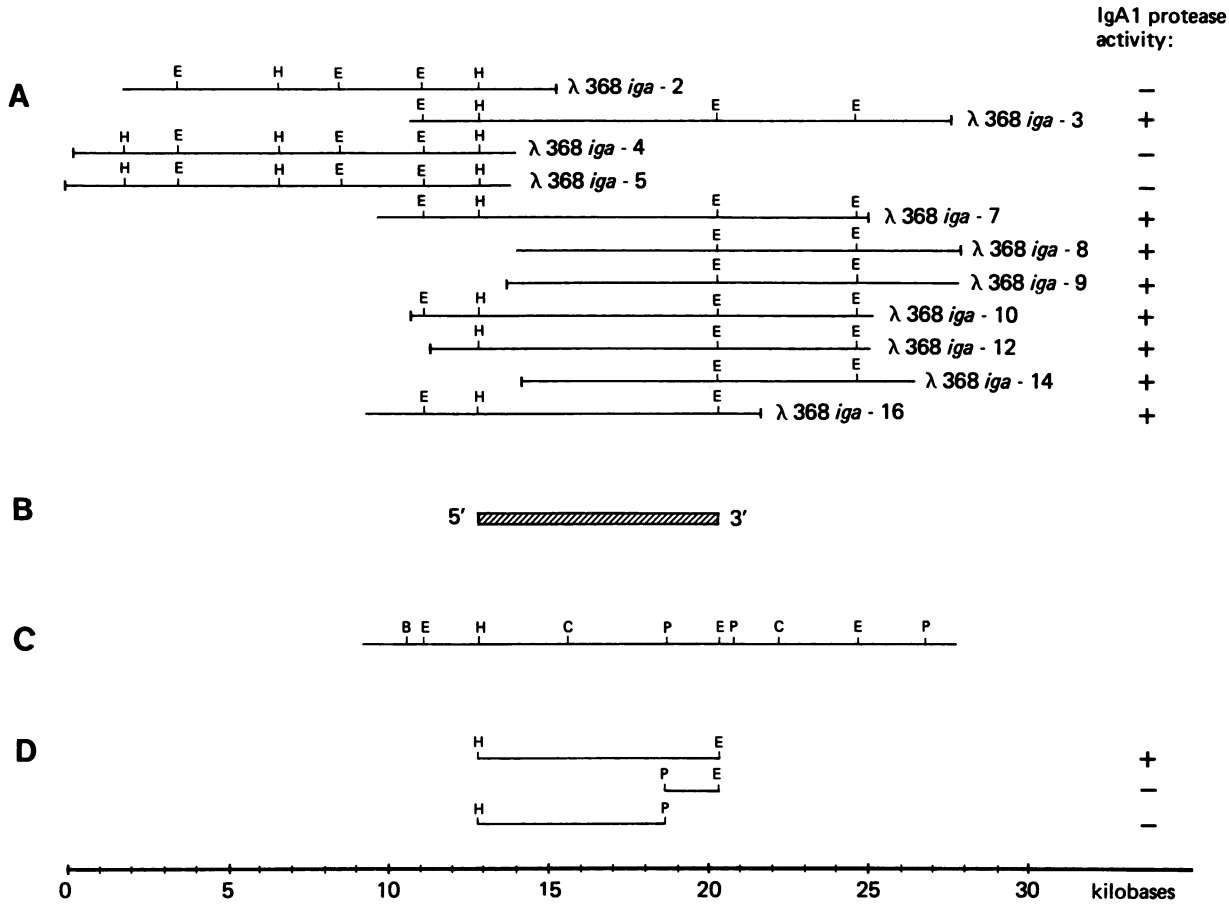

FIG. 2. Structure of the *H. influenzae* serotype b IgA1 protease gene. (A) Restriction maps of the recombinant lambda-phages containing sequences with homology to the *H. influenzae* serotype d *iga* gene probe. Symbols and abbreviations: E, *Eco*RI; H, *Hin*dIII; terminal bars, left arms of the individual phages; + or −, presence or absence, respectively, of human IgA1-cleaving activity in culture supernatants of *E. coli* harboring the individual clones. (B) Restriction fragments from *Eco*RI-*Hin*dIII double digests hybridizing to the two serotype d *iga* gene probes. The deduced orientation of the *iga* gene is indicated by 5′ and 3′. (C) Combined restriction map of λ368*iga*-8 and λ368*iga*-16, an extension of A by including the following restriction enzymes: B, *Bam*HI; C, *Cla*I; P, *Pst*I. (D) Restriction fragments of λ368*iga*-16 subcloned into M13mp19 for sequence analysis.

not detect IgA1-cleaving activity. DNA from these three clones lacked some of the *Sau*3A fragments found in strain HK368 genomic DNA, which hybridized to the 3′ *iga* probe, and therefore only contain the 5′ part of the *iga* gene (Fig. 2). The clone λ368*iga*-8, which encodes IgA1 protease activity, lacked some of the HK368 genomic *Sau*3A fragments hybridizing to the 5′-*iga* probe. This indicates that plasmid pVD116, which was used as probe, contains sequences originating from the *H. influenzae* serotype d genome beyond the *iga* gene.

