Molecular Cloning and Characterization of Porphyromonas gingivalis Lysine-specific Gingipain

A NEW MEMBER OF AN EMERGING FAMILY OF PATHOGENIC BACTERIAL CYSTEINE PROTEINASES

(Received for publication, April 9, 1996, and in revised form, August 27, 1996)

Nadine Pavloff‡, Philip A. Pemberton, Jan Potempa§, Wan-Ching Audrey Chen, Robert N. Pike‡, Vaclav Prochazka, Michael C. Kiefer, James Travis§, and Philip J. Barr

From LXR Biotechnology Inc., Richmond, California 94804, the §Department of Microbiology and Immunology, Institute of Molecular Biology, Jagiellonian University, 31-120 Cracow, Poland, and the ¶Department of Biochemistry, University of Georgia, Athens, Georgia 30602

The proteinases of Porphyromonas gingivalis are key virulence factors in the etiology and progression of periodontal disease. Previous work in our laboratories resulted in the purification of arginine- and lysine-specific cysteine proteinases, designated gingipains, that consist of several tightly associated protein subunits. Recent characterization of arginine-specific gingipain-1 (gingipain R1; RGP-1) revealed that the sequence is unique and that the protein subunits are initially translated as a polyprotein encoding a proteinase domain and multiple adhesin domains (Pavloff, N., Potempa, J., Pike, R. N., Prochazka, V., Kiefer, M. C., Travis, J., and Barr, P. J. (1995) J. Biol. Chem. 270, 1007–1010).

We now show that the lysine-specific gingipain (gingipain K; KGP) is also biosynthesized as a polyprotein precursor that contains a proteinase domain that is 22% homologous to the proteinase domain of RGP-1 and multiple adhesin domains. This precursor is similarly processed at distinct sites to yield active KGP. The key catalytic residues in the proteinase domain of KGP are identical to those found in RGP-1, but there are significant differences elsewhere within this domain that likely contribute to the altered substrate specificity of KGP. Independent expression of the proteinase domain in insect cells has shown that KGP does not require the presence of the adhesin domains for correct folding to confer proteolytic activity.

The anaerobic Gram-negative rod bacterium Porphyromonas gingivalis is implicated in the initiation and progression of certain forms of periodontitis, including juvenile and adult periodontal disease (1–3). Mechanisms by which P. gingivalis evades the host defense response and elicits hard and soft tissue damage in the periodontal pocket are under investigation, with many studies being focused on the proteinases produced by the organism. To date, the combined proteolytic activities of P. gingivalis have been shown to be capable of degrading and inactivating host defense proteins (iron-binding proteins, immunoglobulins, and complement components), structural proteins (collagen, fibronectin, and fibrinogen), and plasma proteinase inhibitors (4–10). In addition, proteinase activity is associated with the ability of the organism to adhere to collagenous substrata and to hemagglutinate and lyse red cells, thus allowing the organism to remain in the periodontal pocket while meeting its nutritional requirements for both heme and peptides (11, 12).

Recent work in our laboratories has resulted in the identification and purification of multiple cysteine proteinases, designated gingipains, some of which comprise several tightly associated protein subunits. The recent complete gene sequence of arginine-specific gingipain-1 (gingipain R1; RGP-1) revealed a unique proteinase that is synthesized initially as a polyprotein encoding a proteinase domain and four adhesin domains (13). The protein subunits are generated by subsequent proteolytic processing to mature RGP-1.

We have also purified a lysine-specific gingipain (gingipain K; (KGP) from P. gingivalis. This enzyme is a 60-kDa cysteine proteinase that is also found associated with adhesin molecules (14, 15). Amino-terminal protein sequencing of the 60-kDa proteinase subunit of KGP showed that it also is a unique proteinase. Here we present the complete DNA sequence of the kgp gene and its encoded protein. This represents the second fully characterized proteinase from an emerging family of cysteine proteinases that have structural features distinct from those of all other known families.

