THE UNIVERSITY
of EDINBURGH

Characterisation of proteins secreted in the outer membrane vesicles of *Bacteroides fragilis*

Maria Theresa Kowal

THE UNIVERSITY *of* EDINBURGH

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

2016

Declaration

All methods and analyses described in this study were performed by the author unless stated otherwise. No part of this thesis has been submitted for any other degree or qualification.

Acknowledgements

There are many people without whom I would never have completed this PhD. Most importantly, I would like to thank my supervisor Dr Garry Blakely for introducing me to the fascinating organism known as *Bacteroides fragilis* and for having faith in me from day one. I would also like to thank my second supervisor, Dr Maurice Gallagher, for his support and advice, especially over the last, and most trying, year of my PhD. Many of the techniques I have learned and used in the course of my PhD were taught to me by third parties, and I would like to thank Dr Sarah Martin, Dr Thierry Le Bihan, John Hopkins, Dr Martin Wear, Dr Liz Blackburn and Dr Jo Stevens for their significant contributions to my work. I am especially grateful to Prof. Sheila Patrick of Queen's University Belfast for her unending support and encouragement and for the opportunity to spend two months conducting research in her lab.

I am grateful to my former lab mates Aparna, David, Kelly, Harriet and Eleana for all the help and company they provided over the years. I give extra special thanks for and to Adam, Alex, Liz, Helen, Matt, Sam, Stephen, Christian, Vic and Rachel for getting me through my undergraduate degree with pub quizzes, and to Alex, Claire, Tom and Steven for getting me through my postgraduate degree with board games and even more pub quizzes. I'm happy to have had the EUBDS and all its members to provide a much-needed distraction, and I have been very lucky to have Caroline Fonseca force her way into my life; I'm not sure I can imagine it without her. I am especially grateful to Dr Job Thijssen, who put up with me, and helped me for much longer than I thought anyone would.

Finally, I would like to thank my wonderful Mum and step father, Tracey and Kevin Cardew for their moral (and financial) support, as well as my father and sister, Anthony and Anna Kowal, who have always believed in me.

I would like to consider this work a dedication to my grandmothers, Theresa Margaret Ella Cave and Adela Mauvina Kowal, both of whom passed away during my PhD. I would have liked for both of them to have had the chance to read it, even if they wouldn't have understood a word.

Contents

List of Figures

List of Tables

Abbreviations

ADP	Adenosine diphosphate
amp	Ampicillin
AMP	Adenosine monophosphate
ANOVA	Analysis of variance
APS	Ammonium persulphate
ATP	Adenosine triphosphate
BHI-S	Brain-heart infusion, with supplements
BLAST	Basic local alignment search tool
Cfu	Colony-forming units
CPS	Capsular polysaccharide
DC	Dendritic cell
DLS	Dynamic light scattering
DM	Defined medium
DNA	Deoxyribonucleic acid
DOMV	Detergent-extracted outer membrane vesicles
DTT	Dithiothreitol
DUB	Deubiquitylating enzyme
EDL	Electron dense layer
EDTA	Ethylenediaminetetraacetic acid
EPEC	Enteropathogenic *Escherichia coli*
ery	Erythromycin
ETBF	Enterotoxigentic *Bacteroides Fragilis*
ETEC	Enterotoxigenic *Escherichia coli*
FITC	Fluorescein isothiocyanate
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
gent	Gentamicin
GF	Germ free (mice)
GG	Double glycine
GI	Gastro-intestinal
HGT	Horizontal gene transfer
HPLC	High performance liquid chromatography
IBD	Inflammatory bowel disease

IM	Inner membrane
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
iTRAQ	Isobaric tag for relative and absolute quantification
LB	Luria-Bertani Broth
LC	Large capsule
LC-MS	Liquid chromatography-mass spectrometry
LPS	Lipopolysaccharide
LT	Labile toxin
LTQ	Linear ion trap
MC	Microcapsule
MIC	Minimum inhibitory concentration
MS	Mass spectrometry
NCBI	National Centre for Biotechnological Information
NOMV	Native outer membrane vesicles
OM	Outer membrane
OMV	Outer membrane vesicles
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PG	Peptidoglycan
PP	Periplasm
PSA	Polysaccharide A
PTAG	Phospho-tag
PVDF	Polyvinylidene fluoride
qPCR	Quantitative polymerase chain reaction
R/M	Restriction/modification
RNA	Ribonucleic acid
SC	Small capsule
SCF	Skp, cullin, F-box complex
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SOMV	Spontaneous outer membrane vesicles
SPR	Surface plasmon resonance
TEM	Transmission electron microscopy

TEMED	Tetramethylethylenediamine
tet	Tetracycline
TFF	Tangential flow filtration
TMT	Tandem mass tag
TNFα	Tumour necrosis factor α
TRITC	Tetramethylrhodamine
UBD	Ubiquitin-binding domain
UBL	Ubiquitin-like protein
UbVME	Ubiquitin vinyl methylester
UDP	Ubiquitin domain protein
ULD	Ubiquitin-like domain
ULF	Ultra-low flow
WT	Wild type
YFP	Yellow fluorescent protein
YY	Double tyrosine
α2m	α-2-Macroglobulin

Lay Summary

Bacteroides fragilis is a bacterial species, found in the gut of most humans, which is known to promote host health and immune development. *B. fragilis*, like many bacteria, produces outer membrane vesicles (OMV), small globules, containing cellular proteins, which protrude from the bacterial surface and are released into the surrounding environment. OMV from various bacteria are proposed to have a range of functions, and *B. fragilis* OMV are thought to be able to deliver their contents to cells of the human host. One of the proteins believed to be packaged into these OMV is BfUbb, which is similar to a human protein named ubiquitin. Ubiquitin plays an important role in many cell processes and the similarity of BfUbb to this protein suggests that BfUbb may interfere with these processes.

This study demonstrates that BfUbb is able to interact with several important ubiquitin-related enzymes and shows particular preference for enzymes which are known targets for anti-cancer therapies. This suggests BfUbb may play a role in preventing colon cancer. BfUbb was also found to inhibit the growth of a known gut pathogen, *Salmonella enterica* Typhimurium, suggesting that *B. fragilis* strains which produce BfUbb-containing OMV may be able to protect against infection in the gut. In addition, this work describes the creation of several DNA constructs which will facilitate the continued study of BfUbb.

Proteomics is the term given to the large scale identification and characterisation of proteins. A proteomic study of the OMV of *B. fragilis* was performed using highly sensitive techniques, and a list of proteins believed to be present in the OMV was produced. From this list, potential functions of the OMV could be inferred. For example, a large number of proteins on this list are known to be involved in the breakdown of complex nutrients, suggesting that OMV may play a role in scavenging for "food" for the bacteria.

Overall, the results of this study provide further evidence that the presence of *B. fragilis* in our gut has a beneficial effect on our health.

Abstract

Bacteroides fragilis is an important, anaerobic commensal of the human gastro-intestinal tract. As a Gram-negative bacterium, *B. fragilis* produces a large number of outer membrane vesicles (OMV), spherical globules consisting of outer membrane and periplasmic material, which have a range of potential functions and which are known to be able to deliver their cargo to host dendritic cells (DCs). One of the proteins believed to be packaged into the OMV of *B. fragilis* is BfUbb (encoded by the *ubb* gene) which shares 63% homology with human ubiquitin. Ubiquitin is a small, common, eukaryotic protein modifier, which is conjugated to target proteins via a series of activating, conjugating and ligating enzymes, and which has known roles in a wide range of eukaryotic cell processes.

Due to key differences between the two proteins, BfUbb has the potential to act as a suicide substrate mimic of ubiquitin. BfUbb was therefore assayed for its ability to interact with ubiquitin E2 conjugating enzymes of the ubiquitylation cascade *in vitro*, and was found to covalently bind the majority of available enzymes in a DTT-sensitive manner. BfUbb showed a preference for three specific E2 enzymes, all of which are involved in the degradation of mitotic check point proteins, suggesting a role for BfUbb in the inhibition of cell cycle progression and, consequently, tumorigenesis. No binding partners of BfUbb were identified outside of the ubiquitylation cascade, however BfUbb was found to form spontaneous multimers *in vitro*, the biological function of which is unknown. This study also describes the construction of two sets of plasmids. The first set will allow the expression of untagged and fluorescently tagged forms of BfUbb for purification and use in biochemical assays. The second set will allow the expression of his-tagged and fluorescently tagged forms of BfUbb in mammalian cells, so that the effects of BfUbb on the host epithelial cells may be studied.

The proteome of the OMV of *B. fragilis* was solved using LTQ-Orbitrap mass spectrometry. The identified proteins indicated several putative roles for *B. fragilis* OMV, including nutrient acquisition and protease inhibition. The suitability of techniques used during the isolation and proteomic analysis of OMV in different studies is discussed.

BfUbb-carrying *B. fragilis* OMV were able to inhibit growth of *Salmonella enterica* Typhimurium, thus indicating a role for BfUbb in the inhibition of competing, pathogenic bacteria in the gastro-intestinal tract.

The conclusions of this study are that the putative roles of both BfUbb and the OMV of *B. fragilis* may promote both survival of the bacterium and the gastro-intestinal health of the host.

1. Introduction

The human gastro-intestinal (GI) tract is host to ~10^{14} residential bacteria, made up of up to 1,000 distinct species (Backhed *et al.,* 2005). The stable colonisation of the gut by microbiota is very important for the general health of an animal. GI bacteria contribute significantly to the digestion of complex carbohydrates and plant polysaccharides, providing energy from otherwise inaccessible sources. They are also essential to proper immunological development, impacting IgA production, T-cell-dependent immune responses and allergy formation. Various studies have also indicated that the indigenous gut microbiome can protect the host against pathogenic bacteria, such as *Clostridium difficile*, and parasitic infections, such as *Plasmodium* sp. and *Toxoplama gondii* (Backhed *et al*., 2005; Berg., 1996; Denny *et al.,* 2016; Wexler., 2007). Changes to the composition and structure of the bacterial communities in the gut can therefore drastically impact the health of the host and leave the GI tract vulnerable to infection and disease (Fujimura *et al.*, 2010).

