INFECTION AND IMMUNITY, Oct. 1996, p. 4000–4007
0019-9567/96/$04.00+0
Vol. 64, No. 10

The Hemagglutinin Gene A (*hagA*) of *Porphyromonas gingivalis* 381 Contains Four Large, Contiguous, Direct Repeats

NAIMING HAN,* JOAN WHITLOCK, AND ANN PROGULSKE-FOX

Department of Oral Biology, University of Florida, Gainesville, Florida 32610-0424

Received 6 December 1995/Returned for modification 17 January 1996/Accepted 9 July 1996

***Porphyromonas gingivalis* is a gram-negative anaerobic bacterial species strongly associated with adult periodontitis. One of its distinguishing characteristics and putative virulence properties is the ability to agglutinate erythrocytes. We have previously reported the cloning of multiple hemagglutinin genes from *P. gingivalis* 381. Subsequent sequencing of clone ST 2 revealed that the cloned fragment contained only an internal portion of the gene which lacked both start and stop codons. We here report the cloning and sequencing of the entire gene, designated *hagA*, as well as its relationship to other genes of this species. By use of inverse PCR technology and the construction of several additional genomic libraries, the complete open reading frame of *hagA* was found to be 7,887 bp in length, encoding a protein of 2,628 amino acids with a molecular mass of 283.3 kDa, which is among the largest genes ever cloned from a prokaryote to date. Within its open reading frame, four large, contiguous, direct repeats (varying from 1,318 to 1,368 bp) were identified. The repeat unit (*HArep*), which is assumed to contain the hemagglutinin domain, is also present in other recently reported protease and hemagglutinin genes in *P. gingivalis*. Thus, we propose that *hagA* and the other genes which share the *HArep* sequence form a multigene family with *hagA* as a central member.**

Porphyromonas gingivalis is a gram-negative anaerobic bacterial species which is isolated primarily from infectious periodontal pockets and considered the major pathogen for adult periodontitis as well as one of the major pathogens for several forms of rapidly progressive periodontitis and refractory periodontitis (36). The presence of a hemagglutinin on the *P. gingivalis* cell surface was first reported in 1974 by Okuda and Takazoe (26). Hemagglutination activity is one of the major phenotypic characteristics which differentiate this species from other oral and nonoral asaccharolytic black-pigmenting species (14).

The production of hemagglutinins is a well-established virulence factor for a number of bacterial species. These include some of the more virulent and troublesome microorganisms that afflict the human host, such as *Vibrio cholerae* (4), *Salmonella* spp. (3, 9), *Bordetella pertussis* (6), and *Escherichia coli* (10). It is thus reasonable to expect that the hemagglutinins of *P. gingivalis* are also involved in virulence.

Hemagglutinins are expressed on the bacterial cell surface in association either with filamentous structures such as fimbriae (fimbrial adhesins) or with nonfimbrial surface components (nonfimbrial adhesins) and are frequently the adhesins through which bacteria attach to mammalian cells. It is likely that the hemagglutinin(s) of *P. gingivalis* also functions as an adhesin in vivo, although this has not yet been proven. In addition, since this species requires heme for growth, we have previously suggested that hemagglutination and subsequent lysis of the bound erythrocytes might provide an efficient means for *P. gingivalis* to acquire hemin (20).

To facilitate the biochemical, genetic, and functional studies of *P. gingivalis* hemagglutinins and ultimately determine their roles in pathogenesis, we have cloned multiple hemagglutinin genes from *P. gingivalis* 381 (20, 30, 31). Among them is clone ST 2, now designated *hagA*, which is functionally expressed in *E. coli* since its expression resulted in strong hemagglutination activity of the transformants. Sequencing this fragment revealed the presence of a 1.0-kb repeat but also that this fragment (3,164 bp in length) contained only an internal portion of the gene which lacked both 5′ and 3′ termini. The purpose of this study was to obtain the complete gene and its sequence to ultimately produce isogenic mutations in the gene which would allow the determination of its importance in virulence. In addition, the cloning of the gene will allow comparisons with other known genes and facilitate the characterization of the gene product, such as identification of the active site. We report here the cloning, sequencing, and characterization of the complete *hagA* gene. Analysis of the sequence revealed several striking features, including the presence of four large, contiguous, direct repeats and an open reading frame (ORF) which encodes a 2,628-amino-acid (aa) protein of 283.3 kDa, among the largest, if not the largest, ORFs found to date in prokaryotes.

MATERIALS AND METHODS

Bacterial strains, plasmids, media, and growth conditions. *P. gingivalis* 381 was grown on Trypticase soy agar (BBL Microbiology Systems, Cockeysville, Md.) supplemented with sheep blood (5%), hemin (5 μg/ml), and menadione (5 μg/ml) in an anaerobic chamber with an atmosphere of 5% CO_2, 10% H_2, and 85% N_2. When broth-grown cells were required, cells were grown in Todd-Hewitt broth (BBL Microbiology Systems) supplemented with hemin (5 μg/ml), menadione (5 μg/ml), and glucose (2 mg/ml).

E. coli JM109 [*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1* Δ(*lac-proAB*) (F′ *traD36 proAB lacI*^{q}ZΔM15)] was used for all cloning studies and cultured aerobically on Luria-Bertani medium consisting of Bacto Tryptone (10 g/liter), Bacto yeast extract (5 g/liter), and NaCl (10 g/liter). *E. coli* JM109 transformants were maintained on Luria-Bertani medium supplemented with ampicillin (50 μg/ml).

Bacterial alkaline phosphatase-treated pUC18 (Pharmacia, Piscataway, N.J.) was used for construction of *Bam*HI- and *Sma*I-generated genomic banks. The plasmids pBluescript II SK(+) (Stratagene Cloning Systems, La Jolla, Calif.) and pUC19 (Bethesda Research Laboratories, Gaithersburg, Md.) were used as vectors for subcloning.

Preparation of DNA. Chromosomal DNA was isolated from *P. gingivalis* 381 with hexadecyltrimethyl ammonium bromide (CTAB)-CsCl ultracentrifugation (37).

Plasmid DNA was routinely isolated by the alkaline lysis method (35) and with the Wizard mini-prep kit (Promega Co., Madison, Wis.). DNA samples for

* Corresponding author. Mailing address: Department of Oral Biology, Box 100424, University of Florida, Gainesville, FL 32610-0424. Phone: (352) 846-0766. Fax: (352) 392-2361. Electronic mail address: nhan@dental.ufl.edu.

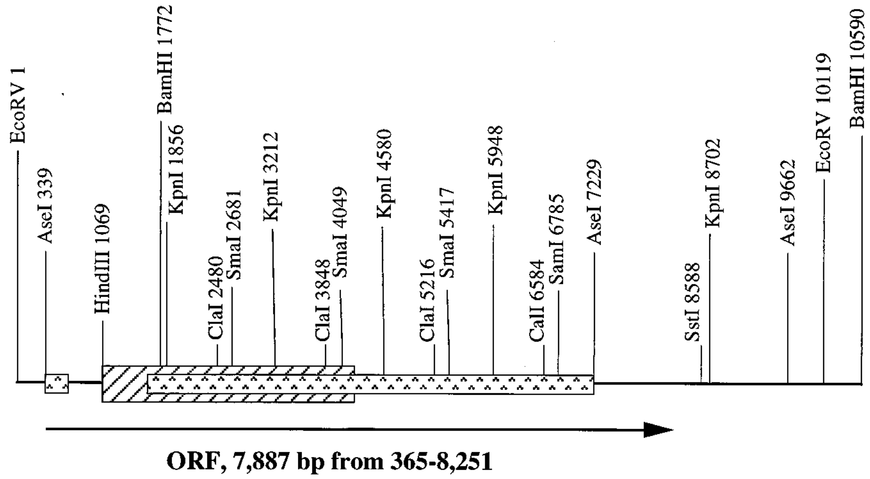

FIG. 1. Restriction enzyme map of cloned fragments of *P. gingivalis* 381. The hatched area designates the originally cloned ST 2 fragment; the stippled area designates the amplified IPCR fragment. Drawn to scale.

sequencing were prepared by alkaline lysis-CsCl ultracentrifugation (35) and alkaline lysis-polyethylene glycol 8000 precipitation (2a) and with the Wizard midi-prep kit (Promega).

Southern blot analysis. For Southern blot analysis, chromosomal DNA samples were digested with various restriction enzymes, including *Acc*I, *Ase*I (Biolabs), *Ava*II, *Bcl*I, *Bgl*II, *Bst*XI, *Dra*I (Bethseda Research Laboratories), *Dra*III (Stratagene), *Eco*RV, *Nru*I (Stratagene), *Pst*I, *Pvu*II, *Sac*I, *Sal*I, *Sph*I, *Ssp*I, *Stu*I, and *Xho*I. Unless specified otherwise, all enzymes were purchased from Promega. The resulting digested fragments were transferred to positively charged nylon membranes (Boehringer Mannheim Co., Indianapolis, Ind.) by the capillary transfer method (35). A region of the first 394 bp of ST 2, which is distant from the repeat sequence region, was labeled with a nonradioactive digoxigenin DNA labeling and detection kit (Boehringer Mannheim) and used as a probe to detect the bound DNA fragments on the nylon membrane. Hybridizing fragments were visualized on X-ray film with the LumiPhos 530 system (Boehringer Mannheim).

IPCR. For inverse PCR (IPCR), a negative primer at 405 nucleotides upstream of the 5′ end of the ST 2 fragment (GGC AAA CCA AAA AGA TTC) and a positive primer at 529 nucleotides upstream of the ST 2 fragment (TTC TTC CAA CGA CTA CAC) were selected and synthesized at the University of Florida DNA Synthesis Core Facility.