The 7.5-kb *Eco*RI-*Hin*dIII fragment of λ368*iga*-16, which contains the *iga* gene, was subcloned into M13mp19. In the supernatant of liquid cultures of *E. coli* JM109 transformed with this recombinant M13 phage, we detected IgA1 protease activity at a level comparable to the IgA1-cleaving activity found in the supernatant of *H. influenzae* HK368 cultures. This result indicates that the secretion mechanisms of the IgA1 protease function in *E. coli*, since the protease was present in the medium although the M13 phages did not lyse the bacterial cells.

Crude extract from a λ368*iga*-16 lysate was used to immunize rabbits. The immunoglobulin fraction of the resulting antisera had a complete inhibitory effect on the IgA1 protease activity from *H. influenzae* HK368 culture supernatant. This shows that *H. influenzae* HK368 does not secrete any IgA1 protease unrelated to the one encoded by the *iga* gene cloned in λ368*iga*-16.

Nucleotide sequence of the *iga* gene. The 7.5-kb *Eco*RI-*Hin*dIII fragment of λ368*iga*-16, which contains the *iga* gene, occasionally was unstable when subcloned into M13mp19 and propagated in *E. coli* JM109. Therefore, in the sequencing strategy, we used the internal *Pst*I site to divide this fragment (Fig. 2). The sticky ends of the 5.9-kb *Hin*dIII-*Pst*I fragment and the 1.6-kb *Pst*I-*Eco*RI fragment of λ368*iga*-16 (Fig. 2) were made blunt ended by the Klenow enzyme and subcloned in both orientations into the *Hin*cII site of M13mp19. Deletion derivatives of these subclones were generated by nuclease *Exo*III digestions for various times of *Sac*I-*Bam*HI-opened replicative-form DNA followed by nuclease S1 treatment and self-ligation. The 5,091-nucleotide (nt) sequence of the 7.5-kb *Eco*RI-*Hin*dIII fragment presented in Fig. 3 was obtained by sequencing a total of 79 such deletion clones, which gave overlapping sequences of both strands in this central part of the fragment. The sequence across the internal *Pst*I site, i.e., at nt 4856 to nt 4861 in Fig. 3, was verified by subcloning into M13mp19 and sequencing 300 nucleotides from the 3′ end of the 1,640-base-pair *Hha*I fragment ranging from nt 3349 to nt 4988.