Many residential bacteria interact with the cells of their host, via a myriad of mechanisms, to mediate host cell processes and to promote both bacterial growth and host health (Backhed *et al.*, 2005). One of these mechanisms is the production of outer membrane vesicles (OMV), which are formed from the surface of most Gram-negative bacteria and are capable of delivering their contents to other cells, either of the host or of other bacteria (Kulkarni & Jagannadham, 2014). One bacterial resident of the GI tract, *Bacteroides fragilis*, is an important and ubiquitous commensal, believed to promote host immune development and quell inflammatory responses in the gut. As it is Gram-negative, *B. fragilis* produces a large number of OMV, which have already been shown to interact with host dendritic cells (DCs) (Shen *et al.*, 2012).

Recently, *B. fragilis* was also found to encode and express a unique bacterial homologue of eukaryotic ubiquitin, an essential protein modifier with a large number of roles in eukaryotic cells (Patrick *et al.*, 2011). The key differences between native ubiquitin and this homologue suggest that this protein may provide *B. fragilis* with a mechanism for interfering with host cell pathways, such as antigen-sensing and the inflammatory response.

1.1 An overview of *Bacteroides fragilis*

1.1.1 *Bacteroides fragilis*

The human body is host to vast numbers of bacterial cells which outnumber eukaryotic cells 10 to 1 (Fujimura *et al.*, 2010; Karlsson *et al.*, 2011). The gut is the largest reservoir for bacteria, containing ~10^{12} microorganisms per gram of intestinal material. The intestinal microbiome is mostly made up of species from the phyla Firmicutes and Bacteroidetes, the latter of which contains *Bacteroides fragilis* (Karlsson *et al.*, 2011; Troy & Kasper, 2010; Wexler, 2007; Wick & Sears, 2010). *B. fragilis* was originally identified in the late 19th century as *Bacillus fragilis* and later reclassified (Patrick *et al.*, 2010). It makes up 4-13% of total Bacteroidetes species (<1% of total microbiota) and colonises the mucus layer of the lower GI tract of most known mammalian species (Gibson *et al.*, 1998; Huang *et al.*, 2011b; Patrick *et al.*, 2010; Troy & Kasper, 2010; Wexler, 2007). *B. fragilis* is Gram-negative, non-sporulating and a strict anaerobe, however it is aerotolerant owing to an oxidative stress response system (Patrick *et al.*, 2010; Sherwood *et al.*, 2011; Troy & Kasper, 2010; Wexler, 2007). The complete genome sequence of *B. fragilis* NCTC9343 identified a singular ~5.2Mbp, circular chromosome, encoding >4,000 putative proteins, as well as a single plasmid, pBF9343 (Cerdeno-Tarraga *et al.*, 2005).

1.1.2 Key features of *B. fragilis*

A large part of the *B. fragilis* genome is devoted to carbohydrate metabolism, i.e. the degradation of dietary polysaccharides and the production of capsular polysaccharides (CPS) (Cerdeno-Tarraga *et al.*, 2005; Troy & Kasper, 2010), which reflects the fact that this organism has the most complex bacterial CPS system yet discovered (Wexler, 2007; Wick & Sears, 2010). Patrick *et al* (1986) described 3 distinct surface capsules produced by *B. fragilis*: a thin electron dense layer (EDL), also known as the microcapsule (MC), a small

fibrous capsule (SC) and a large fibrous capsule (LC). The three types can be separated from a mixed population by discontinuous Percoll density centrifugation, and the LC and SC can be distinguished by light microscopy, while the MC can only be seen by electron microscopy (Figure 1). There are 8 polysaccharide types (A-H) that can be differentially expressed within the MC, as distinuished by immunofluorescence microscopy (Figure 1 c and d), and epitopes of the MC and the LC can be detected simultaneously, suggesting the two are co-expressed. The SC, meanwhile, is antigenically distinct from both the MC and the LC. There is wide, within-strain and within-population variation of CPS, and cultures enriched for one capsular type will revert to a mixed population over time (Lutton *et al.*, 1991; Patrick *et al.*, 1986; Patrick *et al.*, 2009; Patrick *et al.*, 2010).

Variable expression occurs via the inversion of certain DNA segments in the promoter regions of 7 of the polysaccharide biosynthesis loci of the *B. fragilis* NCTC 3943 genome. Invertible regions, designated *fin* (fragilis invertible) regions, are flanked by inverted repeats, referred to as *fix* (fragilis inversion crossover) regions. Two recombinases, the chromosomally encoded FinA and the plasmid encoded FinB, are responsible for the inversion of the fin regions, switching the promoter of the biosynthesis loci between an "ON" position, in which the promoter is active, and an "OFF" position, in which it is inactivate (Cerdeno-Tarraga *et al.*, 2005; Patrick *et al.*, 2003). Between 3 of the sequenced strains of *B. fragilis* (NCTC 9343, YCH46 and 638R) there are a total of 30 biosynthesis loci, only 2 of which are shared between strains, allowing for the expression of 28 different known CPS types (Patrick *et al.*, 2003). This level of variable polysaccharide expression has not been documented before in other bacterial strains; although there are a large number of *Escherichia coli* O-antigen serotypes available, usually no more than one serotype will be expressed by a single strain (Patrick *et al.*, 2009). The ability to express so many CPS variations is believed to confer multiple advantages to *B. fragilis*, including improved colonisation of the gastro-intestinal tract (Mazmanian *et al.*, 2005) and evasion of the host immune system (Wexler, 2007).

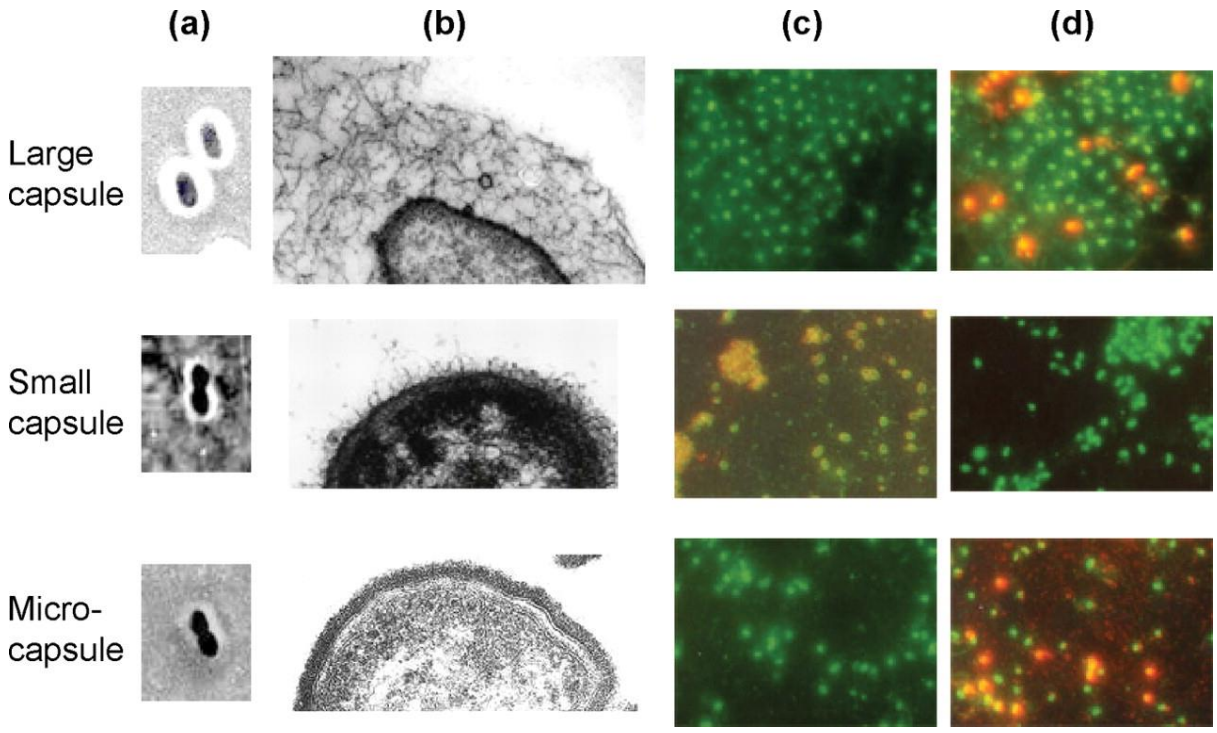


Figure 1 Micrographs of *B. fragilis* cultures enriched for large capsule (LC), small capsule (SC) and microcapsule (MC) via Percoll density centrifugation. **a)** Light micrographs of negative capsule stain: the LC and SC are clearly visible whilst the MC is not. **b)** Transmission electron micrographs: here the MC can be seen. **c and d)** Immunofluorescence micrographs. Green: primary labelling with rabbit anti-*B. fragilis* polyclonal antibody and secondary labelling with anti-rabbit FITC-conjugated antibody. Orange/red: primary labelling with mouse monoclonal anti-QUBf4 **(c)** or anti-QUBf5 **(d)** and secondary labelling with anti-mouse TRITC-conjugated antibody (Patrick *et al.*, 2009).