The detailed procedure for the IPCR was as described by Han and Progulske-Fox (13). Briefly, the *Ase*I-digested fragments were purified with phenol-chloroform and self-ligated at a DNA concentration of 4 ng/μl in the presence of 1 U of T4 DNA ligase (Promega) per 50-μl reaction mixture. IPCRs were performed as follows. First, 10 to 20 ng of the self-ligated DNA sample was heated for 30 min at 94°C in PCR buffer with deoxynucleoside triphosphates (dNTPs) and primers in a PTC-100 programmable thermal controller (MJ Research, Inc., Watertown, Mass.). *Taq* polymerase (Promega) was then added, and the reaction was carried out with 35 cycles of denaturation at 94°C for 1 min, primer annealing at 52°C for 1 min, and extension at 72°C for 6 min. All reactions were performed in a 100-μl volume including 50 mM KCl, 1.5 mM $MgCl_2$, 0.25 μM each primer, 200 μM each dNTP, and 5 to 10 U of *Taq* polymerase. The amplified mixture was extracted with phenol-chloroform, electrophoresed through a 1% low-melting-point agarose gel, treated with agarase (Boehringer Mannheim), and used for direct sequencing. After analysis of the sequence, the IPCR fragments were digested with *Eco*RI and *Bam*HI and ligated into pBluescript II SK(+), and the recombinant plasmids were transformed into *E. coli* JM109 competent cells.

Construction of genomic banks. For construction of additional genomic banks, 10 μg of the *P. gingivalis* chromosomal DNA was digested with either *Bam*HI or *Eco*RV and electrophoresed in a 1% low-melting-point agarose gel. The fragments were cut out from gels and treated with agarase (Boehringer Mannheim). The isolated fragments were then cloned into the bacterial alkaline phosphatase-treated *Bam*HI or *Sma*I sites of pUC18, respectively. The recombined plasmids were electroporated into *E. coli* JM109 cells prepared by standard procedures. The resulting 2,000 to 3,000 transformants of each bank were screened by in situ hybridization, and the positive colonies were confirmed by restriction enzyme digestion of plasmid DNA and DNA sequencing.

Subcloning and sequencing. Subclones for sequencing were constructed with either pBluescript II SK(+) or pUC19, and oligonucleotide primers were synthesized to complete the sequencing of both strands.

Sequencing was carried out at the University of Florida DNA Sequencing Core Facility with the *Taq* Dye Primer and *Taq* Dyedeoxy Terminator Cycle Sequencing Protocol developed by Applied Biosystems, Inc. (Foster City, Calif.) with a fluorescent labeled primer(s) and labeled dideoxynucleotides. The labeled extension products were analyzed on an ABI 373 DNA sequencer (Applied Biosystems).

DNA sequence analysis. Sequence data were analyzed by the Sequence Analysis Software Package of the University of Wisconsin.

Hemagglutination assay. Hemagglutination activity was assayed with sheep erythrocytes by use of a microtiter plate method as described previously by Progulske-Fox et al. (30). In each of the experiments, *P. gingivalis* 381 and *E. coli* containing vector alone were included as positive and negative controls, respectively.

Nucleotide sequence accession number. The accession number for the *hagA* nucleotide sequence in GenBank is U41807.

