Vol. 207, No. 1, 1995 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
February 6, 1995 Pages 424-431

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 HAEMAGGLUTININ DOMAIN

Louis Kirszbaum, Chris Sotiropoulos, Christine Jackson, Steven Cleal, Nada Slakeski and Eric C. Reynolds*

Biochemistry and Molecular Biology Unit, School of Dental Science, 711 Elizabeth Street, Melbourne, Victoria , 3000, Australia

Received December 22, 1994

We have purified from *Porphyromonas gingivalis* W50 a 45 kDa arginine-specific, thiol-activated, EDTA-sensitive endopeptidase, designated prtR. Oligonucleotide probes based on the *N*-terminal amino acid sequence were used to isolate a genomic fragment containing an open reading frame (3654 bp) with the potential to encode a 132 kDa protein including the prtR *N*-terminus. Analysis of this *prtR* gene revealed that the predicted nascent product contains a protease domain followed by a haemagglutinin domain and is post-translationally processed by proteolytic (possibly autolytic) events to produce a 43-54 kDa arginine-specific, thiol protease and a 41-53 kDa haemagglutinin. Comparison of the *prtR* with the *P. gingivalis prtH* gene suggests that the *prtH* gene product also contains protease and haemagglutinin domains but in the reverse order to that in the prtR. An overlapping but shifted reading frame at the 3' end of the *prtR* encodes the 5' region of the *prtH*.

Porphyromonas gingivalis has been implicated as a pathogen in the development of adult periodontitis; a chronic inflammatory disease of the supporting tissues of the teeth that can lead to tooth loss (1,2). A number of virulence factors have been implicated in the pathogenicity of *P. gingivalis*, in particular proteases have received a great deal of attention for their ability to degrade a broad range of host proteins including structural and defence proteins (3). The major proteolytic activities associated with this organism are "trypsin-like" (4) or "collagenase-like" (5). *P. gingivalis* spontaneous mutants with reduced "trypsin-like" activity have been found to be less virulent in animal models (6,7).

Using chloroform extraction of whole cells and FPLC chromatography we have purified from *P. gingivalis* W50 a 45 kDa arginine-specific, thiol-activated, EDTA-sensitive endopeptidase designated prtR (Genbank accession No. L26341). The prtR exists as part of a larger (90 kDa) complex containing a 45 kDa and a 44 kDa protein. The 44 kDa protein has an amino-terminal sequence identical with the 44 kDa haemagglutinin described by Pike *et al.*, (8). The *N*-terminal sequence of the prtR protease has been determined in our laboratory and together with its arginine specificity is similar to gingipain of Chen *et al.*, (9) although these workers reported the protein to be 50 kDa on SDS-PAGE analysis.

A large number of other workers claim to have purified proteases with "trypsin-like" specificity from *P. gingivalis*, these proteases are also thiol proteases and specific for arginine and/or lysine in the P_1 position

*To whom correspondence should be addressed. Fax: 613 341 0339.

0006-291X/95 $5.00

but vary in molecular mass from 44 kDa to 180 kDa (8,9,10,11,12). At present the relationship between most of these protease structures is unclear and whether or not they are processed from a larger product of a single gene or are derived from separate genes is of particular interest.

We have synthesised oligonucleotide probes based on the *N*-terminal sequence of the prtR and in this work we report the molecular characterisation of the gene encoding the prtR protease from *P. gingivalis* W50 and reveal its relationship with other *P. gingivalis* proteins implicated as proteases and haemagglutinins.

Materials and Methods

Strains, media and growth conditions. *Escherichia coli* JM109 (*rec*A1, *sup*E44, *end*A1, *hsd*R17, *gyr*A96, *rel*A1, *thi* Δ (*lac-proAB*)) (13) and *E. coli* LE392 (*sup*E44, *sup*F58, *hsd*R514, *gal*K2, *gal*T22, *met*B1, *trp*R55, *lac*Y1) (14) were grown in LB medium at 37°C. *P. gingivalis* W50 (15) was grown in modified batch culture medium (15) supplemented with 1 μg/ml haemin in an atmosphere of 10% CO_2, 10% H_2 and 80% N_2 at 37°C.