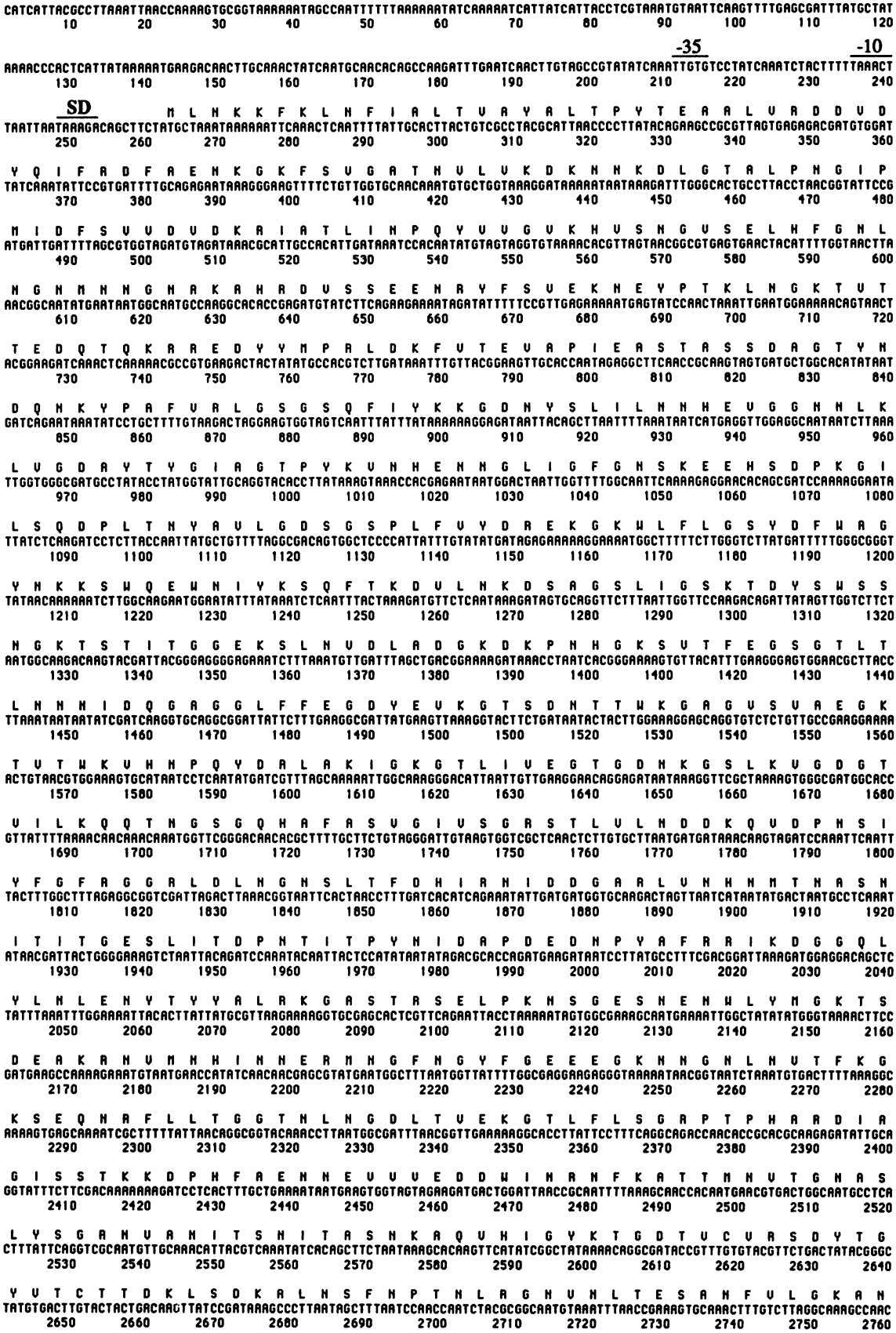

FIG. 3. Nucleotide sequence of the cloned *iga* gene from *H. influenzae* serotype b strain HK368. The sequence shown is derived from a part of the 7.5-kb *Eco*RI-*Hin*dIII fragment of the λ368*iga*-16 insert. The deduced amino acid sequence of the large open reading frame is shown above the nucleotide sequence. The two sequential stop codons are indicated by asterisks. Putative −35 and −10 promoter elements as well as a possible Shine-Dalgarno (SD) sequence are overlined. The possible rho-independent dual terminator is indicated by divergent arrows.