RESULTS

Obtaining the entire *hagA* sequence. Southern blot analysis results (data not shown) indicated that *Ase*I restriction of genomic DNA produced a single 6.9-kb fragment which hybridized to the probe used. Thus, IPCR was used to generate this fragment. Under the conditions used, a 5,963-bp fragment was successfully amplified via IPCR. Surprisingly, restriction of the cloned IPCR fragment with *Acc*65I and analysis by agarose gel electrophoresis demonstrated that the 5,544-bp insert was composed exclusively of several 1.35-kb fragments which were visible as one very dense band on the gel (data not shown) and which suggested the presence of four repeats within the amplified fragment, not two, as previously predicted from the sequence of the ST 2 fragment. The sequence data of the IPCR fragment revealed that this amplified fragment contained an additional 2,997 bp of sequence 3′ to the ST 2 fragment, that the size of the repeats was 1.3 kb, instead of 1.0 kb, and that the repeats are continuous, not interrupted as previously determined from the ST 2 sequence (Fig. 1 to 3, beginning at base 1069 and ending at base 4232). In addition, the start codon was found to be located 720 bp upstream of the 5′ end of the ST 2 fragment. However, this 6,896-bp *Ase*I fragment did not contain a stop codon. To obtain the 3′ end of this gene, a *Bam*HI gene bank was constructed from which a 8,818-bp cloned fragment containing an additional 3,362 bp of previously unknown downstream DNA was obtained. Sequencing this downstream region revealed that the stop codon was located 1,017 bp downstream of the 3′ end of the 6.9-kb *Ase*I fragment (Fig. 2). Since this cloned fragment almost completely overlaps with the amplified IPCR fragment (Fig. 1), any errors produced by PCR were corrected by sequencing this genomic fragment.

```
        EcoRV
    1   GATATCCGGCTCTTCGGCAGAGAATGCGAGAGATTCAGGATATATCGCAACGGCCTTGTCAAGATCGAGGCCTCTTTAGGTCATGGATATAACGTGAGTT
                                                -35                    -10
  101   CGATGTAAGCTTTTCGGCCTTTCCATCATACAATCGATTCGATTCTCTTTGGACTCAATAAAAAATATAAAATACTCAAAGAGTTGGCATATAACTTTGC
  201   CTCAGTGGCGAGTGGGTTTTTCGGCCAATTCCTAAAGAAGAAAATAGCTGTTTGTATCTTTTTGCGAAAAAAGTTTGGCGGATTAAGATTAAAAACATAT
                                            AseI
  301   CTTTCGGGCGATAGTGGTAGAGCACTATCTTGCGAAACATTAATCTTTAATACTTTCAAAAGGTATGAGAAAATTGAATTCTTTATTTTCGCTCGCCGTC
    1                                                             M  R  K  L  N  S  L  F  S  L  A  V
  401   CTATTATCCCTATTGTGTTGGGGACAGACGGCTGCCGCACAGGGAGGGCCGAAGACTGCTCCTTCTGTGACGCACCAAGCGGTGCAGAAAGGTATTCGAA
   13   L  L  S  L  L  C  W  G  Q  T  A  A  A  Q  G  G  P  K  T  A  P  S  V  T  H  Q  A  V  Q  K  G  I  R  T
  501   CATCCAAGGTTAAGGATCTCCGAGATCCGATTCCTGCCGGTATGGCACGAATTATCTTGGAGGCTCACGATGTATGGGAAGACGGCACAGGCTATCAAAT
   47    S  K  V  K  D  L  R  D  P  I  P  A  G  M  A  R  I  I  L  E  A  H  D  V  W  E  D  G  T  G  Y  Q  M
                                                      primer (-)
  601   GCTTTGGGATGCAGATCACAATCAGTACGGCGCATCCATTCCCGAAGAATCTTTTTGGTTTGCCAACGGAACGATCCCGGCCGGTCTTTACGATCCTTTC
   80    L  W  D  A  D  H  N  Q  Y  G  A  S  I  P  E  E  S  F  W  F  A  N  G  T  I  P  A  G  L  Y  D  P  F
  701   GAGTATAAAGTTCCGGTCAATGCCGATGCATCTTTTTCTCCCACGAATTTCGTGCTTGATGGAACAGCATCAGCCGATATTCCTGCCGGCACTTATGACT
  113   E  Y  K  V  P  V  N  A  D  A  S  F  S  P  T  N  F  V  L  D  G  T  A  S  A  D  I  P  A  G  T  Y  D  Y
  801   ATGTAATCATTAACCCCAATCCTGGCATAATATATATAGTAGGAGAGGGTGTCTCCAAAGGTAACGATTATGTGGTAGAGGCCGGTAAGACTTATCATTT
  147    V  I  I  N  P  N  P  G  I  I  Y  I  V  G  E  G  V  S  K  G  N  D  Y  V  V  E  A  G  K  T  Y  H  F
  901   CACTGTCCAACGACAAGGCCCCGGCGATGCTGCGTCCGTTGTAGTGACCGGAGAAGGTGGCAATGAATTCGCTCCCGTACAGAATCTCCAATGGTCTGTA
  180    T  V  Q  R  Q  G  P  G  D  A  A  S  V  V  V  T  G  E  G  G  N  E  F  A  P  V  Q  N  L  Q  W  S  V
                                                                       HindIII
 1001   TCTGGGCAGACAGTGACCCTCACTTGGCAAGCCCCCGCATCCGACAAACGGACTTATGTGTTGAACGAAAGCTTCGATACGCAAACGCTTCCTAACGGCT
  213   S  G  Q  T  V  T  L  T  W  Q  A  P  A  S  D  K  R  T  Y  V  L  N  E  S  F  D  T  Q  T  L  P  N  G  W
 1101   GGACAATGATCGATGCTGATGGTGATGGTCACAATTGGCTATCTACAATAAACGTTTACAACACTGCTACTCATACAGGTGACGGTGCTATGTTTAGCAA
  247    T  M  I  D  A  D  G  D  G  H  N  W  L  S  T  I  N  V  Y  N  T  A  T  H  T  G  D  G  A  M  F  S  K
 1201   ATCATGGACTGCTAGCGGTGGTGCAAAAATTGATTTGAGTCCTGACAACTATTTGGTAACTCCAAAGGTTACGGTTCCTGAGAATGGTAAACTTTCTTAT
  280    S  W  T  A  S  G  G  A  K  I  D  L  S  P  D  N  Y  L  V  T  P  K  V  T  V  P  E  N  G  K  L  S  Y
 1301   TGGGTTTCATCTCAAGTGCCTTGGACTAATGAGCATTATGGAGTGTTCTTGTCCACAACCGGAAACGAGGCTGCAAACTTTACGATAAAGCTACTGGAAG
  313   W  V  S  S  Q  V  P  W  T  N  E  H  Y  G  V  F  L  S  T  T  G  N  E  A  A  N  F  T  I  K  L  L  E  E
 1401   AAACCCTCGGATCCGACAAACCTGCTCCGATGAACTTGGTGAAGAGTGAAGGAGTAAAGCTTCCTGCACCTTATCAGGAAAGAACCATCGATCTCTCTGC
  347    T  L  G  S  D  K  P  A  P  M  N  L  V  K  S  E  G  V  K  L  P  A  P  Y  Q  E  R  T  I  D  L  S  A
 1501   CTATGCCGGACAACAGGTGTACTTGGCATTCCGTCATTTCAACTCTACAGGTATATTCCGTCTTTATCTTGATGATGTGGCTGTTTCTGGTGAAGGTTCT
  380    Y  A  G  Q  Q  V  Y  L  A  F  R  H  F  N  S  T  G  I  F  R  L  Y  L  D  D  V  A  V  S  G  E  G  S
         primer (+)
 1601   TCCAACGACTACACGTACACGGTATATCGTGACAATGTTGTTATTGCCCAGAATCTCGCGGCAACGACATTCAATCAGGAAAATGTAGCTCCCGGCCAGT
  413   S  N  D  Y  T  Y  T  V  Y  R  D  N  V  V  I  A  Q  N  L  A  A  T  T  F  N  Q  E  N  V  A  P  G  Q  Y
                                                                          BamHI
 1701   ATAACTACTGTGTTGAAGTTAAGTACACAGCCGGCGTATCTCCGAAGGTATGTAAAGACGTTACGGTAGAAGGATCCAACGAATTTGCTCATGTACAGAA
  447    N  Y  C  V  E  V  K  Y  T  A  G  V  S  P  K  V  C  K  D  V  T  V  E  G  S  N  E  F  A  H  V  Q  N
                                                            KpnI
 1801   CCTGACCGGTAGTGCAGTAGGTCAGAAAGTAACGCTTAAGTGGGATGCACCTAATGGTACCCCGAATCCGAATCCCGGAACAACAACACTTTCCGAATCA
  480    L  T  G  S  A  V  G  Q  K  V  T  L  K  W  D  A  P  N  G  T  P  N  P  N  P  G  T  T  T  L  S  E  S
 1901   TTCGAAAATGGTATTCCTGCCTCATGGAAGACGATCGATGCAGACGGTGACGGCAACAATTGGACGACGACCCCTCCTCCCGGAGGCACCTCTTTTGCAG
  513   F  E  N  G  I  P  A  S  W  K  T  I  D  A  D  G  D  G  N  N  W  T  T  T  P  P  P  G  G  T  S  F  A  G
 2001   GTCACAACAGTGCAATCTGTGCCTCTTCGGCTTCTTATATCAACTTTGAAGGTCCTCAGAACCCTGATAACTATCTGGTTACACCGGAGCTATCTCTTCC
  547    H  N  S  A  I  C  A  S  S  A  S  Y  I  N  F  E  G  P  Q  N  P  D  N  Y  L  V  T  P  E  L  S  L  P
 2101   TAACGGAGGAACGCTTACTTTCTGGGTATGTGCACAAGATGCCAATTATGCATCAGAGCACTATGCCGTGTACGCATCTTCTACGGGTAACGACGCTTCC
  580    N  G  G  T  L  T  F  W  V  C  A  Q  D  A  N  Y  A  S  E  H  Y  A  V  Y  A  S  S  T  G  N  D  A  S
 2201   AACTTCGCCAACGCTTTGTTGGAAGAAGTGCTGACGGCCAAGACAGTTGTTACGGCACCTGAAGCCATTCGTGGCACTCGTGTTCAGGGCACCTGGTATC
  613   N  F  A  N  A  L  L  E  E  V  L  T  A  K  T  V  V  T  A  P  E  A  I  R  G  T  R  V  Q  G  T  W  Y  Q
 2301   AAAAGACGGTACAGTTGCCTGCGGGTACTAAGTATGTTGCTTTCCGTCACTTCGGCTGTACGGACTTCTTCTGGATTAACCTTGATGATGTTGAGATCAA
  647    K  T  V  Q  L  P  A  G  T  K  Y  V  A  F  R  H  F  G  C  T  D  F  F  W  I  N  L  D  D  V  E  I  K
 2401   GGCCAACGGCAAGCGCGCAGACTTCACGGAAACGTTCGAGTCTTCTACTCATGGAGAGGCACCGGCGGAATGGACTACTATCGATGCCGATGGCGATGGT
  680    A  N  G  K  R  A  D  F  T  E  T  F  E  S  S  T  H  G  E  A  P  A  E  W  T  T  I  D  A  D  G  D  G
 2501   CAGGGTTGGCTCTGTCTGTCTTCCGGACAATTGGACTGGCTGACAGCTCATGGCGGCACCAACGTAGTAGCCTCTTTCTCATGGAATGGAATGGCTTTGA
  713   Q  G  W  L  C  L  S  S  G  Q  L  D  W  L  T  A  H  G  G  T  N  V  V  A  S  F  S  W  N  G  M  A  L  N
 2601   ATCCTGATAACTATCTCATCTCAAAGGATGTTACAGGCGCAACTAAGGTAAAGTACTACTATGCAGTCAACGACGGTTTTCCCGGGGATCACTATGCGGT
  747    P  D  N  Y  L  I  S  K  D  V  T  G  A  T  K  V  K  Y  Y  Y  A  V  N  D  G  F  P  G  D  H  Y  A  V
 2701   GATGATCTCCAAGACGGGCACGAACGCCGGAGACTTCACGGTTGTTTTCGAAGAAACGCCTAACGGAATAAATAAGGGCGGAgcAAGATTCGGTCTTTCC
  780    M  I  S  K  T  G  T  N  A  G  D  F  T  V  V  F  E  E  T  P  N  G  I  N  K  G  G  A  R  F  G  L  S
 2801   ACGGAAGCCGATGGCGCCAAACCTCAAAGTGTATGGATCGAGCGTACGGTAGATTTGCCTGCGGGTACTAAGTATGTTGCTTTCCGTCACTACAATTGCT
  813   T  E  A  D  G  A  K  P  Q  S  V  W  I  E  R  T  V  D  L  P  A  G  T  K  Y  V  A  F  R  H  Y  N  C  S
 2901   CGGATTTGAACTACATTCTTTTGGATGATATTCAGTTCACCATGGGTGGCAGCCCCACCCCGACCGATTATACCTACACGGTGTATCGTGACGGTACGAA
  847    D  L  N  Y  I  L  L  D  D  I  Q  F  T  M  G  G  S  P  T  P  T  D  Y  T  Y  T  V  Y  R  D  G  T  K
 3001   GATCAAGGAAGGTCTGACCGAAACGACCTTCGAAGAAGACGGTGTAGCTACGGGCAACCATGAGTATTGCGTGGAAGTGAAGTACACAGCCGGCGTATCT
  880    I  K  E  G  L  T  E  T  T  F  E  E  D  G  V  A  T  G  N  H  E  Y  C  V  E  V  K  Y  T  A  G  V  S
 3101   CCGAAAGAGTGTGTAAACGTAACTGTTGATCCTGTGCAGTTCAATCCTGTACAGAACCTGACCGGTAGTGCAGTCGGCCAGAAAGTAACGCTTAAGTGGG
  913   P  K  E  C  V  N  V  T  V  D  P  V  Q  F  N  P  V  Q  N  L  T  G  S  A  V  G  Q  K  V  T  L  K  W  D
                  KpnI
 3201   ATGCACCTAATGGTACCCCGAATCCAAATCCAAATCCGAATCCGGGAACAACAACACTTTCCGAATCATTCGAAAATGGTATTCCTGCCTCATGGAAGAC
  947    A  P  N  G  T  P  N  P  N  P  N  P  N  P  G  T  T  T  L  S  E  S  F  E  N  G  I  P  A  S  W  K  T
 3301   GATCGATGCAGACGGTGACGGCAACAATTGGACGACGACCCCTCCTCCCGGAGGCACCTCTTTTGCAGGTCACAACAGTGCGATCTGTGCCTCTTCGGCT
  980    I  D  A  D  G  D  G  N  N  W  T  T  T  P  P  P  G  G  T  S  F  A  G  H  N  S  A  I  C  A  S  S  A
 3401   TCTTATATCAACTTTGAAGGCCCTCAGAACCCTGATAACTATCTGGTTACACCGGAGCTATCTCTTCCTAACGGAGGAACGCTTACTTTCTGGGTATGTG
 1013   S  Y  I  N  F  E  G  P  Q  N  P  D  N  Y  L  V  T  P  E  L  S  L  P  N  G  G  T  L  T  F  W  V  C  A
 3501   CACAAGATGCCAATTATGCATCAGAGCACTATGCCGTGTATGCATCTTCTACGGGTAACGACGCTTCCAACTTCGCCAACGCTTTGTTGGAAGAAGTGCT
 1047    Q  D  A  N  Y  A  S  E  H  Y  A  V  Y  A  S  S  T  G  N  D  A  S  N  F  A  N  A  L  L  E  E  V  L
 3601   GACGGCCAAGACAGTTGTTACGGCACCTGAAGCCATTCGTGGCACTCGTGTTCAGGGCACCTGGTATCAAAAGACGGTACAGTTGCCTGCGGGTACTAAG
 1080    T  A  K  T  V  V  T  A  P  E  A  I  R  G  T  R  V  Q  G  T  W  Y  Q  K  T  V  Q  L  P  A  G  T  K
```

