Author/s:
Veith, PD;Chen, YY;Chen, D;O'Brien-Simpson, NM;Cecil, JD;Holden, JA;Lenzo, JC;Reynolds, EC

Title:
Tannerella forsythia Outer Membrane Vesicles Are Enriched with Substrates of the Type IX Secretion System and TonB-Dependent Receptors

Date:
2015-12-04

Citation:
Veith, P. D., Chen, Y. Y., Chen, D., O'Brien-Simpson, N. M., Cecil, J. D., Holden, J. A., Lenzo, J. C. & Reynolds, E. C. (2015). Tannerella forsythia Outer Membrane Vesicles Are Enriched with Substrates of the Type IX Secretion System and TonB-Dependent Receptors. Journal of Proteome Research, 14 (12), pp.5355-5366. https://doi.org/10.1021/acs.jproteome.5b00878.

Persistent Link:
https://hdl.handle.net/11343/108661

Revised MS pr-2015-00878s

Tannerella forsythia Outer Membrane Vesicles are Enriched with Substrates of the Type IX Secretion System and TonB-dependent Receptors

*Paul D. Veith, Yu-Yen Chen, Dina Chen, Neil M. O'Brien-Simpson, Jessica D. Cecil, James A. Holden, Jason C. Lenzo and Eric C. Reynolds**

Oral Health CRC, Melbourne Dental School, Bio21 Institute, The University of Melbourne, 720 Swanston Street, Melbourne, Victoria, 3010, Australia

KEYWORDS: *Tannerella forsythia;* outer membrane vesicles; proteome; cryo-TEM;, chronic periodontitis; type IX secretion; TonB-dependent receptors

ABSTRACT

Tannerella forsythia, a Gram-negative oral bacterium closely associated with chronic periodontitis naturally produces outer membrane vesicles (OMVs). In this study, OMVs were purified by gradient centrifugation and the proteome was investigated together with cellular fractions using LC-MS/MS analyses of SDS-PAGE fractions, resulting in the identification of 872 proteins including 297 OMV proteins. Comparison of the OMV proteome with the subcellular proteomes led to the localization of 173 proteins to the vesicle membrane and 61 proteins to the vesicle lumen while 27 substrates of the type IX secretion system were assigned to the vesicle surface. These substrates were generally enriched in OMVs, however the stoichiometry of the S-layer proteins, TfsA and TfsB was significantly altered, potentially to accommodate the higher curvature required of the S-layer around OMVs. A vast number of TonB-dependent receptors related to SusC, together with their associated SusD-like lipoproteins were identified and these were also relatively enriched in OMVs. In contrast, other lipoproteins were significantly depleted from the OMVs. This study identified the highest number of membrane associated OMV proteins to date in any bacterium and conclusively demonstrates cargo sorting of particular classes of proteins which may have significant impact on the virulence of OMVs.

INTRODUCTION

Chronic periodontitis is an inflammatory disease of the supporting tissues of the teeth associated with a polymicrobial biofilm (subgingival plaque) accreted to the tooth which results in destruction of the tooth's supporting tissues including the alveolar bone [1]. Periodontitis has been linked to an increased risk of cardiovascular diseases, certain cancers (orogastrointestinal tract and pancreas), pre-term birth, low birth weight and other systemic diseases related to the regular bacteraemia and chronic inflammation associated with the disease [2, 3]. Although chronic periodontitis is associated with a polymicrobial biofilm, the specific bacterial species of the biofilm *Porphyromonas gingivalis, Treponema denticola and Tannerella forsythia* as a consortium have been closely associated with clinical measures of disease [4, 5]. *T. forsythia* is a Gram-negative, anaerobic, fusiform bacterium [6]. It has been demonstrated to cause alveolar bone loss in mice [7] and synergistically caused increased bone loss when co-infected with *P. gingivalis* or *Fusobacterium nucleatum* in rats [8]. *T. forsythia* produces several putative protein virulence factors such as proteinases, a surface layer, BspA and various glycosidases, and these have been reviewed [9]. More recently, a group of six secreted proteases sharing a near identical C-terminal domain ending with the amino acid sequence KLIKK was identified and proposed to be virulence factors [10]. Of these, the metalloproteinases karilysin and mirolysin were shown to inhibit the human complement system by degrading C4, C5 and ficolin [11, 12] while mirolase was shown to degrade human fibrinogen and the antimicrobial peptide LL-37 [13].

A distinctive feature of SDS-PAGE profiles of whole cell extracts of *T. forsythia* is the presence of a major double band greater than 200 kDa. These bands correspond to two glycoproteins designated surface layer protein A (TfsA) and surface layer protein B (TfsB) which are the reported constituents of the surface layer (S-layer), a distinct layer beyond the outer membrane (OM) that can be clearly seen in electron micrographs [14, 15]. The structure of

this 22 nm thick S-layer has been recently probed by both TEM and AFM techniques and is described as a P4 structure, with a lattice constant of 9-10 nm resulting from co-assembly of equimolar amounts of TfsA and TfsB [16, 17]. In our proteomic analysis of the *T. forsythia* OM, we identified 15 putative glycoproteins including TfsA and TfsB which shared a similar C-terminal domain sequence (CTD) that was also common to many surface-associated *P. gingivalis* proteins [18]. This sequence is now understood to be the secretion signal for the type IX secretion system (T9SS, formerly PorSS [19]) and has been indirectly shown to be cleaved and modified by a novel Gram negative sortase, PorU [20, 21 22]. Recently, the *T. forsythia* surface layer proteins have been confirmed to be secreted by the T9SS, since the surface layer was absent in mutants lacking various components of the T9SS [23, 24]. A study comparing the proteomes of *T. forsythia* grown as a biofilm and as a planktonic culture quantitated the relative level of 348 proteins with 44 reported to have changed significantly in abundance [25]. The proteins that were reduced in abundance in the biofilm included several enzymes involved in central metabolism and butyrate production while the proteins that increased in abundance were mainly membrane proteins and proteins involved in stress response.

Gram-negative bacteria naturally produce outer membrane vesicles (OMVs) by "blebbing" of their OM. OMVs are composed of a single membrane that is derived from the outer membrane (OM) and contains OM proteins, lipopolysaccharide, and other lipids, while the vesicle lumen mainly contains periplasmic proteins [26]. OMVs are important virulence factors being involved in bacterial adherence, defence against host factors, and the delivery of a wide range of toxins [26]. For example, OMVs from enterotoxigenic *Escherichia coli* specifically bind to and enter intestinal epithelial cells thereby delivering its cargo of heat-labile enterotoxin [27]. OMVs also play an important role in adhesion in bacterial biofilms including biofilms of *Pseudomonas aeruginosa* and *Myxococcus xanthis* [28]. OMVs of *P. gingivalis* are able to bind to *T. forsythia*, assisting its attachment to epithelial cells [29]. OMVs

of periodontal pathogens have been proposed to have a critical role in the pathogenesis of periodontitis by being released from the bacteria as components of the biofilm accreted on the tooth root and penetrating the host's gingival tissue whereby they induce vascular disruption (bleeding) and immune dysregulation. [30-32] OMVs are also being developed as vaccines for different bacterial species [33].

It is useful to study the proteomes of OMVs for several purposes including the characterisation of OMV-based vaccines, identification of virulence factors, OMV biogenesis and protein localization [34]. Our recent study of the OMV proteome of *P. gingivalis* for example demonstrated that the substrates of the T9SS including toxins such as the gingipain proteinases were enriched in OMVs and identified other enriched proteins that may be involved in OMV biogenesis [32]. Unlike other OMV proteomes the OMV proteome of *P. gingivalis* was very clean with no detectable cytoplasmic contamination and was therefore very helpful as a means of determining the cellular localization of proteins to the outer membrane or periplasm [32]. In a recent study of the *T. forsythia* OMV proteome, 175 proteins were identified and the OMVs were shown to retain the S-layer and stimulate the release of pro-inflammatory mediators from a human monocytic cell line [35]. In the current study we significantly extend the proteomic analysis of *T. forsythia* OMVs identifying 297 proteins, experimentally determining protein localization to the vesicle membrane and lumen and providing strong evidence for cargo sorting of particular protein classes.

METHODS

Isolation and Enrichment of *T. forsythia* OMVs

T. forsythia ATCC 43037 was obtained from the Melbourne Dental School culture collection and cultured in volumes of 400 mL in brain heart infusion broth (BD, Sydney, Australia) with added tryptic soy broth (15 g/L) and yeast extract (10 g/L) supplemented with 1 g/L cysteine, 10 mg/L N-acetylmuramic acid (Sigma-Aldrich), 5 mg/L haemin, 0.5 mg/L menadione (Sigma-Aldrich) and the filtrate of 5% v/v heat inactivated fetal calf serum filtered using a Vivaspin 10 kDa ultrafiltration filter (GE HealthCare Life Sciences). The ultrafiltration step was used to remove most of the serum proteins, which were otherwise found to heavily contaminate the purified OMVs. Bacteria were grown to late exponential phase at 37 °C under anaerobic conditions.

Whole bacterial cells were harvested by centrifugation at 8,000 x g for 30 min at 4 °C. The collected supernatant was filtered through a 0.22-μm filter and then concentrated through a 100-kDa filter using a tangential flow filtration Minimate TFF System (PALL Life Sciences, Melbourne, Australia) according to the manufacturer's instructions. The collected concentrated supernatant was centrifuged at 100,000 x g for 2 hours at 4 °C using a JA-30.50 Ti fixed angle rotor (Beckman Coulter) to pellet crude OMVs. The crude OMVs were then resuspended in 800 μL HEPES buffer (50 mM HEPES, 150 mM NaCl, pH 6.8).

Crude OMVs were further purified by density gradient centrifugation using OptiPrep™ (60% w/v iodixanol in water, Sigma-Aldrich); whereby the crude OMV preparation was resuspended in 3 mL HEPES buffer containing 50% w/v iodixanol and placed in an Ultra-Clear™, 14 mL, 14 x 95 mm tube (Beckman Coulter). A discontinuous iodixanol gradient was achieved by layering successive 1.5 mL volumes of HEPES buffer containing 45%, then 40%, 35%, 30%, 25% and 20% w/v iodixanol. After centrifugation at

150,000 x g for 48 hours at 4 °C using a SW40 Ti rotor (Beckman Coulter) eight 1.5 mL gradient fractions were collected by pipette from top to bottom of the density gradient solution. Analysis of the fractions by SDS-PAGE and MS (see below) revealed that the OMVs were present in fractions 2-5 (data not shown). Fraction #4 containing 35% iodixanol contained the most OMVs and was used for the proteomic analysis. For EM analysis, fractions containing the purified OMVs were pooled and washed with 0.22 µm filtered 10 mM phosphate buffered saline pH 7.4 (PBS), at 150,000 x g for 2 hours at 4 °C (using the SW40 Ti rotor) and resuspended in 100 µL of 0.22 µm filtered PBS and stored at 4 °C.