The sequence shows that KGP is similar to RGP-1 in its structural organization, biosynthesis, and maturational processing. We also show by expression of the KGP proteinase domain in insect cells that KGP proteinase activity is not dependent on the presence of the adhesin domains for correct folding. It is likely that the tightly regulated coexpression of proteinase and adhesin functions is present in other soluble or membrane-bound forms of gingipains that have been described and shown to be key factors in contributing to the pathogenicity of the organism.

EXPERIMENTAL PROCEDURES

Bacterial Strains—P. gingivalis strains HG66 and W50 were obtained from Dr. Roland Arnold (Emory University, Atlanta).

Oligonucleotide Synthesis—Oligonucleotide primers for PCR probes and DNA sequencing were synthesized by the phosphoramidite method on an automated DNA synthesizer (Applied Biosystems Model 394), purified by polyacrylamide gel electrophoresis, and desalted on Sep-Pak C18 cartridges (Millipore Corp.) using standard protocols. Primer MK-9-29 was designed to complement the noncoding strand of kgp DNA corresponding to the six amino-terminal residues of the mature protein (DVYTDH) (14). The sequence of this 29-base primer consisted of 17

The abbreviations used are: RGP-1, arginine-specific gingipain-1 (gingipain R1); KGP, lysine-specific gingipain (gingipain K); HGP, high molecular mass gingipain; PCR, polymerase chain reaction; kb, kilobase(s).
Molecular Cloning of Lys-specific Gingipain

kgp-specific bases, a 6-base EcoRI restriction site, and 6 extra bases at the 5’-end (underlined) as follows: 5’-AGATCTGGAATTCGACTGTGAAC/G/C/T/TA/CT/TAC/G/AG/T/GT/TA/CT/A/C/G/T/GA/CT/C/A/GT/AC/GT/C/A/G/GT/AC/G/TA/A/G/CAT-3’. Primer MK-10-29 was designed to complement the coding strand of kgp DNA corresponding to residues 16–21 of the mature protein (MLVYAC) (14). The sequence of this primer, which comprised 17 bases complementary to DNA encoding the 3’-end of the proteinase domain. The 5’-oligonucleotide contained six extra nucleotides, a BamHI site, the Kozak consensus sequence GCC (19), an initiation codon, and the first six amino acids of the KGP proteinase domain. The 3’-oligonucleotide contained six extra nucleotides, a PsI site, a termination codon, and the last six amino acids of the proteinase domain. PCR fragments were cloned into pBlueBac III, and recombinant plasmids were isolated. Two of them were sequenced as described above and used to generate recombinant viruses by in vivo homologous recombination. Recombinant viruses were used to infect Spodoptera frugiperda clone 9 (Sf9) cells. After 48 h, recombinant viruses were collected, identified by PCR, and further purified. Standard protocols for plasmid cloning were used (16). Standard procedures for selection, screening, and propagation of recombinant baculovirus were performed as described by the supplier (Invitrogen).

RESULTS AND DISCUSSION

Cloning and Sequencing of the kgp Gene—Pike et al. (14) have previously determined the primary structure of the NH2 terminus of Lys-specific gingipain by direct amino acid sequencing. This information was used to prepare a mixture of synthetic oligonucleotides complementary to amino acids 1–6 and 16–21 of the mature protein. These primers were used to amplify, by PCR, kgp gene sequence from P. gingivalis DNA. A single 76-base pair product (P76) was identified, cloned, and sequenced to determine codon usage for the amino-terminal residues of KGP. On the basis of this sequence, a 32P-labeled oligonucleotide probe (Lys-1-33) corresponding to the coding strand of this partial kgp DNA was synthesized and used to screen the ADASH P. gingivalis DNA library. Incorporated radiolabeled nucleotides were separated from unincorporated nucleotides on a Sephadex G-25 column (Boehringer Mannheim).

Genomic DNA Library Synthesis—The P. gingivalis strain HG66 ADASH and P. gingivalis strain W50 ZAP DNA libraries used here have been described previously (13).