FIG. 2. Nucleotide and the deduced amino acid sequences of *hagA* and its product. The potential promoter region of −35 and −10, the start and stop codons, the potential terminator regions, the primers used for IPCR, and some restriction enzymes are underlined and labeled above the sequences. The amino acids at the 5′ and 3′ ends of the four repeat regions are indicated by italic letters and underlining.

```
3701  TATGTTGCTTTCCGTCACTTCGGCTGTACGGACTTCTTCTGGATCAACCTTGATGATGTTGAGATCAAGGCCAACGGCAAGCGCGCAGACTTCACGGAAA
1113   Y  V  A  F  R  H  F  G  C  T  D  F  F  W  I  N  L  D  D  V  E  I  K  A  N  G  K  R  A  D  F  T  E  T
3801  CGTTCGAGTCTTCTACTCATGGAGAGGCACCGGCGGAATGGACTACTATCGATGCCGATGGCGATGGTCAGGGTTGGCTCTGTCTGTCTTCCGGACAATT
1147    F  E  S  S  T  H  G  E  A  P  A  E  W  T  T  I  D  A  D  G  D  G  Q  G  W  L  C  L  S  S  G  Q  L
3901  GGGCTGGCTGACAGCTCATGGCGGCACCAACGTAGTAGCCTCTTTCTCATGGAATGGAATGGCTTTGAATCCTGATAACTATCTCATCTCAAAGGATGTT
1180   G  W  L  T  A  H  G  G  T  N  V  V  A  S  F  S  W  N  G  M  A  L  N  P  D  N  Y  L  I  S  K  D  V
4001  ACAGGCGCAACTAAGGTAAAGTACTACTATGCAGTCAACGACGGTTTTCCCGGGGATCACTATGCGGTGATGATCTCCAAGACGGGCACGAACGCCGGAG
1213  T  G  A  T  K  V  K  Y  Y  Y  A  V  N  D  G  F  P  G  D  H  Y  A  V  M  I  S  K  T  G  T  N  A  G  D
4101  ACTTCACGGTTGTTTTCGAAGAAACGCCTAACGGAATAAATAAGGGCGGAGCAAGATTCGGTCTTTCCACGGAAGCCGATGGCGCCAAACCTCAAAGTGT
1247    F  T  V  V  F  E  E  T  P  N  G  I  N  K  G  G  A  R  F  G  L  S  T  E  A  D  G  A  K  P  Q  S  V
4201  ATGGATCGAGCGTACGGTAGATTTGCCTGCGGGTACTAAGTATGTTGCTTTCCGTCACTACAATTGCTCGGATTTGAACTACATTCTTTTGGATGATATT
1280   W  I  E  R  T  V  D  L  P  A  G  T  K  Y  V  A  F  R  H  Y  N  C  S  D  L  N  Y  I  L  L  D  D  I
4301  CAGTTCACCATGGGTGGCAGCCCCACCCCGACCGATTATACCTACACGGTGTATCGTGACGGTACGAAGATCAAGGAAGGTCTGACCGAAACGACCTTCG
1313  Q  F  T  M  G  G  S  P  T  P  T  D  Y  T  Y  T  V  Y  R  D  G  T  K  I  K  E  G  L  T  E  T  T  F  E
4401  AAGAAGACGGTGTAGCTACGGGCAACCATGAGTATTGCGTGGAAGTGAAGTACACAGCCGGCGTATCTCCGAAAGAGTGTGTAAACGTAACTGTTGATCC
1347    E  D  G  V  A  T  G  N  H  E  Y  C  V  E  V  K  Y  T  A  G  V  S  P  K  E  C  V  N  V  T  V  D  P
                                                                                         KpnI
4501  TGTGCAGTTCAATCCTGTACAGAACCTGACCGGTAGTGCAGTCGGCCAGAAAGTAACGCTTAAGTGGGATGCACCTAATGGTACCCCGAATCCAAATCCA
1380   V  Q  F  N  P  V  Q  N  L  T  G  S  A  V  G  Q  K  V  T  L  K  W  D  A  P  N  G  T  P  N  P  N  P
4601  AATCCGAATCCGGGAACAACAACACTTTCCGAATCATTCGAAAATGGTATTCCTGCCTCATGGAAGACGATCGATGCAGACGGTGACGGCAACAATTGGA
1413  N  P  N  P  G  T  T  T  L  S  E  S  F  E  N  G  I  P  A  S  W  K  T  I  D  A  D  G  D  G  N  N  W  T
4701  CGACGACCCCTCCTCCCGGAGGCACCTCTTTTGCAGGTCACAACAGTGCGATCTGTGCCTCTTCGGCTTCTTATATCAACTTTGAAGGCCCTCAGAACCC
1447    T  T  P  P  P  G  G  T  S  F  A  G  H  N  S  A  I  C  A  S  S  A  S  Y  I  N  F  E  G  P  Q  N  P
4801  TGATAACTATCTGGTTACACCGGAGCTATCTCTTCCTAACGGAGGAACGCTTACTTTCTGGGTATGTGCACAAGATGCCAATTATGCATCAGAGCACTAT
1480   D  N  Y  L  V  T  P  E  L  S  L  P  N  G  G  T  L  T  F  W  V  C  A  Q  D  A  N  Y  A  S  E  H  Y
4901  GCCGTGTATGCATCTTCTACGGGTAACGACGCTTCCAACTTCGCCAACGCTTTGTTGGAAGAAGTGCTGACGGCCAAGACAGTTGTTACGGCACCTGAAG
1513  A  V  Y  A  S  S  T  G  N  D  A  S  N  F  A  N  A  L  L  E  E  V  L  T  A  K  T  V  V  T  A  P  E  A
5001  CCATTCGTGGCACTCGTGTTCAGGGCACCTGGTATCAAAAGACGGTACAGTTGCCTGCGGGTACTAAGTATGTTGCTTTCCGTCACTTCGGCTGTACGGA
1547    I  R  G  T  R  V  Q  G  T  W  Y  Q  K  T  V  Q  L  P  A  G  T  K  Y  V  A  F  R  H  F  G  C  T  D
5101  CTTCTTCTGGATCAACCTTGATGATGTTGAGATCAAGGCCAACGGCAAGCGCGCAGACTTCACGGAAACGTTCGAGTCTTCTACTCATGGAGAGGCACCG
1580   F  F  W  I  N  L  D  D  V  E  I  K  A  N  G  K  R  A  D  F  T  E  T  F  E  S  S  T  H  G  E  A  P
5201  GCGGAATGGACTACTATCGATGCCGATGGCGATGGTCAGGGTTGGCTCTGTCTGTCTTCCGGACAATTGGGCTGGCTGACAGcTCATGGCGGCACCAACG
1613  A  E  W  T  T  I  D  A  D  G  D  G  Q  G  W  L  C  L  S  S  G  Q  L  G  W  L  T  A  H  G  G  T  N  V
5301  TAGTAGCCTCTTTCTCATGGAATGGAATGGCTTTGAATCCTGATAACTATCTCATCTCAAAGGATGTTACAGGCGCAACTAAGGTAAAGTACTACTATGC
1647    V  A  S  F  S  W  N  G  M  A  L  N  P  D  N  Y  L  I  S  K  D  V  T  G  A  T  K  V  K  Y  Y  Y  A
5401  AGTCAACGACGGTTTTCCCGGGGATCACTATGCGGTGATGATCTCCAAGACGGGCACGAACGCCGGAGACTTCACGGTTGTTTTCGAAGAAACGCCTAAC
1680   V  N  D  G  F  P  G  D  H  Y  A  V  M  I  S  K  T  G  T  N  A  G  D  F  T  V  V  F  E  E  T  P  N
5500  GGAATAAATAAGGGCGGAGCAAGATTCGGTCTTTCCACGGAAGCCGATGGCGcCAAACCTCAAAGTGTATGGATCGAGCGTACGGTAGATTTGCCTGCGG
1713  G  I  N  K  G  G  A  R  F  G  L  S  T  E  A  D  G  A  K  P  Q  S  V  W  I  E  R  T  V  D  L  P  A  G
5601  GTACTAAGTATGTTGCTTTCCGACACTACAATTGCTCGGATTTGAACTACATTCTTTTGGATGATATTCAGTTCACCATGGGTGGCAGCCCCACCCCGAC

1747    T  K  Y  V  A  F  R  H  Y  N  C  S  D  L  N  Y  I  L  L  D  D  I  Q  F  T  M  G  G  S  P  T  P  T
5701  CGATTATACCTACACGGTGTATCGTGACGGTACGAAGATCAAGGAAGGTCTGACCGAAACGACCTTCGAAGAAGACGGTGTAGCTACGGGCAACCATGAG
1780   D  Y  T  Y  T  V  Y  R  D  G  T  K  I  K  E  G  L  T  E  T  T  F  E  E  D  G  V  A  T  G  N  H  E
5801  TATTGCGTGGAAGTGAAGTACACAGCCGGCGTATCTCCGAAAGAGTGTGTAAACGTAACTGTTGATCCTGTGCAGTTCAATCCTGTACAGAACCTGACCG
1813  Y  C  V  E  V  K  Y  T  A  G  V  S  P  K  E  C  V  N  V  T  V  D  P  V  Q  F  N  P  V  Q  N  L  T  G
                                              KpnI
5901  GTAGTGCAGTCGGCCAGAAAGTAACGCTTAAGTGGGATGCACCTAATGGTACCCCGAATCCAAATCCAAATCCGAATCCGGGAACAACAACACTTTCCGA
1847    S  A  V  G  Q  K  V  T  L  K  W  D  A  P  N  G  T  P  N  P  N  P  N  P  N  P  G  T  T  T  L  S  E
6001 ATCATTCGAAAATGGTATTCCTGCCTCATGGAAGACGATCGATGCAGACGGTGACGGCAACAATTGGACGACGACCCCTCCTCCCGGAGGCACCTCTTTT
1880    S  F  E  N  G  I  P  A  S  W  K  T  I  D  A  D  G  D  G  N  N  W  T  T  T  P  P  P  G  G  T  S  F
6101  GCAGGTCACAACAGTGCGATCTGTGTCTCTTCGGCTTCTTATATCAACTTTGAAGGCCCTCAGAACCCTGATAACTATCTGGTTACACCGGAGCTATCTC
1913  A  G  H  N  S  A  I  C  V  S  S  A  S  Y  I  N  F  E  G  P  Q  N  P  D  N  Y  L  V  T  P  E  L  S  L
6201  TTCCTGGCGGAGGAACGCTTACTTTCTGGGTATGTGCACAAGATGCCAATTATGCATCAGAGCACTATGCCGTGTATGCATCTTCTACGGGTAACGACGC
1947    P  G  G  G  T  L  T  F  W  V  C  A  Q  D  A  N  Y  A  S  E  H  Y  A  V  Y  A  S  S  T  G  N  D  A
6301  TTCCAACTTCGCCAACGCTTTGTTGGAAGAAGTGCTGACGGCCAAGACAGTTGTTACGGCACCTGAAGCCATTCGTGGCACTCGTGTTCAGGGCACCTGG
1980   S  N  F  A  N  A  L  L  E  E  V  L  T  A  K  T  V  V  T  A  P  E  A  I  R  G  T  R  V  Q  G  T  W
6401  TATCAAAAGACGGTACAGTTGCCTGCGGGTACTAAGTATGTTGCCTTCCGTCACTTCGGCTGTACGGACTTCTTCTGGATCAACCTTGATGAAGTTGAGA
2013  Y  Q  K  T  V  Q  L  P  A  G  T  K  Y  V  A  F  R  H  F  G  C  T  D  F  F  W  I  N  L  D  E  V  E  I
6501  TCAAGGCCAACGGCAAGCGCGCAGACTTCACGGAAACGTTCGAGTCTTCTACTCATGGAGAGGCACCGGCGGAATGGACTACTATCGATGCCGATGGCGA
2047    K  A  N  G  K  R  A  D  F  T  E  T  F  E  S  S  T  H  G  E  A  P  A  E  W  T  T  I  D  A  D  G  D
6601  TGGTCAGGGTTGGCTCTGTCTGTCTTCCGGACAATTGGACTGGCTGACAGCTCATGGCGGCACCAACGTAGTAGCCTCTTTCTCATGGAATGGAATGGCT
2080   G  Q  G  W  L  C  L  S  S  G  Q  L  D  W  L  T  A  H  G  G  T  N  V  V  A  S  F  S  W  N  G  M  A
6701  TTGAATCCTGATAACTATCTCATCTCAAAGGATGTTACAGGCGCAACTAAGGTAAAGTACTACTATGCAGTCAACGACGGTTTTCCCGGGGATCACTATG
2113  L  N  P  D  N  Y  L  I  S  K  D  V  T  G  A  T  K  V  K  Y  Y  Y  A  V  N  D  G  F  P  G  D  H  Y  A
6801  CGGTGATGATCTCCAAGACGGGCACGAACGCCGGAGACTTCACGGTTGTTTTCGAAGAAACGCCTAACGGAATAAATAAGGGCGGAGCAAGATTCGGTCT
2147    V  M  I  S  K  T  G  T  N  A  G  D  F  T  V  V  F  E  E  T  P  N  G  I  N  K  G  G  A  R  F  G  L
6901  TTCCACGGAAGCCGATGGCGCCAAACCTCAAAGTGTATGGATCGAGCGTACGGTAGATTTGcCTGCGGGCACGAAGTATGTTGCTTTCCGTCACTACAAT
2180   S  T  E  A  D  G  A  K  P  Q  S  V  W  I  E  R  T  V  D  L  P  A  G  T  K  Y  V  A  F  R  H  Y  N
7001  TGCTCGGATTTGAACTACATTCTTTTGGATGATATTCAGTTCACCATGGGTGGCAGCCCCACCCCGACCGATTATACCTACACGGTGTATCGTGACGGTA
2213  C  S  D  L  N  Y  I  L  L  D  D  I  Q  F  T  M  G  G  S  P  T  P  T  D  Y  T  Y  T  V  Y  R  D  G  T
7101  CGAAGATCAAGGAAGGTCTGACCGAAACGACCTTCGAAGAAGATGGTGTAGCTACGGGCAATCATGAGTATTGCGTGGAAGTGAAGTACACAGCCGGCGT
2247    K  I  K  E  G  L  T  E  T  T  F  E  E  D  G  V  A  T  G  N  H  E  Y  C  V  E  V  K  Y  T  A  G  V
                                AseI
7201  ATCTCCGAAGGTGTGTGTAAACGTAACTATTAATCCGACTCAGTTCAATCCTGTACAGAACCTGACGGCAGAACAAGCTCCTAACAGCATGGATGCAATC
2280   S  P  K  V  C  V  N  V  T  I  N  P  T  Q  F  N  P  V  Q  N  L  T  A  E  Q  A  P  N  S  M  D  A  I
7301  CTTAAATGGAATGCACCGGCATCTAAGCGTGCGGAAGTTCTGAACGAAGACTTCGAAAATGGTATTCCTTCCTCATGGAAGACGATCGATGCAGACGGGG
2313  L  K  W  N  A  P  A  S  K  R  A  E  V  L  N  E  D  F  E  N  G  I  P  S  S  W  K  T  I  D  A  D  G  D
7401  ACGGCAACAATTGGACGACGACCCCTCCTCCCGGAGGCTCCTCTTTTGCAGGTCACAACAGTGCGATCTGTGTCTCTTCGGCTTCTTATATCAACTTTGA
2347    G  N  N  W  T  T  T  P  P  P  G  G  S  S  F  A  G  H  N  S  A  I  C  V  S  S  A  S  Y  I  N  F  E
7501  AGGTCCTCAGAACCCTGATAACTATCTGGTTACACCGGAGCTTTCTCTTCCTGGCGGAGGAACGCTTACTTTCTGGGTATGTGCACAAGATGCCAATTAT
```