Inverted repeat elements have also been found to be associated with several other genes, including outer membrane proteins and signal transduction proteins, but most notably a type I restriction modification (R/M) system. The presence of 4 invertible regions in the R/M locus allows for the production of 8 different DNA-binding proteins from a single locus, each specific for a different DNA sequence (Cerdeno-Tarraga *et al.*, 2005). This broad range R/M system provides defence against invading bacteriophage and also makes genetic manipulation of *B. fragilis* very difficult (Patrick *et al.*, 2009), however it does not seem to prevent a large amount of horizontal gene transfer (HGT) from foreign species into *B. fragilis* (Patrick *et al.*, 2010).

B. fragilis has a rare bacterial O-glycosylation system which is essential to its growth and survival *in vivo*. This glycosylation occurs in extra-cytoplasmic proteins such as proteases, surface lipoproteins and proteins involved in protein-protein interactions and protein folding (Fletcher *et al.*, 2009). Computational analysis of the predicted proteome of *B. fragilis* suggests that more than half of the proteins exported from the cytoplasm are glycosylated. So far, 20 of these have been confirmed experimentally as glycoproteins (Fletcher *et al.*, 2011).

B. fragilis, and many other anaerobic gut bacteria, have exhibited increasing resistance to antibacterials commonly used in treating anaerobic infections, including some which were presumed to be universally effective, such as metronidazole and the carbapenems (Liu *et al.*, 2008; Seifert *et al.*, 2010; Sherwood *et al.*, 2011). In studies over the last decade 100% of clinically isolated strains were resistant to penicillins, while up to 40% were resistant to ampicillin, cefmetazole and clindamycin. There is also emerging resistance to chloramphenicol and metronidazole (Liu *et al.*, 2008; Seifert *et al.*, 2010). The mechanisms of antibiotic resistance in *B. fragilis* are varied, for example, CepA, a β-lactamase, confers resistance to penicillins and most cephalosporins, the ribosomal *tetQ* mutation prevents the action of tetracycline, and efflux pumps MsrSA and BexA remove erythromycin and

fluoroquinolones respectively (the latter also confers some resistance to moxifloxacin) (Eitel *et al.*, 2013; Meggersee & Abratt, 2015; Seifert *et al.*, 2010; Sherwood *et al.*, 2011; Wexler, 2007). Antibiotics that still remain largely active against *B. fragilis* include metronidazole, ertapenem, moxifloxacin and linezolid. This increasing level of antibiotic resistance is a serious concern for the treatment of *B. fragilis*-related infections which have the potential to be fatal (Seifert *et al.*, 2010; Sherwood *et al.*, 2011). Interestingly, clinical isolates of *B. fragilis* were found to produce a form of bacteriocin, a broad range antibiotic of bacterial origin, and were resistant to bacteriocin produced by other *B. fragilis* strains (Avelar *et al.*, 1999).

The rapid spread of antibiotic genes amongst *B. fragilis* is due to the presence of these genes on mobile genetic elements, such as conjugative plasmids and transposons (Eitel *et al.*, 2013). These antibiotic genetic elements are surprisingly stable, even in the absence of antibiotic selective pressure. Due to increasing rates of resistance amongst both commensals and pathogens, it has been hypothesised that *B. fragilis* and other residential gut bacteria may act as resevoirs for antibiotic resistance genes which can then be conferred to other pathogens and commensals in the gut (Salyers *et al.*, 2004; Wexler, 2007).

1.1.3 *B. fragilis* as a symbiote of the gastro-intestinal tract

The majority of host-bacterial interactions are not pathogenic; in fact, many gut bacteria have a symbiotic relationship with the host (Mazmanian *et al.*, 2005; Mazmanian *et al.*, 2008; Troy & Kasper, 2010; Wexler, 2007). This relationship is usually described as a commensalism, in which one organism benefits from the other without affecting it, since most residential gastro-intestinal bacteria are provided with a stable, protected and nutrient-rich environment in which to flourish, and do not cause disease in the host (Huang *et al.*, 2011b; Karlsson *et al.*, 2011; Wexler, 2007). However, some have argued that many gut bacteria have a mutualistic relationship with their host. Mutualism is a coevolutionary process in which the fitness of the

two participants becomes inter-dependent (Backhed *et al.*, 2005). For example, some bacteria, such as *B. thetaiotaomicron*, metabolise complex carbohydrates in the host diet and secrete molecules (e.g. simple carbohydrates, amino acids, vitamins) that can be utilised by the host (Huang *et al.*, 2011b; Karlsson *et al.*, 2011; Wexler, 2007). Other species are known to promote development of the host immune system, aid host health and homeostasis, and protect against colonisation by pathogenic bacteria (Huang *et al.*, 2011b; Mazmanian *et al.*, 2005; Mazmanian *et al.*, 2008; Wexler, 2007).

B. fragilis contributes significantly to host immune development. *B. fragilis* polysaccharide A (PSA), one of the many possible CPS, is able to correct immune deficiencies in germ free (GF) mice. GF mice are born sterile and remain uncolonised by the normal gut microbiota, and consequently they develop several immune deficiencies, including a lack of helper T cell differentiation. Helper T cells 1 (T_H1) and T_H2 (also known as $CD4^+$T cells) are important players in the adaptive immune system. They have different functions and cytokine profiles and both are necessary for a proper immune response. GF mice are biased towards producing T_H2 cells, creating a cytokine imbalance. However, when colonised with *B. fragilis*, or treated with purified PSA, the proper levels of T_H1 and T_H2 cells are produced and cytokine balance in the GF mice is restored (Mazmanian *et al.*, 2005). PSA has a zwitterionic structure, meaning it contains both positive and negative charges in each repeating unit. Such structures are already known to mediate $CD4^+$T cell proliferation (Tzianabos & Kasper, 2002). PSA is taken up by DCs *in vitro* and, in addition to promoting T_H1 differentiation, it has been shown to stimulate production of the anti-inflammatory cytokine IL-10. IL-10 suppresses T-cell response to antigen-sensing and thus inhibits inflammation in the gut by this pathway. Furthermore, either colonisation with *B. fragilis* or treatment with purified PSA is able to protect against experimentally-induced colitis in immunocompromised mice. This suggests that *B. fragilis* PSA may be sufficient to alleviate or even prevent inflammatory bowel diseases (IBD), such as Crohn's disease, in humans (Mazmanian *et al.*, 2008).

1.1.4 *B. fragilis* as a pathogen

As well as benefiting the host, *B. fragilis* can cause serious, sometimes fatal, disease. Enterotoxigenic *B. fragilis* (ETBF) strains are able to cause disease in the gastro-intestinal tract. These strains have a single, enterotoxigenic virulence factor: *B. fragilis* enterotoxin (Bft), a 20kDa, chromosomally encoded metalloprotease, specifially a matrixin (Moncrief *et al.*, 1995; Sears *et al.*, 2008; Sears, 2009; Wick & Sears, 2010). Bft has multiple mechanisms of action, including cleaving E-cadherin, a structurally important linkage in the zonula adherens tight junctions between epithelial cells. As a result, the junctions become leaky, allowing immune response cells into the gut lumen and exposing the sub-mucosa to antigens of other bacterial species (Moncrief *et al.*, 1995; Sears, 2009; Wexler, 2007; Wick & Sears, 2010; Wu *et al.*, 2007). Additionally, the cleavage of E-cadherin promotes T-cell proliferation (Wu *et al.*, 2007), and Bft uptake by the cell stimulates the production of the pro-inflammatory cytokine IL-8 (Sears *et al.*, 2008; Sears, 2009). All of these factors contribute to an inflammatory response in the gut stronger than that caused by enterotoxigenic *E. coli* (ETEC) or *Vibrio cholerae*, resulting in intestinal colitis (Sears *et al.*, 2008). Colonisation with ETBF can be asymptomatic (~4-20% of cases) (Sears, 2009) but when colitis does occur it is usually associated with self-limiting watery diarrhoea, abdominal pain and tenesmus. Other, rarely seen, symptoms can include recurring diarrhoea (for up to several years), a high fever, and faecal blood. ETBF has been implicated in the development of IBD and colorectal carcinomas, however direct causality has yet to be established (Sears *et al.*, 2008; Sears, 2009; Wick & Sears, 2010). *B. fragilis* may also contribute to disease in the gut by secreting enzymes that digest molecules associated with the mucosal layer, such as neuramidase and hylauronidase (Huang *et al.*, 2011b; Vieira *et al.*, 2005). The destruction of such an important barrier between the lumen and the epithelial layer exposes host cells to attack by other pathogens as well as increasing the likelihood of bacteria-induced inflammation.