Isolation of Cellular Fractions

To obtain the membrane fraction, harvested cells (from above) were washed once in buffer (150 mM NaCl, 5 mM $MgCl_2$ and 50 mM sodium phosphate, pH 7.4) and resuspended with 20 mL of 20% buffer (1/5 dilution in water) in a 50-mL centrifuge tube (JA-20 tube, Beckman Coulter). The tube was placed in a beaker containing an ice/water mixture and sonicated using an ultrasonic processor (model CPX 750, Cole Parmer) fitted with a 6.5 mm tapered microtip. The amplitude was set to 40%, pulser to 1 s on, 2 s off for a total of 30 min. Proteinase inhibitor cocktail (10 µL/mL, Sigma-Aldrich) was added immediately prior to and after sonication. The membranes were collected by centrifugation at 48400 ×g for 20 min, and the supernatant was retained as the "soluble" fraction containing both cytoplasm and periplasm. The membranes were washed with 20 mL and 5 mL of buffer. Resuspension of the membrane fraction into buffer (for washing) was aided by sonication using a 3 mm stepped microtip at 21% amplitude. The periplasm was prepared exactly as described previously [36].

SDS-PAGE and In-Gel Digestion

The OMV sample and the soluble fractions were first concentrated by precipitation with 12% trichloroacetic acid and the pellet washed with ice-cold acetone. All samples were solubilised and denatured by heating in LDS sample solution (Life Technologies, Melbourne, Australia) and 50 mM dithiothreitol at 100° C for 5 min. SDS-PAGE was conducted using NuPAGE Novex 10% Bis-Tris gels with MOPS running buffer (Life Technologies, Melbourne, Australia). Each gel was excised into approximately 20 contiguous segments with a scalpel blade. In-gel digestion using trypsin was performed after reduction with dithiothreitol and alkylation with iodoacetamide as previously published [36]. For quantitation of cargo sorting, fresh cell membrane and OMV samples from biological replicates were treated the same way except that gels were only run for 8 min at 126 V (2 min after all the tracking dye had clearly entered the gel). After Coomassie Blue staining, the stained portion of each lane was excised and digested as above. These samples were run in triplicate.

Mass Spectrometry

Tryptic digests were acidified with trifluoroacetic acid to 0.1% before online LC–MS/MS analyses. An UltiMate 3000 system (Thermo Scientific) was used with a precolumn of PepMap C18, 300 μm i.d. × 5 mm (Thermo Scientific) and an analytical column of PepMap C18, 180 μm i.d. × 15 cm (Thermo Scientific). Buffer A was 2% (v/v) acetonitrile, 98% H_2O, 0.1% (v/v) formic acid, and buffer B was 98% (v/v) acetonitrile, 2% H_2O, 0.1% (v/v) formic acid. Digested peptides were initially loaded and desalted onto the precolumn in buffer A at a flow rate of 30 μL/min for 5 min. The peptides were eluted using a linear gradient of 2–40% acetonitrile over 45 min (or 65 min for the unfractionated samples used for quantitation of cargo sorting) at a flow rate of 2 μL/min directly into an Esquire HCT Ultra ion trap mass spectrometer via a 50 μm ESI needle (Bruker Daltonics, Bremen, Germany). The ion trap was operated via EsquireControl v6.1 in the positive ion mode at an MS scan speed of 8100 m/z/s

over an m/z range of 300–1200 and a fast Ultra Scan of 26000 m/z/s for MS/MS analysis over an m/z range of 100–2800. The drying gas (N_2) was set to 6 L/min and 300 °C. The peptides were fragmented using auto-MS/MS with the SmartFrag option on with up to nine precursor ions between m/z 400 and 1200 for each MS scan. Spectra were smoothed, deconvoluted, labelled, and exported using DataAnalysis v3.4 (Bruker Daltonics).

Protein Sequence Databases and MS/MS search parameters

Proteins were identified by MS/MS ions search using Mascot v2.2 (Matrix Science, UK) originally against an annotated *T. forsythia* strain 92A2 database of 3034 protein sequences obtained from the Los Alamos National Laboratory (LANL) website (http://www.oralgen.lanl.gov) in May 2006 (the original source of this data is the J. Craig Venter Institute http://www.jcvi.org). It should be noted that the genome sequence of the 92A2 strain was deposited at NCBI with the sequence number CP003191 (Bioproject PRJNA83157) but incorrectly attributed to the type strain ATCC 43037 [37]. The true genome sequence of the type strain (the strain used in this study) recently became available [37], and the corresponding NCBI-annotated protein sequence database then became the primary database for MS/MS searches. This database was however not publicly available in a suitable format with the "Tanf" locus numbers. It was therefore constructed in July 2015 by Valentin Friedrich by first downloading the entire genbank file from http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=JUET01#contigs and then the FASTA-formatted database containing 2692 protein sequences was synthesised by extracting the locus number, comments and amino acid sequence with the aid of a Genbank-FASTA conversion tool (http://rocaplab.ocean.washington.edu/tools/genbank_to_fasta). The sequences of human keratins and porcine trypsin were also added to the database. It was found that this sequence database contained many errors due to incomplete assembly of the

genome and incorrect annotation. Therefore, for protein sequence entries noted in this study to have an accurate reference, the entries were corrected. These corrected entries together with the methods of correction are listed in Table S-2.4. Search parameters were as follows. Enzyme = trypsin, MS tolerance = 1.5 Da, MS/MS tolerance = 0.5 Da, missed cleavages = 1, fixed modifications = carbamidomethyl (Cys), optional modifications = oxidation (Met). Additional searches with missed cleavages = 2, Enzyme = semi-trypsin, and additional variable modifications = Gln-pyro-Glu (N-term Q, E) were conducted to enable the identification of signal peptide cleavage sites.

Protein Identification

Proteins were initially considered identified when the protein score was greater than 25 which corresponds to the ions score threshold of 25 for individual peptides ($p < 0.05$). The false discovery rate as calculated by automatic decoy searches performed by Mascot was determined to be 1.1% and 0.7% for OMV replicate 1 and 2 respectively. This was achieved by summing the number of peptide matches above the identity threshold across every run in the replicate for both the true database and the decoy database as reported by Mascot. The FDR was then calculated in the usual way (decoy hits / true database hits). Furthermore, for all proteins identified from a single peptide, the Mascot score had to be at least 30 in one of the replicates. OMV proteins identified from a single peptide with a Mascot score between 30 and 35 were manually checked. The identification data presented in Table S-1 and Table S-2.3 uses “Total Mascot” scores. When proteins were identified from more than one SDS-PAGE band, the “Total Mascot” score was calculated by summing the Mascot scores from the individual bands. Spectral counting was done as follows and only applied to proteins passing the above criteria. In order to count all spectra, spectra were extracted from the MS data without averaging them over a retention time window. All peptides identified with

redundant modification position were removed (keeping highest scoring hit only). Peptides that were shared between multiple proteins were scored as only 0.5 spectral counts. Proteins identified with only these shared peptides were removed altogether. Spectra that matched to two different peptides were deleted if the assignment was in doubt as follows. Only peptides ranked #1 were accepted, and if a lower ranked peptide matched to a protein of higher rank, the #1 ranked peptide was only accepted if its Mascot score was at least 20 units higher. Only peptides with a Mascot score of greater than 20 were counted. The spectra were then summed across gel bands to give a total spectral count for each protein. The number of "countable" peptides presented in Table S-2.3 represents the number of non-redundant peptide sequences identified for each protein.

Electron Microscopy

Cryo-transmission electron microscopy (Cryo-TEM) was performed using an FEI Tecnai G2 F30 instrument (FEI company) as previously described [36].

Bioinformatics and Statistics

Outer membrane beta-barrel predictive scores were obtained using the Freeman Wimley analysis tool obtained from the Wimley Laboratory (http://www.tulane.edu/~biochem/WW/Barrel.html). The program was used without considering the SignalP predictions [38]. The Pfam data were obtained by performing a batch sequence search using all *T. forsythia* protein sequences against the Pfam 27.0 database at pfam.xfam.org. The top Pfam-A hit for each protein is included but those considered insignificant according to Pfam are marked with an asterisk in Table S-1. Type I and II (lipoprotein) signal peptides were predicted based on published criteria [39] and by comparison with the experimentally determined cleavage sites of *T. forsythia* and *P. gingivalis* proteins

determined in this and previous studies [18, 32]. Inner membrane proteins (IMPs) were predicted by homology and with the help of the TMHMM Server v. 2.0 (www.cbs.dtu.dk/services/TMHMM-2.0/) for the purpose of assessing OMV purity (see Figure 2B). Only helices predicted beyond the N-terminal 60 aa residues were considered, since those predicted within the first 60 aa may just be signal peptides. The following IMPs were predicted based on homology to known IMPs: Tanf_02365 & Tanf_02370 (homology to PorL & PorM [19]); Tanf_03330 & Tanf_03335 (fumarate reductase [40]); Tanf_06985, Tanf_11185 & Tanf_02050 (membrane fusion proteins [41]); Tanf_12470 & Tanf_12505 (TonB complex [42]); Tanf_05740 & Tanf_05755 (ATP synthase [40]); Tanf_00655, Tanf_00640 & Tanf_00630 (NADH:ubiquinone oxidoreductase [40]); and the general secretion system (SEC) proteins Tanf_07690 & Tanf_09235. Predictions of T9SS substrates were taken from Veith et al (2013) [21]. Metabolic pathways were mapped with the aid of the KEGG pathway database (http://www.genome.jp/kegg/pathway.html) using "tfo" as the organism prefix. T-tests were performed on the observed cargo sorting trends using the T-test function in Excel, with tails = 2, and type = 2 (homoscedastic), see Table S-3.3.