Screening of Genomic Libraries—Approximately 2 × 109 phage were grown on 5 × 150-mm plates, lifted in duplicate onto supported nitrocellulose transfer membranes (BAS-N, Schleicher & Schuell), and hybridized with the 32P-labeled Lys-1-33 probe. Hybridizations were performed overnight at 42 °C in 2 × Denhardt’s solution, 6 × SSC (15 mM NaCl and 150 mM NaCitrate), 0.5% (w/v) SDS, and 500 mg/ml fish sperm DNA. Filters were washed in 2 × SSC containing 0.5% (w/v) SDS at 48 °C. The DNA from positive plaques was purified and subjected to Southern analysis (see below). A 3.8-kb BamHI fragment and 3.4-kb PsI fragment were identified, excised, and cloned into pBluescript SK(−). The 3.4-kb PsI fragment and a 0.9-kb PsI/BamHI-generated 3’-end fragment of the 3.8-kb BamHI fragment were cloned into M13mp18 and M13mp19 vectors and sequenced. Standard protocols for cDNA library screening, λ phage purification, agarose gel electrophoresis, and plasmid cloning were employed (16). To clone the 3’-end of the kgp gene, PsI/BamHI- and HindIII-digested DNAs were size-selected on 1% agarose, and the regions at 0.2 and 4.5 kilobase pairs, respectively, were cloned into pBluescript SK(−). Positive clones were identified by probe hybridization, and smaller fragments were subcloned into M13 for DNA sequencing (see Fig. 1B).

Southern Blot Analysis—Membranes were washed as described above. BamHI-, HindIII-, or PsI-digested P. gingivalis genomic DNAs were hybridized with 32P-labeled Lys-1-33.

DNA Sequencing—Double-stranded DNA cloned into pBluescript SK(−) and single-stranded DNA cloned into M13mp18 and M13mp19 vectors were sequenced by the dideoxy terminator method (17) using sequencing kits purchased from U. S. Biochemical Corp. (Sequenase Version 2.0). DNA was sequenced using the M13 universal primer, reverse sequencing primer, and internal primers (see Fig. 1B).

Active-site Titration and Labeling—KGP from P. gingivalis HG66 was purified and titered, and the active site was labeled as described previously (13). Biotinylated KGP (20 nmol) was denatured in guanidine HCl, equilibrated with 2-mercaptoethanol, and 2-mercaptoethylated by the method of Howke and Yuan (18). The sample was desalted by dialysis against 25 mM ammonium bicarbonate, pH 7.8. Polypeptide Chain Fragmentation and Analysis—The derivatized protein (25 nmol) was digested in 25 mM ammonium bicarbonate buffer, pH 7.8, with trypsin at 37 °C for 16 h (1:50 enzyme/substrate weight ratio). Each digest was analyzed as described earlier (13). Biotinylated peptides were analyzed for amino-terminal sequence, amino acid composition, and molecular mass.

Construction of Recombinant Baculovirus—The KGP proteinase domain was cloned into the pBlueBac III vector (Invitrogen). A PCR primer corresponding to DNA encoding the 5’-end of the proteinase domain was used in conjunction with kgp genomic DNA encoding the 3’-end of the proteinase domain. The 5’-oligonucleotide contained six extra nucleotides, a BamHI site, the Kozak consensus sequence GCC (19), an initiation codon, and the first six amino acids of the KGP proteinase domain. The 3’-oligonucleotide contained six extra nucleotides, a PsI site, a termination codon, and the last six amino acids of the proteinase domain. PCR fragments were cloned into pBlueBac III, and recombinant plasmids were isolated. Two of them were sequenced as described above and used to generate recombinant viruses by in vivo homologous recombination. Recombinant viruses were used to infect Spodoptera frugiperda clone 9 (Sf9) cells. After 48 h, recombinant viruses were collected, identified by PCR, and further purified. Standard protocols for plasmid cloning were used (16). Standard procedures for selection, screening, and propagation of recombinant baculovirus were performed as described by the supplier (Invitrogen).

