Major outer membrane proteins and proteolytic processing of RgpA and Kgp of *Porphyromonas gingivalis* W50

Paul D. VEITH, Gert H. TALBO, Nada SLAKESKI, Stuart G. DASHPER, Caroline MOORE, Rita A. PAOLINI and Eric C. REYNOLDS[1]
School of Dental Science, The University of Melbourne, 711 Elizabeth Street, Melbourne, Victoria, 3000, Australia

Porphyromonas gingivalis is an anaerobic, asaccharolytic Gram-negative rod associated with chronic periodontitis. We have undertaken a proteomic study of the outer membrane of *P. gingivalis* strain W50 using two-dimensional gel electrophoresis and peptide mass fingerprinting. Proteins were identified by reference to the pre-release genomic sequence of *P. gingivalis* available from The Institute for Genomic Research. Out of 39 proteins identified, five were TonB-linked outer membrane receptors, ten others were putative integral outer membrane proteins and four were putative lipoproteins. Pyroglutamate was found to be the N-terminal residue of seven of the proteins, and was predicted to be the N-terminal residue of 13 additional proteins. The RgpA, Kgp and HagA polyproteins were identified as fully processed domains in outer membranes prepared in the presence of proteinase inhibitors. Several domains were found to be C-terminally truncated 16–57 residues upstream from the N-terminus of the following domain, at a residue penultimate to a lysine. This pattern of C-terminal processing was not detected in a W50 strain isogenic mutant lacking the lysine-specific proteinase Kgp. Construction of another W50 isogenic mutant lacking the arginine-specific proteinases indicated that RgpB and/or RgpA were also involved in domain processing. The C-terminal adhesin of RgpA, designated RgpA27, together with RgpB and two newly identified proteins designated P27 and P59 were found to migrate on two-dimensional gels as vertical streaks at a molecular mass 13–42 kDa higher than that calculated from their gene sequences. The electrophoretic behaviour of these proteins, together with their immunoreactivity with a monoclonal antibody that recognizes lipopolysaccharide, is consistent with a modification that could anchor the proteins to the outer membrane.

Key words: lipopolysaccharide, mass spectrometry, peptide mass fingerprinting, two-dimensional PAGE.

INTRODUCTION

Porphyromonas gingivalis is a Gram-negative anaerobic pathogen associated with chronic periodontitis, an inflammatory disease of the supporting tissues of the teeth leading to tooth loss [1]. Considerable attention has been given to the characterization of secreted and surface-associated proteins of this bacterium and examination of their contribution to virulence. In particular, the properties of fimbriae [2] and the cysteine proteinases RgpA, RgpB and Kgp [3] have been examined extensively. Of these, the proteinases have been linked directly to disease pathogenesis through their ability to degrade host protein, induce pro-inflammatory cytokines involved in tissue destruction and alveolar bone resorption, and dysregulate the immune response [3,4]. Several other surface-associated proteins have been described in limited detail, including haemagglutinins, in particular HagA [5], the putative TonB-linked outer membrane receptors (TOMRs) designated Tlr [6], HmuR [7], IhtA [8] and RagA [9], and the putative outer membrane (OM) lipoproteins IhtB [8] and RagB [9]. Some of these proteins have only been identified from the DNA sequence.

The availability of the *P. gingivalis* W83 genomic DNA sequence from the website of The Institute for Genomic Research (TIGR; http://www.tigr.org), together with developments in protein analytical techniques such as two-dimensional (2D)-PAGE and mass spectrometry now enables the rapid and comprehensive analysis of the OM proteome of *P. gingivalis*. *P. gingivalis* contains 2.3 Mb of chromosomal DNA that encodes approximately 2200 predicted proteins (TIGR). From a bioinformatic analysis of the deduced primary protein structures from the TIGR *P. gingivalis* genomic database using FASTA [10] and PSORT [11], over 100 potential candidates for cell-surface location have been identified [12]. This number, although only a small fraction of the proteome, is substantially greater than the number of surface-exposed proteins of *P. gingivalis* that have been identified to date.

Of particular interest in the study of the *P. gingivalis* OM is the mature form of the major virulence factors, RgpA, RgpB and Kgp, and their mode of attachment to the cell surface. RgpA is a polyprotein consisting of an N-terminal Arg-Xaa-specific cysteine proteinase RgpA45 and four C-terminal adhesin domains, RgpA44, RgpA15, RgpA17 and RgpA27 [13]. Similarly, Kgp consists of the Lys-Xaa-specific proteinase Kgp48 and three C-terminal adhesins, Kgp39, Kgp15 and Kgp44 [14]. Each of the Kgp domains exhibits sequence similarity to its equivalent RgpA domain, with the degree of similarity ranging from 28 % sequence identity for Kgp44 and RgpA27 to 100 % sequence identity for RgpA15 and KgpA15. After proteolytic processing of the polyprotein into individual domains, it has been proposed that the domains aggregate via the presence of an adhesin-binding

Abbreviations used: ABM, adhesin-binding motif; BHI, Brain Heart Infusion broth; Cat, chloramphenicol acetyltransferase; DTT, dithiothreitol; HBA, horse blood agar; IPG, immobilized pH gradient; LPS, lipopolysaccharide; mAb, monoclonal antibody; OM, outer membrane(s); Omp, OM protein; PMF, peptide mass fingerprinting; TIGR, The Institute for Genomic Research; TOMR, TonB-linked outer membrane receptor; Tos-Lys-CH_2Cl, tosyl-L-lysylchloromethane; 2D, two-dimensional.

[1] To whom correspondence should be addressed (e-mail e.reynolds@unimelb.edu.au).

motif (ABM) to form non-covalently associated cell-surface complexes [13,15].

RgpB is a second Arg-Xaa-specific cysteine proteinase that has been isolated as a 50 kDa protein from the culture supernatant, and as a 70–80 kDa membrane-associated isoform [15,16]. This proteinase is not found associated with adhesins, and the *rgpB* gene does not encode the ABM that is present in the RgpA45 and Kgp48 catalytic domains [15]. The deduced amino acid sequence of RgpB is identical with that of RgpA45, except for the C-terminal 70 residues of RgpB, which are not found in RgpA45. This C-terminal segment of RgpB shares 51 % sequence identity with the C-terminal 70 residues of the last RgpA domain, RgpA27 [15]. Membrane-associated RgpB migrates on SDS/PAGE as a diffuse 70–80 kDa band, and is reactive with a monoclonal antibody (mAb 1B5) that recognizes lipopolysaccharide (LPS) [17]. This modification is thought to be responsible for the attachment of the protein to the membrane.

Recently, we reported the initial findings of a study of the *P. gingivalis* OM proteome, and described the identification of two abundant OmpA-like OM proteins (Omps), designated Omp40 and Omp41, which were heterodimeric under non-reducing conditions [18]. In the present study, we report the identification of an additional 37 proteins, most of which have not been identified at the protein level before. The complete list includes 15 putative integral Omps, 12 proteins derived from RgpA, RgpB, Kgp and HagA, four putative OM lipoproteins, and four modified proteins that were found to be reactive with mAb 1B5.

MATERIALS AND METHODS

Bacterial strains and batch culture growth conditions

Freeze-dried cultures of *P. gingivalis* W50 (A.T.C.C. 53978) and the mutant W50 strain K1A (*kgp*::*erm*), lacking the Lys-Xaa-specific proteinase Kgp [19], were obtained from the culture collection of the School of Dental Science, The University of Melbourne. Strains W50, K1A and W50AB (see below) were grown anaerobically and maintained (< 10 passages) on horse blood agar (HBA; Oxoid Blood Agar Base 2, Oxoid, Basingstoke, U.K.) supplemented with 10 % (v/v) defibrinated horse blood, with weekly subculturing. Batch cultures of *P. gingivalis* were grown at 37 °C in 3.7 % (w/v) Brain Heart Infusion Broth (BHI) (Oxoid), supplemented with 0.5 g/l cysteine/HCl and 5 mg/l bovine haemin (Sigma, St Louis, MO, U.S.A.). Culture purity was assessed regularly by Gram-staining and colony morphology.

Continuous culture conditions

P. gingivalis W50 was also grown in continuous culture using a model C-30 BioFlo chemostat (New Brunswick Scientific, Edison, NJ, U.S.A.) with a working volume of 365 ml. The culture vessel was continuously gassed with a CO_2/N_2 mixture (1:9). Measurement of the redox potential using an ORP electrode on an HI8424 meter (Hanna Instruments, Padova, Italy) showed that it remained in the −240 to −260 mV range. The temperature was maintained at 37 °C, and pH was maintained at 7.2 ± 0.1. The dilution rate was 0.09/h, which is equivalent to a mean generation time of 7.9 h. The culture medium used was 3.7 % (w/v) BHI, supplemented with 0.5 g/l cysteine/HCl and 0.4 mg/l bovine haemin. The BHI medium and haemin were sterilized by autoclaving, and filter-sterilized cysteine/HCl was then added. The culture was allowed to reach steady-state (at least ten generations) before sampling. The cell density of the chemostat culture was monitored daily by measuring attenuance at 650 nm. Culture purity was assessed regularly by Gram-staining and colony morphology. Chemostat-grown cells to be used for OM preparation were collected from the overflow at 4 °C.