Cloning and DNA manipulation. *E. coli* strains containing plasmids were grown in LB broth supplemented with 100 μg/ml ampicillin. DNA was isolated from *E. coli* by standard techniques (16). Chromosomal DNA was isolated from *P. gingivalis* W50 as described previously (17). Lambda bacteriophage DNA was isolated from lytic culture supernatants using the procedure of Benson and Taylor (18). A genomic library of *P. gingivalis* W50 was constructed in the vector lambda GEM-12 as described (16). Briefly, *P. gingivalis* W50 DNA was partially digested with *Eco*RI and then ligated into *Eco*RI half site arms of λGEM-12. The ligated DNA was packaged using Packagene (Promega, U.S.A.) and used to infect *E. coli* LE392 cells. Plaque lifts and hybridisation was performed as described (16). Genomic DNA recovered from probe-reactive clones was mapped and subcloned into pUC vectors (host, *E. coli* JM109) for subsequent manipulation and sequencing. The degenerate primers used to isolate the gene were derived from the *N*-terminal sequence information of the purified prtR protease. The synthetic oligonucleotides were as follows: sense primer (5'AA(AG)GA(TC)TT(TC)GT(TCAG)GA(TC)TGGAA (AG)AA(TC)CA-3') that codes for the amino acid sequence [KDFVDWKNQ] and antisense primer (5'TC(CT)TT (AGT)AT(AG)TC(TCAG) CC(CT)TC(AG)TA-3') that is complementary to the sequence that encodes the amino acid sequence [YEGDIKD].

DNA and protein sequence analyses. DNA sequence was determined by the dideoxy termination method using [α-^{35}S] dATP (19), with modified T7 polymerase (20) and double-stranded DNA as a template. Sequence data were analysed by the program suites and nucleic and protein databases accessed via the Australian National Genomic Information Service.

Results and Discussion

The use of redundant oligonucleotides based on the *N*-terminal sequence data of the prtR resulted in the isolation of a 13 kb *Eco*RI fragment from a probe-reactive λGEM-12 clone. Once mapped, subclones were generated in pUC vectors for subsequent manipulation. Using a primer extension sequencing strategy the probe reactive region of the genomic fragment from *P. gingivalis* was sequenced (both strands) and 4.6 kb of the nucleotide sequence is presented in Figure 1. This genomic region contains a single complete open reading frame 3654 bases between the 5' stop codon beginning at base 34 and the 3' stop codon beginning at base 3691. The deduced amino acid sequence of the *prtR* gene product encoded by the open reading frame is shown under the nucleotide sequence (Fig. 1) and contains the *N*-terminal sequence determined for the 45 kDa prtR protease (YTPVEEK...) and confirms that this gene encodes the prtR of *P. gingivalis*.