Upon perforation of the gut, gastro-intestinal bacterial species are able to escape into the peritoneal cavity and eventually the bloodstream. Despite making up only 4-13% of the Bacteroides species in the gut, *B. fragilis* has been isolated from 63-80% of extra-intestinal Bacteroides infections (Patrick *et al.*, 2010; Wexler, 2007). Meanwhile, *B. thetaiotaomicron* makes up 15-29% of Bacteroides species in the gut and is only isolated from 13-17% of extra-intestinal infections (Cerdeno-Tarraga *et al.*, 2005). The transmission of *B. fragilis* outside the gastro-intestinal tract is associated with intra-abdominal and gynaecological sepsis, soft tissue infections, bacteraemia and abscesses in the abdomen, brain, liver and lungs (Gibson *et al.*, 1998; Meggersee & Abratt, 2015; Patrick *et al.*, 2010; Sherwood *et al.*, 2011; Wexler, 2007). *B. fragilis* that escape the gut usually end up in the peritoneal cavity where they adhere to the peritoneal wall, stimulating the formation of abscesses. When pathogens invade an epithelial layer, the immune response recruits polymorphonuclear leukocytes, neutrophils and fibroblasts, and a fibrous network is formed over the infection site, trapping the dead epithelial cells and invading bacteria (Gibson *et al.*, 1998; Wexler, 2007). Untreated abscesses can expand, causing obstructions, or rupture, allowing dissemination of the bacteria into the bloodstream, i.e. bacteraemia (Wexler, 2007). *B. fragilis*-associated bacteraemia is fatal in 19.3% of treated cases and up to 45% of untreated cases (Redondo *et al.*, 1995; Vieira *et al.*, 2005). The major virulence determinant of extra-intestinal *B. fragilis* infections appears to be the CPS. Firstly, the ability to express a variety of surface antigens provides a means for escaping detection and clearance by the immune system (Liu *et al.*, 2008; Wexler, 2007). Encapsulated *B. fragilis* are able to resist both phagocytosis and complement-mediated killing (Patrick *et al.*, 2009; Reid & Patrick, 1984). Secondly, the CPS is directly responsible for the induction of peritoneal abscess formation by stimulating the TNFα (tumour necrosis factor α) pathway, which recruits the appropriate immune response cells (Gibson *et al.*, 1998; Wexler, 2007). In addition to the CPS, *B. fragilis* expresses 2 hemolysins, HlyA and HlyB, which lyse erythocytes (potentially causing hemolytic anaemia) (Vieira *et al.*, 2005; Wexler, 2007), and is also capable of evading killing by macrophages via an unknown mechanism (Vieira *et al.*, 2005).

Due to the severity of disease caused by *B. fragilis*, the increase in antibiotic resistance amongst clinically isolated strains is a major concern. In 2009 7% of tested ETBF strains were found to be metronidazole resistant (Sears, 2009) and in a particularly serious case of extra-intestinal *B. fragilis* infection the responsible strain was resistant to all antibiotics tested except moxifloxacin and linezolid (Sherwood *et al.*, 2011). It has therefore been stated that periodic monitoring of antibiotic susceptibility in clinical isolates and species identification and susceptibility testing prior to prescription of antibiotics for treatment of infection are vitally important in tackling this problem (Seifert *et al.*, 2010).

1.2 Outer membrane vesicles of Gram-negative bacteria

1.2.1 Discovery of OMV

In 1959, De noted that the cell-free supernatant of a *V. cholerae* culture was capable of enterotoxicity and supposed that the enterotoxigenic OM components of the bacteria were being released into the medium (De, 1959). Later, a lipid-dense material, designated 'extracellular lipoglycopeptide', was identified in the cell-free filtrate of a lysine-requiring strain of *E. coli*. This material could not 'be said to originate from any one cellular fraction' (Bishop & Work, 1965) and was later visualised in *E. coli* and *V. cholerae* by electron microscopy as globules of roughly 40-100nm in diameter which appeared to bud from the cell surface (Figure 2) (Chatterj.Sn & Das, 1967; Work *et al.*, 1966). These globules are complex proteoliposomes, now known as outer membrane vesicles, which are produced by all Gram-negative bacteria and some Gram-positive bacteria (Manning & Kuehn, 2013).

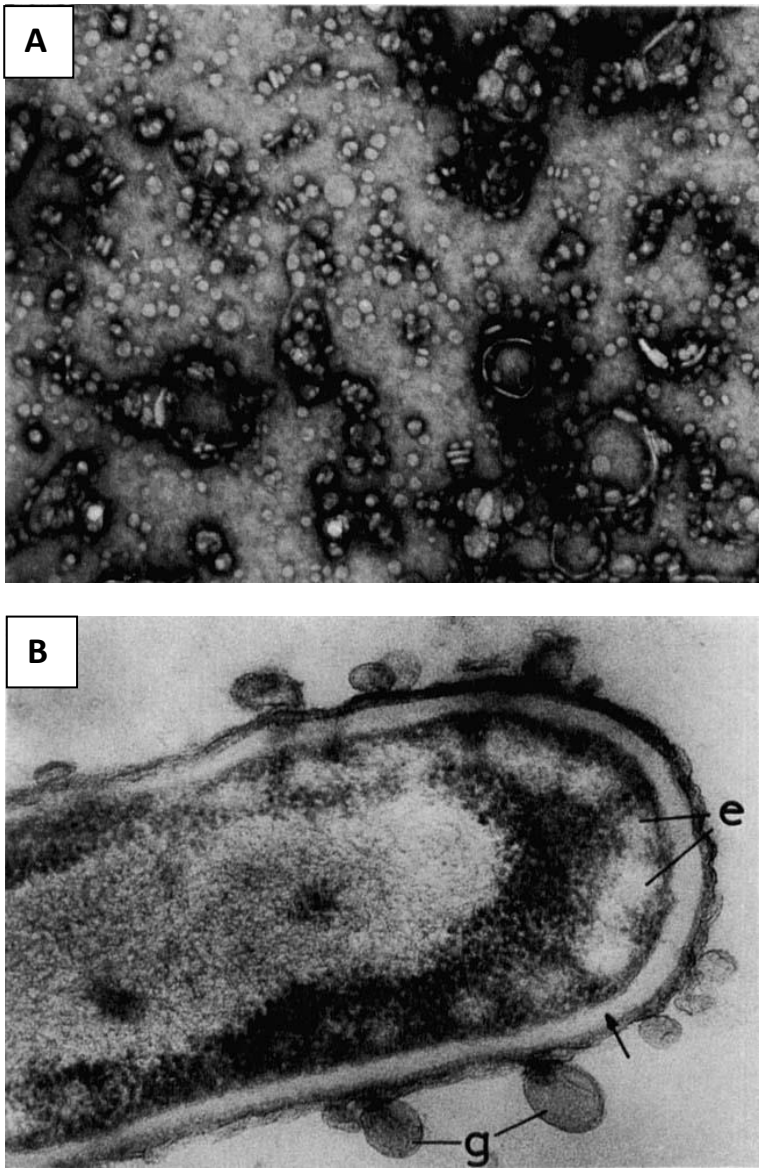


Figure 2 Electron micrographs showing the first documented outer membrane vesicles. **A)** Cell-free supernatant of *V. cholerae* was centrifuged at 144,000xg for 1 hour, negatively stained with phosphostungstate and observed under an electron microscope (magnification 63,000x). Globules of varying sizes can be clearly distinguished. **B)** Lysine limited *V. cholerae* cells were grown for 18 hours, negatively stained and observed using an electron microscope (magnification 88,000x). Work *et al* (1966) noted the area of low electron density between the outer and cytoplasmic membranes (**arrow**) that we now know as the periplasm, as well areas of low electron density in the cytoplasm (**e**). They were also able to observe multiple globules protruding from the cell surface (**g**) (Work *et al.*, 1966).

1.2.2 Structure and contents

Outer membrane vesicles of Gram-negative bacteria have been studied extensively in recent years and are now known to be stable structures, produced in all growth phases (Kadurugamuwa & Beveridge, 1999; Post *et al.*, 2005), at sizes ranging from 20 to 250nm in diameter. The basic structure consists of a spherical, outer membrane-derived lipid bilayer enclosing a mainly periplasmic lumen (Bomberger *et al.*, 2009; Grenier & Mayrand, 1987; Kulkarni & Jagannadham, 2014; Kulp & Kuehn, 2010; Patrick *et al.*, 1996; Perez-Cruz *et al.*, 2013) (Figure 3). Consequently, the majority of OMV contents are expected to be of OM and periplasmic origin, i.e. outer membrane lipids and proteins, lipopolysaccharide (LPS) and periplasmic proteins. Interestingly, the majority of OMV have also been found to contain inner membrane and cytoplasmic proteins, nucleic acids, small signalling molecules, metabolites, and ions. More than 200 proteins have been consistently identified in OMV proteomic studies, across a range of bacterial species, and these include porins, nutrient binding outer membrane proteins, virulence factors, toxins, murein hydrolases, efflux pumps and transporters (Aguilera *et al.*, 2014; Altindis *et al.*, 2014; Elhenawy *et al.*, 2014; Jang *et al.*, 2014; Kulkarni & Jagannadham, 2014; Kulp & Kuehn, 2010; Lee *et al.*, 2008; Perez-Cruz *et al.*, 2013; van de Waterbeemd *et al.*, 2013). The variety of proteins and other molecules from all subcellular compartments may be indicative of protein sorting and diverse functions of OMV, however some studies have observed alternative mechanisms of vesicle formation that may explain the presence of proteins from the cytoplasm and inner membrane (McCaig *et al.*, 2013; Perez-Cruz *et al.*, 2013) (Section 1.2.4). Whilst the protein complement of the OMV reflects that of the parent cell under any one environmental condition (Altindis *et al.*, 2014), the protein profile of the OMV is distinct from that of any other subcellular fraction (Kulkarni & Jagannadham, 2014). In multiple proteomics studies, OMV have been found to contain low abundance periplasmic or OM proteins and yet do not contain some high abundance proteins from those fractions, e.g. Lpp, a common OM protein, is virtually absent from the OMV (Kulkarni & Jagannadham, 2014; Kulp & Kuehn, 2010). The same is true of

other molecules found in OMV. LPS is commonly found on the outer surface of OMV, however the relative abundance of LPS chain types differs from OM to OMV (Li *et al.*, 1996), for example OMV of *Pseudomonas aeruginosa* contain mostly Band B LPS while containing very little Band A LPS (Kadurugamuwa & Beveridge, 1995). Phospholipids have also been observed at different levels of relative abundance in OM and OMV (Kulkarni & Jagannadham, 2014). DNA and RNA have been found in OMV from various bacterial species (Kulkarni & Jagannadham, 2014; Kulp & Kuehn, 2010; Lee *et al.*, 2008; Perez-Cruz *et al.*, 2013). Due to a certain level of DNase sensitivity some of the DNA is thought to be bound to the OMV surface, however as some of the DNA was protected there is still evidence that nucleic acids are carried in the lumen of the vesicles (Renelli *et al.*, 2004). The DNA in these OMV can be of chromosomal, plasmid or phage origin, either linear or circular and may be packaged into the OMV during biogenesis or taken up from the environment. The phenomenon of RNA-carrying OMV has not been studied in detail (Kulkarni & Jagannadham, 2014). The consistent population of enriched and excluded proteins in OMV points to an active sorting and biogenesis mechanism.