RESULTS

OMV Purity

OMVs were prepared from *T. forsythia* using gradient centrifugation and the proteins were separated by SDS-PAGE (Figure 1A). Two identical gel lanes were excised into 22 segments, digested with trypsin and analyzed by LC-MS/MS resulting in the identification of 297 non-redundant proteins (Table S-1). To enable an assessment of the OMV purity and to determine whether proteins were likely to be present in the vesicle membrane or lumen, cellular fractions were also prepared and analyzed in a similar manner (Figure 1B). From the duplicate cell membrane samples, 638 non-redundant proteins were identified while 481 and 355 proteins were identified from the soluble and periplasm fractions respectively. In total, 872 non-redundant proteins were identified representing 32% of the theoretical proteome (Table S-2.3). OMV purity was assessed by checking for the presence of abundant cytoplasmic and inner membrane proteins. Abundant cytoplasmic proteins were identified by their high spectral counts in the soluble fraction and by the absence of signal peptides. Of the 25 most strongly identified cytoplasmic proteins, none were identified in the OMV preparations whereas all except one were also identified in both the cell membrane and periplasm fractions (Figure 2A). Similarly, of the 25 most strongly identified known or predicted inner membrane proteins identified, only three were identified in OMVs (Figure 2B). These were Tanf_12470 (TonB), Tanf_12490 (ExbB) and Tanf_11185 (HlyD), all of which are known to include domains that are localized to the periplasm. Indeed each of these three proteins was also identified in the periplasmic fraction suggesting that these proteins had dissociated from the IM to some degree enabling their incorporation into the OMV lumen. Alternatively, each of these proteins also binds to Omps (ExbB-TonB-TDR and HlyD-OEP) and could conceivably be dragged into the OMV lumen during their formation. However, we cannot rule out that these proteins are also inserted into the vesicle membrane

and therefore these three proteins were considered to be of uncertain localization within the OMVs (Table S-1). The absence of cytoplasmic proteins in OMVs indicates that the level of whole cell or cell debris contamination was below our detection limit, while the absence of most of the abundant IM proteins suggests that the vesicles purified were solely derived from the OM.

Localization of Proteins to the Vesicle Membrane and Lumen

The localization of the 297 OMV proteins to the vesicle lumen, vesicle membrane or vesicle surface was differentiated based on their simultaneous identification in the corresponding cellular fraction. If a protein is present in both OMVs and cell membranes, then the localization of that protein within OMVs is likely to be the membrane, and its localization within cells is likely to be the OM. Similarly, if a protein is present in both OMVs and the soluble fractions (soluble fraction or periplasmic fraction) its localization within OMVs is likely to be the lumen, while its cellular localization would be the periplasm. To determine the localization, the spectral counts obtained for the cell membrane (M) were divided by spectral counts from the periplasmic (P) fraction to give the M/P ratio. For OMV proteins with M/P ratios of greater than 2.5, the protein was classified as being localized to the vesicle membrane, whereas if the ratio was less than 0.4 it was classified as being localized to the vesicle lumen (Table S-1). For intermediate M/P ratios, some proteins were localized to the vesicle membrane based on being a member of an OM-specific protein family (Pfam), or by exhibiting a very high propensity for trans-membrane beta-barrel (TMBB) formation (score > 0.8). Proteins exhibiting a type II signal peptide were predicted to be lipoproteins and hence tethered to the vesicle membrane. Proteins containing the C-terminal signal (CTD) for type IX secretion as previously predicted [21] were classified as “cell surface” (OMV surface). In this way, 173 proteins were assigned to the vesicle membrane, 61 to the lumen, 27 to the

OMV surface and 36 uncertain (Table S-1). One of the proteins, Tanf_03675 containing a thioredoxin domain was initially predicted to be on the OMV surface but was found to have an extremely low M/P ratio of 0.03, usually indicative of localization to the vesicle lumen. Its cell surface prediction was the weakest of the group with an E-value of 0.003 (Table S-2.3, "CTD E-value"), and when checked it was observed that the thioredoxin domain overlapped the proposed C-terminal signal (data not shown). Furthermore Tanf_03675 was identified with its C-terminal region intact in band 15, just below the 50 kDa marker, in agreement with its calculated MW of 44 kDa, whereas T9SS substrates are generally more elevated in MW through their modification and have their C-terminal signal removed. For these reasons Tanf_03675 was considered not to be a substrate of the T9SS and was localized to the vesicle lumen.

The experimental localization based on the M/P ratio was deemed significant since the known or strongly predicted outer membrane proteins exhibited high M/P ratios (Table S-1). Of the 65 proteins that exhibited a valid M/P ratio and were members of an OM-specific protein family (LptE, OEP, OmpA-like, OMP_b-brl, TonB_dep_Rec, Plug, Porin_O_P, Toluene_X, SusD, SusD-like, YfiO) 64 had an M/P ratio greater than 1 (Table S-1).

Despite the known localization of T9SS substrates including the surface layer proteins to the cell surface, many exhibited low M/P ratios (Table S-1). Their notable presence in the periplasmic fraction suggested that either the precursor forms of these proteins were relatively abundant in the periplasm or that they had dissociated from the cell surface to some extent during the osmotic shock procedure.

Cargo Sorting

To determine whether OMVs are selective in the protein cargo they carry, proteins localized to the vesicle membrane and vesicle surface region were compared between the OMV and

cell membrane fractions to give a measure of relative enrichment. The CTD proteins and TonB-dependent receptors were generally found to have higher spectral counts in the OMVs than the cell membranes suggesting they are relatively enriched in the OMVs, while most lipoproteins were preferentially retained in the cell membranes (Table S-2.3, "O/M ratio"). Other proteins strongly retained by the cell included the OM efflux protein Tanf_02045 and the type IX secretion protein, PorN (Tanf_02375). The retention of these proteins can be explained by their predicted association with trans-envelope spanning complexes. Consistent with this, four other OM efflux proteins (Tanf_06990, Tanf_07785, Tanf_08785, Tanf_11225) were only found in the cell membrane fraction (Table S-2.3). Since there may have been bias in the gel cutting and MS analyses between the OMV and cell membrane fractions (Figure 1), a second, more quantitative analysis using new samples was conducted where the OMV and cell membrane fractions were analyzed side by side on an SDS-PAGE gel that was run for 8 min only, so that a single band representing each fraction could be excised, digested and analyzed by LC-MS/MS. This analysis was done in triplicate. For the OMV fraction, 99 proteins were identified, while 208 proteins were identified from the cell membrane fraction. In total, 79 proteins were quantified that had been localized to the vesicle membrane or vesicle surface previously (Table S-1). Of these 79 proteins identified in the second experiment, 50 were found in both fractions, 15 were found in the OMVs only and 14 were found in the cell membrane fraction only (Table S-3.3). The CTD proteins (T9SS substrates) were again found to be the most enriched group in the OMVs (Figure 3, yellow markers), followed by the TonB-dependent receptors, together with their associated SusD-like lipoproteins (Figure 3, red and black markers). As with the first analysis, predicted lipoproteins that were not associated with TDRs were strongly retained in the cell membranes (Figure 3, light purple markers). T-test analysis of these groups showed that the OMV/Membranes spectral count ratios for CTD proteins and TDRs were significantly

different to the lipoproteins ($p=2.3x10^{-4}$ and $1.3x10^{-3}$ respectively) indicating that the observed enrichment and depletion of these groups were extremely unlikely to be random (Table S-3.3).

While the majority of non-TDR-associated lipoproteins were depleted in the vesicle membrane, at least four appeared to be highly enriched and were examined further (Figure 3). It was found that the genes of some of these lipoproteins were in fact located in TDR-containing loci, despite not being predicted to be SusD-like homologs. These were HmuY (Tanf_09520), Tanf_13105, and Tanf_09950 (Figure 3). Of note, the lipoprotein Tanf_03910 also appeared to be very enriched however this protein exhibited a low M/P ratio of 0.83 suggesting that it may also be localized to the vesicle lumen. Examination of its signal peptide indicated that in addition to the type II (lipoprotein) signal peptide cleavage site (…MGIFSG/C), there was also a potential type I cleavage site (…MGIFS/GC), consistent with dual localization to the vesicle membrane and vesicle lumen (Table S-2.3). The spectral count in the OMVs was considerably higher than that of the periplasm fraction (Table S-2.3) indicating that regardless of the distribution between the vesicle membrane and lumen, Tanf_03910 was considerably enriched in OMVs.

Signal Peptides

Of the 297 OMV proteins, 294 exhibited a predicted cleavable signal peptide (Table S-2.3). Of these, 99 were predicted to be lipoproteins due to the presence of a potential lipoprotein signal (type II) and 195 exhibited predicted type I signal peptides. Of these 195, 143 (73%) were predicted to have a cleavage site N-terminal to a Gln residue suggesting that the N-terminus of the mature protein would be pyroglutamate. Of these, 33 were confirmed by direct MS/MS identification of semi-tryptic peptides corresponding to their predicted mature N-termini (Table S-2.5). Some signal peptides exhibited both type I and type II predicted

cleavage sites potentially giving rise to dual localizations. It should be noted that the lipoprotein Tanf_11185 with a type II signal peptide is a predicted IMP (HlyD) and therefore its localization within OMVs was considered uncertain (see also above). Some of the signal peptides were not immediately evident in the protein sequences. In such cases, the sequences were examined both internally and within the 5' intragenic region to find them.

Electron Microscopy of OMVs

The OMVs were also analyzed by cryo-TEM. The OMVs appeared in various shapes with some being close to spherical but others being quite elongated and irregular (Figure 4). The size also varied considerably, with a diameter of 125 nm being typical. Most, but not all possessed a surface layer which projected approximately 30 nm from the membrane surface (Figure 4).

Metabolic pathways

Since the metabolic pathways of *T. forsythia* have not been studied in detail, the major pathways that could be connected to central metabolism were mapped out in accordance with the identified proteins. In this way, 62 enzymes were mapped to pathways that included glycolysis, the TCA cycle, and amino acid fermentation (Figure S-1).