FIG. 2—*Continued.*

```
2380  G  P  Q  N  P  D  N  Y  L  V  T  P  E  L  S  L  P  G  G  G  T  L  T  F  W  V  C  A  Q  D  A  N  Y
7601 GCATCAGAGCACTATGCCGTGTATGCATCTTCTACGGGTAACGACGCTTCCAACTTCGCCAACGCTTTGTTGGAAGAAGTGCTGACGGCCAAGACAGTTG
2413 A  S  E  H  Y  A  V  Y  A  S  S  T  G  N  D  A  S  N  F  A  N  A  L  L  E  E  V  L  T  A  K  T  V  V
7701 TTACGGCGCCTGAAGCCATTCGTGGCACTCGTGTTCAGGGCACCTGGTATCAAAAGACGGTACAGTTGCCTGCGGGTACTAAGTATGTTGCCTTCCGTCA
2447   T  A  P  E  A  I  R  G  T  R  V  Q  G  T  W  Y  Q  K  T  V  Q  L  P  A  G  T  K  Y  V  A  F  R  H
7801 CTTCGGCTGTACGGACTTCTTCTGGATCAACCTTGATGATGTTGTAATCACTTCAGGGAACGCTCCGTCTTACACCTATACGATCTATCGTAATAATACA
2480  F  G  C  T  D  F  F  W  I  N  L  D  D  V  V  I  T  S  G  N  A  P  S  Y  T  Y  T  I  Y  R  N  N  T
7901 CAGATAGCATCAGGCGTAACGGAGACTACTTACCGAGATCCGGACTTGGCTACCGGTTTTTACACGTACGGTGTTAAGGTTGTTTACCCGAACGGAGAAT
2513 Q  I  A  S  G  V  T  E  T  T  Y  R  D  P  D  L  A  T  G  F  Y  T  Y  G  V  K  V  V  Y  P  N  G  E  S
8001 CAGCTATCGAAACTGCTACGTTGAATATCACTTCGTTGGCAGACGTAACGGCTCAGAAGCCTTACACGCTGACAGTTGTAGGAAAGACGATCACGGTAAC
2547   A  I  E  T  A  T  L  N  I  T  S  L  A  D  V  T  A  Q  K  P  Y  T  L  T  V  V  G  K  T  I  T  V  T
8101 TTGCCAAGGCGAAGCTATGATCTACGACATGAACGGTCGTCGTCTGGCAGCCGGTCGCAACACGGTTGTTTACACGGCTCAGGGCGGCCACTATGCAGTC
2580  C  Q  G  E  A  M  I  Y  D  M  N  G  R  R  L  A  A  G  R  N  T  V  V  Y  T  A  Q  G  G  H  Y  A  V
8201 ATGGTTGTCGTTGACGGCAAGTCCTACGTAGAGAAACTCGCTGTAAAGTAACGAGATGATTATTTTCGATCGGTATGCTCTACCAACCGATCGCTTTAAT
2613 M  V  V  V  D  G  K  S  Y  V  E  K  L  A  V  K  *
8301 CGGTCGCCCGGCTTCCATAAAAAGGAGTCGGGCGACTCTTTTACTCCAACCAAATAAGCATTGTTTTATAGCCTTTCGGAATATACTCCGGAAGGGGGTC
8401 GAGCTACGCCCTACAGCGACTCGGGCTACGCCGTAGAGCGTACCGAGCTGCGCTCTACGGCTCTTCGAGCTACGCTGTAGGGCTCACTGCGCCAAGCTCT
                                                                                                SstI
8501 ACGGCTCAGCTCGGCCACCTCTACGGCTCCCGGAGCGGAACTCTACGGCTCGGCTCGCTACGCTGTAGAGCGTACCTACGCCGAGCTC
```