Recombinant KGP Activity Assay—Sf9 cells were infected at a multiplicity of infection of 5 with either wild-type baculovirus or baculovirus engineered to encode the KGP proteinase domain. At 24, 48, and 72 h, cells and media were collected and centrifuged (5000 rpm, 3 min). Supernatants (2 ml) were collected and resuspended in RIPA assay buffer were assayed for lysine-specific amidolytic activity as follows. Aliquots (200 μl) of the supernatant or resuspended pellet were added to 800 μl of assay buffer (0.2 M Tris, 0.1 M NaCl, 10 mM t-cysteine, and 5 mM CaCl2, pH 7.6) and mixed for 1 min. Substrate (S-2251, Chromogenix) was added to a final concentration of 100 μM, and amidolytic activity was measured at 405 nm. The amount of active recombinant KGP proteinase domain expressed was quantitated by comparison with the activity of a fixed amount of KGP purified from P. gingivalis.

RESULTS AND DISCUSSION

Cloning and Sequencing of the kgp Gene—Pike et al. (14) have previously determined the primary structure of the NH2 terminus of Lys-specific gingipain by direct amino acid sequencing. This information was used to prepare a mixture of synthetic oligonucleotides complementary to amino acids 1–6 and 16–21 of the mature protein. These primers were used to amplify, by PCR, kgp gene sequence from P. gingivalis DNA. A single 76-base pair product (P76) was identified, cloned, and sequenced to determine codon usage for the amino-terminal residues of KGP. On the basis of this sequence, a 32P-labeled oligonucleotide probe (Lys-1-33) corresponding to the coding strand of this partial kgp DNA was synthesized and used to screen the ADASH P. gingivalis DNA library.

DNA from positive clones was extracted, purified, and subjected to restriction enzyme analysis. All clones gave an 3.8-kb BamHI fragment and an 3.4-kb PsI fragment. Similar results were obtained by Southern analysis of P. gingivalis total genomic DNA (data not shown). We therefore concentrated on one clone designated A2. The 3.8-kb BamHI and 3.4-kb PsI fragments from clone A2 were cloned into pBlueScript SK(−). The 3.4-kb PsI fragment and a 0.9-kb PsI/BamHI-generated 3’-end fragment of the 3.8-kb BamHI fragment were cloned into M13mp18 and M13mp19 vectors and sequenced. Standard protocols for cDNA library screening, λ phage purification, agarose gel electrophoresis, and plasmid cloning were employed (16). To clone the region of the kgp gene containing the termination codon, overlapping clones were isolated from size-selected PsI/BamHI- and HindIII plasmid libraries using oligonucleotide probes (see “Experimental Procedures”). Using this procedure, several 0.7- and 5.7-kb fragment containing clones were obtained, respectively. In total, 8.1 kb of genomic DNA, from PsI to HindIII sites (Fig. 1B), was isolated and cloned as described.

The composite 7.304-kb PsI/NcoI fragment of this genomic DNA (Fig. 1B) was fully sequenced in both directions and is described here.
The exact ATG codon used for initiation of translation is currently unknown, but the presence of a consensus TATA box (ATAAATT) at nucleotides 635–641 and a 15-amino acid signal peptide sequence immediately following the 5'-most ATG codon suggests it to be the strongest candidate (amino acids 222 to 237; Fig. 1A).