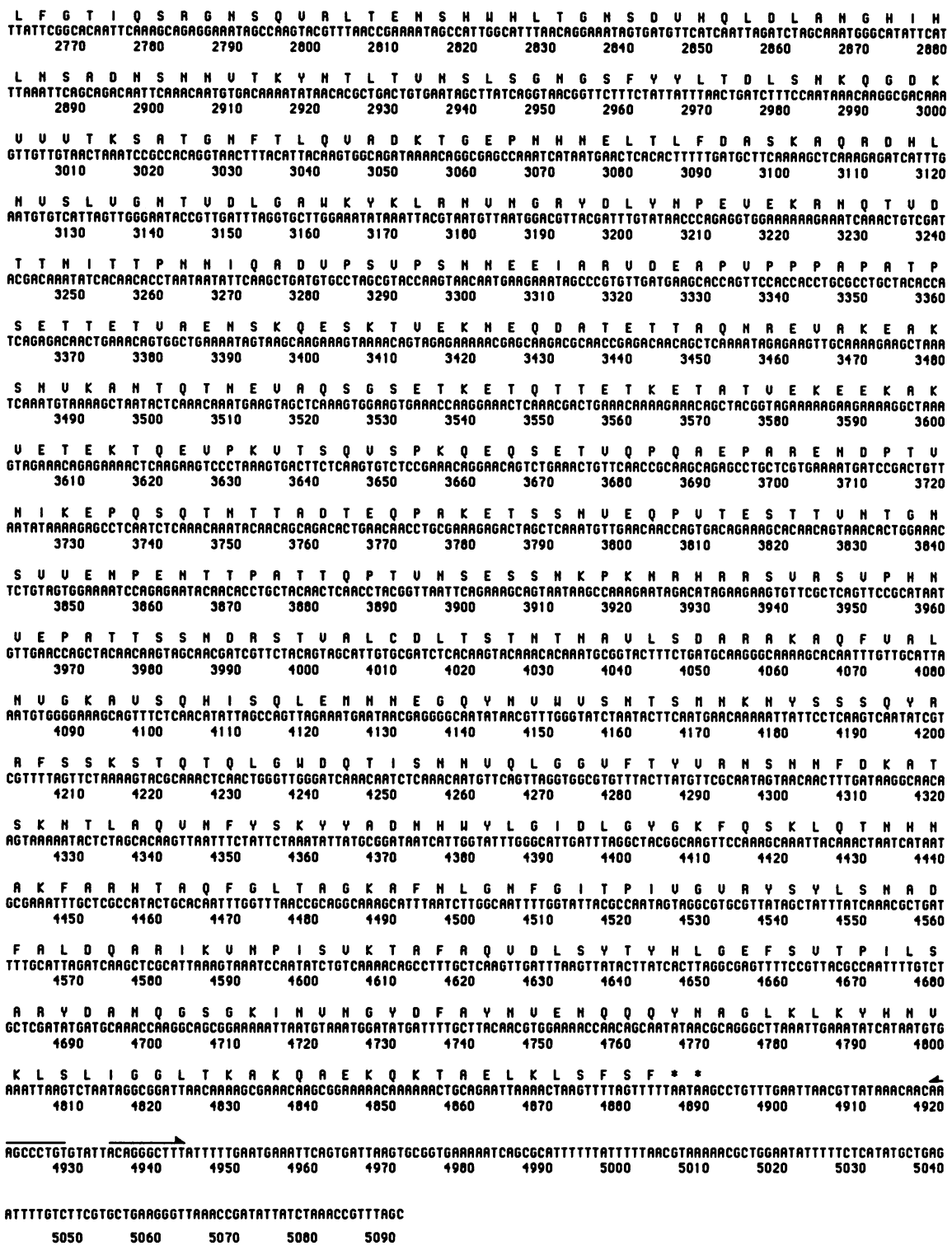

FIG 3—*Continued.*

The sequence revealed a large open reading frame with homology to the *N. gonorrhoeae iga* gene sequence previously published (31). The suggested *iga* gene from *H. influenzae* HK368 starting at the first ATG in the open reading frame as shown in Fig. 3 consists of 4,646 nt, including the TAA stop codon, encoding a deduced protein of 1,541 amino acids. This primary translation product has a deduced molecular mass of 169 kilodaltons (kDa), in contrast to the estimated size of the mature IgA1 protease of approximately 100 kDa (11). In the Discussion we propose a scheme for posttranslational modifications of the preprotein to account for this discrepancy.

TABLE 1. Codon usage of the *H. influenzae iga* gene

Amino acid	Codon	No. of times used[a]	Amino acid	Codon	No. of times used	Amino acid	Codon	No. of times used	Amino acid	Codon	No. of times used
F	TTT	45	S	TCT	30	Y	TAT	52	C	TGT	2
F	TTC	8	S	TCC	9	Y	TAC	8	C	TGC	1
L	TTA	53	S	TCA	26	*[b]	TAA	1	*	TGA	0
L	TTG	11	S	TCG	3	*	TAG	0	W	TGG	14
L	CTT	18	P	CCT	21	H	CAT	20	R	CGT	15
L	CTC	5	P	CCC	0	H	CAC	10	R	CGC	11
L	CTA	15	P	CCA	26	Q	CAA	57	R	CGA	4
L	CTG	3	P	CCG	6	Q	CAG	7	R	CGG	1
I	ATT	34	T	ACT	46	N	AAT	127	S	AGT	36
I	ATC	7	T	ACC	17	N	AAC	39	S	AGC	19
I	ATA	11	T	ACA	60	K	AAA	98	R	AGA	18
M	ATG	11	T	ACG	18	K	AAG	13	R	AGG	1
V	GTT	43	A	GCT	30	D	GAT	71	G	GGT	36
V	GTC	6	A	GCC	16	D	GAC	13	G	GGC	45
V	GTA	36	A	GCA	43	E	GAA	62	G	GGA	26
V	GTG	29	A	GCG	11	E	GAG	25	G	GGG	13

[a] Number of times the codon is used to code for the amino acid.
[b] Asterisks represent the stop codons.

DISCUSSION

In this study we cloned and sequenced the *iga* gene encoding the IgA1 protease from the *H. influenzae* serotype b strain HK368 to characterize the protein further and thereby to achieve a better understanding of its role during bacterial infection. The *iga* nucleotide sequence presented in Fig. 3 and the deduced primary structure of the IgA1 protease protein reveal several interesting features.

Potential start and stop signals for transcription and translation. The observed expression in *E. coli* implies that the *H. influenzae iga* gene must have transcription and translation signals that are recognized by the *E. coli* cell. We found a potential ribosome-binding Shine-Dalgarno sequence 5'-TAAAGA-3' (nt 248 to nt 253 in Fig. 3) eight nucleotides upstream from the putative translation start at the first ATG in the open reading frame (A at nt 262 in Fig. 3). This sequence has a homology of five identical positions out of six to the complementary sequence of the 3' end of the *E. coli* 16S rRNA (36). In addition, it is located in an acceptable position according to the rules for translational initiation suggested by Stormo et al. (37). The sequences 5'-TAAACT-3' (nt 235 to nt 240) and 5'-TTGTG-3' (nt 211 to nt 215) might constitute the transcription signals at positions −10 and −35, respectively. The proposed position −10 sequence is identical to the *E. coli* consensus sequence at the four best-conserved positions (13). Located 19 base pairs upstream from this, the −35 sequence is within the allowed spacing and has a homology of three of five positions identical to the *E. coli* consensus sequence (13). The identification of these three putative promoter elements strongly supports our suggestion of translational initiation at nt 262, although this can only be determined with certainty by direct analysis of the transcriptional start and the N terminus of the mature enzyme.