FIG. 2—*Continued.*

To obtain an intact *hagA* gene from the *P. gingivalis* 381 chromosome, an *Eco*RV bank was constructed since it was determined from the sequence that an *Eco*RV site existed near the 3′ end of the cloned *Bam*HI fragment and from Southern blot analysis (data not shown) that such a fragment should include the entire *hagA* ORF. Thus a 10,119-bp *Eco*RV fragment, which included an additional 338 bp upstream sequence of the *Ase*I fragment, was cloned. The complete ORF began at base 365 and ended at base 8251 of this fragment, resulting in an ORF 7,877 bp in length (Fig. 1 and 2). This ORF is calculated to encode a 2,628-aa protein with a molecular weight of 283.3 kDa.

Analysis of the sequence revealed potential −10 and −35 consensus sequences located at bases 168 and 143, respectively (Fig. 2). However, no *E. coli*-like ribosome binding site could be found upstream of the start codon except for AGG at the −4 to −2 positions. Two potential stem-loop structures, forming 14- and 9-bp-long inverted repeats, were identified 51 and 101 bp downstream of the stop codon, respectively (Fig. 2). Residues 5 to 21 are consistent with a typical, hydrophobic leader or signal sequence according to the Chou-Fasman prediction (16).

The repeat region was found to begin immediately after the first *Kpn*I site at base 1862 and to end at base 7265, making the entire repeat region 5,404 bp in length without a single gap (Fig. 2). The first repeat unit (*HArep* 1) is 1,350 bp and has 99.5% identity to the second repeat unit. The second and third repeat units (*HArep* 2 and *HArep* 3) are 1,368 bp in length and are 99.9% identical to each other. The fourth repeat unit (*HArep* 4) is 1,318 bp in length and has 98.6% identity to both *HArep* 2 and *HArep* 3. The beginning amino acid sequence of the *HArep* 1 product is PNPNPGTTT, while that of the other three repeat products is GTPNPNPNPNPGTTT. Thus, the beginning sequences of the products of *HArep* 2 to 4 contain 6 aa more than the product of *HArep* 1. This difference is due to the product of *HArep* 1 containing two fewer repeats of the PN sequence since the GT is present before the sequence of PNP NPGTTT in the product of *HArep* 1 (Fig. 2). The Chou-Fasman rules predict these beginning amino acids of the product of *HArep* to be very antigenic and hydrophilic.

Comparison of *hagA* sequence with the other cloned genes of *P. gingivalis*. Others have recently reported the cloning of protease genes from various strains of *P. gingivalis* (5, 12, 17, 18, 21, 25, 27). When the sequences of these genes were compared with that of *hagA*, several of them were found to contain one copy of the *HArep* sequence (Fig. 3 and 4). For example, the product of *prtH*, a gene encoding a C3 protease cloned from strain W83 (12), has a region of 270 aa with 95.6% homology to the *hagA* product. The product of *rgp1*, the arginine-specific cysteine protease/hemagglutinin gene cloned from strain H66 (27), contains a 522-aa region with 93.1% homology, as do the products of *prtR* (18), cloned from strain W50 by Reynolds et al. (33a), *agp* (25), cloned from strain 381 by Okamoto et al. (25), and *prpR1*, cloned from strain W50 also by Curtis et al. (8a), genes identical to *rgp1* isolated from different strains, all of which contain one *HArep* sequence of *hagA*. In addition, the product of *prtP*, a cysteine protease/hemagglutinin gene cloned from strain W12 jointly by M. Lantz and us, has a 849-aa C-terminal region which has 94.4% homology to the product of *hagA*, with the last 171 aa being absolutely identical. This homologous region accounts for almost half of the length of the *prtP* gene. The product of *tla*, another protease gene cloned from strain W50 by Curtis et al., has a 789-aa C-terminal region with 95.2% homology to the *hagA* product, with the last 171 aa completely identical (8a). This region constitutes almost three-fourths the length of this gene. In addition, the product of *hagD*, a fourth hemagglutinin gene cloned from strain 381 by us, has a 523-aa region with 92.7% homology to the *hagA* product. The product of *hagE*, an additional hemagglutinin

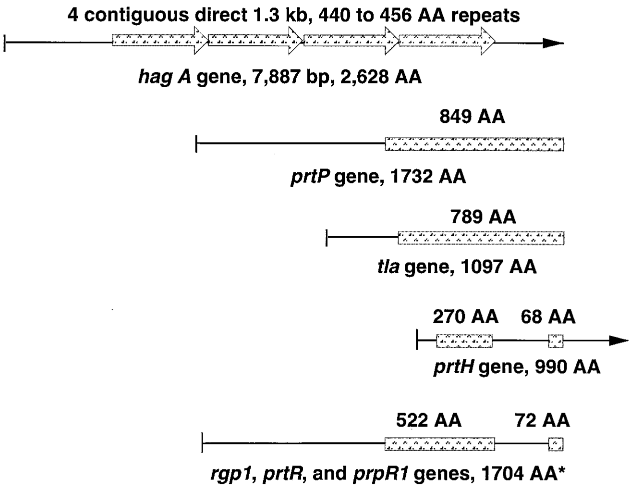
4 contiguous direct 1.3 kb, 440 to 456 AA repeats
hag A gene, 7,887 bp, 2,628 AA
849 AA
prtP gene, 1732 AA
789 AA
tla gene, 1097 AA
270 AA
68 AA
prtH gene, 990 AA
522 AA
72 AA
rgp1, prtR, and prpR1 genes, 1704 AA*

FIG. 3. Comparison of *hagA* product with products of other cloned genes. The stippled areas represent areas of high homology (see details in Results). An asterisk indicates that the products of these genes vary in length by 2 aa. Drawn to scale.

```
1780 DYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTVDPV.QFNPVQNLTGSAVGQKVTLKWDAPNGT..PNPNPNPNPGTTT
     ||||||||||||||||||.|||||||||||||||||||||||||||| | :|||:.  :|.||||||||||||||||||||||||  |||||||||||||
 865 DYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKWDAPNGTPNPNPNPNPNPGTTT

1877 LSESFENGIPASWKTIDADGDGNNWTTTPPPGGTSFAGHNSAICVSSASY.INFEGPQNPDNYLVTPELSLPGGGTLTFWVCAQDANYASEHYAVYASST
     ||||||||||||||||||||||:.|... :||   :||.||. || |.|: :.  |. .|||||:||.|.||.||.||||||||||||||||||||||||
 965 LSESFENGIPASWKTIDADGDGHGWKPGNAPG...IAGYNSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNGGKLTFWVCAQDANYASEHYAVYASST

1976 GNDASNFANALLEEVLTAKTVVTAPEAIRGTRVQGTWYQKTVQLPAGTKYVAFRHFGCTDFFWINLDEVEIKANGKRADFTETFESSTHGEAPAEWTTID
     |||||||.||||||.:|||. | .|||||| |:|||| ||||:|||||||||||||.:||:|:|:|||||||||||||||||||||||||||||||||||
1062 GNDASNFTNALLEETITAKG.VRSPEAIRG.RIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVEIKANGKRADFTETFESSTHGEAPAEWTTID

2076 ADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGA
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1160 ADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDGFPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGA

2176 RFGLSTEADGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKY
     ||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1260 RFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPTPTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKY

2276 TAGVSPKVCVNVTINPTQFNPVQNLTA 2302
     ||||||| ||||||||||||||.||.|
1360 TAGVSPKECVNVTINPTQFNPVKNLKA 1386


2557 SLADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGHYAVMVVVDGKSYVEKLAVK 2628
     ::||||||||||||||||||||||||||||||||||||||||||||||||||.|||||||||||||||||||
1633 GVADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGYYAVMVVVDGKSYVEKLAVK 1704
```

FIG. 4. Comparison of the *hagA* product amino acid sequence with that of the *rgp1* product (stippled area in Fig. 3). The upper row is the *hagA* product sequence, and the lower row is the *rgp1* product sequence. The homology between these two sequences is 93%, and the identity is 90%. A comparison of other protease gene products with the *hagA* product in the same region indicates that the homology or identity among the products of other genes is equal to or greater than that of the *rgp1* and *hagA* products. For example, the identity between the *prtP* and *rgp1* products is 97.8% in this 522-aa region with only one gap, while the identity between the *prtP* and *tla* products is 99.5% in a span of 795-aa without any gap. The last two lines provide a comparison of the last 72 aa of the two gene products. The homology is 98.6%, and the identity is 95.8%.

gene also cloned from strain 381 in our lab, contains a 523-aa region with 93% homology to the *hagA* product. Without exception, these high-homology regions of each of these genes are within or extend from the repeat region of *hagA*.

In addition, the product of each of these genes contains a 72-aa C terminus in common with the *hagA* product (90 to 100% homology), except for the *prtH* product, in which this region is located in the middle of the sequence (Fig. 3 and 4).