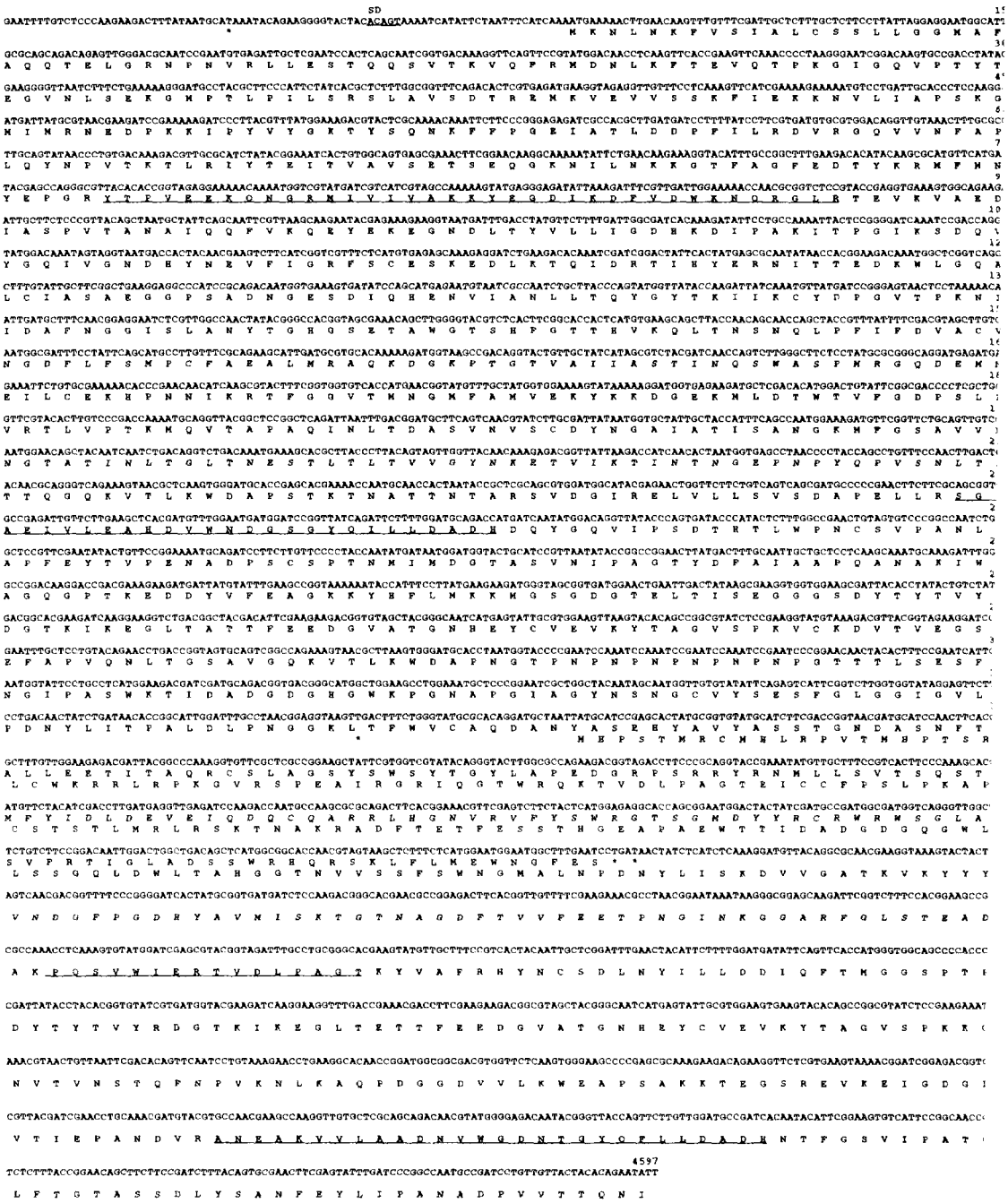

Figure 1. Nucleotide sequence of the cloned protease gene (*prtR*) from *P. gingivalis* W50. The sequence is derived from an *Eco*RI-*Eco*RI fragment of λER1. The deduced amino acid sequence of the open reading frame is indicated below the nucleotide sequence. A putative Shine-Dalgarno (SD) sequence is underlined 25 bp upstream from an initiation codon. The *prtR* stop codon (TGA) is indicated by an asterisk. *N*-terminal sequences previously identified are underlined. The shifted reading frame encoding the prtH (26) is indicated as the second deduced amino acid sequence overlapping the carboxyl end of the *prtR* gene product.

The amino terminus of the 45 kDa endopeptidase is encoded 730 bp downstream of the 5'stop codon (which begins at base 34, Fig. 1). Several possible start codons are present between the 5' stop codon and the *N*-terminus determined for the mature prtR, however a possible Shine-Dalgarno sequence (bases 55-59) was observed 25 bases upstream of the start codon which begins at base 85 (Fig. 1). The putative Shine-Dalgarno sequence is based on sequence complementarity to the 3' end of the 16S rRNA of *P. gingivalis* (21). The 44 kDa haemagglutinin which copurifies with the 45 kDa prtR is also encoded by the *prtR* gene, with its *N*-terminus (SGQAEIV...) located 2206 bp from the 5' stop codon and 1476 bases downstream of the prtR *N*-terminal codon (Fig. 1). The *prtR* gene therefore has the potential to encode a nascent translation product of 132 kDa molecular mass containing two domains with sequence homology to the 45 kDa protease and the 44 kDa haemagglutinin. These proteins therefore appear to be the product of post-translational modification which takes the form of proteolytic (possibly autolytic) cleavage of the larger precursor molecule into the smaller protease and haemagglutinin molecules that remain complexed to each other. The protease has been characterised by other workers including Chen *et al.*, (9) who purified a 50 kDa arginine-specific, thiol protease designated gingipain that has the same *N*-terminal sequence as our 45 kDa arginine-specific protease. The 44 kDa haemagglutinin has been purified and characterised by a number of workers including Pike *et al.*, (8), Nishikata and Yoshimura (22) and Hayashi *et al.*, (23).