Generation of the *P. gingivalis* W50 *rgpA rgpB* double mutant

Plasmid pGEM-TEasy::*rgpB* containing a PCR insert of *P. gingivalis* W50 *rgpB* promoter and coding region has been described previously [20]. A 2.65 kb *Eco*ICRI fragment encoding the *Bacteroides thetaiotamicron* tetracycline-resistance (Tet^r) gene (*tetQ*) isolated from plasmid pNJR12 [21] was ligated into the *Eco*72I site of pGEM-TEasy::*rgpB*, and then transformed into *Escherichia coli* JM109 cells by heat shock. The *rgpB*::*tetQ* insert from a positive clone was amplified by PCR, and electroporated into *P. gingivalis* W50. The procedure for preparation and transformation of cells was essentially the same as that described by Fletcher et al. [22], except that transformed cells were grown to an attenuance (D_{650}) of 0.1 and selected on HBA containing 1 μg/ml tetracycline by incubation at 37 °C for 7–10 days under anaerobic conditions. The RgpB mutant, designated W50B and generated as a result of homologous recombination, was then used to generate an RgpARgpB double mutant designated W50AB by insertional inactivation of *rgpA* using a chloramphenicol acetyltransferase (Cat) gene as follows. Plasmid pNS3.8 containing a *Bam*HI fragment encoding the 5′ region of *rgpA* in pUC18 [23] was digested with *Kpn*I, treated with T4 DNA polymerase to generate blunt ends and re-ligated to disrupt the *Sma*I site within the polylinker to generate pNS3.8Sma. A promoterless Cat DNA block (0.75 kb) was excised from pKD352 (kindly provided by Dr Koji Nakayama, Department of Microbiology, Faculty of Dentistry, Kyushu University, Japan) using *Hin*dIII. The Cat block was end-filled and inserted (in the forward orientation) into the *Sma*I site within *rgpA* in pNS3.8Sma, transformed into *E. coli* JM109 and selected on Luria Broth agar containing 100 μg/ml ampicillin. As chloramphenicol resistance in *P. gingivalis* requires at least two *cat* genes, plasmid (*rgpA*::*cat*) was subsequently linearized with *Eco*RV, treated with alkaline phosphatase, and a second Cat block was inserted to generate p*rgpA*::$cat_{(2)}$. This construct was verified by DNA sequence analysis. Purified p*rgpA*::$cat_{(2)}$ was linearized using *Eco*RI, electroporated into *P. gingivalis* W50B and selected on HBA containing 1 μg/ml tetracycline and 10 μg/ml chloramphenicol. Insertional inactivation of *rgpA* and *rgpB* in the mutant W50AB was verified by Southern blot analyses, and by assaying whole-cell Arg-Xaa proteolytic activity using a chromogenic substrate, as described previously [20]. For the Southern blot analyses, genomic DNA was purified from *P. gingivalis* W50 and mutant W50AB grown in BHI following the procedure described by Chen and Kuo [24]. The 2.7 kb *Eco*RI pGEM-TEasy::*rgpB* insert, the 2.6 kb *Eco*ICRI *tetQ* and the 0.75 kb *Bam*HI *cat* fragments used as probes were labelled with [α-^{32}P]dATP using the Prime-a-gene® labelling system (Promega, Madison, WI, U.S.A.). Triplicate Southern blots of *Hin*dIII-digested W50 and W50AB genomic DNA were hybridized with the probes at 65 °C, as described previously [23], and were washed extensively in a solution of 0.2 × SSC (where 1 × SSC is 150 mM NaCl/15 mM sodium citrate, pH 8.0) containing 0.01 % (w/v) SDS at 65 °C, before exposure to X-ray film at room temperature for 30 min.

OM preparation

P. gingivalis OMs were prepared from batch-grown cells by the sarkosyl method [25] in the presence of the proteinase inhibitor tosyl-L-lysylchloromethane (Tos-Lys-CH_2Cl; 'TLCK') as described previously [18]. Cells from continuous culture were obtained from the overflow for 1 h at 4 °C, and treated im-

mediately with 5 mM Tos-Lys-CH_2Cl. The OM was prepared as for batch-grown cells, with the exception that the Tos-Lys-CH_2Cl-treated cells were sonicated without prior harvesting of cells by centrifugation.

2D-PAGE

2D-PAGE was conducted as described previously [18], with minor modifications. Each *P. gingivalis* OM preparation was solubilized in 500 μl of immobilized pH gradient (IPG) sample solution [7 M urea/2 M thiourea/4 % (w/v) CHAPS/0.5 % (w/v) Triton X-100/50 mM dithiothreitol (DTT)/0.4 % (v/v) carrier ampholytes (Pharmalytes 3–10)/40 mM Tris/5 mM EDTA/0.005 % Bromophenol Blue] and applied to 17-cm-long, pH 3–10 linear IPG strips (Bio-Rad, Hercules, CA, U.S.A.) using the in-gel sample rehydration technique [26], before focusing for approx. 70 kVh. IPG strips were prepared for the second dimension by reduction with DTT and alkylation with iodoacetamide in equilibration solution [50 mM Tris/HCl (pH 8.8) 6 M urea/30 % (w/v) glycerol/2 % (w/v) SDS] [27]. SDS/PAGE was conducted using the wide-format Protean II xl system (Bio-Rad) with 1 mm spacers. Gels were stained with Coomassie Brilliant Blue G250 [28]. At least three 2D gels were prepared from each different OM sample to ensure that the pattern of spots was reproducible.

Electrophoretic transfer on to PVDF membranes and N-terminal sequencing

Immediately after the second dimension of the 2D procedure, SDS gels were soaked in buffer [10 % (v/v) methanol/10 mM 3-(cyclohexylamino)propane-1-sulphonic acid ('Caps') (pH 11)] for 5 min and electroblotted on to a PVDF membrane pre-wetted with methanol. Electroblotting was conducted for 2 h at 60 V using a Transblott Cell (Bio-Rad) filled with buffer. Blots were stained with 0.2 % (w/v) Coomassie Blue R250 in 30 % (v/v) ethanol/aq. 0.5 % (v/v) acetic acid for 5 min, destained in aq. 50 % (v/v) methanol, and selected spots were then subjected to N-terminal Edman sequencing using a Hewlett–Packard G1000A protein sequencer.

Western blotting

Western blots were performed according to Dashper et al. [29]. The primary antibody mAb 1B5 [17] was used at a dilution of 1:40, and goat anti-mouse horseradish peroxidase conjugate was used as a secondary antibody at a dilution of 1:500.

Peptide mass fingerprinting (PMF)

In-gel digestion and the subsequent extraction of peptides were performed as described previously [18]. Peptide mass analyses were performed using a Voyager DE™ matrix-assisted laser-desorption ionization–time-of-flight ('MALDI-TOF') mass spectrometer (Applied Biosystems, Foster City, CA, U.S.A.) in linear mode using the nitrocellulose method [18].

Protein identification

Preliminary DNA sequence of the *P. gingivalis* W83 genome was obtained from TIGR's website at http://www.tigr.org. The DNA sequence was translated to protein in all six reading frames, and a database of more than 128000 putative proteins was produced. Proteins were identified from peptide mass data by using the program General Protein Mass Analysis for Windows ('GPMAW'; Lighthouse data, Odense, Denmark). After a protein was identified, the observed peptide masses were standardized by using linear regression against the calculated mass values. Using the standardized values of the observed masses, the number of peptides matching to within 0.05 % of the calculated values was recorded. More than 95 % of the matches were within an error of 0.03 %.

Bioinformatic analyses

Database searches were conducted using the program BLASTP [30]. The subcellular location of proteins was predicted from their amino acid sequences using the program PSORT [11]. Multiple alignments of amino acid sequences were constructed with the aid of the program CLUSTAL W [31].

RESULTS

2D-PAGE and identification of *P. gingivalis* Omps

OM preparations from haem-limited chemostat-grown cultures were subjected to 2D-PAGE, and the resultant 2D gels were stained with Coomassie Blue G250. The pattern of spots obtained (Figure 1) was reproducible, and included proteins represented by single spots and many proteins that were represented by a train of spots, reflecting pI heterogeneity. More than 30 spots were excised from a gel (Figure 1), incubated with trypsin and identified using PMF (Tables 1 and 2). All proteins with a molecular mass above 30 kDa were identified by matching at least ten peptides to within a mass error of 0.05 %. Several smaller proteins were identified by matching fewer peptides; however, in these cases, nearly all of the expected peptides were observed. Processed domains derived from the RgpA, Kgp and HagA polyproteins were identified by both PMF and N-terminal sequence analysis of PVDF blots of the 2D gels (Table 2). The N-terminal sequence of P27 (Table 1) was ^{38}EEATIQVXGELA^{49}. With the exception of RgpB, RgpA27, P59 and P27, the calculated pI (not shown) and mass of the predicted mature protein were similar to that observed on the gels (Tables 1 and 2). Most of the proteins identified in Table 1 have not been characterized functionally, and were therefore designated Pxx, where 'xx' corresponds to the calculated molecular mass in kDa.