Comparison of the *hagA* sequence with other sequences in the databases. A search through the National Center for Biotechnology Information database with the GENINFO Experimental Blast Network Service revealed no significant homology of the *hagA* sequence to any other sequences in the databases except for the *Mycoplasma gallisepticum* hemagglutinin genes (*pMGA*) (22) and the circumsporozoite protein (CS) genes of *Plasmodium falciparum* (7). The products of these genes were found to have partial homology to the product of *hagA* in very short regions (11 of 14 aa for *pMGA* of *M. gallisepticum*, i.e., **PNGTPNPNPNPNPG** corresponding to positions 948, 1404, and 1860 of the *hagA* product; 9 of 13 aa for the CS genes of *P. falciparum*, i.e., **PNGTPNPNPNPNP** corresponding to positions 948, 1404, and 1860 of the *hagA* product [bold letters indicate common amino acids]).

Hemagglutination results. Clone pNH 201, which contains the entire 10.1-kb *Eco*RV fragment, clone pNH 1, which contains the first two *HArep* units from base 339 to 4838, and clone pNH 9, which contains a single *HArep* unit constructed from the ST 2 fragment with *Acc*65I, all conferred hemagglutination activity at a titer of 1/8 (corresponding to an optical density of 2.5 to 3 at 660 nm), while *E. coli* cells containing vector alone showed no hemagglutination activity even at a titer of 1/2 (corresponding to an optical density of 12 at 660 nm). The wild-type *P. gingivalis* strain 381 had a titer of 1/64 (corresponding to an optical density of 0.125 at 660 nm).

DISCUSSION

The cloning and sequencing of the entire *hagA* ORF from strain 381 proved to be a daunting task because of the presence of four large, contiguous, direct repeats within this gene and the enormous size of the ORF. The presence of the repeats made the construction of appropriate subclones and the designing of useful oligonucleotide primers difficult since most enzymes which cut in one repeat region would cut in any other repeat unit and the oligonucleotide primers which bind in one repeat region would bind to any other repeat unit. In addition, the length of the repeats was problematic since the length of one repeat is longer than what could be sequenced in one sequencing reaction (at most, 300 to 400 bp) at the time sequencing was completed. Thus, the subclones and oligonucleotide primers were designed to be unambiguous with respect to the repeat units. The number and length of repeats were established and verified from IPCR products and multiple genomic banks.

The extremely large size of *hagA* is rare in prokaryotes. Recently, Reuven et al. (33) reported the cloning of the *lhr* gene, a member of the helicase superfamily II, which encodes a 1,538-aa protein. They claimed it is the longest gene found to date in *E. coli*. However, *lhr* is only slightly more than half (58.5%) the length of *hagA*. We thus suspect that *hagA* is among the longest genes ever cloned from a prokaryote.

It is likely that the *HArep* sequence contains the functional hemagglutinin domain since a clone of a single repeat unit demonstrated strong hemagglutination activity. The sequence PGPNPNPNPNPG, which begins each of the products of the *HArep* units, is very similar to a region of *M. gallisepticum* hemagglutinin gene products. The common PN repeat sequence among *P. gingivalis* gene products listed above and short peptide sequences in *M. gallisepticum* in which there is an abundance of P and N amino acids may indicate that this region is involved in erythrocyte binding or some other common function. Interestingly, this region is predicted to be very antigenic by the Chou-Fasman prediction methodology (16).

The fact that clone pNH 201, which contains the entire *hagA* ORF including the putative promoter region, demonstrated strong hemagglutination activity suggests that the *hagA* gene, in spite of the fact that no *E. coli*-like ribosome binding sequence was identified, is functionally expressed from its own promoter since this fragment was cloned in the opposite direction of the *lac* promoter in pUC 18.

The motif of several direct repeats within an ORF has been reported for virulence-associated genes of various other pathogenic species, including *Mycoplasma hyorhinis* (38), *Mycoplasma hominis* (19), group A streptococci (11), *Streptococcus gordonii* (23), *Rickettsia rickettsii* (2), and *Anaplasma marginale* (1). Interestingly, without exception, this motif is found in genes encoding major surface antigens which are directly involved in virulence. For example, in the case of *M. hyorhinis* and *A. marginale*, the repeats are the mechanism whereby the surface antigen undergoes antigenic or size variation (1, 38). Other gene products in which these repeats occur include the M protein of group A streptococci (11) and the products of genes encoding adhesins such as the *cshA* gene of *S. gordonii* (23) and the group A streptococcus serum opacity factor gene (32) in which the repeats encode the fibronectin binding domain. The repeated sequences in these genes vary from 42 to 471 bp in length, with as few as 2 to as many as 13 repeats of each sequence. Thus, the *hagA* repeats of 1,318 to 1,368 bp are exceptionally large compared with these. Like *hagA*, the repeat domains of several of these genes constitute more than 50% of the ORF. The presence of the *HArep* units, establishing a motif for *hagA* similar to these other surface virulence factors, and the fact that *hagA* has adherence functions suggest that *hagA* is a virulence factor with functions like those virulence genes listed above. In addition, the presence of multiple repeat units may provide a means for *hagA* to undergo rearrangements (duplications or deletions of a repeat unit) and thus antigenic variation. This possibility is presently under investigation.

Surprisingly, the *HArep* sequence of *hagA* is also found in most protease genes of this species except for *tpr*, a thiol protease gene isolated from strain W83 (5), *prtC*, a collagenase gene from strain ATCC 53977 (17), and *prtT*, a cysteine protease/hemagglutinin gene from strain ATCC 53977 (21). The fact that *HArep* is found in *hagD, hagE*, and several protease genes suggests that these genes which contain *HArep* sequences form a *HArep* multigene family which functions in virulence and interacts with host tissues. Given the number of *HArep* sequences present in *hagA, hagA* likely is a central gene in this family. The other genes of the family may be the result of recombination events which involved a *HArep* unit of *hagA*, or they may all be derived from a common ancestral gene during earlier evolution of this species.

It has been reported that several types of large repeat sequences (>1 kb) are present in *E. coli* and *Salmonella typhimurium* genomes, including *rrn* loci, *rhs* loci (*E. coli*), *rtl-atl/gat* alternation of alleles, and insertion sequences (34). These elements are believed to contribute to restructuring the chromosome on which they reside in the form of duplications, deletions, transpositions, and inversions. Whether the *HArep* sequences may function similarly is not yet known. The analysis of *hagA* and other genes of this family in various strains would provide information as to this possibility.

A second distinguishing characteristic of the *hagA* multigene family is the presence of a 72-aa sequence normally at the extreme carboxyl terminus of the proteins. This region is hydrophobic according to the Chou-Fasman prediction and may serve to anchor the proteins in the outer membrane or in some other common recognition function.

It has been suggested previously that protease and hemagglutination activities of *P. gingivalis* are related. In a study of the trypsin-like protease activity of *P. gingivalis*, Hoover et al. (15) reported that mutant strains of *P. gingivalis* deficient in trypsin-like protease activity had markedly reduced hemagglutination activity. Nishikata and Yoshimura (24) have reported that a 44-kDa purified outer membrane hemagglutinin has been further characterized as a cysteine protease. We have previously reported that a cloned fibrinogen-binding cysteine protease gene (presently designated *prtP*) from *P. gingivalis* W12 has a 2.0-kb region with greater than 90% homology to *hagA* (29). Recent studies of the Arg-Cys or Lys-Cys proteases have demonstrated that protease activities are always accompanied by hemagglutination activity, and subsequently, these authors reported that the Arg-Cys protease and hemagglutinins were encoded by a single gene (27, 28). Collectively, these data strongly support the idea that the *P. gingivalis* hemagglutinins are involved with proteases and might be an important virulence factor in the initiation and progression of periodontal disease. We do not yet know whether the *hagA* product possesses protease activity.

In addition to the in vitro data indicating that the *hagA* product is a hemagglutinin and likely has virulence-related properties, recent in vivo data suggest that the *hagA* product may be involved in colonization in humans. Curtis and coworkers (8) have demonstrated that a monoclonal antibody which inhibits hemagglutination and confers passive immunization to *P. gingivalis* recolonization for up to 9 months recognizes colonization determinants of the PrpRI β fragment which is the adhesin domain of the *prpRI* gene product and which has high homology to the *hagA* product. Indeed, the 25-residue sequence (GVSPK VCKDV TVEGS NEFAP VQNLT) encoding the antigenic epitope is also present in the *hagA* product (in the *hagA* product, residue 20 is H) immediately before the beginning of the *HArep* product sequences and once in each of the products of the *HArep* units, with the first 5 and last 6 aa being identical. This clearly supports the hypothesis that the *hagA* product plays a pivotal role in *P. gingivalis* colonization of humans and thus in periodontal infections.

In summary, the entire *hagA* gene was cloned and sequenced. Analysis of the sequence revealed the presence of four large, contiguous, direct repeats which make it a most interesting gene for multiple reasons. The presence of the *HArep* in other genes of this species suggests that the genes which contain *HArep* form a multigene family with important virulence functions. In addition, the unusual size and the motif of the *hagA* protein as provided by the presence of the multiple repeats, similar to virulence-associated surface proteins of other species, indicate that this protein may have multiple roles in the biology and virulence properties of *P. gingivalis*.

ACKNOWLEDGMENTS

This study was supported by NIH grants DE 07496 and DE 00336 to Ann Progulske-Fox. Charles I. Hoover, Department of Stomatology, University of California, San Francisco, provided a portion of the *P. gingivalis* 381 genomic DNA used in this study.

The assistance of the University of Florida Periodontal Disease Research Center and Marilyn Lantz is appreciated.