The most striking feature of the deduced protein sequence is the presence of multiple homologous sequences immediately carboxyl-terminal to the proteinase-coding domain (Fig. 1B), leading to a calculated molecular mass of 186.8 kDa for the encoded polyprotein. As described for RGP-1 (13), within these sequences can be found peptides identified by Pike et al. (14) as the components of 95-kDa gingipain R (high molecular mass gingipain (HGP)) that likely confer adhesion activity on the high molecular mass KGP complex. The polyprotein sequence deduced from the gene sequence now allows exact delineation of the primary sequence of the mature KGP proteinase. The amino terminus is derived from proteolytic processing at an arginine residue (Fig. 1B). The carboxyl terminus of the proteinase domain is derived by processing at Arg-509, which also releases the amino terminus of a novel subunit that is a hybrid of the 27-kDa (HGP27) and 44-kDa (HGP44) polypeptide chains described recently in the rgg-1 gene structure (Fig. 1B) (13). Previously, we identified three large repeats of homologous sequence located within three of the HGP cleavage products (13). It is interesting to note that the hybrid HGP27/
HGP44 appears to result from fusion within this repeated region. The molecular mass of this hybrid protein is 44.7 kDa. The first 147 amino acids are identical to HGP27. The remaining 271 amino acids show 99% identity to HGP44, including one amino acid change and two amino acid deletions (98% identity at the nucleotide level). Okamoto et al. (20) have cloned and sequenced RGP-1 from P. gingivalis strain 381. The encoded protein also contains a hybrid of the HGP44/HGP27 molecules that results from fusion within this homologous sequence. It is, however, in the reverse combination to that observed for KGP. The remainder of the processing events are identical to those described for RGP-1. The alignment of each structure is shown in Fig. 1B. Processing at Arg-927 gives rise to HGP15; processing at Lys-1062 gives rise to HGP17 (with two amino acid changes compared with HGP17/RGP-1 (13); and processing at Arg-1220 gives rise to HGP27. The homology to the rgp-1 gene is also evident in the first 120 nucleotides of the 3′-noncoding sequences.

The proteinase domain of KGP comprises 509 amino acids with a calculated molecular mass of 55.9 kDa. This is lower than the 70-kDa molecular mass of Lys-gingivain first described by Scott et al. (21), but corresponds more closely to the 60-kDa molecular mass determined by Pike et al. (14). A comparison of the deduced amino acid sequences of the proteinase domains of RGP-1 and KGP indicates 22% identity using the Macaw alignment program (22). This includes a long stretch of homologous sequences in the carboxyl terminus that incorporates the motif LTATT present in all HGP repeat sequences (Fig. 2, A and B) (13). This sequence includes Asn-463, which aligns with Asn-442 in RGP and is the putative asparagine in the catalytic triad. Similarly, His-216 in KGP and His-211 in RGP aligns with Asn-442 in RGP and is the putative asparagine in the catalytic triad. Similarly, His-216 in KGP and His-211 in RGP-1. This prompted us to re-examine the active site of RGP-1. This methodology, which utilizes a more sensitive reduction and pyridylethylation step, identified Cys-249 as the actual active-site cysteine of RGP-1, and not Cys-185 as reported previously (13). Cys-249 has been shown to be a residue of the active site.
Molecular Cloning of Lys-specific Gingipain

Fibrinogen is found in high concentrations in blood plasma, where it plays roles in blood coagulation and wound-healing processes. Its deposition on foreign bodies and at sites of trauma allow it to serve as a substrate for microbial adhesion. Sequence alignments of the proteinase domains of RGP-1 and KGP and of HGP17 and HGP44 reveal a region of high homology that might function as an adhesin domain (Fig. 2A). The ability of RGP-1, KGP, and other membrane-associated \textit{P. gingivalis} gingipains to adhere to matrix molecules indicates similarity to the fibrinogen-binding microbial surface components recognizing adhesive matrix molecules (25, 26). Recently, two \textit{P. gingivalis} cell-surface cysteine proteases, named porphypains 1 and 2, were purified and shown to bind and degrade fibrinogen (27). These likely correspond to two forms of high molecular mass complexes of gingipains.

The availability of RGP-1 and KGP DNA sequences will allow the further study of erythrocyte hemolysis mediated by the hemagglutinin/adhesin activities of this family of proteolytic polypeptides. In addition, recombinant polypeptides derived from these sequences may also prove useful in the development of potential immunoprophylactic and therapeutic agents against this human pathogen.

Acknowledgments—The technical contributions of D. Wong, T. Rigel, and I. Beletskaya are gratefully acknowledged.

Note Added in Proof—During the review of this manuscript, two papers were published describing the cloning and sequencing of kgp-like genes from \textit{P. gingivalis} strains 381 and W12 (Barkocy-Gallagher, G. A., Hae, N., Patti, J. M., Whitlock, J., Progulske-Fox, A., and Lantz, M. S. (1996) \textit{J. Bacteriol.} 178, 2734–2741; Okamoto, K., Kadowaki, T., Nakayama, K., and Yamamoto, K. (1996) \textit{J. Biochem. (Tokyo)} 120, 398–406).