The predicted amino acid sequence of each identified protein was examined to assess the existence and type of signal sequence. The probability of the protein being located in the OM was assessed further using the program PSORT, and by searching for proteins exhibiting sequence similarity using the BLAST algorithm. These analyses resulted in nearly all of the proteins identified in the 2D gel (Figure 1) being predicted to be associated with the OM or underlying cell wall.

In a previous study [18], the signal peptides of Omp40 and Omp41 were found to be cleaved at Ala–Gln (AQ). The mature N-terminal residue of these proteins was found to be pyroglutamate. Similarly, in this study, of the 23 identified proteins that contained a signal sequence with a putative signal peptidase I cleavage site (Figure 2, groups A, B and C), 20 were predicted to be cleaved at an identical 'AQ' cleavage site (Figure 2, groups A and B). Pyroglutamate was confirmed to be the mature N-terminal residue of five of these proteins by PMF analyses. These five, together with Omp40 and Omp41, are listed in group A of Figure 2, to separate them from the proteins listed in group B, which have not yet been confirmed experimentally to contain N-terminal pyroglutamate. Only three proteins, P92, Tlr and P64, were found to be lacking a suitable AQ cleavage site (group C). The signal sequences of another four proteins (RagB, IhtB, P15 and P61) were predicted to be of the lipoprotein type, since they conformed with the consensus sequence (Prosite accession no. PDOC00013). Interestingly, all four proteins were found to

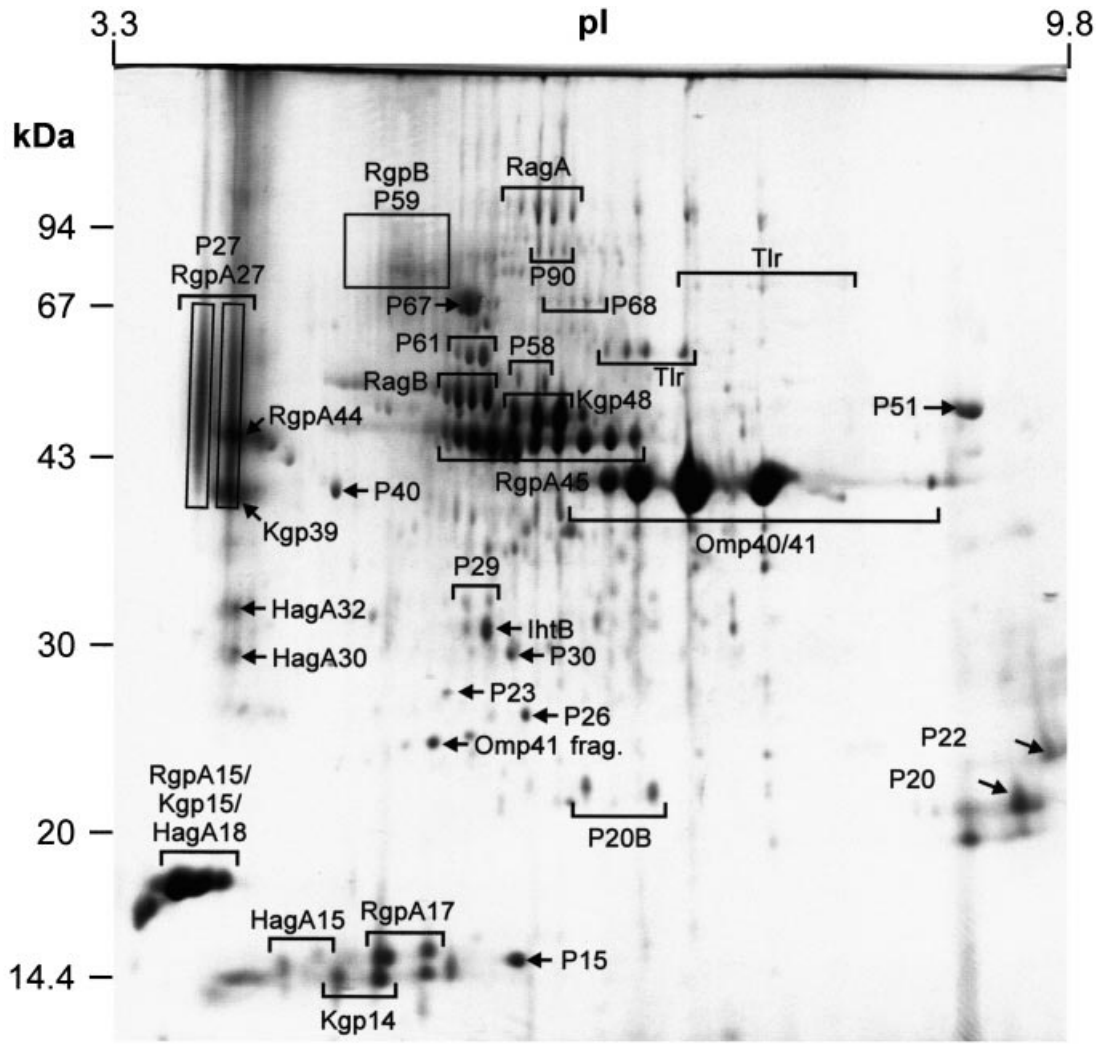

Figure 1 Coomassie-Blue-stained 2D gel of *P. gingivalis* OM

OMs were prepared from haem-limited chemostat-grown cells, and an amount corresponding to 24 mg of the cells (dry weight) was subjected to 2D-PAGE and stained with Coomassie Blue G250. Spots or trains of spots were excised from the gel and identified by PMF (see Table 1).

contain a serine residue immediately N-terminal to their predicted cleavage site (Figure 2, group D). Only four proteins were found to be lacking a signal peptide: P20B, P29, P67 and P68 (Table 1).

Proteolytic processing of RgpA, Kgp and HagA

RgpA and Kgp were identified from 2D gels only as fully processed domains (Figure 1, Table 2). As the proteolytic processing sites of RgpA and Kgp had previously been determined by N-terminal sequencing, each of the domains was thoroughly analysed by PMF in order to locate their respective C-termini. In each spectra derived from the three most N-terminal domains of RgpA (RgpA45, RgpA44 and RgpA15) and Kgp (Kgp48, Kgp39 and Kgp15), a peptide was observed that could not be matched to a tryptic digest of the parent domain. These peptides were therefore identified by matching their mass to the calculated mass of a non-tryptic peptide. The best match for each observed peptide mass was a peptide that exhibited an N-terminal tryptic cleavage site, and a C-terminal non-tryptic cleavage site at a residue penultimate to a lysine (Table 2). The mass accuracy of these matches was well within error, at less than 0.02 %. Because these peptides were the most C-terminal peptides observed of the domain being analysed, the C-terminal residue of each of these peptides was concluded to be the C-terminus of the processed domain. The site of C-terminal processing of each domain was 16–57 residues upstream from the N-terminus of the following domain (Figure 3). In addition to the processing sites determined experimentally, putative C-terminal processing sites were identified for RgpA17, Kgp14 and the putative domain Kgp13 on the basis of similarity to the known sites (Table 2).

Three HagA domains were identified from the gel shown in Figure 1, and these were designated HagA15, HagA30 and HagA32 (Table 2). HagA18 could not be differentiated from RgpA15 and Kgp15, since their amino acid sequences are 99 % identical. HagA32, RgpA17 and Kgp44, which exhibit identical

Table 1 Identification data for all proteins except RgpA, Kgp and HagA

The theoretical mass of each protein was calculated from its predicted mature form. PSORT predicts proteins to be located in the periplasmic space (PS), the OM, the inner membrane (IM) or the cytoplasm (CP). The accession numbers containing the prefix 'PG' are gene identification numbers from the Oral Pathogen Sequence Databases obtained from http://www.oralgen.lanl.gov. ID, identity; res., residues.