DISCUSSION

The OMV proteome of *T. forsythia* as detected in this study includes 173 proteins localized to the vesicle membrane and 27 further proteins predicted to be on the vesicle surface (Table S-1) representing the highest number of OM-associated OMV proteins identified from a single bacterium. This is in part due to the high number of TonB-dependent receptors and their associated lipoproteins that are present in this bacterium. In comparison, 109 OM and secreted (cell surface and extracellular) proteins were identified in *P. gingivalis* OMVs [32] and approximately 100 in *Pseudomonas putida* OMVs [43]. In most studies, less than 100 of the identified proteins were predicted to be OM or secreted proteins (see Table 2 in Veith et al (2014) [32]). Another reason for the high number of identified membrane proteins is the relative purity of the OMV sample. Apart from the OMV proteome of *P. gingivalis* and *Bacteroides fragilis* [44], the proteomes of other OMVs generally contain significant numbers of cytosolic proteins despite rigorous purification regimes [32]. Recently, Friedrich et al [35] also reported the OMV proteome of *T. forsythia* where they identified 175 proteins and on the basis of bioinformatic prediction (CELLO) suggested that 61, 53, 22 and 39 proteins may be OM, periplasmic, extracellular and cytoplasmic respectively. Of their 175 identified proteins, 160 proteins were common to our OMV proteome (Table S-1, last column). Despite the bioinformatic prediction of 39 cytoplasmic proteins, none of their identified proteins matched to cytoplasmic proteins identified in our study demonstrating the limitations of bioinformatic predictions. It appears that *T. forsythia*, *P. gingivalis* and *B. fragilis* naturally form purer OMVs, and this may prove true for other members of the *Bacteroidia* class to which they all belong.

The OMVs of *T. forsythia* are highly complex since they contain at least 297 different proteins. Using spectral counts as estimates of protein group abundance, it can be seen that about 31% of protein mass is derived from T9SS substrates, 54% is localized to the

membrane, 12% to the lumen and 3% uncertain (Figure 5A). Since large proteins produce more peptides after digestion, the spectral count is an estimate of total protein amount (mass) rather than the number of protein molecules (moles). Using the more quantitative data generated for the cargo sorting experiment, it was found that the proportion of T9SS substrates was actually much higher at 49% while the proportion of protein localized to the vesicle lumen was only 3% (Figure 5B). The reason for the increase in proportion of T9SS substrates is likely due to the extreme abundance of the surface layer proteins in their isolated positions in the high MW region of the SDS-PAGE gel of the first experiment (Figure 1A). When analyzed together with all other proteins in the sample (second experiment), their true relative abundance was more easily estimated. So while the first experiment enabled a large number of proteins to be identified, localized and tentatively assessed for OMV enrichment, the second experiment enabled a better quantitation of OMV enrichment and OMV content, albeit for a subset of proteins.

The Type IX Secretion System

As a group, the T9SS substrates were enriched in the OMVs (Figure 3). These substrates are generally found to be extensively modified/glycosylated as shown by proteomic analyses of several diverse species including *P. gingivalis, T. forsythia, Prevotella intermedia, Parabacteroides distasonis* and *Cytophaga hutchinsonii* [18, 21]. Indeed, it was shown that when GFP was fused to a type IX secretion signal and expressed in *P. gingivalis*, it became extensively modified/glycosylated [45]. It is understood that the modification which probably includes a lipid moiety, anchors the T9SS substrate to the cell surface [21, 22]. In *P. gingivalis*, the outside layer visible in cryo-TEM has been described as the electron dense surface layer (EDSL). The EDSL is quite diffuse in appearance which contrasts with the typical S-layers of other bacteria which are defined as having a regular (para-crystalline) structure usually

composed of a single glycoprotein [46, 47]. The EDSL is reduced in intensity and thickness in strains lacking major T9SS substrates [22] and may therefore represent an amorphous mixture of T9SS substrates displayed on the cell surface. The enrichment of T9SS substrates in *P. gingivalis* OMVs was therefore explained by the necessary increase in the EDSL:OM volume ratio required to completely cover the surface of the smaller OMVs relative to the much larger cell[32].

In *T. forsythia*, the S-layer has been shown to comprise a regular structure composed of equimolar TfsA (Tanf_03370) and TfsB (Tanf_03375) and in mutant strains lacking just one of these proteins, the S-layer is not properly formed [48]. The ratio of TfsA:TfsB on the cell surface as estimated by spectral counting was very close to 1:1 with 71 ± 11 and 70 ± 10 being identified for TfsA and TfsB respectively (Table S-3.3). The calculated 1:1 ratio is consistent with analyses of the S-layer using AFM and TEM techniques [16, 17]. However, the stoichiometry of these proteins in the OMVs was significantly different. The MS analysis resulted in the identification of 195 ± 20 and 57 ± 5 respectively at a TfsA:TfsB ratio of approximately 3:1 (Table S-3.3). This change in stoichiometry can also be seen in the OMV gel bands where the band mostly consisting of TfsA (Figure 1A, gel segment 4) was significantly more intense than the band mostly consisting of TfsB (Figure 1A, gel segment 3). In contrast, the intensities of the corresponding bands in the cell membrane sample were similar (Figure 1B, gel segments 2 and 3). Overall, TfsA (Tanf_03370) was relatively enriched in the OMVs, while TfsB (Tanf_03375) was relatively depleted (Figure 3). Since S-layers are highly structured, to produce OMVs with a much greater curvature than the cell, the structure of the S-layer would need to change. The enrichment of TfsA in OMVs suggests the curvature of the S-layer may be controlled by varying the TfsA:TfsB stoichiometry, with TfsA favoring regions of higher curvature. Recently, a square S-layer lattice was reported for *T. forsythia* OMVs based on negatively stained TEM images, suggesting that the S-layer

remained intact in the OMVs, however their western blot results support our finding that TfsA is relatively more abundant than TfsB in OMVs compared to cells [35]. Together, the results suggest that the S-layer of OMVs remains intact and shares the same 2D geometry as the cells, but uses a higher TfsA:TfsB ratio to achieve a more tightly curved 3D structure. Further work is warranted on this topic to confirm the change in surface layer protein stoichiometry and to understand how S-layer structure may vary to accommodate the range of sizes and shapes of OMVs observed under cryo-TEM (Figure 4).

In addition to the 27 T9SS substrates identified in the OMV fraction, one further substrate (Tanf_03310) was identified in the membrane fraction only. Excluding Tanf_03675, for reasons described in the "Results" section, only 10 predicted substrates [21] were not identified in this study. However, two further proteins, both proteinases, were found to have CTD-like domains that were not previously predicted [21]. These were the S8 peptidase Tanf_00440 (mirolase [13]), and karilysin (Tanf_06550), each identified with a low Mascot score from single fractions (Table S-2.3), making it difficult to confirm their localization and modification status. While their CTD sequences were not predicted by our Hidden Markov Model, they are predicted by the TIGR04183 "Por_Secr_tail" TIGRfam model. These two proteinases belong to a group of six proteinases which share an almost identical CTD ending with the amino acid sequence KLIKK [10]. Since all six of these proteinases were either not detected or only detected at low levels, it is interesting to speculate that the expression of this group of potential virulence factors is increased in specific growth conditions or in disease. Indeed, two of these proteinases, forsilysin (Tanf_06225) and miropsin-2 (Tanf_06530) were identified in OMVs produced by bacteria grown in serum-free media [35] but not in our study which included filtered serum and tryptic soy. The expression of six of these proteinases was confirmed by the detection of transcripts in gingival crevicular fluid obtained from periodontitis patients [10], however none of these proteins were detected in biofilm growth [25].

In contrast, the surface layer proteins, TfsA & TfsB were found to have increased in abundance by more than two-fold when *T. forsythia* was grown as a biofilm compared to planktonic culture [25]. Furthermore the T9SS substrate and internalin-like protein, TF0193, was reported to be up regulated in the biofilm, whereas it was not identified at all in our study using planktonic culture.

In addition to identifying many substrates of the T9SS, many components of the system itself were identified (Table 1). The orthologs of PorK, L, M, & N were amongst the most abundant proteins identified in the membrane fraction (Table S-2.3) and of these, only the N protein (Tanf_02375) was also identified in the OMVs. The results are consistent with the K-L-M-N proteins forming a large complex that may span both the IM and OM as shown previously in *P. gingivalis* [19]. PorP (Tanf_02355) was identified with a very low score in both the membrane and OMV fractions suggesting it may be an accessory protein rather than a core structural component. Interestingly, just as the T9SS protein PorV (LptO [36]) was highly enriched in OMVs of *P. gingivalis* [32], in *T. forsythia*, the PorV ortholog Tanf_04220 was also highly enriched (Figure 3). Of the known components, only the Sov protein (Tanf_04410) was not identified.

TonB-Dependent Transport

Of the proteins localized to the vesicle membrane, 70% (32% of total) were attributable to TDRs and their associated lipoproteins (Figure 5B). A total of 36 TDRs were identified in OMVs together with 31 TDR-associated lipoproteins. In comparison, 46 TDRs and 28 associated lipoproteins were identified from our previous study of the outer membrane proteome [18]. Combining the datasets, a total of 52 TDRs and 35 TDR-associated lipoproteins were identified out of 74 proteins matching the “Plug” or “TonB_dep_Rec” Pfam families and 85 proteins matching the SusD, SusD-like and SusE Pfams. As a whole, the TDRs were

relatively enriched in OMVs (Figure 3). The reason for *T. forsythia* expressing such a large number of TDRs is puzzling as noted previously [18]. This enigma is deepened by their enrichment in OMVs, since in OMVs, transport of solute would not be possible. TonB-dependent transport is dependent on the TonB protein, which transduces energy obtained from the inner membrane. In the absence of an energy source, TDRs in OMVs would only be able to bind to solutes. Transport of solute across the membrane would presumably be only possible if solute-loaded OMVs could fuse with cells, and the TDRs regain their ability to interact with cellular TonB. Alternatively OMVs loaded with bound-solutes may be washed out of inflamed gingival tissue by the exudate (gingival crevicular fluid) to be re-united with the bacterial biofilm and facilitate solute exchange with cell-bound receptors. In *P. gingivalis* OMVs, all six TDRs identified (besides HmuR which was only barely identified) were underrepresented in OMVs [32]; this trend was also observed in *B. fragilis* and *B. thetaiotaomicron* [44] and was explained by the role of TDRs in acquiring nutrients for the cell, a role presumed to be more suited to cells than OMVs. Since the observed enrichment is relative, it is uncertain whether the enrichment is active, implying a mechanism to target TDRs to the OMVs, or whether it is a result of active depletion of other proteins. The distribution of the TDR-associated lipoproteins generally followed their cognate TDRs with a similar level of enrichment unlike the depletion of other lipoproteins (Figure 3). The difference in behavior of these two groups of lipoproteins may reflect the expected presence of the TDR-associated lipoproteins at the surface of the OM rather than the periplasmic side of the OM. These lipoproteins are mostly related to SusD and are believed to assist with ligand binding [49].