REFERENCES

1. Johnson, N. W. (ed) (1991) \textit{Markers of Disease Susceptibility and Activity for Periodontal Diseases}, Cambridge University Press, Cambridge, England

2. Holt, S. C., Ebersole, J., Felton, J., Brunsvold, M., and Kornman, K. S. (1988) \textit{Science} 239, 55–57

3. Slots, J., and Listgarten, M. A. (1988) \textit{J. Clin. Periodontol.} 15, 85–93

4. Carlsson, J., Hoflung, J. F., and Sundqvist, K. G. (1984) \textit{J. Med. Microbiol.} 18, 40–46

5. Sundqvist, G., Carlsson, J., Herrmann, B., and Tarnvik, A. (1985) \textit{J. Med. Microbiol.} 19, 85–94

6. Wingrove, J. A., DiScipio, R. G., Chen, Z., Potempa, J., Travis, J., and Hugli, T. E. (1992) \textit{J. Biol. Chem.} 267, 18902–18907

7. Kilian, M. (1988) \textit{Infect. Immun.} 56, 1368–1371

8. Smalley, J. W., Biris, A. J., and Shutttleworth, C. A. (1988) \textit{Arch. Oral Biol.} 33, 323–329

9. Nilsson, T., Carlsson, J., and Sundqvist, G. (1985) \textit{Infect. Immun.} 50, 467–471

10. Carlsson, J., Herrmann, B. F., Hofling, J. E., and Sundqvist, K. G. (1984) \textit{Infect. Immun.} 43, 644–648

11. Shah, H. N., and Gharbia, S. E. (1989) \textit{FEMS Microbiol. Lett.} 25, 213–217

12. Nishikata, M., and Yoshimura, F. (1991) \textit{Biochem. Biophys. Res. Commun.} 178, 336–342

13. Pavloff, N., Potempa, J., Pike, R. N., Peczlakza, V., Kiefer, M. C., Travis, J., and Barr, P. J. (1995) \textit{J. Biol. Chem.} 270, 1007–1010

14. Pike, R., McGraw, W., Potempa, J., and Travis, J. (1995) \textit{J. Biol. Chem.} 269, 406–411

15. Potempa, J., Pike, R., and Travis, J. (1995) \textit{Infect. Immun.} 63, 1176–1182
Molecular Cloning of Lys-specific Gingipain

16. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

17. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* 74, 5463–5467.

18. Howke, D., and Yuan, P. (1987) *Applied Biosystems Protein Sequencer User Bulletin* 28, 1–4, Foster City, CA.

19. Kozak, M. (1984) *Nucleic Acids Res.* 12, 857–872.

20. Okamoto, K., Misumi, Y., Kadowaki, T., Yoneda, M., Yamamoto, K., and Ikehara, Y. (1995) *Arch. Biochem. Biophys.* 316, 917–925.

21. Scott, C. F., Whitaker, E. J., Hammond, B. F., and Colman, R. W. (1993) *J. Biol. Chem.* 268, 7935–7942.

22. Schuler, G. D., Altschul, S. F., and Lipman, D. J. (1991) *Proteins Struct. Funct. Genet.* 9, 180–190.

23. Westerlund, B., and Korhonen, T. K. (1993) *Mol. Microbiol.* 9, 687–694.

24. Sansonetti, P. J. (1993) *Med. Microbiol. Immunol.* 182, 223–232.

25. Patti, J. M., and Hook, M. (1994) *Curr. Opin. Cell Biol.* 6, 752–758.

26. Patti, J. M., Allen, B. L., McCavin, M. J., and Hook, M. (1994) *Annu. Rev. Microbiol.* 48, 585–617.

27. Ciborowski, P., Nishikata, M., Allen, R. D., and Lantz, M. S. (1994) *J. Bacteriol.* 176, 4549–4557.

28. Rawlings, N. D., and Barrett, A. J. (1994) *Methods Enzymol.* 244, 461–486.