1.2.3 Biogenesis of OMV

The mechanism behind OMV biogenesis remains unknown, although there is a general understanding of the process based on observations of vesiculation in other circumstances (Kulp & Kuehn, 2010). There is evidence to support both a stochastic mechanism, in which OMV are produced as a side effect of regular cell processes and stress responses, and active biogenesis, in which specific proteins and other components are packaged into OMV that are released for a purpose using specific machinery (Schwechheimer *et al.*, 2013). The simplest and most passive model of vesiculation suggests that OMV are formed during routine cell wall turnover. The OM is tethered to the peptigolycan (PG) layer by various proteins. As PG is recycled and OM biogenesis occurs areas of excess OM begin to bleb

and pinch off, forming OMV (Figure 3 a). Deletion of genes involved in peptidoglycan turnover, and others that might result in excess OM, causes hypervesiculation in some Gram-negative bacteria (Kulkarni & Jagannadham, 2014; Schwechheimer *et al.*, 2013) and this may also explain why more OMV are produced during exponential growth than during stationary phase (Kulp & Kuehn, 2010). Many authors have noted that physical and chemical stress also lead to membrane disruption and an increase in ruffling and OMV release (Kulkarni & Jagannadham, 2014; Lee *et al.*, 2008; McBroom & Kuehn, 2007; Schwechheimer *et al.*, 2013). McBroom and Keuhn (2007) noted that cells under temperature stress demonstrated hypervesiculation and hypothesised that OMV biogenesis was required to deal with an accumulation of misfolded and potentially toxic proteins in the periplasm (McBroom & Kuehn, 2007). Similarly, the mutation of *V. cholerae degP*, which encodes a protein responsible for regulation of waste proteins, causes hypervesiculation. It is thought that in the absence of the usual mechanism the excess waste proteins are removed via OMV (Schwechheimer *et al.*, 2013). In support of the stress response model, it was found that SOS response genes are upregulated during times of hypervesiculation (Kulkarni & Jagannadham, 2014; Kulp & Kuehn, 2010) and that hypervesiculation mutants were better able to resist damage caused by temperature stress (McBroom & Kuehn, 2007). Tashiro *et al* (2012) suggested that the OMV were involved in the removal of phospholipids which would increase fluidity of the membrane. Whilst increased fluidity and hypervesiculation both correlate with growth at higher temperatures (Kulkarni & Jagannadham, 2014), this does not seem like a complete model for OMV biogenesis. In fact, neither the passive nor stress-response models can account for the fact that while OMV production does not cause envelope stress, it is a resource-depleting process that occurs in non-stressful environments and in every growth phase and therefore likely has a purpose beyond those suggested above (Kulkarni & Jagannadham, 2014; Kulp & Kuehn, 2010).

Although there is no direct evidence of active OMV biogenesis, there are a large number of experimental observations which suggest that the process is more than a passive result of growth and stress. The higher membrane curvature that leads to blebbing and budding prior to vesicle formation requires the rearranging or breaking of linkages between the OM and the IM, such as those formed by Tol-Pal (a multi-protein structure anchored in the IM) and lipoprotein (an OM protein), allowing the OM to ruffle (Kulkarni & Jagannadham, 2014; Kulp & Kuehn, 2010; Schwechheimer *et al.*, 2013) (Figure 3 a). An accumulation of periplasmic proteins at the OM might exert enough pressure on the membrane to cause blebbing and may be the mechanism by which misfolded proteins are released in OMV during the stress response (Figure 3 b) (Kulp & Kuehn, 2010; Schwechheimer *et al.*, 2013). Alternatively, a localised increase in the concentration of phospholipids and fatty acyl chains would increase the curvature of the OM, causing it to bulge outwards from the cell (Figure 3 c). This rearranging of OM components is consistent with the finding that OMV are not produced uniformly across the whole cell surface but originate from certain 'hot spots' (Kulp & Kuehn, 2010). A more direct mechanism for vesiculation was observed in *P. aeruginosa*, where the quorum sensing molecule PQS inserts itself into the OM, forcing an increase in curvature that leads to budding of a vesicle (Kulp & Kuehn, 2010; Mashburn-Warren *et al.*, 2009). As both PQS and OMV are known to stimulate biofilm formation, this is an attractive model, however due to the *P. aeruginosa*-specific nature of PQS, this model for biogenesis cannot be extrapolated to all Gram-negative bacteria. A similar model, again focused on *P. aeruginosa*, theorises that repulsion between highly concentrated electronegative Band B LPS molecules could also cause a localised shape change in the OM, resulting in vesiculation (Kadurugamuwa & Beveridge, 1995).

Several proteins have been implicated in OMV biogenesis. As mentioned above, OM tethering proteins such as Lpp are lacking at biogenesis sites, as are lipoproteins (Kulkarni & Jagannadham, 2014). *P. aeruginosa* has fewer lipoproteins than *E. coli* and also produces more OMV than the latter (Martin *et al.*, 1972), suggesting that lipoproteins inhibit vesiculation. *Salmonella enterica* outer membrane proteins OmpX and PagC are both found in OMV and both accelerate OMV biogenesis when overexpressed, whilst NlpA, an IM protein found in *E. coli* is known to play a role in OMV formation via an unknown mechanism (Kulkarni & Jagannadham, 2014). It is of course possible that several of the hypothesised and observed mechanisms for OMV biogenesis act together or at different times during the cell cycle (Kulp & Kuehn, 2010). It has also been suggested that the initial membrane bulging occurs at random but triggers a response within the cell that leads to the severing of tethers and the rearrangement of localised proteins and components, which in turn prevents resorption into the membrane (Schwechheimer *et al.*, 2013). The final step in OMV production is fission, in which the vesicle pinches off from the membrane. In other, similar cases, such as viral budding, fission is an active, energy dependent process. However, as there is no energy source in the periplasm (PP), the fission of OMV must involve an alternative active mechanism, possibly conformational changes in OM proteins which cause separation of the cellular and vesicular membranes (Kulp & Kuehn, 2010).

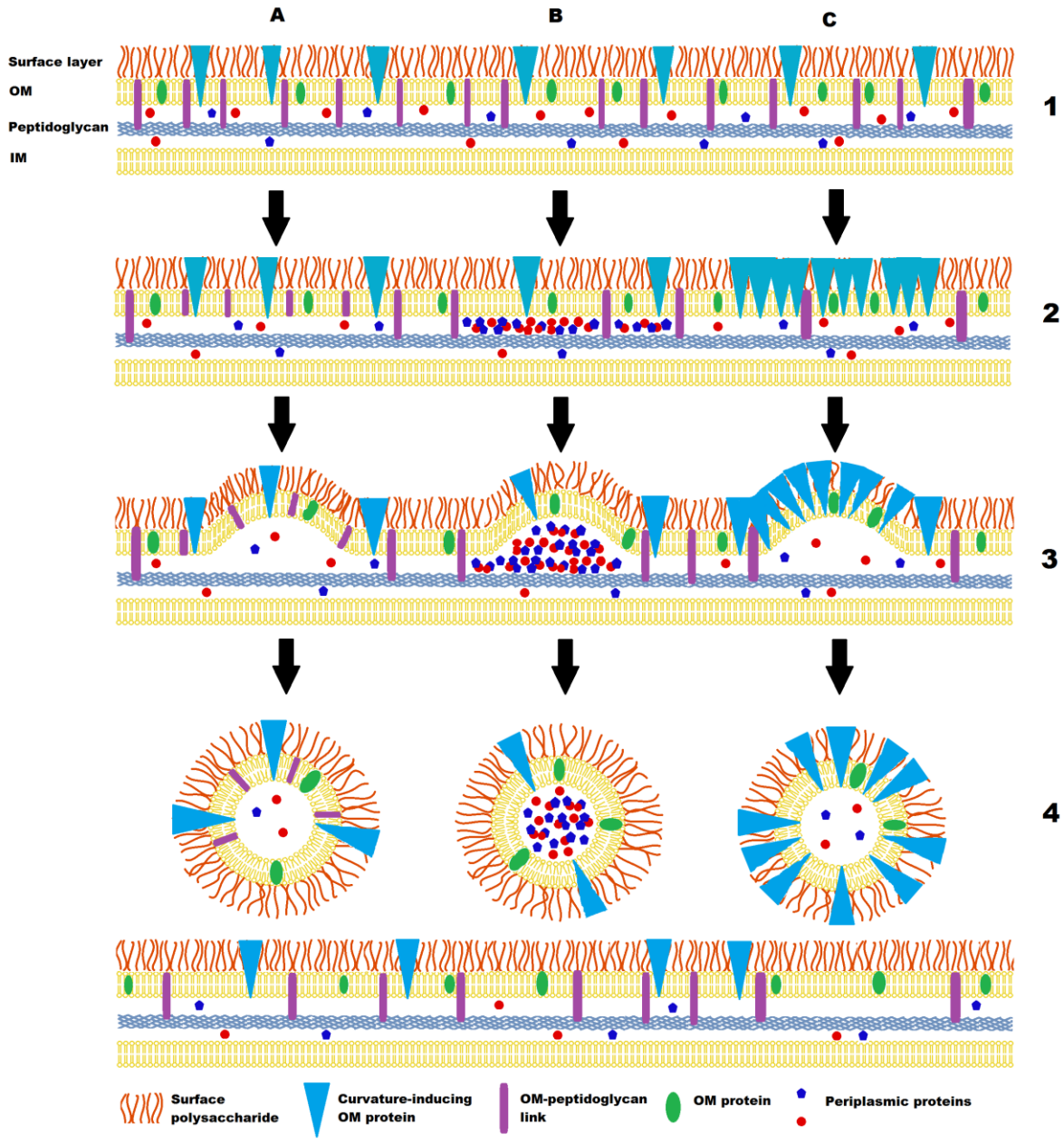


Figure 3 Examples of potential OMV biogenesis mechanisms. **A)** Rearrangement or severing of OM-peptidoglycan links (Step 1-2), allowing natural bulging of the membrane (Step 3) and release of a vesicle which is representative of the OM and periplasm at time of biogenesis (Step 4). **B)** The accumulation of periplasmic proteins at the OM (Step 1-2) exerts pressure on the membrane, causing bulging (Step 3) and eventual release of a vesicle which is enriched with periplasmic proteins (Step 4). **C)** The localised enrichment of curvature inducing proteins in the membranes (Step 1-2) causes the membrane to bulge outwards (Step 3) releasing a vesicle (Step 4). Based on Kulp and Kuehn (2010).