At least two proteolytic events are required to produce the two mature proteins encoded by the *prtR* gene (Fig. 1) and the arginyl- and lysyl-specific proteolytic activities associated with *P. gingivalis* could generate many of the molecular size permutations of these molecules which have been reported in the literature (24). Each of the mature *N*-termini is preceded by an arginyl residue and therefore both can be generated by cleavage of the polypeptide by an endopeptidase with a specificity for arginine in the P_1 position ($R^{\downarrow}$YTPVEEK... and $R^{\downarrow}$SGQAEIV...). The two single cleavage events would produce a mature protease with a molecular mass of 54 kDa, extending to the amino terminus of the haemagglutinin which in turn would extend to the end of the translated molecule with a molecular mass of 53 kDa. The 53 kDa predicted for the haemagglutinin is not consistent with the 44 kDa molecular mass observed in our laboratory and by other workers but cleavage at one of the many arginyl or lysyl residues in the carboxyl region of the haemagglutinin domain could account for the 44 kDa molecular mass. At 54 kDa the predicted size of the protease domain is consistent with the 50 kDa gingipain of Chen *et al.*, (9) and also the 50 kDa component of the high-molecular-mass Arg-gingipain which is a complex of several proteins exhibiting both arginine-specific proteolytic and haemagglutination activities purified by Pike *et al.*, (8). However, as with the 53 kDa haemagglutinin, several other arginyl and lysyl residues are present in the carboxyl region of the prtR protease domain and *C*-terminal truncation of the prtR protease could give rise to molecules 43-54 kDa in size being consistent with our size of 45 kDa for the arginine-specific protease It is interesting to speculate that the processing of an inactive precursor molecule into smaller active fragments is part of a mechanism which may act to prevent premature activation within the cell. However it is also possible that the proteolytic processing is part of a mechanism to evade host defenses.

Analysis of the deduced amino acid sequence of the *prtR* revealed two regions of homology consistent with its arginine-specific endopeptidase activity. Firstly a segment of the prtR sequence (beginning at base 1999, Fig. 1) within the arginine-specific, endopeptidase domain exhibited homology with other cysteine proteases in particular around a portion of the active site of papaya papain. Residues identical to the prtR sequence are boxed, Fig. 2 (adapted from Bourgeau *et al.*, (25)) and may well reflect a region of the

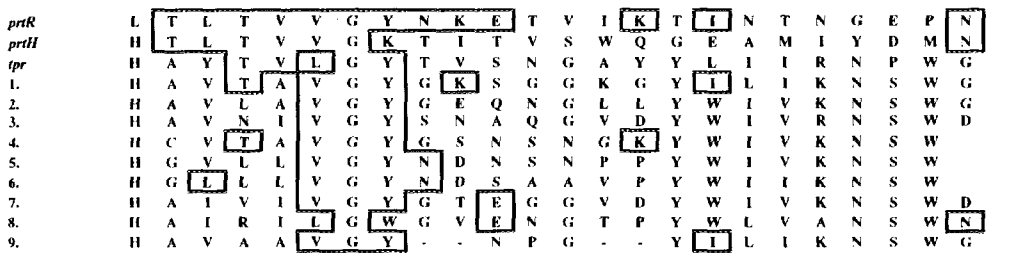

Figure 2. Homology between the prtR protease domain and other cysteine proteases. Residues identical with prtR are boxed. *prtR*, *P. gingivalis* W50 protease; *prtH*, *P. gingivalis* W83 protease (26); *tpr*, *P. gingivalis* W83 protease (25); 1, *Carica papaya*; 2, rat cathepsin; 3, *Dermatophagoides pteronyssinus*; 4, *Entamoeba histolytica*; 5, *Trypanosoma brucei*; 6, *Trypanosoma cruzi*; 7, Chinese gooseberry actinidin; 8, human cathepsin B; 9, papaya papain. This diagram was adapted from Bourgeau *et al.* (25).

molecule critical for the activity of the endopeptidase. Included in this montage are segments of the predicted amino acid sequences from the cloned and expressed *P. gingivalis* protease genes *prtH* (26) and *tpr* (25).