REFERENCES

1. **Allred, D. R., T. C. McGuire, G. H. Palmer, S. R. Leib, T. M. Harkins, T. F. McElwain, and A. T. Barbet.** 1990. Molecular basis for surface antigen size polymorphisms and conservation of a neutralization-sensitive epitope in *Anaplasm marginale*. Proc. Natl. Acad. Sci. USA **87:**3220–3224.
2. **Anderson, B. E., G. A. McDonald, D. C. Jones, and R. L. Regnery.** 1990. A

protective protein antigen of *Ricketsia ricketsii* has tandemly repeated, nearly identical sequences. Infect. Immun. **58:**2760–2769.

2a. **Applied Biosystems, Inc.** 1991. High-quality template DNA for Taq cycle sequencing using DyeDeoxy terminators: an improved preparation procedure. User bulletin 18. Applied Biosystems, Inc., Foster City, Calif.

3. **Aslanzadeh, J., and L. J. Paulissen.** 1990. Adherence and pathogenesis of *Salmonella enteritidis* in mice. Microbiol. Immunol. **34:**885–893.

4. **Booth, B. A., M. Boesman-Finkelstein, and R. A. Finkelstein.** 1984. *Vibrio cholerae* hemagglutinin/protease nicks cholera enterotoxin. Infect. Immun. **45:**558–560.

5. **Bourgeau, G., H. Lapointe, P. Peloquin, and D. Mayrand.** 1992. Cloning, expression, and sequencing of a protease gene (*tpr*) from *Porphyromonas gingivalis* W83 in *Escherichia coli*. Infect. Immun. **60:**3186–3192.

6. **Brennan, M. J., Z.-M. Li, L. J. Cowell, M. E. Bisher, A. C. Steven, P. Novotny, and C. R. Manclark.** 1988. Identification of a 69-kilodalton nonfimbrial protein as an agglutinogen of *Bordetella pertussis*. Infect. Immun. **56:**3189–3195.

7. **Caspers, P., R. Gentz, H. Matile, R. Pink, and F. Sinigaglia.** 1989. The circumsporozoite protein gene from NF54, a *Plasmodium falciparum* isolate used in malaria vaccine trials. Mol. Biochem. Parasitol. **35:**185–190.

8. **Curtis, M. A., et al.** Analysis of the protease and adhesin domains of the prpRI of *P. gingivalis*. J. Dent. Res., in press.

8a. **Curtis, M. A., et al.** Personal communication.

9. **Duguid, J. P., E. S. Anderson, and I. Campbell.** 1966. Fimbriae and adhesive properties in *Salmonella*. J. Pathol. Bacteriol. **92:**107–138.

10. **Duguid, J. P., S. Clegg, and J. I. Wilson.** 1979. The fimbrial and non-fimbrial hemagglutination of *Escherichia coli*. J. Med. Microbiol. **12:**213–227.

11. **Fischetti, V.** 1991. Streptococcal M protein. Sci. Am. **264:**58–65.

12. **Fletcher, H. M., H. A. Schenkein, and F. L. Macrina.** 1994. Cloning and characterization of a new protease gene (*prtH*) from *Porphyromonas gingivalis*. Infect. Immun. **62:**4279–4286.

13. **Han, N., and A. Progulske-Fox.** Obtaining a 6-kb sequence of *hagA* from *P. gingivalis* by inverse polymerase chain reaction. Submitted for publication.

14. **Holdman, L. V., R. W. Kelly, and W. E. C. Moore.** 1984. Anaerobic gram-negative straight, curved and helical rods. Family I. *Bacteroidaceae*, p. 602–662. *In* N. R. Krieg and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 1. The Williams & Wilkins Co., Baltimore.

15. **Hoover, C. I., C. Y. Ng, and J. R. Felton.** 1992. Correlation of hemagglutination activity with trypsin-like protease activity of *Porphyromonas gingivalis*. Arch. Oral Biol. **37:**515–520.

16. **Jameson, B. A., and H. Wolf.** 1988. The antigenic index: a novel algorithm for predicting antigenic determinants. Comput. Appl. Biosci. **4:**181–186.

17. **Kato, T., N. Takashi, and H. K. Kuramitsu.** 1992. Sequence analysis and characterization of the *Porphyromonas gingivalis prtC* gene, which expresses a novel collagenase activity. J. Bacteriol. **174:**3889–3895.

18. **Kirszbaum, L., C. Sotiropolos, C. Jackson, S. Cleal, N. Slakeski, and E. C. Reynolds.** 1995. Complete nucleotide sequence of a gene *prtR* of *Porphyromonas gingivalis* W50 encoding a 132 kDa protein that contains an arginine-specific thiol endopeptidase domain and a hemagglutinin domain. Biochem. Biophys. Res. Commun. **207:**424–431.

19. **Ladefoged, S. A., S. Birkelund, S. Hauge, B. Brock, L. T. Jensen, and G. Christiansen.** 1995. A 135-kilodalton surface antigen of *Mycoplasma hominis* PG21 contains multiple directly repeated sequences. Infect. Immun. **63:**212–223.

20. **Lepine, G., and A. Progulske-Fox.** 1996. Duplication and differential expression of hemagglutinin genes in *Porphyromonas gingivalis*. Oral Microbiol. Immunol. **11:**65–78.

21. **Madden, T. E., V. L. Clark., and H. K. Kuramitsu.** 1995. Revised sequence of the *Porphyromonas gingivalis prtT* cysteine protease/hemagglutinin gene: homology with streptococcal pyrogenic exotoxin B/streptococcal proteinase. Infect. Immun. **63:**238–247.

22. **Markham, P. F., M. D. Glew, K. G. Whithear, and I. D. Walker.** 1993. Molecular cloning of the gene family that encodes pMGA, a hemagglutinin of *Mycoplasma gallisepticum*. Infect. Immun. **61:**903–909.

23. **McNab, R., H. Jenkinson, D. M. Loach, and G. W. Tannock.** 1994. Cell-surface-associated polypeptides *CshA* and *CshB* of high molecular mass are colonization determinants in the oral bacterium *Streptococcus gordonii*. Mol. Microbiol. **14:**743–754.

24. **Nishikata, M., and F. Yoshimura.** 1991. Characterization of *Porphyromonas (Bacteroides) gingivalis* hemagglutinin as a protease. Biochem. Biophys. Res. Commun. **178:**336–342.

25. **Okamoto, K., Y. Misumi, T. Kadowaki, M. Yoneda, K. Yamamoto, and Y. Ikehara.** 1995. Structural characterization of argingipain, a novel arginine-specific cysteine protease as a major periodontal pathogenic factor from *Porphyromonas gingivalis*. Arch. Biochem. Biophys. **316:**917–925.

26. **Okuda, K., and I. Takazoe.** 1974. Hemagglutination activity of *Bacteroides melaninogenicus*. Arch. Oral. Biol. **19:**415.

27. **Pavloff, N., J. Potempa, R. N. Pike, V. Prochazka, M. C. Kiefer, J. Travis, and P. J. Barr.** 1995. Molecular cloning and structural characterization of the arg-gingipain protease of *Porphyromonas gingivalis*. J. Biol. Chem. **270:**1007–1010.

28. **Pike, R., W. McGraw, J. Potempa, and J. Travis.** 1994. Lysine- and arginine-specific proteases from *Porphyromonas gingivalis*. J. Biol. Chem. **269:**406–411.

29. **Progulske-Fox, A., V. Rao, N. Han, G. Lepine, J. Whitlock, and M. Lantz.** 1993. Molecular characterization of hemagglutinin genes of periodontopathic bacteria. J. Periodontal Res. **28:**473–474.

30. **Progulske-Fox, A., S. Tumwasorn, and S. C. Holt.** 1989. The expression and function of a *Bacteroides gingivalis* hemagglutinin gene in *Escherichia coli*. Oral Microbiol. Immunol. **4:**121–131.

31. **Progulske-Fox, A., S. Tumwasorn, G. Lepine, J. Whitlock, D. Savett, J. J. Ferretti, and J. A. Banas.** 1995. The cloning, expression and sequence analysis of a second *Porphyromonas gingivalis* gene that codes for a protein involved in hemagglutination. Oral Microbiol. Immunol. **10:**311–318.

32. **Rakonjac, J. V., J. C. Robbins, and V. A. Fischetti.** 1995. DNA sequence of the serum opacity factor of group A streptococci: identification of a fibronectin-binding repeat domain. Infect. Immun. **63:**622–631.

33. **Reuven, N. B., E. V. Koonin, K. E. Rudd, and M. P. Deutscher.** 1995. The gene for the longest known *Escherichia coli* protein is a member of helicase superfamily II. J. Bacteriol. **177:**5393–5400.

33a. **Reynolds, X., et al.** Personal communication.

34. **Riley, M., and S. Krawiec.** 1987. Genome organization, p. 967–981. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.

35. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

36. **Shah, H. N., D. Mayrand, and R. J. Genco (ed.).** 1993. Biology of the species *Porphyromonas gingivalis*. CRC Press, Inc., Boca Raton, Fla.

37. **Wilson, K.** 1994. Preparation of genomic DNA from bacteria, p. 2.4.1–2.4.5. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology, suppl. 13. John Wiley & Sons, Inc., New York.

38. **Yogev, D., R. Rosengarten, R. Warson-McKown, and K. S. Wise.** 1991. Molecular basis of *Mycoplasma* surface antigenic variation: a novel set of divergent genes undergo spontaneous mutation of periodic coding regions and 5′ regulatory sequences. EMBO J. **10:**4069–4079.

Editor: J. R. McGhee