In further support of active biogenesis, it has become apparent from proteomic studies that proteins and other molecules are actively sorted into OMV. The presence of IM and cytoplasmic proteins in the OMV, discrepancies between SDS-PAGE gels of OMV, OM and periplasmic extracts and significant differences in the proteomes of the same, as determined by mass spectrometry, all point to an active secretion system (Elhenawy *et al.*, 2014; Kulkarni & Jagannadham, 2014; Lee *et al.*, 2008). Several hypotheses have been put forward to explain the presence of DNA in the periplasmic lumen of the OMV. One suggests that DNA-containing OMV are formed during cell lysis. When a cell is lysed the membranes naturally reform into spherical structures, potentially encapsulating the spilled cytoplasmic and periplasmic components. This is important to note as cell lysis is a significant risk during experimental purification of OMV, and detection of DNA in these studies could therefore be considered an artefact. A second theory supposes that DNA is passed through the inner membrane into the periplasm, via transport proteins for packaging into the OMV and would therefore be encompassed in the active sorting hypothesis. Finally, a model based on observations of OMV biogenesis in *Shewanella vesiculosa* involves a localised weakening of peptidoglycan, allowing the formation of a dual bilayer, encompassing OM, periplasm, IM and cytoplasm, known as complex outer membrane vesicles (Perez-Cruz *et al.*, 2013). The biological reason for exportation of DNA via OMV is unknown (discussed in section 1.2.5). Additionally, as no true vesiculation mutants have yet been identified, the correct model for OMV biogenesis also remains unsolved (Kulkarni & Jagannadham, 2014; Kulp & Kuehn, 2010).

1.2.4 Structural variations in OMV

Until recently, proposed models of vesiculation have not allowed for the inclusion of the inner membrane or cytoplasmic components, assuming a single lipid bilayer with a periplasmic lumen. Not only are OMV now thought to contain proteins from all subcellular fractions, but

new types of vesicles have been observed that require an entirely different model. *Shewanella vesiculosa* M7^T is a hypervesiculating strain that produces 2 distinct types of OMV: simple OMV, of the kind described in the previous sections, and complex OMV. Complex OMV are 100-250nm in diameter, consisting of two lipid bilayers, derived from the outer and inner membranes. The innermost lumen has a high electron density, similar to that of the cytoplasm when post-stained with aqueous uranyl acetate and lead citrate. These OMV are rare, making up only 0.1% of all OMV produced by *S. vesiculosa* M7^T, presumably as complex vesicle biogenesis appears to require localised weakening of peptidoglycan. Producing large amounts of these structures would therefore compromise cell stability and viability (Perez-Cruz *et al.*, 2013).

Another unusual form of vesiculation is associated with *Delftia* spp. Cs1-4 and *Francisella novidica*. These bacteria have been observed with long membranous tubes protruding from the cell surface from which OMV are pinched off and released into the environment. The OMV themselves are not structurally different from those previously described (McCaig *et al.*, 2013; Shetty *et al.*, 2011), however this type of biogenesis is rare and may be an improved method of long distance delivery in sparse environments (Sanchez, 2011).

1.2.5 Biological functions of OMV

OMV have been associated with a wide range of biological functions but they should primarily be considered as a type of secretion system (Bomberger *et al.*, 2009; Kulp & Kuehn, 2010; Lee *et al.*, 2008; Shen *et al.*, 2012). Gram-negative bacteria can posses a range of secretion systems, most of which span both the inner and outer membranes (Types I, II, III, IV and VI) and some of which operate entirely in the outer membrane (Type V, curli and pili). These systems allow long range interaction between bacteria and their

environment; transporting substrates either into the extracellular medium or directly into a target cell (Costa *et al.*, 2015; Kulp & Kuehn, 2010).

Vesicles hold distinct advantages over several of the other bacterial secretion systems. Firstly, the lipid bilayer provides a protective environment for their contents, which would usually be susceptible to proteases and other environmental forces, allowing them to travel long extracellular distances to their target (Kesty & Kuehn, 2004; Kulp & Kuehn, 2010). The OMV structure also allows transport of insoluble membrane proteins and co-transport of proteins that work together as part of a system (Altindis *et al.*, 2014; Kulkarni & Jagannadham, 2014; Kulp & Kuehn, 2010). Thirdly, although OMV secretion may not be as efficient as the Type 3 Secretion System (T3SS), which injects substrates directly into host epithelial cells (Costa *et al.*, 2015), vesicles do allow for delivery of proteins at high local concentrations. For example, OMV may undergo autolysis near target cells, be taken up by the target cell via endocytosis, or fuse with the target cell membrane, thus "emptying" their contents in the target cell cytoplasm (Kesty & Kuehn, 2004; Kulp & Kuehn, 2010). This allows for delivery of effectors without the need to be directly adjacent to the host target cell (Bomberger *et al.*, 2009) and the small size of OMV may allow them to pass anatomical barriers that prevent the movement of whole bacterial cells (Grenier & Mayrand, 1987). Due to these factors, the efficiency of OMV as a secretion system justifies the supposed energy expenditure of vesiculation (Kulkarni & Jagannadham, 2014).

As with other secretion systems, OMV can deliver virulence factors (Altindis *et al.*, 2014; Kuehn & Kesty, 2005). For example, ETEC vesicles were able to deliver labile toxin (LT) to host epithelial cells (Horstman & Kuehn, 2000; Kesty & Kuehn, 2004) while *Helicobacter pylori* vesicles were found to contain the vacuolating cytotoxin VacA (Fiocca *et al.*, 1999), and those of *Actinobacillus actinomycetemcomitaus* exhibit leukotoxic activity (Kato *et al.*, 2002). The enrichment of such proteins in OMV suggests that this secretion system could be important in virulence and disease, however OMV from commensal bacteria may also

benefit the host by boosting the native immune system (Kulp & Kuehn, 2010; Shen *et al.*, 2012).

OMV are thought to assist with biofilm formation, an important factor in the virulence of many species (Kolodziejek *et al.*, 2013), by promoting coaggregation of cells (Inagaki *et al.*, 2006; Schooling & Beveridge, 2006). Vesicles of *Porphyromonas gingivalis* are inherently aggregative and are able to promote aggregation between cells of different species, causing them to flocculate (Grenier & Mayrand, 1987), and thus have the potential to initiate biofilm formation and cooperation or communication between species. The exact mechanism behind the aggregative properties of OMV remains unclear (Kulkarni & Jagannadham, 2014).

OMV can interact with bacteria in other ways. For example, the DNA transfer via OMV allows for the exchange of genetic material within and between species of bacteria (Kuehn & Kesty, 2005; Kulkarni & Jagannadham, 2014; Kulp & Kuehn, 2010; Perez-Cruz *et al.*, 2013; Yaron *et al.*, 2000), thus providing a mechanism of HGT for species that lack the machinery for conjugation. OMV can also be bactericidal; *Bacteroides uniformis* vesicles contain a bacteriocin (Austinprather & Booth, 1984), whilst other OMV carry autolysins capable of breaking down peptidoglycan (PG), so that by fusing with other bacterial membranes they can cause lysis (Kadurugamuwa & Beveridge, 1996; Li *et al.*, 1996; Perez-Cruz *et al.*, 2013). These antibacterial properties confer an advantage when competing against other bacterial species for nutrients.

Most OMV contain a large number of nutrient- and metal-binding proteins, proteases and other degradative proteins, and in some species vesiculation increases when nutrients are sparse. It has therefore been hypothesised that they can be used by bacteria to scavenge important resources and break down complex molecules in the environment (Grenier & Mayrand, 1987; Kulkarni & Jagannadham, 2014; Kulp & Kuehn, 2010; Lee *et al.*, 2007).

It is thought that OMV biogensis is primarily a stress response mechanism. OMV are able to relieve pressure from unstable membranes, remove misfolded proteins after heat shock and dispose of antibiotics and toxins present in the periplasm. Some have also suggested that OMV in the extracellular environment may act as decoys for molecules and predators that attack the cell, such as antibiotics, antibodies and bacteriophage (Kulkarni & Jagannadham, 2014; Kulp & Kuehn, 2010; Lee *et al.*, 2008; Manning & Kuehn, 2011; McBroom & Kuehn, 2007; Schwechheimer *et al.*, 2013). Interestingly, OMV of one species can protect the cells of another from antibiotic attack (Mashburn & Whiteley, 2005; Schaar *et al.*, 2014).