Secondly, a 79 amino acid segment of the prtR sequence (beginning at base 961, Fig. 1) within the arginine-specific endopeptidase domain exhibited a 34% identity with a domain of the catalytically active subunit of the arginine-specific, human carboxypeptidase N (30). The motif GNXHXNE, including the zinc binding residues H and E, of carboxypeptidase N is also found in the prtR sequence (beginning at base 1066, Fig. 1). As the prtR protease was not inhibited by *o*-phenanthroline it is unlikely to be a zinc-binding metalloprotease, but its stabilisation by Ca^{2+} may suggest that this region is important in Ca^{2+} binding. However, its similarity with carboxypeptidase N may also relate to arginine specificity.

A comparison of the *prtR* nucleotide sequence with those of other cloned genes from *P. gingivalis* in particular the cloned protease genes, *prtC* (27), *tpr* (25), *prtT* (28) and *prtH* (26) revealed that the 3' region of the *prtR* genomic fragment (from base 2843, Fig. 1) is identical to the 5'region of the nucleotide sequence containing the *prtH* gene. The *N*-terminal region of the *prtH* gene product is encoded by a shifted reading frame that overlaps the *prtR* near the 3' end of that gene (Fig. 1). The initiation codon that marks the beginning of the open reading frame for the *prtH* occurs 3200 bp 3' of the start of the open reading frame of *prtR* and 455 bp 5' of the *prtR* translation termination codon (Fig. 1). Examination of the predicted gene product of the *prtH* revealed a similar situation to the protein encoded by the *prtR* in that both appear to encode a larger molecule which undergoes proteolytic cleavage to give rise to smaller proteins which have been characterised by *N*-terminal sequence analysis. The *N*-terminal sequence of a 17 kDa protein (PQFTEIFRQVDL...), associated with high molecular mass complexes purified from *P. gingivalis* culture supernatants by Pike *et al.* (8), shows significant homology to a region in the prtH protein. These high molecular mass complexes exhibit haemagglutination activity as well as lysine-specific and arginine-specific proteolytic activity and have been designated Lys-gingipain and high-molecular-mass Arg-gingipain respectively by Pike *et al.* (8). The sequence PQSVWIERTVDLPAGT is encoded by the prtH beginning at base 3907 (Fig. 1) and contains identical residues (underlined) to the sequence determined at the protein level by Pike *et al.* (8). The differences in the protein's *N*-terminal sequence and that of the deduced *prtH* translation product may reflect *P. gingivalis* strain differences and/or incorrect assignments during *N*-terminal sequence analysis. The prtH also shows homology to

another *N*-terminal sequence of a protein found in the large protease-haemagglutinin complexes purified by Pike *et al.* (8). The prtH sequence, ANEAKVV... (underlined, Fig. 1, beginning at base 4382), has been identified as the *N*-terminal sequence of a 44, 30 and 27 kDa component of Lys-gingipain, a 27 kDa component of high-molecular-mass Arg-gingipain (8) and as a 43 kDa proteolytic fragment of 120 kDa and 150 kDa cell-associated proteases, designated porphypain 2 and 1 respectively, with specificity for both arginine and lysine in the P_1 position, purified from *P. gingivalis* by Ciborowski *et al.* (12). It appears therefore that Lys-gingipain and the porphypains are derived, or at least have components derived, from the prtH and that the large protease-haemagglutinin complex designated high-molecular-mass Arg-gingipain purified by Pike *et al.* (8) contains components derived from both the prtR and prtH. It is interesting to speculate that the prtH, like the prtR contains both protease and haemagglutinin domains although in the reverse order with the haemagglutinin domain first in the prtH. Three subdomains, based on sequence homology, can be identified within the prtR haemagglutinin domain and have been designated I, II and III (Fig. 3). These three subdomains are also present in the prtH but in the reverse order, that is, III, II and I, in a putative haemagglutinin domain *N*-terminal of the protease domain identified by Fletcher *et al.*, (26) which is beyond the 3' end of our obtained sequence (Fig. 1). Further, the prtH contains a motif -PTPT- (beginning at base 4040, Fig. 1) within the proposed haemagglutinin