Protein	Observed mass (kDa)	Calculated mass (kDa)	Peptides matched/expected§	Total peptides matched ‖	PSORT prediction	Description (reference)	Accession no.
RgpB	71–98	55.6	13/17	18	0.83 PS	Proteinase [15]	AF007124
RagA	98	110	30/38	34	0.94 OM	TOMR [9]	AJ130872
RagB	52	54.4	19/19	46	0.64 PS	OM lipoprotein [9]	AJ130872
Tlr	76	76.7	16/34	18	0.94 OM	TOMR [6]	AF155223
IhtB	31	30.7	12/12	21	0.93 PS	OM lipoprotein [8]	AAF03904
Omp40	43	40.4	16/16	40	0.81 OM	Putative porin [18]	PG0626
Omp41	43	41.3	16/16	40	0.62 OM	Putative porin [18]	PG0627
P15*	15	15.6	4/4	16	0.93 PS	OM lipoprotein	PG1177
P61*	57	60.5	20/23	42	0.93 PS	OM lipoprotein	PG0916
P90*	86	90.4	16/34	24	0.94 OM	TOMR	PG1752
P92*†	91	92	33/36	38	0.94 OM	TOMR	PG0637
P93*‡	86	92.6	16/37	16	0.94 OM	TOMR	PG1242
P49*†	49	48.7	16/19	18	0.84 PS	TolC, 20% ID, 350 res.	PG1455
P27*	43–64	26.9	4/8	6	0.54 PS	Surface protein	PG1570
P59*	71–98	58.4	15/21	27	0.85 OM	Surface protein	PG1838
P20*	21	19.9	9/9	16	0.93 OM	Integral Omp	PG1592
P22*	24	22.0	8/9	20	0.64 OM	Integral Omp	PG1840
P23*	27	23.0	5/7	6	0.85 OM	Integral Omp	PG0551
P26*	26	25.6	8/9	10	0.93 OM	Integral Omp	PG0836
P30*	30	30.2	12/13	22	0.93 OM	Integral Omp	PG1562
P40*	43	40.2	13/17	16	0.93 OM	Integral Omp	PG0023
P64*†	57	64.1	18/23	20	0.82 OM	Integral Omp	PG1442
P51*	50	51.4	17/21	27	0.93 PS	Contains OmpA motif	PG1893
P58*	54	57.5	16/20	17	0.22 IM	–	PG1424
P20B*	22	20.1	8/8	10	0.15 IM	Oxidoreductase, 31% ID	PG1581
P29*	33	28.5	11/11	19	0.35 CP	Electron transfer flavoprotein, 60% ID	PG0959
P67*	68	67.5	20/21	32	0.22 CP	Ribosomal protein, 44% ID	PG1140
P68*	69	67.5	13/23	14	0.20 IM	Oxidoreductase, 30% ID	PG0393

* Proteins designated Pxx, where 'xx' is the calculated molecular mass.

† P49, P64 and P92 were identified from a 2D gel of an OM preparation derived from batch-grown cells in the stationary phase.

‡ P93 was identified from a 2D gel similar to that in Figure 1, except that a greater amount of sample was loaded.

§ The number of fully cleaved peptides accounted for within a mass range of 1000–5000 Da relative to the total number of fully cleaved peptides in the range 1000–5000 Da in the theoretical tryptic digest.

‖ Total number of peptides matched, including partially digested fragments.

N-terminal sequences, were differentiated by the presence of unique peptides in the PMF data. Inspection of the amino acid sequences of the HagA domains also revealed the presence of putative C-terminal processing sites (Table 2). The identification of these domains and putative processing sites allowed us to propose a domain structure for HagA (Figure 3). Comparison of the domain structures of RgpB, RgpA, Kgp and HagA with sequence blocks that are conserved across the four proteins (Figure 3) revealed that the C-terminal processing sites were either located immediately C-terminal to the ABM or were approx. 115 residues into the conserved sequence designated CS2 (Figure 3).

The amino acid residue immediately after each C-terminal processing site was lysine, suggesting the involvement of the lysine-specific proteinase (Kgp) in C-terminal processing. The residue preceding the N-terminus of most of the domains is arginine, suggesting the involvement of the arginine-specific proteinases, RgpA and RgpB, in N-terminal processing. To investigate the involvement of these enzymes in N- and C-terminal processing, OMs were prepared from mutant strains lacking either a functional Kgp (K1A) [19] or both RgpA and RgpB in a double mutant (W50AB) that was constructed by insertional inactivation for this study. The correct insertion of the antibiotic-resistance cassettes into the *rgpA* and *rgpB* genes in W50AB was confirmed by Southern blot analysis (results not shown).

The OM preparations from mutants K1A and W50AB were analysed by 2D-PAGE, PMF and N-terminal sequence analysis. The spots corresponding to the RgpA domains were absent from the 2D gel of the W50AB OM preparation (results not shown). The 71–98 kDa vertical streaks containing RgpB and P59 in the wild-type strain were found to contain only P59 in W50AB. The results therefore confirm the insertional inactivation of *rgpA* and *rgpB* in the double mutant. Furthermore, this mutant did not display whole-cell Arg-Xaa-specific proteolytic activity, being consistent with the lack of both RgpA and RgpB. PMF analyses of Kgp48, Kgp39 and Kgp15 from the W50AB mutant revealed that the C-terminal processing of these domains from the Kgp polyprotein was unchanged. Edman sequencing, however, revealed that the N-termini of these same domains were extended and often ragged (Figure 4), confirming that RgpA and/or RgpB are involved in the N-terminal processing of these domains, and demonstrating that there are multiple processing sites in the interdomain segments.

Analysis of OM preparations of the Kgp mutant K1A by 2D-PAGE revealed that all Kgp domains were absent (results not shown), confirming the previous characterization of this mutant

Table 2 Identification data for processed domains of RgpA, Kgp and HagA

The identified or proposed processed domains of RgpA, Kgp and HagA, as presented diagrammatically in Figure 3, are listed together with their observed and calculated molecular masses (Mass obs. and Mass calc. respectively), the number of peptides matched by PMF analysis, and their observed or predicted N- and C-terminal sequences. The proteins are arranged in two groups according to the similarity of their observed or predicted C-terminal peptides. The amino acid sequences of RgpA and Kgp were obtained from Slakeski et al. [23] (Genbank® accession no. AAC18876) and Slakeski et al [14] (Genbank accession U75366) respectively. The sequence of HagA was obtained from the Oral Pathogen Sequence Databases at http://www.oralgen.lanl.gov, where it has the gene identification number PG1602.

Processed domain (residues numbered from initial methionine)	Mass obs. (kDa)/Mass calc. (kDa)	Peptides matched/expected**	Total peptides matched††	N-terminal sequence§	Identified or proposed C-terminal peptide		
					Sequence§	Calc. mass (mean; Da)	Measured mass (Da)
RgpA45 (228–688)*	47/51	12/14	19	R-YTPVEEKQ	K-TIN…VTLKWDAPST-K	3773.1	3773.0
Kgp48 (229–710)*	50/53	13/14	20	R-DVYTDHGD	K-QIQ…VTLKWEAPSA-K	3655.1	3655.3
RgpA17 (1274–1404)*	15/15	4/5	6	K-PQSVWIER	K-AQP…VVLKWEAPSA-K	‡	‡
Kgp14 (1292–1424)†	14/15	4/4	5	K-PQSVWIER	K-CVN…AILKWNAPAS-K	‡	‡
HagA15 (1724–1856)†	15/15	4/4	4	K-PQSVWIER	K-ECV…AILKWNAPAS-K	‡	‡
RgpA15 (1139–1257)*							
Kgp15 (1157–1275)*							
HagA18 (685–803), (1137–1255), (1589–1707)†	18/13	2/2	4	R-ADFTETFE	K-TGT…VFEETPNGIN-K	2083.2	2083.5
RgpA44 (720–1081)*	46/38	7/7	22	R-SGQAEIVL	K-LTF…NALLEETITA-K	4617.9	4617.8
Kgp39 (738–1099)*	42/38	7/7	17	R-ANEAKVVL	K-LTF…NALLEETITA-K	4617.9	4617.7
HagA30 (366–625)†	30/28	6/6	7	K-APAPYQER	K-TID…NALLEEVLTA-K	‡	‡
HagA32 (820–1077), (1272–1529)†	32/28	5/5	6	K-PQSVWIER	K-TID…NALLEEVLTA-K	‡	‡
Kgp13 (1427–1546)‖							
HagA13 (1859–1978)‖	–/13	–	–	R-AEVLNED¶	K-TID…NALLEEVLTA-K	‡	‡
HagA35 (26–351)‖	–/35	–	–	ZGGPKTAPS¶	K-LSY…KLLEETLGSG-K	‡	‡
RgpA27 (1432–1706)*	(43–64)/30	5/8	7	R-ANEAKVVL	–	–	–
Kgp20 (1548–1732)‖							
HagA20 (1980–2164)‖	–/20	–	–	K-TVVTAPE¶	–	–	–

* Nomenclature taken from Bhogal et al. [13] and O'Brien-Simpson et al. [20].
† Nomenclature based on the name of protein and molecular mass observed.
‡ C-terminal processing site predicted by sequence similarity to known processing site shown above.
§ Hyphen indicates the cleavage site.
‖ Domain proposed from presence of predicted processing site.
¶ Proposed N-terminus.
** The number of fully cleaved peptides accounted for within a mass range of 1000–5000 Da relative to the total number of fully cleaved peptides in the range 1000–5000 Da.
†† Total number of peptides matched including partially digested fragments.