The open reading frame is ended by two sequential TAA stop codons at nt 4885 to nt 4890. A sequence similar to a typical rho-independent transcription terminator is found at nt 4919 to nt 4943 and contains a perfect inverted repeat with the potential of forming a hairpin structure, having a loop of 5 nt and a stem of 10 nt including five pairs of C·G (Fig. 3). Like the typical *E. coli* terminators, this structure is followed by a stretch rich in T residues (34).

Codon usage. Table 1 shows the codon usage of the *H. influenzae iga* gene. There is a striking tendency for the triplets to end preferentially with A or T. This may be a simple reflection of the high A+T content of the genome (62 mol% A+T) (15). The codon frequencies in the *iga* gene are similar to those observed in four other *H. influenzae* genes for which the nucleotide sequence is known (6, 20, 28).

We found 63% A+T for the 4,626-nt *H. influenzae iga* gene sequenced, which is close to the published estimate of 62% A+T for the whole genome (15). According to the observations of Bibb et al. (2), genes of this genome should have an A+T content of 51, 65, and 69%, respectively, for the first, second, and third position within codons. We found 53, 61, and 75% A+T, respectively, at these positions.

Based on the preference of T over C in the third base of duet codons and the relatively low quartet-to-duet frequency ratios among sextet codons encoding Arg, Leu, and Ser, the *iga* gene should be weakly expressed according to the rules of Grantham et al. (9). This agrees well with the relatively low amount of IgA1 protease protein produced by *H. influenzae*.

This bias toward A and T residues of the third position of degenerate codons in the *H. influenzae iga* gene is in contrast to the markedly unbiased codon usage for the *iga* gene from *N. gonorrhoeae*, which has a genomic A+T content of 51% (4). These two genes most likely have a common ancestor, indicating that apparently different constraints in codon usage have worked during the evolution of these two organisms.

Comparison of *H. influenzae* and *N. gonorrhoeae* protease sequences. Koomey and Falkow (18) have previously shown that the cloned type 2 *N. gonorrhoeae iga* gene hybridizes to the *H. influenzae* type 1 *iga* gene. The present comparison of the nucleotide sequences of these two *iga* genes revealed regions with a high degree of homology as well as areas with essentially no homology. The maximum homology between the corresponding deduced amino acid sequences of the two