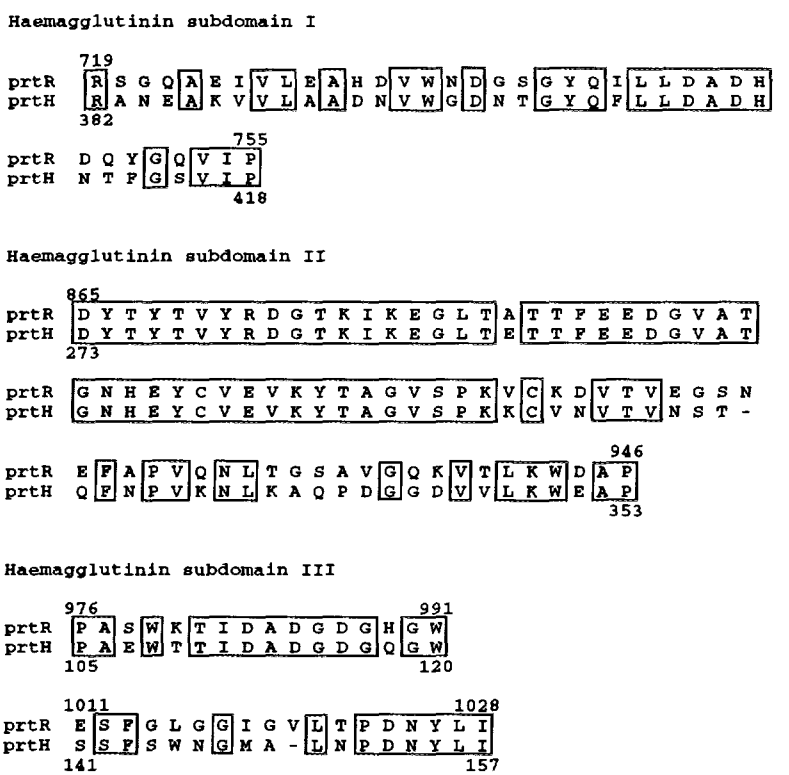

Figure 3. Homology between the prtR haemagglutinin domain and the proposed haemagglutinin domain of the prtH. The haemagglutinin domain of the prtR has been divided into three subdomains I, II and III which occur in the reverse order in the prtH. Identical residues are boxed and amino acid residues are numbered from the initiation codon residue methionine of both gene products (see Fig. 1).

domain similar to the PNPTPTPN motif of the haemagglutinin from *Mycoplasma gallisepticum* (29) which is homologous with the PNPNPNPN motif (beginning at base 2932, Fig. 1) of the haemagglutinin domain of the prtR.

The cloning and sequencing of the *prtR* gene, which encodes a 132 kDa protein that contains both protease and haemagglutinin domains that are post-translationally processed to a 45-50 kDa arginine-specific thiol protease and a 44 kDa haemagglutinin, now confirms that these proteins are derived from a single gene. The close structural relationship of the *prtR* and *prtH* genes and the homology observed between the haemagglutinin domains of their products strongly suggest a common ancestral origin. The presence of these genes in *P. gingivalis* may reflect the importance of the availability of haemin and amino acids/peptides for growth, bioenergetics and virulence of this micro-organism (15). It is possible that the close association of the protease and haemagglutinin activities would be of considerable benefit to the organism in terms of sequential binding and lysis of red blood cells.

Acknowledgments

This work was supported by Australian National Health and Medical Research Council project grant No. 930687 and The Australian Dental Research Fund incorporated.