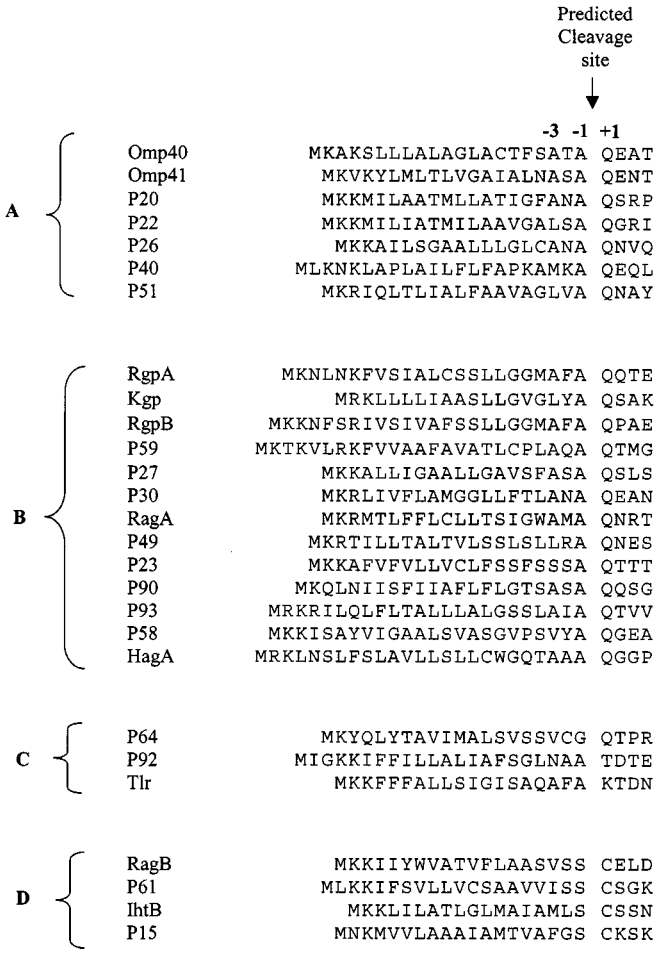

Figure 2 Alignment of signal peptides showing predicted cleavage sites

The predicted amino acid sequences of all identified proteins were inspected for the presence and type of signal peptide using established rules [44]. Group A: signal peptides that contain an AQ cleavage site, which was verified by the identification of a pyroglutamyl N-terminal peptide; Group B, signal peptides that contain a predicted AQ cleavage site; Group C, signal peptides that were predicted to be cleaved at sites other than AQ; Group D, signal peptides predicted to be of the lipoprotein type.

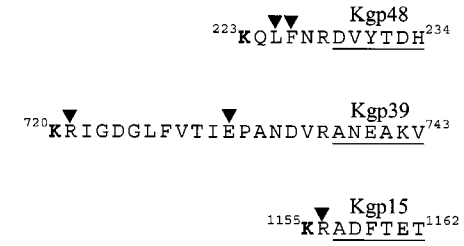

Figure 4 N-terminal processing of Kgp domains in the RgpA⁻ RgpB⁻ mutant, W50AB

The W50AB OM preparation was subjected to 2D-PAGE. The gel was blotted on to a PVDF membrane and stained with Coomassie Blue. The spots that corresponded to Kgp48, Kgp39 and Kgp15 were excised, and subjected to six cycles of Edman sequencing. The triangles mark each of the N-terminal amino acids. The N-terminal sequences obtained for the wild type are underlined.

[20]. The RgpA domains, RgpA45, RgpA44 and RgpA15 from this mutant were analysed by PMF. None of the C-terminal peptides observed in the wild-type W50 (Table 2) were present, and, in RgpA15, the C-terminal peptide measuring 2083.5 Da (Table 2) was replaced by a peptide measuring 2211.8 Da. This mass was best matched to the peptide TGT…VFEETPNGINK (where '…' reflects the intervening amino acid residues), which is identical with the 2083.5 Da peptide except that it contains a C-terminal lysine residue, thus constituting a tryptic peptide.

Identification of post-translationally modified proteins

RgpA27 was identified from two vertical streaks spanning a molecular mass range of 43–64 kDa using PMF analyses and N-terminal sequencing (Figure 1 and Table 2). A second major protein, P27, was also identified from these streaks (Figure 1 and Table 1). When OM preparations from batch-grown *P. gingivalis* were analysed by 2D-PAGE, a series of intensely stained vertical streaks was consistently observed spanning a molecular mass range of 71–98 kDa (results not shown). Interestingly, the streaks were faint in gels derived from the OM of haem-limited chemostat-grown cells (Figure 1, boxed area). PMF analyses of the streaks (of both preparations) revealed the presence of two proteins, RgpB and P59 (Table 1). Despite the close sequence

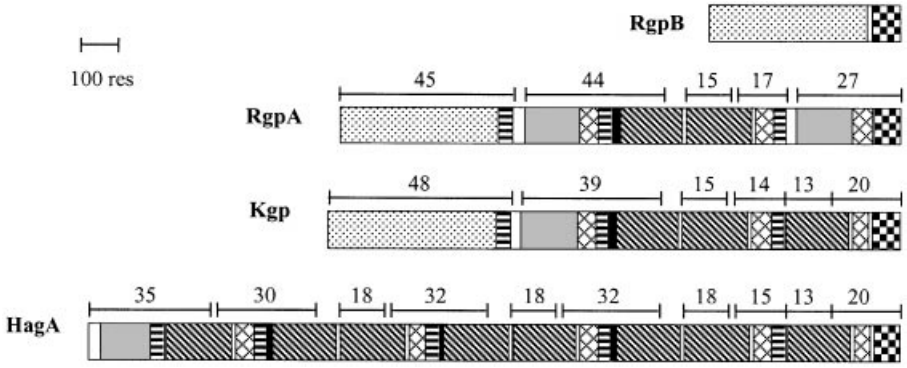

Figure 3 Domain structure of RgpB, RgpA, Kgp

The amino acid sequences of RgpB, RgpA, Kgp and HagA were compared with each other by BLAST analyses in order to identify conserved sequences. These conserved sequences are presented as variously shaded rectangles. The unshaded (white) regions are not well conserved. The locations of the identified or proposed processed domains are presented above the rectangles as solid lines with vertical bars denoting the N- and C-termini. The domains are labelled according to Table 2. ▨, Catalytic domain; ▬, ABM; ▩, CS1; ▧, CS2; ■, CS3; ▨, CS4; ▪, C-terminal sequence. res, residues.

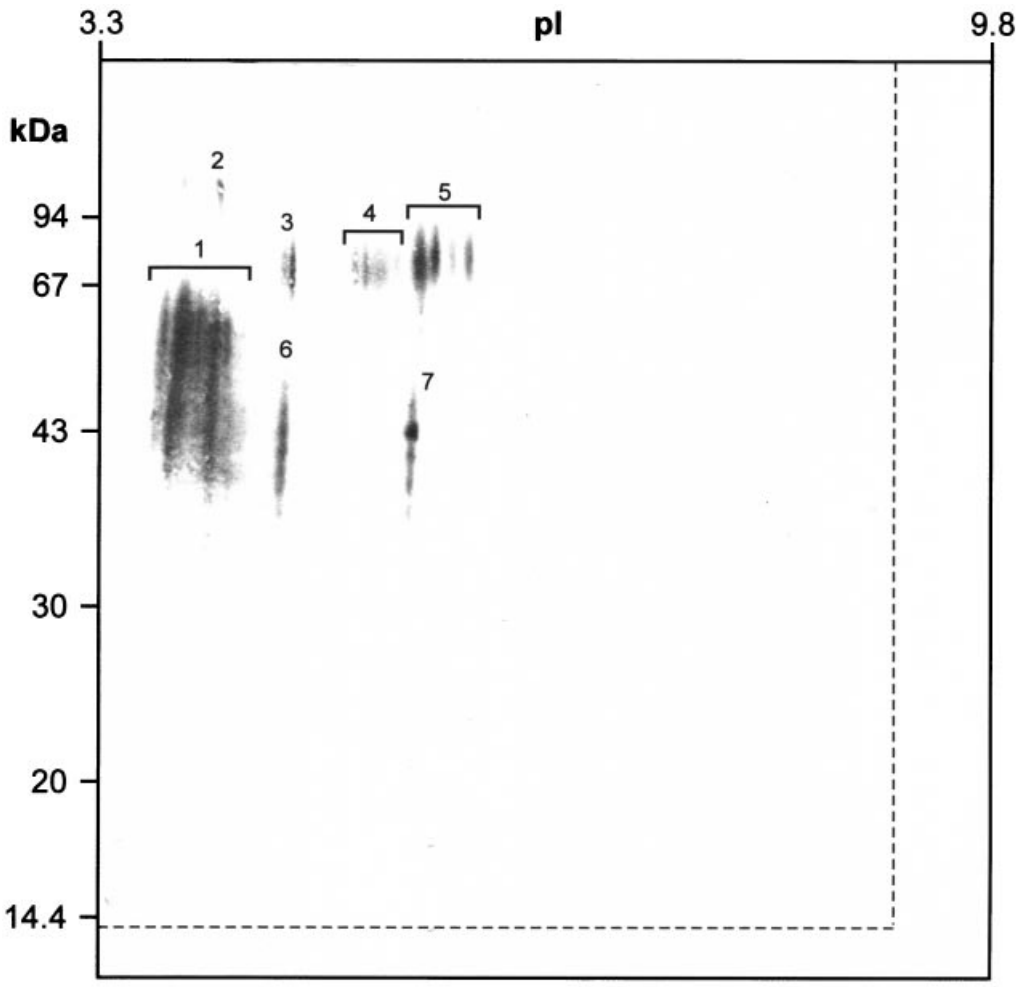

Figure 5 Western blot of a *P. gingivalis* OM 2D gel

A 2D gel of OM was performed exactly the same as for Figure 1, and was Western-blotted using mAb 1B5. An overlaying of the gels demonstrated that blot series #1 (shown in the upper left-hand region of the Figure) coincides with RgpA27/P27, and series #4 coincides with RgpB/P59. The broken lines represent the edges of the blot.

similarity between RgpB and RgpA45, RgpB was distinguished from RgpA45 by the presence of many RgpB-specific peptides and the lack of RgpA45-specific peptides.