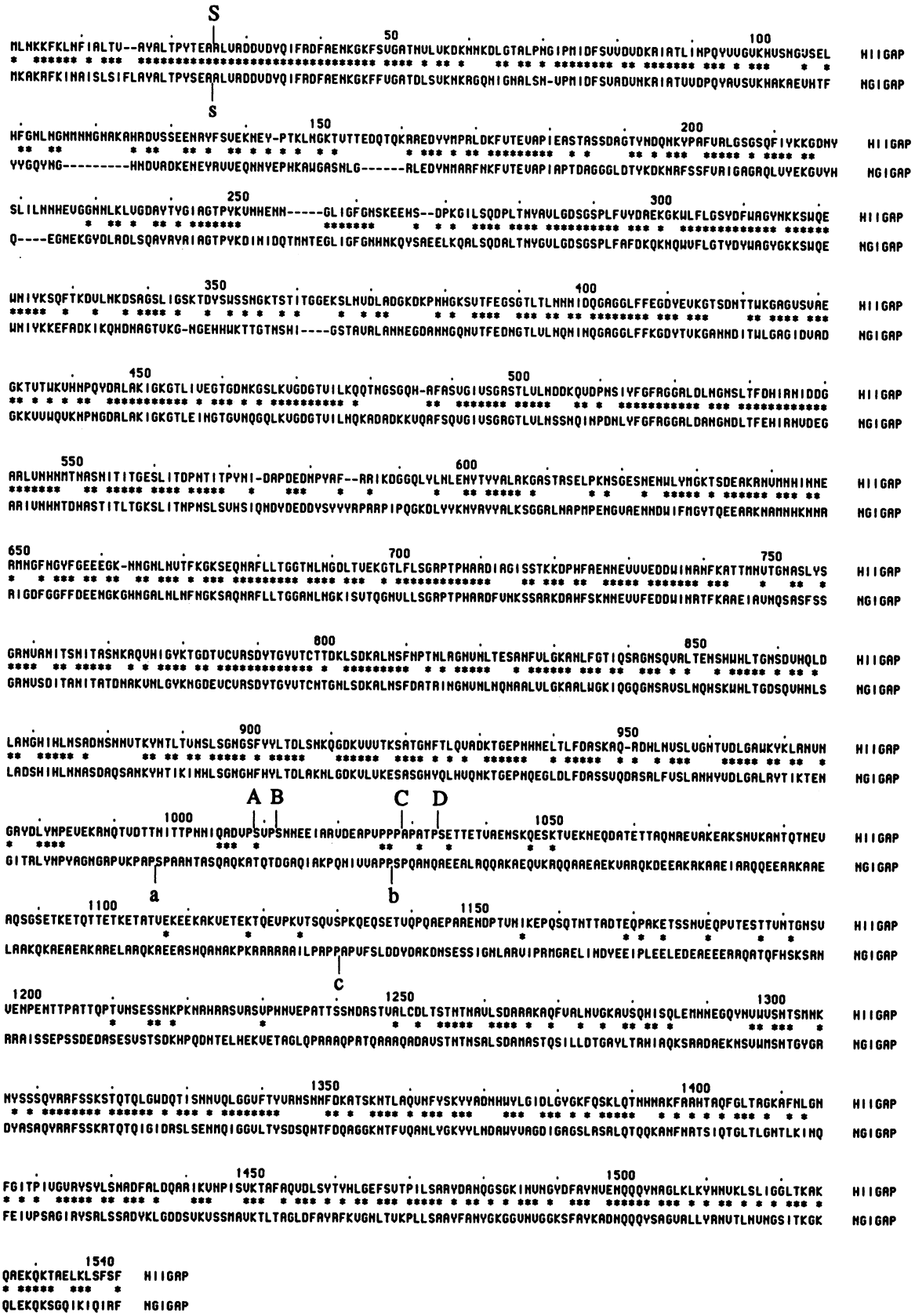

FIG. 4. Amino acid homologies between the deduced *H. influenzae iga* gene product (HIIGAP) and the *N. gonorrhoeae iga* gene product (NGIGAP). Sequence data of *N. gonorrhoeae iga* was derived from Pohlner et al. (31). Asterisks denote identical or functionally equivalent residues. Gaps indicated by dashes are introduced in the two sequences to obtain maximal homology. s, Cleavage site for the amino-terminal signal peptide of the *N. gonorrhoeae* protease; S, proposed equivalent cleavage site of the *H. influenzae* protease precursor. Positions a, b, and c represent autoproteolytic sites for the *N. gonorrhoeae* protease; positions A, B, C, and D indicate similar autoproteolytic sites suggested for the *H. influenzae* protease.

IgA1 protease proteins was obtained by aligning them as shown in Fig. 4.

Pohlner et al. (31) found that the precursor of the *N. gonorrhoeae* IgA1 protease contains three functional domains; the amino-terminal signal peptide, the central IgA1 protease, and the carboxy-terminal helper domain. The leader peptide is released during translocation of the preprotein into the periplasmic space, whereas the helper domain is assumed to create a pore in the outer membrane for excretion of the protease domain and remains associated with the membrane upon autoproteolytic cleavage.

The primary translation product, deduced from the *H. influenzae iga* sequence presented in Fig. 4, has a molecular mass of 169 kDa. The mature type 1 IgA1 protease from *H. influenzae* has been estimated to be about 100 kDa (11). Grundy et al. (10) found that a 2.2- to 3.1-kb region of the 3′ end of a type 1 IgA1 protease gene from an *H. influenzae* serotype d strain is necessary for the secretion of the protease but not for its activity. They suggest that this region is cleaved off during maturation of the protease. Based on these observations and the amino acid sequence homologies shown in Fig. 4, we suggest that the *H. influenzae* IgA1 protease is secreted by a mechanism similar to the one proposed by Polner et al. (31) for the *N. gonorrhoeae* IgA1 protease.

The amino-terminal part of the *H. influenzae* IgA1 protease contains positively charged lysines at amino acid positions 4, 5, and 7, followed by a hydrophobic stretch including the helix-breaking proline at position 21, four residues before an alanine (Fig. 4). These features are characteristic of signal sequences that are cleaved off and released with alanine as the C-terminal amino acid (38). The two protease sequences aligned as shown in Fig. 4 are identical around the position of the cleavage site of the *N. gonorrhoeae* protease leader peptide determined by Pohlner et al. (31). Therefore, we propose that the 25 amino-terminal amino acids of the *H. influenzae* IgA1 protease preprotein constitute the signal peptide.