The abundance of TDRs and associated SusD homologs in the OM of *T. forsythia* has been discussed previously, and it was noted that this was usually associated with the ability to utilize a wide range of polysaccharides [18]. The presence of such TDRs on the surface of

OMVs may extend their adhesive properties by facilitating binding to a wider range of substrates. Although most of the TDR-containing loci did not appear to encode glycosyl hydrolases, some such as Tanf_13470-13490 and Tanf_13685-13725 did include these enzymes. Tanf_13470-13490 encodes a starch utilization locus with homology to the SusB locus in *Bacteroides thetaiotaomicron* [49] (Sharma book reference, 2011). In addition to the TDR (Tanf_13475, SusC) and associated lipoproteins which are all associated with the OM (Tanf_13480, SusD; Tanf_13485, SusE; Tanf_13490, SusF) we identified the α glucosidase SusB (Tanf_13470) and localized it to the vesicle lumen/periplasm (Table S-2.3). Tanf_13685-13725 has recently been characterized as a sialic acid utilization locus, and sialic acid was shown to stimulate biofilm growth [50, 51]. In this study we identified the SusD (Tanf_13705) and SusC (Tanf_13710) homologs, together with enzymes involved in the catabolism of sialic acid (Tanf_13720-13725), and the sialidase (Tanf_13700) which was strongly localized to the vesicle lumen/periplasm. Utilization of sialic acid would appear particularly relevant for *T. forsythia* since it is a common modification of host proteins present in the oral cavity [51]. The importance of carbohydrate utilization for *T. forsythia* was further supported by the identification of all components of the glycolysis pathway (except glucokinase) (Figure S-1) and 12 predicted glycosyl hydrolases (Table S-2.3) including three predicted β-galactosidases and two mannosidases. A recently described fucosidase (Tanf_06770) was also identified [52].

Lipoproteins

Of the 173 proteins localized to the vesicle membrane, 98 were predicted lipoproteins (Table S-1). Besides those lipoproteins associated with TDRs (see above), most lipoproteins were strongly retained in the cell outer membrane (Figure 3). Examples of these included enzymes such as peptidases (Tanf_07525 & Tanf_08225), glycosyl hydrolase (Tanf_04515) and

peptidyl-prolyl cis-trans isomerase (Tanf_00225) (Figure 3). Non-lipoproteins of similar predicted functions were localized to the vesicle lumen, and therefore it is likely that these lipoprotein enzymes are anchored to the luminal side of the vesicle membrane. In contrast, lipoproteins known to be anchored to the outer leaflet of the cell outer membrane such as those associated with TDRs (see above) and including the HmuY [32] homolog Tanf_09520, were not depleted in OMVs (Figure 3). It is interesting to speculate that all of the lipoproteins anchored on the periplasmic (luminal) side of the cell outer membrane are selectively retained by the cell either through binding of these proteins to the peptidoglycan or other fixed periplasmic components or else via an OMV biogenesis mechanism that selectively incorporates lipids and proteins from the outer leaflet of the outer membrane and partially excludes the inner leaflet along with lipoproteins anchored to the inner leaflet.

In conclusion, we have characterized the proteome of *T. forsythia* OMVs and provided strong evidence for the localization of the proteins identified. The OMVs are enriched with substrates of the T9SS which are likely to represent the most important cargo with respect to virulence properties.

FIGURE LEGENDS

Figure 1. SDS-PAGE of *T. forsythia* fractions used for proteome analysis. (A) SDS-PAGE of the OMV samples showing how the gel lanes were excised into 22 sections. (B) SDS-PAGE of cellular fractions showing how the gel lanes were excised into 16 sections. V1 & V2, OMV replicates 1 & 2; P, Periplasm fraction; M1 & M2, Membrane fraction replicates 1 & 2; S, Soluble fraction containing both cytoplasm and periplasm.

Figure 2. Purity of *T. forsythia* OMV preparations. (A) The top scoring 25 predicted cytoplasmic proteins from the soluble fraction are plotted showing the total spectral counts obtained for these proteins from all four fractions. Notably, none were found in the OMV fraction. (B) The top scoring 25 predicted inner membrane proteins are also plotted for all four fractions (see 'Methods'). Of these, only Tanf_11185, Tanf_12470 and Tanf_12490 were detected in the OMV fraction.

Figure 3. Selective cargo sorting to the OMV membrane in *T. forsythia*. Proteomic analyses were conducted using new samples that were separated by SDS-PAGE for a short time such that the whole stained gel portion could be digested and analysed as one band for each fraction. The spectral count for proteins from the OMV fraction were compared to the cell membrane fraction for proteins previously localized to the OMV membrane. Each protein group is shown in a different color according to the legend (inset). The dashed trendline represents all proteins such that proteins above the trendline are relatively enriched in OMVs while proteins below the trendline are relatively depleted. Proteins mentioned in the text are marked with their respective 'Tanf' locus numbers or common names. The data for this figure can be found in Table S-3.3. Please note the log scale.

Figure 4. Cryo-TEM analysis of *T. forsythia* OMVs. The image shows OMVs of various shapes and sizes. Most OMVs are fully covered with a surface layer of around 30 nm in thickness.

Figure 5. *T. forsythia* OMVs comprise mostly of proteins localized to the membrane and vesicle surface region. The total average spectral counts obtained for proteins within each OMV localization were summed to give an estimate of the total protein content (by mass) of that locale. (A) Spectral count data from initial experiment found in Table S-2.3. (B) Spectral count data from cargo sorting experiment found in Table S-3.3. VM, vesicle membrane; VL, vesicle lumen; VS, vesicle surface (T9SS substrates); ?, uncertain location; TDRs, TonB-dependent receptors; LPs, lipoproteins.

ASSOCIATED CONTENT

Supporting Information

Supplementary Tables is a single Excel file comprising nine tables (sheets). Table S-1 is a summary table listing the proteins of the OMV proteome grouped according to localization with key identification and bioinformatic data included. Table S-2 consists of five worksheets labeled S-2.1 to S-2.5. Sheet 2.1 contains the complete protein identification data for the main experiment while sheet 2.2 contains the complete filtered peptide identification data. Sheet 2.3 contains the complete processed data used for both OMV and cellular proteomes, OMV purity and OMV localization. Sheet 2.4 contains the protein sequences that were used to replace incorrect database entries while sheet 2.5 contains the MS identification data for signal peptides. Table S-3 contains three tables (S-3.1 – S-3.3) that detail the separate analyses used for determining cargo sorting. Sheet 3.1 and 3.2 contain the complete protein and peptide data respectively, while sheet 3.3 contains the processed data used to create Figure 3.

Figure S-1 is a pdf file showing identified enzymes within metabolic pathways.

This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

Professor Eric C. Reynolds, Melbourne Dental School, The University of Melbourne, 720 Swanston Street, Melbourne, Victoria, 3010, Australia. Email: e.reynolds@unimelb.edu.au. Phone +61 3 9341 1547. Fax +61 3 9341 1597

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

This study was funded by the Oral Health Cooperative Research Centre, Australia.

ACKNOWLEDGEMENTS

We thank Valentin Friedrich, Department of NanoBiotechnology, NanoGlycobiology Unit, Universität für Bodenkultur Wien, Vienna, Austria for providing the protein sequence database for the *T. forsythia* ATCC 43037 strain. Dr. Eric Hanssen, Dr. Sergey Rubanov and Mr. Roger Curtain of the Bio21 EM facility are thanked for their advice and assistance with EM.

ABBREVIATIONS

OM, outer membrane; OMVs, outer membrane vesicles; IM, inner membrane; IMPs, inner membrane proteins; T9SS, type IX secretion system; CTD, C-terminal domain; TMBB, trans-membrane beta-barrel; M/P, membrane:periplasm ratio of spectral counts; M/S, membrane:soluble ratio of spectral counts; EDSL, electron dense surface layer; TDR, TonB-dependent receptor; LP, lipoprotein.

REFERENCES

(1) Wiebe, C. B.; Putnins, E. E. The periodontal disease classification system of the American Academy of Periodontology--an update. *J. Can. Dent. Assoc.* **2000,** *66*, 594-597.

(2) Linden, G. J.; Lyons, A.; Scannapieco, F. A. Periodontal systemic associations: review of the evidence. *J. Periodontol.* **2013,** *84*, S8-S19.

(3) Tonetti, M. S.; Van Dyke, T. E.; working group 1 of the joint, E. F. P. A. A. P. w. Periodontitis and atherosclerotic cardiovascular disease: consensus report of the Joint EFP/AAP Workshop on Periodontitis and Systemic Diseases. *J. Periodontol.* **2013,** *84*, S24-29.

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

(5) Haffajee, A. D.; Cugini, M. A.; Tanner, A.; Pollack, R. P.; Smith, C.; Kent, R. L., Jr.; Socransky, S. S. Subgingival microbiota in healthy, well-maintained elder and periodontitis subjects. *J. Clin. Periodontol.* **1998,** *25*, 346-353.

(6) Tanner, A. C. Characterization of *Wolinella spp*., *Campylobacter concisus*, *Bacteroides gracilis*, and *Eikenella corrodens* by polyacrylamide gel electrophoresis. *J. Clin. Microbiol.* **1986,** *24*, 562-565.

(7) Sharma, A.; Inagaki, S.; Honma, K.; Sfintescu, C.; Baker, P. J.; Evans, R. T. *Tannerella forsythia*-induced alveolar bone loss in mice involves leucine-rich-repeat BspA protein. *J. Dent. Res.* **2005,** *84*, 462-467.

(8) Kesavalu, L.; Sathishkumar, S.; Bakthavatchalu, V.; Matthews, C.; Dawson, D.; Steffen, M.; Ebersole, J. L. Rat model of polymicrobial infection, immunity, and alveolar bone resorption in periodontal disease. *Infect. Immun.* **2007,** *75*, 1704-1712.

(9) Sharma, A. Virulence mechanisms of *Tannerella forsythia*. *Periodontol. 2000* **2010,** *54*, 106-116.

(10) Ksiazek, M.; Mizgalska, D.; Eick, S.; Thogersen, I. B.; Enghild, J. J.; Potempa, J. KLIKK proteases of *Tannerella forsythia:* putative virulence factors with a unique domain structure. *Front. Microbiol.* **2015,** *6*, 312.