The two sets of vertical streaks comprising the four proteins RgpB, P59, RgpA27 and P27 encompassed a molecular mass range that was approximately 13–42 kDa higher than the predicted mass of these proteins, suggesting that they are post-translationally modified. RgpB has been previously shown to react with the antibody mAb 1B5, which recognizes LPS [17]. Western blots of OM preparations were therefore performed from 2D-gels and probed with mAb 1B5 to determine whether RgpA27, P27 and P59 were also reactive with this antibody (Figure 5). The Western blot pattern consisted of seven groups of vertical streaks. Alignment of the blot with the gel shown in Figure 1 revealed that the first group coincided with the vertical streaks containing RgpA27 and P27, whereas the fourth group coincided with RgpB and P59 (Figure 3). The other groups (Figure 5) did not coincide with any of the identified proteins.

The amino acid sequences of the four proteins found to be present in these vertical streaks, together with the sequence-related proteins Kgp and HagA, were compared with each other by BLAST analyses and by multiple alignment in search of a common motif that might be involved in modification. The alignment resulted in the identification of several highly conserved residues in the C-terminal 50-amino-acid residues of each protein. A search of the *P. gingivalis* protein database for additional proteins that share these conserved residues provided a more extensive multiple alignment (Figure 6). The similarity across the 11 sequences included five invariant residues and a further eight highly conserved residues. A consensus sequence was constructed from these residues (Figure 6) and used to perform a Prosite Scan of the Swiss-Prot and TREMBL databases. The only proteins in the databases that were found to contain the consensus sequence were the proteins listed in Figure 6 and related proteins from other *P. gingivalis* strains.

Interestingly, the peptide mass data for the four proteins (RgpB, RgpA27, P27, P59) were similar in that a high proportion of expected peptides were found for the N-terminal segment of each sequence, yet no peptides were found for the C-terminal 70 residues of each protein. For the smaller proteins, RgpA27 and P27, all (five out of five) expected peptides (m/z 1000–7000) were observed for the N-terminal segment of the sequence compared with none out of three and none out of four peptides observed respectively for the C-terminal 70 residues.

```
RgpB      694IFDMNGRRVATA-------KNRM-VFEAQNGVYAVRIAT-EGKTYTEKVIVK736
RgpA     1664IYDMNGRRLAAG-------RNTV-VYTAQGGYYAVMVVV-DGKSYVEKLAIK1706
Kgp      1690IYDMNGRRLAAG-------RNTV-VYTAQGGHYAVMVVV-DGKSYVEKLAVK1732
HagA     2122IYDMNGRRLAAG-------RNTV-VYTAQGGHYAVMVVV-DGKSYVEKLAVK2164
PG1922    251LYDINGRTVAVA-------PNRLE-YMAQTGFYAVRFDVGN-KHHVSKIQVR293
P27       236LFDMNGKVVYTG-------VSET-IAAPQKGMYILRVGA---KSI--KLAI-273
P59       492VYDMQGRVVAEASFSGNEYR--LNVQHLAKGTYILKVVS-DTERFVEKLIVE540
PG1836    506VYDMQGRIVEKVSFSGKEYK--LNVQYLSKGTYMLKVVA-DTEYFVEKIIVE554
PG1572    357MYDLQGKLIRDCALSGDKVE--MGVGSLTKGTYLLKVNT-DQGAFVRKVVIR405
PG0212    773LYDINGRVVLKTTP--NQLRSTVDLSILPEGIYTINIKT-EKSARTEKIHIG821
PG0592    341LFDMNGKLIRATELTGDlAIIGV--ASLPRGTYIAEITAANSKTIRAKVSLR390
            ::*::*: :               .     * * . .   .      *: :
```

CONSENSUS SEQUENCE: [LIVM]-[YF]-D-[LIVM]-[NQ]-G-[RK]-X-[LIV]-X(12,27)-G-X-Y-X(8,15)-K-[LIV]-X-[LIV]

Figure 6 Multiple alignment of proteins containing a C-terminal motif implicated in LPS attachment

The amino acid sequences of RgpA27, RgpB, P27 and P59, together with HagA and several hypothetical *P. gingivalis* proteins, were analysed by the multiple-alignment tool CLUSTAL W. Positions in the sequence that contained identical residues are marked with an asterisk (*). Highly conserved positions are marked with a colon (:). Less highly conserved positions are marked with a period (.). Residues located in identical or highly conserved positions are shown in bold, and were used to construct the consensus sequence. The protein labels containing the prefix 'PG' are gene identification numbers from the Oral Pathogen Sequence Databases obtained from http://www.oralgen.lanl.gov/.

DISCUSSION

P. gingivalis major Omps

In the present study, 39 major proteins derived from 31 distinct genes were identified in *P. gingivalis* OM preparations. Many proteins were represented on 2D gels as a train of spots, reflecting pI heterogeneity. It was not determined whether these multiple-charged forms reflect an important biological property of these proteins, such as post-translational modification, or whether they reflect artefactual modification or conformational equilibria resulting from incomplete denaturation, as we have determined recently for the train of Omp40/Omp41 spots [18]. The proteins identified can be differentially classified according to the nature of their association with the OM. The first group, RgpB, RgpA, Kgp and HagA, have all been described as surface proteins, due to the proteolytic and haemagglutination activity of whole cells [20,32], the presence of these proteinases in OM preparations [20] and their release into the culture supernatant [33]. The second group, RagB, IhtB, P15 and P61, have been classified as lipoproteins, because they exhibit a signal peptide of the lipoprotein type (Prosite accession no. PDOC00013). RagB [9] and IhtB [8] are thought to be anchored to the exterior side of the OM, since they have been classified as receptors and are recognized by sera from patients infected with *P. gingivalis* [34,35]. P61, in contrast, is likely to be on the periplasmic side of the OM, since it contains the OmpA motif associated with the strong binding of OmpA and peptidoglycan-associated lipoprotein ('Pal') to the cell wall [36]. The third group of proteins identified were predicted to be integral Omps. Of these, five (Tlr [6], RagA [9], P90, P92 and P93) were classified as TOMRs, due to their overall sequence similarity to known TOMRs and the presence of the TonB box III motif [9,12,37]. TOMRs contain a large β-barrel domain that is integral to the OM [38]. The TOMRs were successfully predicted by PSORT to be integral Omps, and they each contain C-terminal phenylalanine, which is also typical of integral Omps [39]. P90 was found to share 40 % sequence identity with P92 over their full lengths, suggesting that they may share common functionality. Omp40 and Omp41 share many similarities with OmpA and OprF porins, as discussed previously [18]. P49 exhibited significant sequence similarity (20 % identity, 42 % similarity over 350 residues) to TolC proteins involved in protein export and multidrug efflux, including *E. coli* TolC, which is a trimer of 471-residue-containing protomers. Each protomer contains a small β-sheet OM domain and a large α-helical periplasmic domain [40]. Seven proteins were identified (P20, P22, P23, P26, P30, P40, and P64) that did not share significant sequence similarity with any protein of known function, but exhibited a high PSORT score for OM location. Consistent with being integral Omps, six of these proteins contained at most a single cysteine residue, and P20, P30 and P64 contained a C-terminal phenylalanine. P26 and P40B contained a phenylalanine penultimate to their C-terminal residues. P22 was found to share 33 % sequence identity with P20.

P51 was predicted by PSORT to be located in the periplasm, consistent with it containing the same peptidoglycan-binding motif mentioned above for P61. P58 was predicted by PSORT to be located in the inner membrane, due to the presence of a single transmembrane segment. However, this protein contains C-terminal phenylalanine and no cysteine residues, consistent with OM location. The remaining proteins (P20B, P29, P67 and P68) lacked the presence of a signal peptide and exhibited sequence similarity to proteins that are generally found in the cytoplasm. The signal sequences of 20 identified proteins were predicted to be cleaved at AQ (Figure 2). This site was confirmed for seven of these proteins, where the mature N-terminal residue was pyroglutamate. The high frequency of AQ cleavage suggests that the *P. gingivalis* signal peptidase may prefer this site. Signal peptidase I from *E. coli* also favours alanine in the -1 position, but has no preference for glutamine in the $+1$ position [41]. The *P. gingivalis* lipoprotein signal peptidase may prefer serine in the -1 position, since all four proteins predicted to be N-terminally attached to lipid had serine in this position (Figure 2).