Sequences rich in prolines, much like the type 1 *H. influenzae* IgA1 protease target site in the hinge region of the human IgA1 molecule (Fig. 1), are exclusively found at three positions in the *H. influenzae* protease sequence (A, B, and D in Fig. 4). A single sequence with homology to the type 2 *H. influenzae* IgA1 protease target site (Fig. 1) is found in the same region of the protease (C in Fig. 4). We propose that one or more of these four sequence elements are functionally equivalent to the autoproteolytic sites a, b, and c in the *N. gonorrhoeae* protease (Fig. 4).

When an N-terminal signal sequence of 25 amino acids is taken into account, autoproteolytic cleavage at position A, B, C, or D would result in excreted proteases of deduced molecular masses of 107.8, 108.1, 109.9, or 110.3 kDa, respectively. Each of these values is in acceptable agreement with the estimated molecular mass of 100 kDa.

Strikingly, a stretch of 32 amino acids (amino acids 16 to 47) is identical in the two proteases except for a single conservative substitution. This shows that the N-terminal part of the mature IgA1 protease has been evolutionarily conserved, suggesting that it is essential to the translocation, enzymatic function, or specificity of the protease molecule.

Another notably well-conserved sequence is found at amino acids 785 to 797. It contains the two conserved cysteines that Pohlner et al. (31) proposed to be part of the active site in the *N. gonorrhoeae* protease. A third cysteine is found in the helper domain of the *H. influenzae* protease preprotein at amino acid position 1250.

In striking contrast to the rest of the *H. influenzae* IgA1 protease, the region from amino acids 980 to 1240 has no significant sequence homology to the *N. gonorrhoeae* protease (Fig. 4). This stretch is proposed to constitute the N-terminal part of the helper domain. It is very hydrophilic in both proteins, suggesting that it has a common function in the secretion of the two proteases. No deletions or insertions had to be introduced in this divergent area of the two protease sequences to obtain the flanking homologies (Fig. 4). This observation suggests that this region functions as a spacer to properly orient flanking domains.

The cloning and sequencing of the IgA1 protease gene from *H. influenzae* serotype b provide a means of producing this protein in large quantities for further structural and functional analyses.

ACKNOWLEDGMENTS

We thank J. Michael Koomey for providing the *E. coli* plasmid pVD116, Margaret Clark for critical review of the manuscript, and Dorthe Eggertsen for typing the manuscript.

This study was supported by the Danish Medical Research Council (grants 12-6924 and 12-7639) and by the Direktør Ib Henriksens Fond.

LITERATURE CITED

1. **Benton, W. D., and R. W. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180–182.
2. **Bibb, M. J., P. R. Findlay, and M. W. Johnson.** 1984. The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. Gene **30**:157–166.
3. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
4. **Bøvre, K.** 1984. Family VIII. Neisseriaceae Prevot 1933, 119, p. 288–290. *In* N. R. Krieg and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 1. The Williams & Wilkins Co., Baltimore.
5. **Bricker, J., M. H. Mulks, A. G. Plaut, E. R. Moxon, and A. Wright.** 1983. IgA proteases of *Haemophilus influenzae*: cloning and characterization in *Escherichia coli* K-12. Proc. Natl. Acad. Sci. USA **80**:2681–2685.
6. **Chandrasegaran, S., K. D. Lunnen, H. O. Smith, and G. G. Wilson.** 1988. Cloning and sequencing the *Hin* fl restriction and modification genes. Gene **70**:387–392.
7. **Fishman, Y., J. Bricker, J. V. Gilbert, A. G. Plaut, and A. Wright.** 1985. Cloning of the type 1 immunoglobulin A1 protease from *Neisseria gonorrhoeae* and secretion of the enzyme from *Escherichia coli*, p. 164–168. *In* G. K. Schoolnik (ed.), The pathogenic neisseriae. American Society for Microbiology, Washington, D.C.
8. **Gilbert, J. V., A. G. Plaut, Y. Fishman, and A. Wright.** 1988. Cloning of the gene encoding streptococcal immunglobulin A protease and its expression in *Escherichia coli*. Infect. Immun. **56**:1961–1966.
9. **Grantham, R., C. Gautier, M. Gouy, M. Jacobzone, and R. Mercier.** 1981. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res. **9**:r43–r74.
10. **Grundy, F. J., A. Plaut, and A. Wright.** 1987. *Haemophilus influenzae* immunoglobulin A1 protease genes: cloning by plasmid integration-excision, comparative analysis, and localization of secretion determinants. J. Bacteriol. **169**:4442–4450.
11. **Grundy, F. J., A. G. Plaut, and A. Wright.** 1987. *Haemophilus influenzae* IgA1 protease: determinants of cleavage site specificity and secretion. Adv. Exp. Med. Biol. **261B**:1251–1260.
12. **Halter, R., J. Pohlner, and T. F. Meyer.** 1984. IgA protease of *Neisseria gonorrhoeae*: isolation and characterization of the gene and its extracellular product. EMBO J. **3**:1595–1601.
13. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and

analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2255.
14. **Kett, K., P. Brandtzaeg, J. Radl, and J. J. Haaijman.** 1986. Different subclass distribution of IgA-producing cells in human lymphoid organs and various secretory tissues. J. Immunol. **136**:3631–3635.
15. **Kilian, M.** 1976. A taxonomic study of the genus *Haemophilus*, with the proposal of a new species. J. Gen. Microbiol. **93**:9–62.
16. **Kilian, M., and J. Reinholdt.** 1986. Interference with IgA defence mechanisms by extracellular bacterial enzymes, p. 173–208. *In* C. S. F. Easmon and J. Jeljaszewics (ed.), Medical microbiology, vol. 5. Academic Press, Inc., London.
17. **Kilian, M., B. Thomsen, T. E. Petersen, and H. Bleeg.** 1983. Molecular biology of *Haemophilus influenzae* IgA1 proteases. Mol. Immunol. **20**:1051–1058.
18. **Koomey, J. M., and S. Falkow.** 1984. Nucleotide sequence homology between the immunoglobulin A1 protease genes of *Neisseria gonorrhoeae*, *Neisseria meningitidis*, and *Haemophilus influenzae*. Infect. Immun. **43**:101–107.
19. **Koomey, J. M., R. E. Gill, and S. Falkow.** 1982. Genetic and biochemical analysis of gonococcal IgA protease: cloning in *Escherichia coli* and construction of mutants of gonococci that fail to produce the activity. Proc. Natl. Acad. Sci. USA **79**: 7881–7885.
20. **Kroll, J. S., I. Hopkins, and E. R. Moxon.** 1988. Capsule loss in *H. influenzae* type b occurs by recombination-mediated disruption of a gene essential for polysaccharide export. Cell **53**: 347–356.
21. **Larson, R., and J. Messing.** 1982. Apple II software for M13 shotgun DNA sequencing. Nucleic Acids Res. **10**:39–49.
22. **Lee, J. J., and H. O. Smith.** 1988. Sizing of the *Haemophilus influenzae* Rd genome by pulsed-field agarose gel electrophoresis. J. Bacteriol. **170**:4402–4405.
23. **Leffers, H., J. Kjems, L. Østergaard, N. Larsen, and R. A. Garrett.** 1987. Evolutionary relationships amongst Archaebacteria. A comparative study of 23S ribosomal RNA's of a sulphur-dependent extreme thermophile, an extreme halophile and a thermophilic methanogen. J. Mol. Biol. **195**:43–61.
24. **Loenen, W. A. M., and W. J. Brammar.** 1980. A bacteriophage lambda vector for cloning large DNA fragments made with several restriction enzymes. Gene **10**:249–259.
25. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Mikkelsen, B. M., M. E. Clark, G. Christiansen, O. M. Klintebaek, J. T. Nielsen, K. K. Thomsen, and J. P. Hjorth.** 1985. The structure of two distinct pancreatic amylase genes in mouse strain YBR. Biochem. Genet. **23**:511–524.
27. **Mulks, M. H.** 1985. Microbial IgA proteases, p. 81–104. *In* I. A. Holder (ed.), Bacterial enzymes and virulence. CRC Press, Boca Raton, Fla.
28. **Munson, R., Jr., and S. Grass.** 1988. Purification, cloning, and sequence of outer membrane protein P1 of *Haemophilus influenzae* type b. Infect. Immun. **56**:2235–2242.
29. **Norrander, J., T. Kempe, and J. Messing.** 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene **26**:101–106.
30. **Plaut, A. G.** 1983. The IgA1 proteases of pathogenic bacteria. Annu. Rev. Microbiol. **37**:603–622.
31. **Pohlner, J., R. Halter, K. Beyreuther, and T. F. Meyer.** 1987. Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. Nature (London) **325**:458–462.
32. **Poulsen, K., J. P. Hjorth, and M. Kilian.** 1988. Limited diversity of the Immunoglobulin A1 protease gene (*iga*) among *Haemophilus influenzae* serotype b strains. Infect. Immun. **56**:987–992.
33. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
34. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
35. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
36. **Shine, J., and L. Dalgarno.** 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
37. **Stormo, G. D., T. D. Schneider, and L. M. Gold.** 1982. Characterization of translational initiation sites in *E. coli*. Nucleic Acids Res. **10**:2971–2996.
38. **Watson, M. E. E.** 1984. Compilation of published signal sequences. Nucleic Acids Res. **12**:5145–5164.
39. **Wood, W. B.** 1966. Host specificity of DNA produced by *Escherichia coli*: bacterial mutations affected the restriction and modification of DNA. J. Mol. Biol. **16**:118–133.
40. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.