(11) Jusko, M.; Potempa, J.; Karim, A. Y.; Ksiazek, M.; Riesbeck, K.; Garred, P.; Eick, S.; Blom, A. M. A metalloproteinase karilysin present in the majority of *Tannerella forsythia* isolates inhibits all pathways of the complement system. *J. Immunol.* **2012,** *188*, 2338-2349.

(12) Jusko, M.; Potempa, J.; Mizgalska, D.; Bielecka, E.; Ksiazek, M.; Riesbeck, K.; Garred, P.; Eick, S.; Blom, A. M. A Metalloproteinase Mirolysin of Tannerella forsythia Inhibits All Pathways of the Complement System. *J. Immunol.* **2015,** *195*, 2231-2240.

(13) Ksiazek, M.; Karim, A. Y.; Bryzek, D.; Enghild, J. J.; Thogersen, I. B.; Koziel, J.; Potempa, J. Mirolase, a novel subtilisin-like serine protease from the periodontopathogen *Tannerella forsythia*. *Biol. Chem.* **2015,** *396*, 261-275.

(14) Higuchi, N.; Murakami, Y.; Moriguchi, K.; Ohno, N.; Nakamura, H.; Yoshimura, F. Localization of major, high molecular weight proteins in *Bacteroides forsythus*. *Microbiol. Immunol.* **2000,** *44*, 777-780.

(15) Lee, S. W.; Sabet, M.; Um, H. S.; Yang, J.; Kim, H. C.; Zhu, W. Identification and characterization of the genes encoding a unique surface (S-) layer of *Tannerella forsythia*. *Gene* **2006,** *371*, 102-111.

(16) Sekot, G.; Posch, G.; Oh, Y. J.; Zayni, S.; Mayer, H. F.; Pum, D.; Messner, P.; Hinterdorfer, P.; Schaffer, C. Analysis of the cell surface layer ultrastructure of the oral pathogen *Tannerella forsythia*. *Arch. Microbiol.* **2012,** *194*, 525-539.

(17) Oh, Y. J.; Sekot, G.; Duman, M.; Chtcheglova, L.; Messner, P.; Peterlik, H.; Schaffer, C.; Hinterdorfer, P. Characterizing the S-layer structure and anti-S-layer antibody recognition on intact *Tannerella forsythia* cells by scanning probe microscopy and small angle X-ray scattering. *J. Mol. Recognit.* **2013,** *26*, 542-549.

(18) Veith, P. D.; O'Brien-Simpson, N. M.; Tan, Y.; Djatmiko, D. C.; Dashper, S. G.; Reynolds, E. C. Outer membrane proteome and antigens of *Tannerella forsythia*. *J. Proteome Res.* **2009,** *8*, 4279-4292.

(19) Sato, K.; Naito, M.; Yukitake, H.; Hirakawa, H.; Shoji, M.; McBride, M. J.; Rhodes, R. G.; Nakayama, K. A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107*, 276-281.

(20) Glew, M. D.; Veith, P. D.; Peng, B.; Chen, Y. Y.; Gorasia, D. G.; Yang, Q.; Slakeski, N.; Chen, D.; Moore, C.; Crawford, S.; Reynolds, E. C. PG0026 is the C-terminal signal peptidase of a novel secretion system of *Porphyromonas gingivalis*. *J. Biol. Chem.* **2012,** *287*, 24605-24617.

(21) Veith, P. D.; Nor Muhammad, N. A.; Dashper, S. G.; Likic, V. A.; Gorasia, D. G.; Chen, D.; Byrne, S. J.; Catmull, D. V.; Reynolds, E. C. Protein substrates of a novel secretion system are numerous in the Bacteroidetes phylum and have in common a cleavable C-terminal secretion signal, extensive post-translational modification, and cell-surface attachment. *J. Proteome Res.* **2013,** *12*, 4449-4461.

(22) Gorasia, D. G.; Veith, P. D.; Chen, D.; Seers, C. A.; Mitchell, H. A.; Chen, Y. Y.; Glew, M. D.; Dashper, S. G.; Reynolds, E. C. *Porphyromonas gingivalis* type IX secretion substrates are cleaved and modified by a sortase-like mechanism. *PLoS Pathog* **2015,** *11*, e1005152.

(23) Narita, Y.; Sato, K.; Yukitake, H.; Shoji, M.; Nakane, D.; Nagano, K.; Yoshimura, F.; Naito, M.; Nakayama, K. Lack of a surface layer in *Tannerella forsythia* mutants deficient in the type IX secretion system. *Microbiology* **2014,** *160*, 2295-2303.

(24) Tomek, M. B.; Neumann, L.; Nimeth, I.; Koerdt, A.; Andesner, P.; Messner, P.; Mach, L.; Potempa, J. S.; Schaffer, C. The S-layer proteins of *Tannerella forsythia* are secreted via a type IX secretion system that is decoupled from protein O-glycosylation. *Mol. Oral Microbiol.* **2014,** *29*, 307-320.

(25) Pham, T. K.; Roy, S.; Noirel, J.; Douglas, I.; Wright, P. C.; Stafford, G. P. A quantitative proteomic analysis of biofilm adaptation by the periodontal pathogen *Tannerella forsythia*. *Proteomics* **2010,** *10*, 3130-3141.

(26) Ellis, T. N.; Kuehn, M. J. Virulence and immunomodulatory roles of bacterial outer membrane vesicles. *Microbiol. Mol. Biol. Rev.* **2010,** *74*, 81-94.

(27) Kesty, N. C.; Mason, K. M.; Reedy, M.; Miller, S. E.; Kuehn, M. J. Enterotoxigenic *Escherichia coli* vesicles target toxin delivery into mammalian cells. *EMBO J.* **2004,** *23*, 4538-4549.

(28) Schooling, S. R.; Beveridge, T. J. Membrane vesicles: an overlooked component of the matrices of biofilms. *J. Bacteriol.* **2006,** *188*, 5945-5957.

(29) Inagaki, S.; Onishi, S.; Kuramitsu, H. K.; Sharma, A. *Porphyromonas gingivalis* vesicles enhance attachment, and the leucine-rich repeat BspA protein is required for invasion of epithelial cells by "*Tannerella forsythia*". *Infect. Immun.* **2006,** *74*, 5023-5028.

(30) O'Brien-Simpson, N. M.; Pathirana, R. D.; Walker, G. D.; Reynolds, E. C. *Porphyromonas gingivalis* RgpA-Kgp proteinase-adhesin complexes penetrate gingival tissue and induce proinflammatory cytokines or apoptosis in a concentration-dependent manner. *Infect. Immun.* **2009,** *77*, 1246-1261.

(31) O'Brien-Simpson, N. M.; Veith, P. D.; Dashper, S.; Reynolds, E. *Porphyromonas gingivalis* gingipains: the molecular teeth of a microbial vampire. *Curr. Protein Peptide Sci.* **2003,** *4*, 409-426.

(32) Veith, P. D.; Chen, Y. Y.; Gorasia, D. G.; Chen, D.; Glew, M. D.; O'Brien-Simpson, N. M.; Cecil, J. D.; Holden, J. A.; Reynolds, E. C. *Porphyromonas gingivalis* outer membrane vesicles exclusively contain outer membrane and periplasmic proteins and carry a cargo enriched with virulence factors. *J. Proteome Res.* **2014,** *13*, 2420-2432.

(33) Acevedo, R.; Fernandez, S.; Zayas, C.; Acosta, A.; Sarmiento, M. E.; Ferro, V. A.; Rosenqvist, E.; Campa, C.; Cardoso, D.; Garcia, L.; Perez, J. L. Bacterial outer membrane vesicles and vaccine applications. *Front. Immunol.* **2014,** *5*, 121.

(34) Lee, E. Y.; Choi, D. S.; Kim, K. P.; Gho, Y. S. Proteomics in gram-negative bacterial outer membrane vesicles. *Mass Spectrom. Rev.* **2008,** *27*, 535-555.

(35) Friedrich, V.; Gruber, C.; Nimeth, I.; Pabinger, S.; Sekot, G.; Posch, G.; Altmann, F.; Messner, P.; Andrukhov, O.; Schaffer, C. Outer membrane vesicles of *Tannerella forsythia*: Biogenesis, composition, and virulence. *Mol. Oral Microbiol.* **2015,** *doi: 10.1111/omi.12104*.

(36) Chen, Y. Y.; Peng, B.; Yang, Q.; Glew, M. D.; Veith, P. D.; Cross, K. J.; Goldie, K. N.; Chen, D.; O'Brien-Simpson, N.; Dashper, S. G.; Reynolds, E. C. The outer membrane protein LptO is essential for the O-deacylation of LPS and the co-ordinated secretion and attachment of A-LPS and CTD proteins in *Porphyromonas gingivalis*. *Mol. Microbiol.* **2011,** *79*, 1380-1401.

(37) Friedrich, V.; Pabinger, S.; Chen, T.; Messner, P.; Dewhirst, F. E.; Schaffer, C. Draft genome sequence of *Tannerella forsythia* type strain ATCC 43037. *Genome Annoc.* **2015,** *3*, e00660-00615.

(38) Freeman, T. C., Jr.; Landry, S. J.; Wimley, W. C. The prediction and characterization of YshA, an unknown outer-membrane protein from *Salmonella typhimurium*. *Biochim. Biophys. Acta* **2011,** *1808*, 287-297.

(39) Pugsley, A. P. The complete general secretory pathway in gram-negative bacteria. *Microbiol. Rev.* **1993,** *57*, 50-108.

(40) Meuric, V.; Rouillon, A.; Chandad, F.; Bonnaure-Mallet, M. Putative respiratory chain of *Porphyromonas gingivalis*. *Future Microbiol.* **2010,** *5*, 717-734.

(41) Thomas, S.; Holland, I. B.; Schmitt, L. The Type 1 secretion pathway - the hemolysin system and beyond. *Biochim. Biophys. Acta* **2014,** *1843*, 1629-1641.

(42) Noinaj, N.; Guillier, M.; Barnard, T. J.; Buchanan, S. K. TonB-dependent transporters: regulation, structure, and function. *Annu. Rev. Microbiol.* **2010,** *64*, 43-60.