Proteolytic processing of RgpA, Kgp and HagA

In this study, we prepared OMs in the presence of the proteinase inhibitor Tos-Lys-CH_2Cl, and conducted 2D-PAGE in the presence of EDTA, which also inhibits RgpA and RgpB proteinase activity. RgpA, Kgp and HagA, however, were observed on 2D gels as fully processed domains, due to arginine- and lysine-specific proteolytic cleavage of the polyproteins. RgpA was represented by five domains, designated RgpA45, RgpA44, RgpA15, RgpA17 and RgpA27, as described previously [13], and Kgp was represented by four domains, Kgp48, Kgp39, Kgp15, and Kgp14. Kgp14 is an N-terminal fragment of Kgp44 [13,14]. As Kgp14 contains a putative C-terminal processing site that is consistent with the observed molecular-mass data (Table 2), we now suggest that Kgp44 consists of two or three separate domains (Figure 3). Various domains of HagA were also found and designated HagA32, HagA30, HagA15 and HagA18. HagA18 could not be differentiated from RgpA15 and Kgp15. The observed domains of RgpA, Kgp and HagA appear to be the product of complete and specific proteolytic processing of the parent polyproteins. No 2D gel spots were found to contain higher-molecular-mass, unprocessed forms of RgpA, Kgp or HagA, and no 2D gel spots were found to contain ragged N- or C-termini, as only a single N-terminal amino acid sequence was obtained for each domain, and the C-terminal processing observed was also complete. The specificity and completeness of these proteolytic events marks them as natural processing events, rather then artefactual cleavage due to residual proteolytic activity during OM preparation and 2D-PAGE. N-terminal sequencing and PMF analyses of these domains derived from mutants lacking Kgp and both RgpA and RgpB implicated the involvement of RgpA and/or RgpB in N-terminal processing, and Kgp in C-terminal processing. As the observed C-terminal peptides were found to be lacking a lysine residue at their C-terminus, it is possible that an additional processing enzyme with lysine-specific carboxypeptidase activity is involved in processing. We have recently purified such an enzyme from a *P. gingivalis* culture supernatant (Y. Chen, K. Cross, G. Talbo, R. Paolini, J. Fielding, N. Slakeski and E. C. Reynolds, unpublished work).

LPS modification of surface proteins

RgpB has been identified previously to exist in two forms: a 50 kDa protein in the culture supernatant and a 70–80 kDa membrane-associated isoform [15,16]. The 70–80 kDa form, but not the 50 kDa isoform, is immunoreactive with mAb 1B5, which recognizes LPS [17]. It has been proposed [20] that the modification recognized by mAb 1B5 is located in the C-terminal segment of RgpB, since the 50 kDa isoform, which does not react with mAb 1B5, is C-terminally truncated by processing upstream of the last 70 residues [42]. In the present study, seven sets of vertical streaks were observed in 2D gels that were reactive with mAb 1B5. RgpB and P59 were present in one set of streaks; RgpA27 and P27 were present in another. Since the electrophoretic mobility of all four proteins indicated that their molecular masses were 13–42 kDa higher than their calculated molecular masses, and the mobility of P59 and P27 was independent of RgpB and RgpA27, as shown by 2D gels of the W50AB mutant lacking RgpB and RgpA27, then this suggests that the increased molecular mass of all four proteins can be attributed to a similar modification. The PMF analyses of these proteins showed that no peptides were recovered from the C-terminal 70 amino acid residues of each sequence, consistent with the presence of a bulky modification that may prevent tryptic digestion or peptide recovery. These data, together with the data on RgpB published previously [16,42], suggest that RgpB, RgpA27, P27 and P59 are all modified by LPS attachment to the conserved C-terminal segment.

In the present study, the RgpA catalytic domain (RgpA45) was only observed as a series of spots at a molecular mass of approx. 47 kDa, and the intense vertical streaks spanning a molecular mass range of 71–98 kDa were found to contain only RgpB and P59. This is consistent with the recent work of O'Brien-Simpson et al. [20], which analysed W50 and two isogenic mutants lacking either RgpB or RgpA by Western blot analyses with an antibody specific for the catalytic domain of the proteinases. In this analysis, RgpA45 appeared as a 45–47 kDa protein, and RgpB appeared as a 70–80 kDa diffuse band. The 70–80 kDa immunoreactive band was absent in the Western blot of the mutant not expressing *rgpB*. Under the growth conditions employed in this study, the RgpA proteinase and all of its adhesin domains were present in the OM preparations. The abundance of these surface proteins in the OM preparation implies that they are firmly attached to the surface of the OM. The domains of RgpA and Kgp have been shown previously to associate to form large non-covalent complexes, which were released from the surface of cells by means of sonication [13]. More recently, it was shown that an ABM is located in the proteinase and adhesin domains of RgpA and Kgp, but not in RgpB (Figure 3) [15]. The demonstrated specific binding of this ABM to the RgpA–Kgp complexes provided an explanation for the association of the RgpA and Kgp domains, but not RgpB, into non-covalent complexes. The attachment of these complexes to the cell surface has been proposed to be via the C-terminal adhesin domains of RgpA and Kgp, due to the sequence similarity of the last 70 residues with the membrane-associated, LPS-modified RgpB [20]. The identification of RgpA27, the C-terminal adhesin of RgpA, as a diffuse 43–64 kDa streak recognized by an LPS-reactive mAb in this study is consistent with this proposal. LPS may be attached to the protein via its sugar moieties [17], possibly at the conserved Tyr^{1688} residue identified in the consensus sequence (Figure 6), since tyrosine is a known residue modified by sugar in bacterial-surface proteins [43]. This also now provides a mechanism for the localization of the processed HagA domains at the cell surface, because the C-terminal HagA domain also contains the proposed LPS-attachment consensus sequence, and the processed HagA domains contain an ABM, similar to that in the RgpA and Kgp polyproteins [15] (Figure 3).

In conclusion, we have identified many novel OM and surface proteins of *P. gingivalis* by 2D-PAGE and PMF, and identified several new proteolytic processing sites of RgpA and Kgp. The finding that RgpA27 is modified and reactive with an LPS-recognizing mAb suggests a mechanism for the attachment of the RgpA and Kgp complexes to the cell surface.

This project was funded by NHMRC project grant no. 114203. We thank Michael Curtis for providing mAb 1B5 and the *P. gingivalis* W50 mutant KIA. The technical assistance of Mr Peter Riley (N-terminal sequencing) and Mr Chris Owen (photography and Figure preparation) is gratefully acknowledged. Sequencing of *P. gingivalis* by TIGR and the Forsyth Dental Center database was accomplished with support from the National Institute of Dental and Craniofacial Research (NIDCR).

REFERENCES

1 Socransky, S. S., Haffajee, A. D., Cugini, M. A., Smith, C. and Kent, Jr, R. L. (1998) Microbial complexes in subgingival plaque. J. Clin. Periodontol. **25**, 134–144

2 Amano, A., Nakamura, T., Kimura, S., Morisaki, I., Nakagawa, I., Kawabata, S. and Hamada, S. (1999) Molecular interactions of *Porphyromonas gingivalis* fimbriae with host proteins: kinetic analyses based on surface plasmon resonance. Infect. Immun. **67**, 2399–2405

3 Travis, J., Pike, R., Imamura, T. and Potempa, J. (1997) *Porphyromonas gingivalis* proteinases as virulence factors in the development of periodontitis. J. Periodontal Res. **32**, 120–125

4 Curtis, M. A., Kurimitsu, H. K., Lantz, M., Macrina, F. L., Nakayama, K., Potempa, J., Reynolds, E. C. and Aduse-Opoku, J. (1999) Molecular genetics and nomenclature of proteases of *Porphyromonas gingivalis*. J. Periodontal Res. **34**, 464–472

5 Han, N., Whitlock, J. and Progulske-Fox, A. (1996) The hemagglutinin gene (*hagA*) of *Porphyromonas gingivalis* 381 contains four large contiguous, direct repeats. Infect. Immun. **64**, 4000–4007

6 Slakeski, N., Dashper, S. G., Cook, P., Poon, C., Moore, C. and Reynolds, E. C. (2000) A *Porphyromonas gingivalis* genetic locus encoding a heme transport system. Oral Microbiol. Immunol. **15**, 388–391

7 Simpson, W., Olczak, T. and Genco, C. A. (2000) Characterization and expression of HmuR, a TonB-dependent hemoglobin receptor of *Porphyromonas gingivalis*. J. Bacteriol. **182**, 5737–5748

8 Dashper, S. G., Hendtlass, A., Slakeski, N., Jackson, C., Cross, K. J., Brownfield, L., Hamilton, R., Barr, I. and Reynolds, E. C. (2000) Characterization of a novel outer membrane hemin-binding protein of *Porphyromonas gingivalis*. J. Bacteriol. **182**, 6456–6462