1.2.6 Biotechnological applications of OMV

OMV have drawn great interest due to their potential for biotechnological application and in particular they have been investigated and developed as novel vaccines. The advantage to using OMV in this case is that they share the same antigens as the strain from which they are derived but can be more easily altered so that they have less potential to cause disease (Altindis *et al.*, 2014; Kolodziejek *et al.*, 2013; van de Waterbeemd *et al.*, 2013). One successful case is that of *Neissera meningitidis* sergroup B, in which individuals were protected against meningitis using a vaccine consisting of OMV extracted with detergent to strip endotoxin and reduce their inflammatory effects. The vaccine depended on the PorA antigen, which is particular to each strain. Later a recombinant PorA was used to create a broad-range vaccine and a mutant was identified which produces attenuated endotoxin, abolishing the need for detergent extraction (van de Waterbeemd *et al.*, 2013). This case demonstrates that, in principle, OMV-based vaccines could be applied to other disease causing organisms, particularly those for which no vaccine has yet been developed, e.g. *Yersinia pestis* (Kolodziejek *et al.*, 2013).

OMV have also been investigated for their potential to act as parcels for the delivery of known antibiotics. For example, *Pseudomonas aeruginosa* OMV are capable of packaging and delivering gentamicin to both Gram-positive and Gram-negative bacteria (Kadurugamuwa & Beveridge, 1996). Many established and novel antibiotics would be more effective and less detrimental to the health of the patient if they were delivered only to bacteria and not the host cells. Similarly, gene therapy, a novel treatment for diseases such as cancer, requires a non-toxic, durable, targeted method for delivery to a particular organ or cell type within the body that could potentially be achieved using OMV.

1.3 Ubiquitin

Ubiquitin is an important, ubiquitous protein modifier, previously thought to be exclusive to eukaryotes. Recently, ubiquitin homologues have been identified in several bacteria, including *B. fragilis*, which are distinct from previously described bacterial ubiquitin analogues such as Pup (Patrick *et al.*, 2011; van den Ent & Lowe, 2005). The function of these homologues is unknown; however, they might play a role in the disruption of eukaryotic host cells. This chapter describes the known roles and mechanism of action of eukaryotic ubiquitin and its known homologues.

1.3.1 Characteristics of ubiquitin

Ubiquitin is a 76aa, 8kDa protein modifier (Grabbe & Dikic, 2009; Komander & Rape, 2012; Wickliffe *et al.*, 2011), originally believed to be present in all living organisms. It was later found to be exclusive to and ubiquitous in eukaryotes and is incredibly well conserved across all species, with yeast ubiquitin differing by only 3 residues from human ubiquitin (Grabbe & Dikic, 2009; Komander & Rape, 2012). The highly thermodynamically stable 3D structure consists of a β-grasp fold (4-5 β sheets surrounding a single α helix) with a 6aa, flexible C-terminal tail (Vijaykumar *et al.*, 1987). Ubiquitin has multiple hydrophobic surface patches as well as 7 conserved lysines, all pointing in different directions, all of which are functionally vital. The ubiquitin C-terminus consists of a double glycine residue that is essential for ubiquitylation. Each eukaryotic cell contains a high abundance of ubiquitin, with ~10^8 molecules per cell (Grabbe & Dikic, 2009; Komander & Rape, 2012; Wickliffe *et al.*, 2011).

1.3.2 Ubiquitylation

Ubiquitin has low affinity for itself and for its substrates (Komander & Rape, 2012) and therefore ubiquitylation requires a series of enzymatic reactions (Figure 4). Firstly, an E1 activating enzyme catalyses the adenylation of the C-terminal glycine of a ubiquitin molecule. A thioester bond is then formed between the ubiquitin C-terminal carboxyl group and the sulfhydryl group of the active site cysteine of the E1 via an ATP-dependent reaction. The E1 transfers the ubiquitin onto the active site cysteine residue of an E2 conjugating enzyme, which works with an E3 ligase to form an isopeptide bond between the C-terminus of the ubiquitin and the ε-group of a lysine residue in the target substrate (Corn & Vucic, 2014; Dye & Schulman, 2007; Herrmann *et al.*, 2007; Wickliffe *et al.*, 2011).

The hydrophobic patches on the surface of ubiquitin contribute to the stability of its interactions with the E2 and E3 enzymes and with the substrate, essentially holding the C-terminal tail in place for nucleophilic attack by the substrate lysine (Wickliffe *et al.*, 2011). There are only a few E1 activating enzyme molecules per cell, whilst there are nearly 50 varieties of E2 conjugating enzymes and a wide range of substrate-specific E3 ligases (Corn & Vucic, 2014; Herrmann *et al.*, 2007). The latter come in 2 main types: the HECT types, which "manually" transfer the ubiquitin molecule from the E2 to the substrate, and the RING types, which bind both the E2 and the substrate to bring the ubiquitin close to the target lysine (Ashida *et al.*, 2014; Dye & Schulman, 2007; Herrmann *et al.*, 2007). Novel E3 ligases have been identified which do not fall into either category. Additionally, ubiquitin can be conjugated to cysteine residues via a thioester bond or to serine/threonine residues, via an oxyester bond, however these are far less common than lysine conjugation (Ashida *et al.*, 2014).

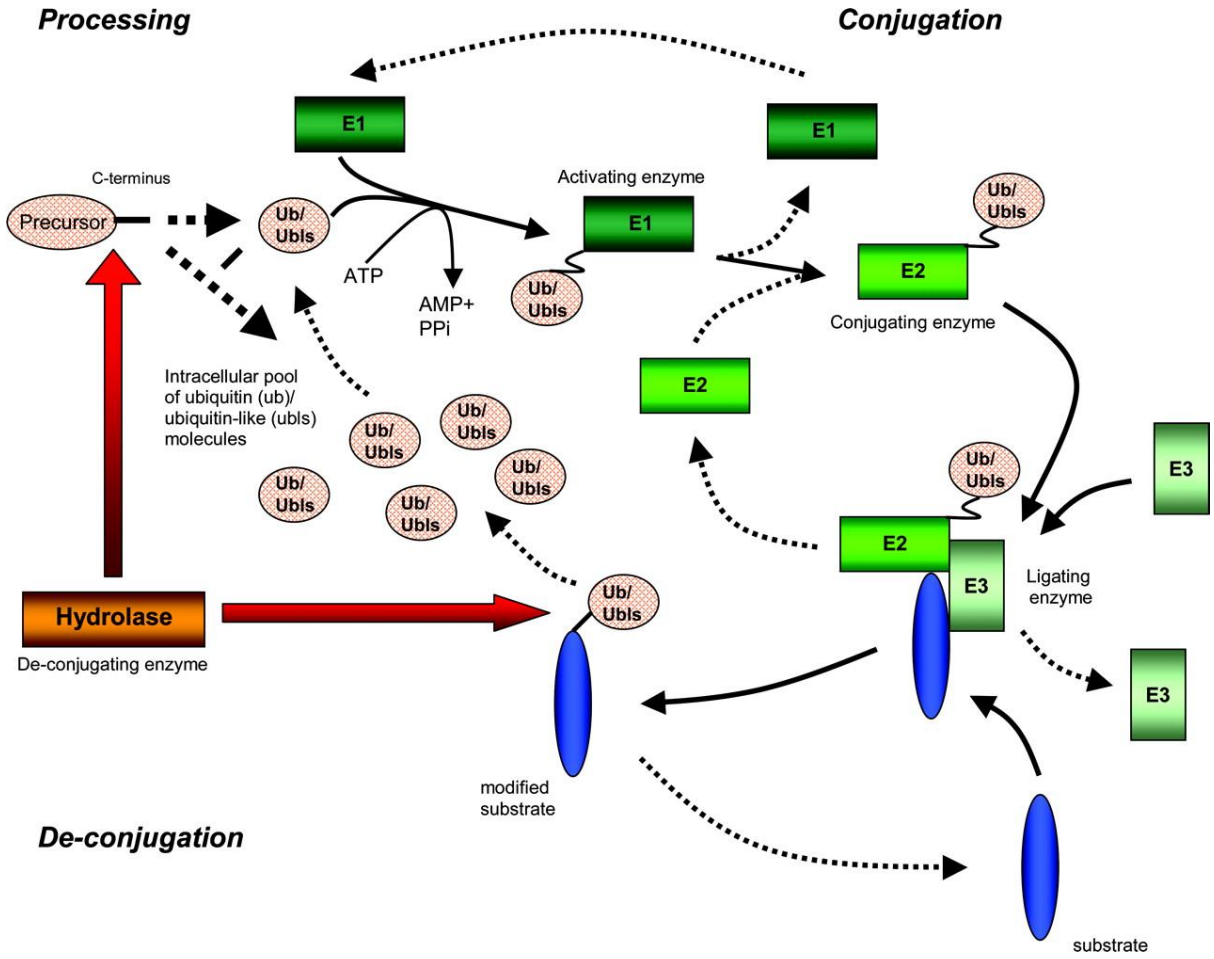


Figure 4 The basic ubiquitylation cycle. Ubiquitin (or ubiquitin-like molecules, see section 1.3.4) are expressed as precursors and require processing by hydrolases to expose the C-terminal double glycine residue. An E1 activating enzyme adenylates the C-terminal glycine of a free ubiquitin monomer and then forms a thioester bond with the C-terminal carboxyl group of the ubiquitin via an ATP-dependent reaction. The E1 then transfers the ubiquitin to the active site cysteine of an E2 conjugating enzyme. The E2 interacts with a substrate-specific E3 ligase to transfer the ubiquitin onto a lysine residue in the target protein. De-conjugating enzymes (deubiquitylases) can also remove the ubiquitin molecule from the target, recycling the ubiquitin into the cytoplasm once more (Herrmann *et al.*, 2007).