(43) Choi, C. W.; Park, E. C.; Yun, S. H.; Lee, S. Y.; Lee, Y. G.; Hong, Y.; Park, K. R.; Kim, S. H.; Kim, G. H.; Kim, S. I. Proteomic characterization of the outer membrane vesicle of *pseudomonas putida* KT2440. *J. Proteome Res.* **2014,** *13*, 4298-4309.

(44) Elhenawy, W.; Debelyy, M. O.; Feldman, M. F. Preferential packing of acidic glycosidases and proteases into Bacteroides outer membrane vesicles. *MBio* **2014,** *5*, e00909-00914.

(45) Shoji, M.; Sato, K.; Yukitake, H.; Kondo, Y.; Narita, Y.; Kadowaki, T.; Naito, M.; Nakayama, K. Por secretion system-dependent secretion and glycosylation of *Porphyromonas gingivalis* hemin-binding protein 35. *PLoS One* **2011,** *6*, e21372.

(46) Messner, P.; Schaffer, C.; Kosma, P. Bacterial cell-envelope glycoconjugates. *Adv. Carbohydr. Chem. Biochem.* **2013,** *69*, 209-272.

(47) Fagan, R. P.; Fairweather, N. F. Biogenesis and functions of bacterial S-layers. *Nat. Rev. Microbiol.* **2014,** *12*, 211-222.

(48) Sakakibara, J.; Nagano, K.; Murakami, Y.; Higuchi, N.; Nakamura, H.; Shimozato, K.; Yoshimura, F. Loss of adherence ability to human gingival epithelial cells in S-layer protein-deficient mutants of *Tannerella forsythensis*. *Microbiology* **2007,** *153*, 866-876.

(49) Cho, K. H.; Salyers, A. A. Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by *Bacteroides thetaiotaomicron*. *J. Bacteriol.* **2001,** *183*, 7224-7230.

(50) Phansopa, C.; Roy, S.; Rafferty, J. B.; Douglas, C. W.; Pandhal, J.; Wright, P. C.; Kelly, D. J.; Stafford, G. P. Structural and functional characterization of NanU, a novel high-affinity sialic acid-inducible binding protein of oral and gut-dwelling *Bacteroidetes* species. *Biochem. J.* **2014,** *458*, 499-511.

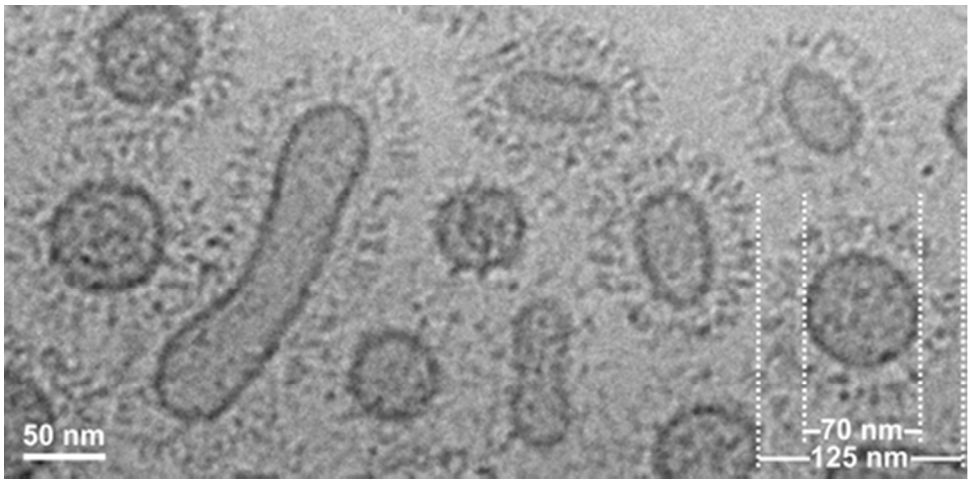
(51) Roy, S.; Douglas, C. W.; Stafford, G. P. A novel sialic acid utilization and uptake system in the periodontal pathogen *Tannerella forsythia*. *J. Bacteriol.* **2010,** *192*, 2285-2293.

(52) Megson, Z. A.; Koerdt, A.; Schuster, H.; Ludwig, R.; Janesch, B.; Frey, A.; Naylor, K.; Wilson, I.; Stafford, G. P.; Messner, P.; Schaffer, C. Characterization of an alpha-l-fucosidase from the periodontal pathogen *Tannerella forsythia*. *Virulence* **2015,** *6*, 282-292.

Table 1. Identified components of the Type IX Secretion System in *T. forsythia*

T9SS Protein	PG Locus	Tanf Locus	Mascot Score	Best Fraction[a]	Localization
PorK	PG0288	Tanf_02360	442	Membrane	OM [19]
PorL	PG0289	Tanf_02365	623	Membrane	IM [19]
PorM	PG0290	Tanf_02370	597	Membrane	IM [19]
PorN	PG0291	Tanf_02375	394	Membrane	OM [19]
PorP	PG0287	Tanf_02355	64	Membrane	OM [19, 32]
PorQ	PG0602	Tanf_12465	112	OMV	OM [32]
PorT	PG0751	Tanf_10520	53	Membrane	OM [32, 36]
PorU	PG0026	Tanf_02580	232	Membrane	OM/EC [20, 32]
PorV, LptO	PG0027	Tanf_04220	372	Membrane	OM [32, 36]
PorW	PG1947	Tanf_00060	33	Membrane	unknown
Sov	PG0809	Tanf_04410	-	not identified	OM (predicted)

[a] The "best fraction" is the fraction where the protein was identified with the highest Mascot score.



TOC graphic
40x19mm (300 x 300 DPI)

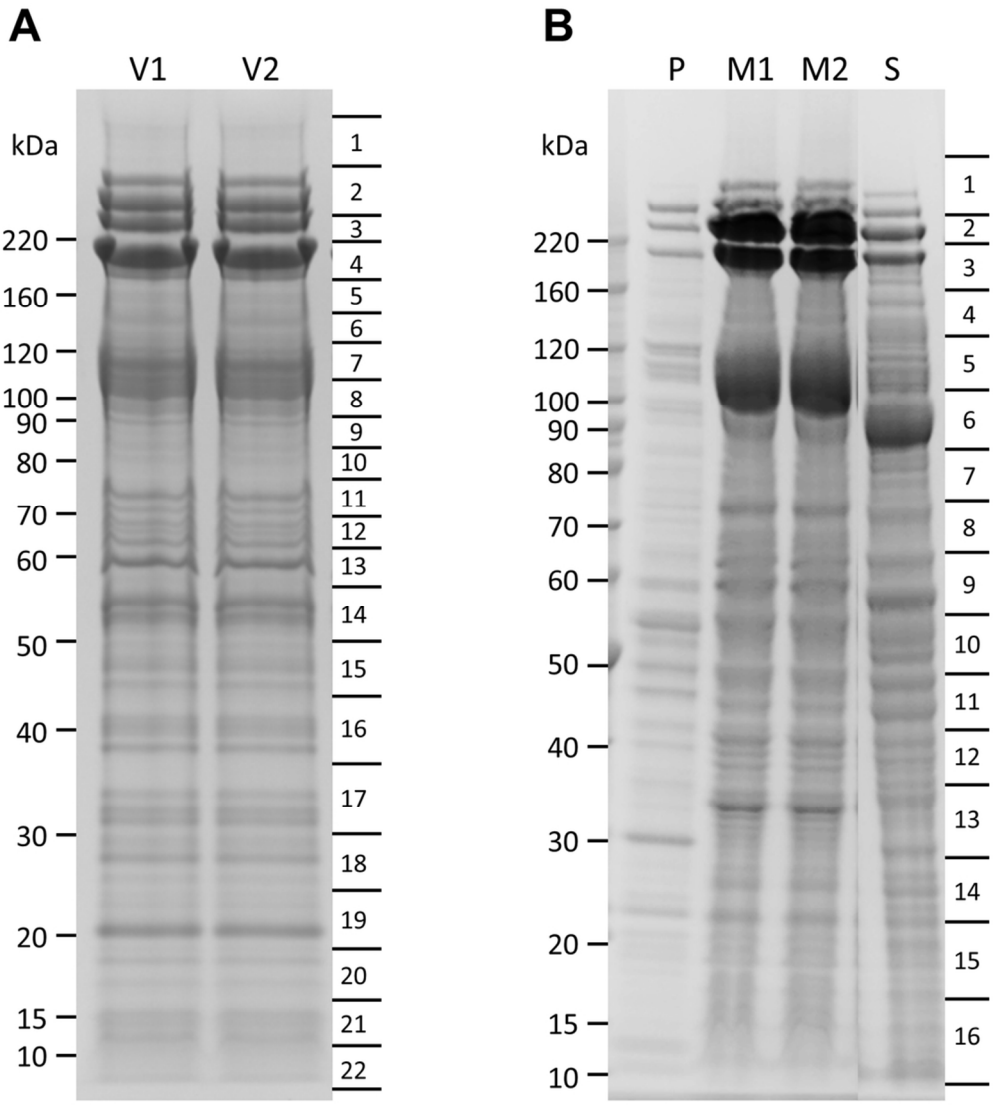


Figure 1. SDS-PAGE of T. forsythia fractions used for proteome analysis. (A) SDS-PAGE of the OMV samples showing how the gel lanes were excised into 22 sections. (B) SDS-PAGE of cellular fractions showing how the gel lanes were excised into 16 sections. V1 & V2, OMV replicates 1 & 2; P, Periplasm fraction; M1 & M2, Membrane fraction replicates 1 & 2; S, Soluble fraction containing both cytoplasm and periplasm.
91x101mm (300 x 300 DPI)