References

1. Slots, J. (1982) Importance of black-pigmented Bacteroides in human periodontal disease p.27-45. Host parasite interactions in periodontal diseases. American Society for Microbiology.
2. Christersson, L.A., Zambon, J.J., Dunford, R.G., Grossi, S.G., and Genco, R.J. (1989) J. Dent. Res. 68, 1633-1639.
3. Mayrand, D., and Holt, S.C. (1988) Microbiol. Rev. 52, 134-152.
4. Yoshimura, F., Nishikata, M., Suzuki, T., Hoover, I., and Newbrun, E. (1984) Arch. Oral Biol. 29, 559-568.
5. Toda, K., Otsuka, M., Ishikawa, Y., Sato, M., Yamamoto, Y., Nakamura, R. (1984) J. Periodont. Res. 19, 372-381.
6. McKee, A.S., McDermid, A.S., Wait, R., Baskerville, A., and Marsh, P.D. (1988) J. Med. Microbiol. 27, 59-64.
7. Smalley, J.W., Birss, A.J., Kay, H.M., McKee, A.S., and Marsh, P.D. (1989) Oral. Microbiol. Immunol. 4, 178-181.
8. Pike, R., McGraw, W., Potempa, J., and Travis, J. (1994) J. Biol. Chem. 269, 406-411.
9. Chen, Z., Potempa, J., Palanowski, A., Wikstrom, M., and Travis, J. (1992) J. Biol. Chem. 267, 18896-18901.
10. Scott, C.F., Whitaker, E.J., Hammond, B.F., and Colman, R.W. (1993) J. Biol. Chem. 268, 7935-7942.
11. Kadowaki, T., Yoneda, M., Okamoto, K., Maeda, K., and Yamamoto, K. (1994) J. Biol. Chem. 269, 21371-21378.
12. Ciborowski, P., Nishikata, M., Allen, D., and Lantz, M.S. (1994) J. Bacteriol. 176, 4549-4557.
13. Yannisch-Perron, C., Veira, J., and Messing, J. (1985) Gene 33, 103-119.
14. Murray, N.E., Brammar, W.J., and Murray, K. (1977) Mol. Gen. Genet. 150, 53-61.
15. McKee, A.S., and McDermid, A.S., Wait, R., Baskerville, A., Dowsett, A.B., Ellwood, D.C., and Marsh, P.D. (1986) Infect. Immun. 52, 349-355.
16. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
17. Smith, G.L.F., Sockransky, S.S., and Smith, C.M. (1989). Oral. Microbiol. Immunol. 4, 47-51.
18. Benson, S.A., and Taylor, R.K. (1984). Biotechniques 4, 126-127.

19. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 60-564.
20. Tabor, S. and Richardson, C.C. (1987) Proc. Natl. Acad. Sci. U.S.A. 84, 4767-4771.
21. Paster, B.J., Dewhirst, F.E., Olsen, I., and Fraser, G.J. (1994) J. Bacteriol. 176, 725-732.
22. Nishikata, M., and Yoshimura, F. (1991) Biochem. Biophys. Res. Comm. 178, 336-342.
23. Hayashi, H., Nagata, A., Hinode, D., Sato, M., and Nakamura, R. (1992) Oral. Microbiol. Immunol. 7, 204-211.
24. Grenier, D., and Mayrand, D. (1993) In Biology Of The Species *Porphyromonas gingivalis* (H.N. Shah, Ed.), Chapter 13, pp. 227-244. CRC Press, Boca Raton, Florida.
25. Bourgeau, G., Lapointe, H., Peloquin, P., and Mayrand, D. (1992) Infect. Immun. 60, 3186-3192.
26. Fletcher, H.M., Schenkein, H.A., and Macrina, F.L. (1994) Infect. Immun. 62, 4279-4286.
27. Kato, T., Takahashi, N., and Kuramitsu, H.K. (1992) J. Bacteriol. 174, 3889-3895.
28. Otogoto, J.I., and Kuramitsu, H.K. (1993) Infect. Immun. 61, 117-123.
29. Markham, P.F., Glew, M.D.,Whithear, K.G., and Walker, I.D. (1993) Infect. Immun. 61, 903-909.
30. Gebhard, C., Schube, M., and Eulitz, M. (1989) Biochem. 178, 603-607.