9 Hanley, S. A., Aduse-Opoku, J. and Curtis, M. A. (1999) A 55-kilodalton immunodominant antigen of *Porphyromonas gingivalis* W50 has arisen via horizontal gene transfer. Infect. Immun. **67**, 1157–1171

10 Lipman, D. J. and Pearson, W. R. (1985) Rapid and sensitive protein similarity searches. Science (Washington, D.C.) **227**, 1435–1441

11 Nakai, K. and Kanehisa, M. (1991) Expert system for predicting protein localization sites in Gram-negative bacteria. Proteins **11**, 95–110

12 Ross, B. C., Czajkowski, L., Hocking, D., Margetts, M., Webb, E., Rothel, L., Patterson, M., Agius, C., Camuglia, S., Reynolds, E. C. et al. (2001) Identification of vaccine candidate antigens from a genomic analysis of *Porphyromonas gingivalis*. Vaccine **19**, 4135–4142

13 Bhogal, P. S., Slakeski, N. and Reynolds, E. C. (1997) A cell-associated protein complex of *Porphyromonas gingivalis* W50 composed of Arg- and Lys-specific cysteine proteinases and adhesins. Microbiology **143**, 2485–2495

14 Slakeski, N., Cleal, S. M., Bhogal, P. S. and Reynolds, E. C. (1999) Characterization of a *Porphyromonas gingivalis* gene prtK that encodes a lysine-specific proteinase and three sequence-related adhesins. Oral. Microbiol. Immunol. **14**, 92–97

15 Slakeski, N., Bhogal, P. S., O'Brien-Simpson, N. M. and Reynolds, E. C. (1998) Characterization of a second cell-associated Arg-specific cysteine proteinase of *Porphyromonas gingivalis* and identification of an adhesin-binding motif involved in association of the PrtR and PrtK proteinases and adhesins into large complexes. Microbiology **144**, 1583–1592

16 Rangarajan, M., Smith, S. J. M., U, S. and Curtis, M. A. (1997) Biochemical characterization of the arginine-specific proteases of *Porphyromonas gingivalis* W50 suggests a common precursor. Biochem. J. **323**, 701–709

17 Curtis, M. A., Thickett, A., Slaney, J. M., Rangarajan, M., Aduse-Opoku, J., Shepherd, P., Paramonov, N. and Hounsell, E. (1999) Variable carbohydrate modifications to the catalytic chains of the RgpA and RgpB proteases of *Porphyromonas gingivalis* W50. Infect. Immun. **67**, 3816–3823

18 Veith, P. D., Talbo, G. H., Slakeski, N. and Reynolds, E. C. (2001) Identification of a novel heterodimeric outer membrane protein of *Porphyromonas gingivalis* by two-dimensional gel electrophoresis and peptide mass fingerprinting. Eur. J. Biochem. **268**, 4748–4757

19 Aduse-Opoku, J., Davies, N. N., Gallagher, A., Hashim, A., Evans, H. E. A., Rangarajan, M., Slaney, J. M. and Curtis, M. A. (2000) Generation of lys-gingipain protease activity in *Porphyromonas gingivalis* W50 is independent of Arg-gingipain protease activities. Microbiology **146**, 1933–1940

20 O'Brien-Simpson, N. M., Paolini, R. A., Hoffmann, B., Slakeski, N., Dashper, S. G. and Reynolds, E. C. (2001) Role of RgpA, RgpB and Kgp proteinases in virulence of *Porphyromonas gingivalis* W50 in the murine lesion model. Infect. Immun. **69**, 7527–7534

21 Maley, J., Shoemaker, N. B. and Roberts, I. S. (1992) The introduction of colonic-*Bacteriodes* shuttle plasmids into *Porphyromonas gingivalis*: Identification of a putative *P. gingivalis* insertion-sequence element. FEMS Microbiol. Lett. **93**, 75–82

22 Fletcher, H. M., Schenkein, H. A., Morgan, R. M., Bailey, K. A., Berry, C. R. and Macrina, F. L. (1995) Virulence of a *Porphyromonas gingivalis* W83 mutant defective in the *prtH* gene. Infect. Immun. **63**, 1521–1528

23 Slakeski, N., Cleal, S. M. and Reynolds, E. C. (1996) Characterization of a *Porphyromonas gingivalis* Gene *prtR* that encodes an arginine-specific thiol proteinase and multiple adhesins. Biochem. Biophys. Res. Commun. **224**, 605–610

24 Chen, W. and Kuo, T. (1993) A simple and rapid method for the preparation of gram-negative bacterial genomic DNA. Nucleic Acids Res. **21**, 2260

25 Filip, C., Fletcher, G., Wulff, J. L. and Earhart, C. F. (1973) Solubilization of the cytoplasmic membrane of *Escherichia coli* by the ionic detergent sodium-lauryl sarcosinate. J. Bacteriol. **115**, 717–722

26 Sanchez, J. C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M., Wilkins, M. R. and Hochstrasser, D. F. (1997) Improved and simplified in-gel sample application using reswelling of dry immobilized pH gradients. Electrophoresis **18**, 324–327

27 Görg, A., Boguth, G., Obermaier, C., Posch, A. and Weiss, W. (1995) Two-dimensional polyacrylamide gel electrophoresis with immobilized pH gradients in the first dimension (IPG-DALT): the state of the art and the controversy of vertical versus horizontal systems. Electrophoresis **16**, 1079–1086

28 Molloy, M. P., Herbert, B. R., Slade, M. B., Rabilloud, T., Nouwens, A. S., Williams, K. L. and Gooley, A. A. (2000) Proteomic analysis of the *Escherichia coli* outer membrane. Eur. J. Biochem. **267**, 2871–2881

29 Dashper, S. G., O'Brien-Simpson, N. M., Bhogal, P. S., Franzmann, A. D. and Reynolds, E. C. (1998) Purification and characterization of a putative fimbrial protein/receptor of *Porphyromonas gingivalis*. Aust. Dent. J. **43**, 99–104

30 Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J. H., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**, 3389–3402

31 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**, 4673–4680

32 Okuda, K. and Takazoe, I. (1974) Haemagglutinating activity of *Bacteroides melaninogenicus*. Arch. Oral Biol. **19**, 415–416

33 Pike, R., McGraw, W., Potempa, J. and Travis, J. (1994) Lysine- and arginine-specific proteinases from *Porphyromonas gingivalis*. Isolation, characterization, and evidence for the existence of complexes with hemagglutinins. J. Biol. Chem. **269**, 406–411

34 Hendtlass, A., Dashper, S. G. and Reynolds, E. C. (2000) Identification of an antigenic protein Pga30 from *Porphyromonas gingivalis* W50. Oral Microbiol. Immunol. **15**, 383–387

35 Curtis, M. A., Slaney, J. M., Carman, R. J. and Johnson, N. W. (1991) Identification of the major surface protein antigens of *Porphyromonas gingivalis* using IgG antibody reactivity of periodontal case-control serum. Oral Microbiol. Immunol. **6**, 321–326

36 Koebnik, R. (1995) Proposal for a peptidoglycan-associating alpha-helical motif in the C-terminal regions of some bacterial cell-surface proteins. Mol. Microbiol. **16**, 1269–1270

37 Aduse-Opoku, J., Slaney, J. M., Rangarajan, M., Muir, J., Young, K. A. and Curtis, M. A. (1997) The Tla protein of *Porphyromonas gingivalis* W50: a homolog of the RI protease precursor (PrpRI) is an outer membrane receptor required for growth on low levels of hemin. J. Bacteriol. **179**, 4778–4788

38 Koebnik, R., Locher, K. P. and Van Gelder, P. (2000) Structure and function of bacterial outer membrane proteins: barrels in a nutshell. Mol. Microbiol. **37**, 239–253

39 Struyvé, M., Moons, M. and Tommassen, J. (1991) Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. J. Mol. Biol. **218**, 141–148

40 Koronakis, V., Sharff, A., Koronakis, E., Luisi, B. and Hughes, C. (2000) Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. Nature (London) **405**, 914–919

41 Karamyshev, A. L., Karamysheva, Z. N., Kajava, A. V., Ksenzenko, V. N. and Nesmeyanova, M. A. (1998) Processing of *Escherichia coli* alkaline phosphatase: role of the primary structure of the signal peptide cleavage region. J. Mol. Biol. **277**, 859–870

42 Potempa, J., Mikolajczyk-Pawlinska, J., Brassell, D., Nelson, D., Thøgersen, I. B., Enghild, J. J. and Travis, J. (1998) Comparative properties of two cysteine proteinases (gingipains R), the products of two related but individual genes of *Porphyromonas gingivalis*. J. Biol. Chem. **273**, 21648–21657

43 Messner, P. (1997) Bacterial glycoproteins. Glycoconj. J. **14**, 3–11

44 Izard, J. W. and Kendall, D. A. (1994) Signal peptides: exquisitely designed transport promoters. Mol. Microbiol. **13**, 765–773

Received 3 November 2001/19 December 2001; accepted 25 January 2002