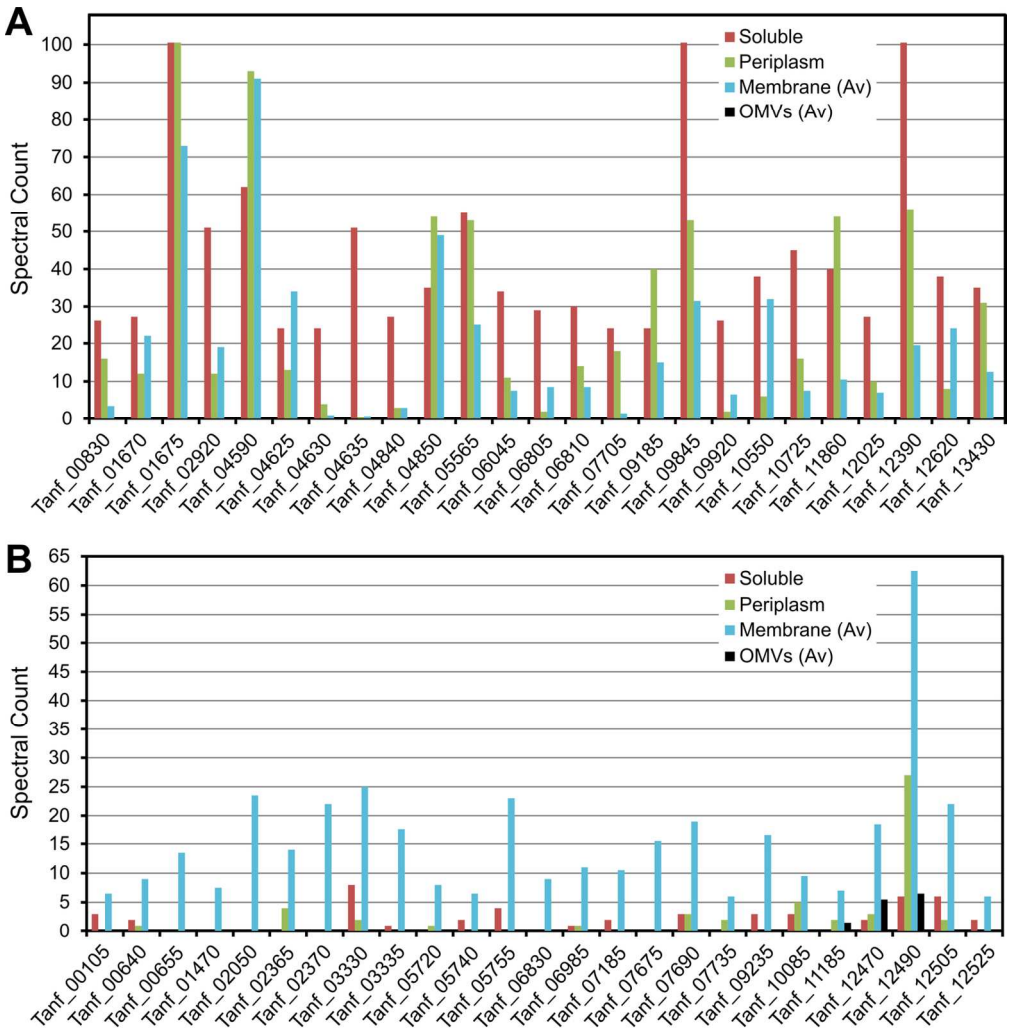


Figure 2. Purity of T. forsythia OMV preparations. (A) The top scoring 25 predicted cytoplasmic proteins from the soluble fraction are plotted showing the total spectral counts obtained for these proteins from all four fractions. Notably, none were found in the OMV fraction. (B) The top scoring 25 predicted inner membrane proteins are also plotted for all four fractions (see 'Methods'). Of these, only Tanf_11185, Tanf_12470 and Tanf_12490 were detected in the OMV fraction.
143x146mm (300 x 300 DPI)

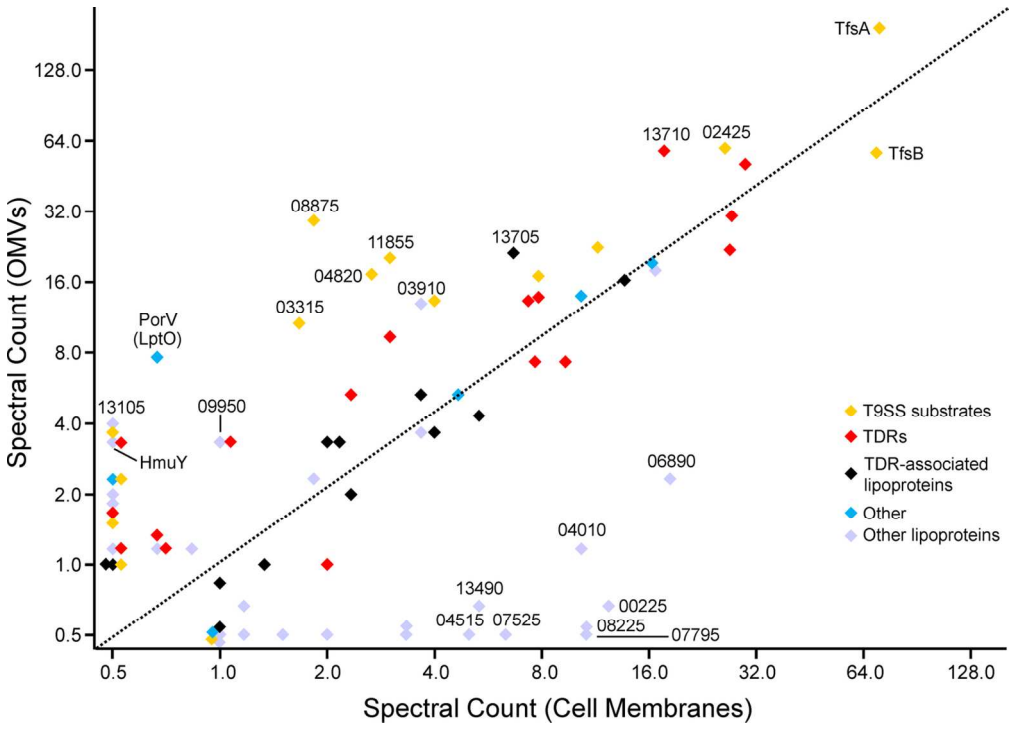


Figure 3. Selective cargo sorting to the OMV membrane in T. forsythia. Proteomic analyses were conducted using new samples that were separated by SDS-PAGE for a short time such that the whole stained gel portion could be digested and analysed as one band for each fraction. The spectral count for proteins from the OMV fraction were compared to the cell membrane fraction for proteins previously localized to the OMV membrane. Each protein group is shown in a different color according to the legend (inset). The dashed trendline represents all proteins such that proteins above the trendline are relatively enriched in OMVs while proteins below the trendline are relatively depleted. Proteins mentioned in the text are marked with their respective 'Tanf' locus numbers or common names. The data for this figure can be found in Table S-3.3. Please note the log scale.
127x90mm (300 x 300 DPI)

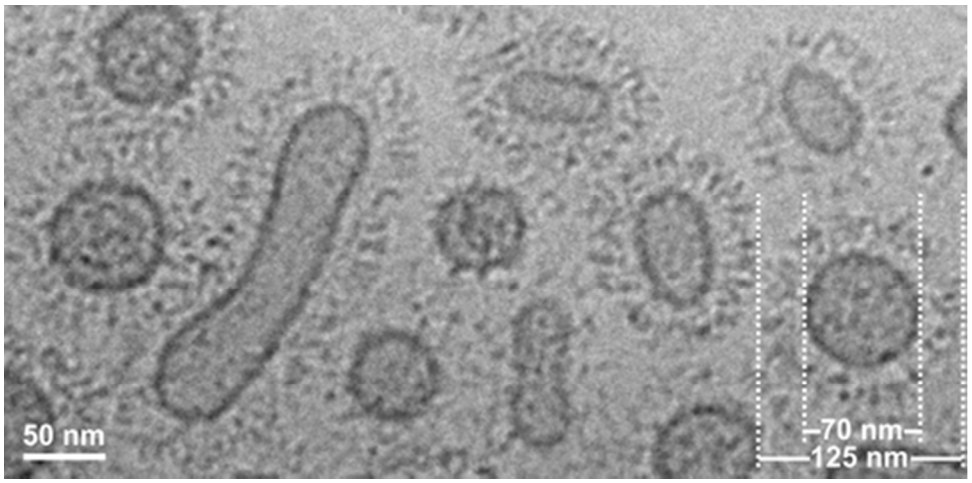


Figure 4. Cryo-TEM analysis of T. forsythia OMVs. The image shows OMVs of various shapes and sizes. Most OMVs are fully covered with a surface layer of around 30 nm in thickness.
40x19mm (300 x 300 DPI)

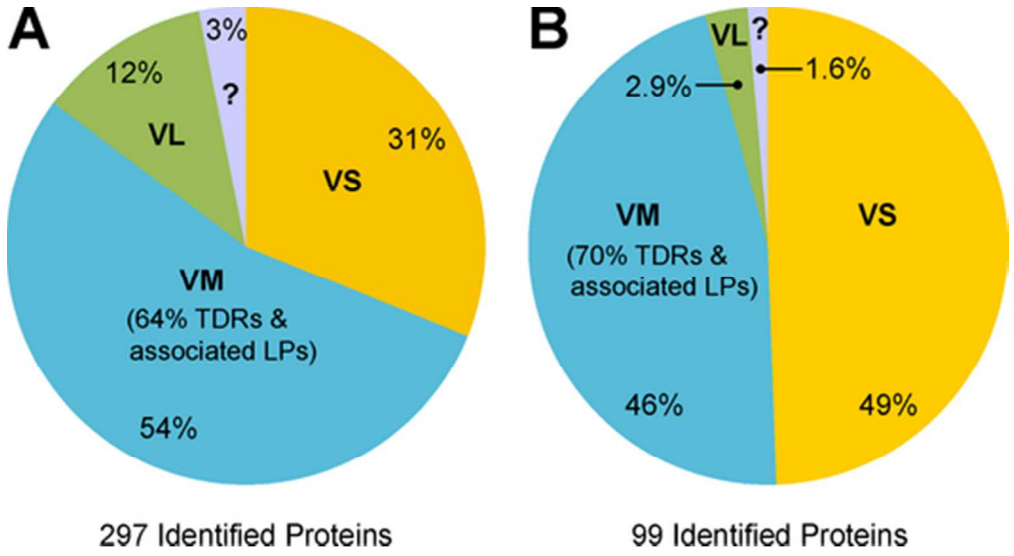


Figure 5. T. forsythia OMVs comprise mostly of proteins localized to the membrane and vesicle surface region. The total average spectral counts obtained for proteins within each OMV localization were summed to give an estimate of the total protein content (by mass) of that locale. (A) Spectral count data from initial experiment found in Table S-2.3. (B) Spectral count data from cargo sorting experiment found in Table S-3.3. VM, vesicle membrane; VL, vesicle lumen; VS, vesicle surface (T9SS substrates); ?, uncertain location; TDRs, TonB-dependent receptors; LPs, lipoproteins.
44x24mm (300 x 300 DPI)