More than one ubiquitin can be conjugated to a substrate, either by monomeric ubiquitylation at multiple lysines or by forming a ubiquitin chain. A second ubiquitin can be added to the first at any of its 7 lysines (K6, K11, K27, K29, K33 K48 or K63) or at its N-terminal methionine (M1) (Ashida *et al.*, 2014; Komander & Rape, 2012; Valkevich *et al.*, 2014). During chain elongation, a free "donor" ubiquitin C-terminus is brought close to the lysine or M1 of a substrate-bound "acceptor" ubiquitin. Nucleophillic attack of the donor by the acceptor results in the formation of an isopeptide bond between the two (Wickliffe *et al.*, 2011). These chains can extend up to 10 molecules and be linear or branched, loose or compact, using different lysine residues (or M1) to create different conformations (Figure 5) (Corn & Vucic, 2014; Komander & Rape, 2012). The rate of formation and the specific linkages of a chain seem to depend on the specific E2 and E3 enzymes involved in elongation (Valkevich *et al.*, 2014). All 8 chain linkages have been detected *in vivo* (Xu *et al.*, 2009), indicating that each is biologically viable and in some way useful to the cell.

Ubiquitin molecules are removed from substrate proteins by deubiquitinases which, like E3 ligases, are usually specific to a substrate. Deubiquitinases (DUBs) can either remove all ubiquitin molecules or merely trim back the chain to allow for chain "editing" (Grabbe & Dikic, 2009; Komander & Rape, 2012).

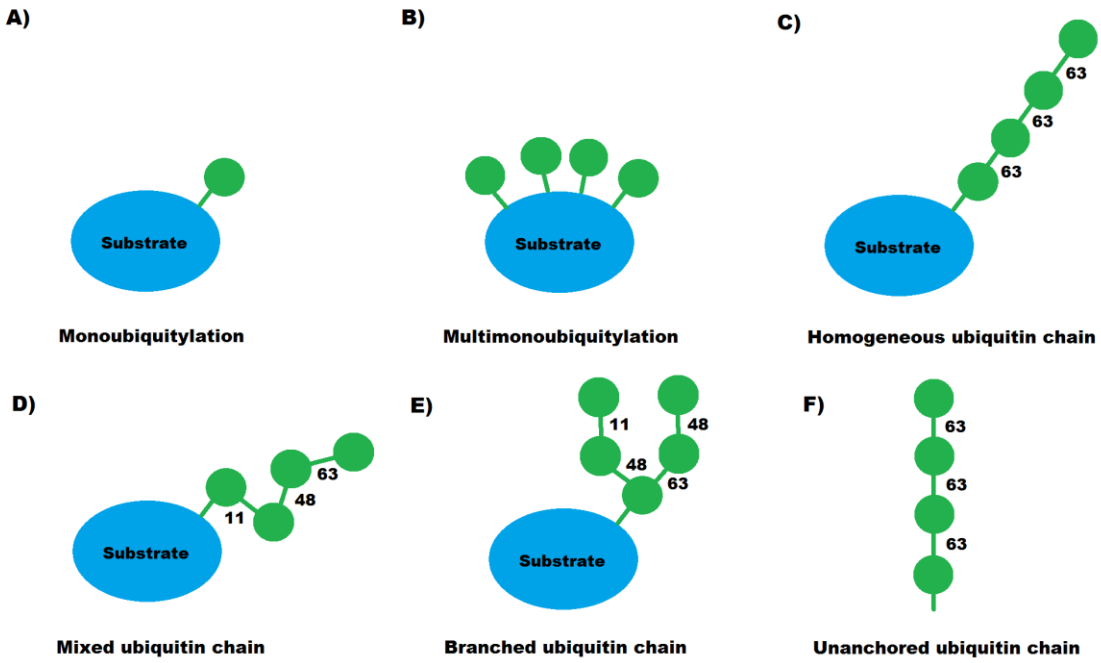


Figure 5 Diagrams showing the different types of ubiquitin chain conformation. **A)** Monoubiquitylation, a single ubiquitin monomer attached to a single lysine. **B)** Multiple ubiquitin monomers attached to different lysines. **C)** Homogeneous ubiquitin chain in which each ubiquitin monomer is connected to the previous monomer via the same lysine. **D)** Mixed ubiquitin chain in which different lysine linkages are used between different ubiquitin monomers. **E)** Branched ubiquitin chain in which one or more ubiquitin monomers has multiple other ubiquitin residues attached via different lysines. **F)** Unanchored ubiquitin chains which exist freely in the cytoplasm. Note that K48 forms a more compact chain than K63. Based on Komander & Rape, 2012.

1.3.3 Roles of ubiquitin in the eukaryotic cell

Ubiquitin is known to play roles in a very large number of eukaryotic cell pathways. The best studied of these is the 26S proteasome pathway, which degrades malfunctioning proteins or those which are no longer needed to maintain homeostasis, thus triggering protein cascades in response to environmental triggers (Finley, 2009; Herrmann *et al.*, 2007). Ubiquitin contains a proteasomal targeting motif common to all proteasome-recruited proteins (Grabbe & Dikic, 2009). Consequently, certain ubiquitylated proteins are brought to the proteasome where ubiquitin is cleaved off and returned to the cytosol before the target is degraded (Finley, 2009; Herrmann *et al.*, 2007). Ubiquitin also plays roles in antigen processing, activating immune and inflammatory responses, DNA repair and cell cycle progression, amongst others (Ashida *et al.*, 2014; Corn & Vucic, 2014; Valkevich *et al.*, 2014). Usually these roles involve modification of protein function, promotion or inhibition of protein-protein interactions and relocalisation of ubiquitylated proteins (Komander & Rape, 2012).

Interestingly, specific ubiquitin chain conformations have been associated with specific cell functions. Chains formed via the K48 linkage are known to mark a protein for degradation by the 26S proteasome pathway (Corn & Vucic, 2014; Herrmann *et al.*, 2007; Komander & Rape, 2012; Valkevich *et al.*, 2014; Wickliffe *et al.*, 2011) and K63 linkages have been implicated in DNA damage repair, cytokine signalling, lysosomal degradation and the assembly of protein complexes (Komander & Rape, 2012; Valkevich *et al.*, 2014; Wickliffe *et al.*, 2011). Inhibition of K11 linkages halts mitotic progression whilst untimely assembly of K11 linked chains results in rapid cell cycle progression (Jin *et al.*, 2008). K11, K63 and M1 linkages all appear to be involved in inflammatory signalling complexes, and several inflammatory pathways, e.g. tumour necrosis factor receptor (TNFR) and toll-like receptor (TLR) pathways, are known to heavily rely on ubiquitylation (Corn & Vucic, 2014). According to Komander and Rape (2012), the complex chains formed via ubiquitylation can be seen as a sort of code that stores and transmits information in the cell. The proteins that

read and respond to this code will have appropriately spaced and numbered ubiquitin-binding domains and will be able to distinguish between structurally similar chains based on flexibility and linkage (Komander & Rape, 2012).

1.3.4 Eukaryotic ubiquitin-like proteins

The "ubiquitin superfamily" is a term used to refer to the large number of proteins that are structurally or functionally related to ubiquitin and the ubiquitylation cycle. Ubiquitin-like proteins (UBLs) are small protein modifiers with >10% identity to mammalian ubiquitin (Larsen & Wang, 2002) and which resemble ubiquitin in tertiary or higher order structures. They are attached to a substrate lysine, using machinery similar to the E1/E2/E3 enzymes, in a way that alters protein conformation, function or localisation. Due to the similarities between ubiquitin and UBLs, the different systems compete for substrate lysines and experience cross talk. The best known examples of UBLs are SUMO, Nedd8 (Dye & Schulman, 2007; Grabbe & Dikic, 2009) and ISG15, which is actually comprised of two UBLs connected by a hinge region, and is active exclusively in the interferon-induced inflammatory pathway (Bade *et al.*, 2012). As they are all conjugated to a substrate lysine, UBLs require a C-terminal double glycine in order to function. Only one characterised UBL, the Yeast protein Hub1, does not possess a GG motif. If this UBL is covalently conjugated to other proteins, it must use hitherto undocumented machinery (Dittmar *et al.*, 2002).

Ubiquitin-like domains (ULD) are domains of 45-80 amino acids that exist within another protein, most frequently as an N-terminal fusion. These domains strongly resemble the ubiquitin tertiary structure and can appear more than once in the same protein; in fact, some proteins contain both a ULD and different UBL domain, e.g. a SUMO-like domain. Whilst ULDs are commonly associated with proteasomal shuttles and signal transduction proteins (Grabbe & Dikic, 2009) they may have other functions, such as providing a thermostable

nucleation or folding template for large proteins. They are also known to serve as nuclei for ubiquitin chain formation (Larsen & Wang, 2002).

Ubiquitin-binding domains (UBD) vary in size (20-150 amino acids) and structure, although most are α-helical. They bind ubiquitin with a low affinity (K_d = 2-500μM), however they have strict patterns of recognition and where multiple UBDs are present on the same substrate the affinity of each interaction increases with the number of ubiquitin molecules that are bound. Certain proteins contain both a ULD and a UBD in the same peptide (Grabbe and Dikic., 2009). Such proteins are deemed to be integral players in 26S proteasomal degradation system, as shuttle factors capable of binding a polyubiquitin chain via their UBD and the proteasome via their ULD (Ng *et al*., 2002), and in cell cycle progression (Masutani *et al*., 1994; Grabbe and Dikic., 2009).

1.3.5 Bacterial ubiquitin-like proteins

The best characterised bacterial ubiquitin-like system is the Pup system, found exclusively in Actinobacteria. Possibly the result of horizontal gene transfer from eukaryotes, Pup is a small protein modifier which bears no sequential or structural similarity to ubiquitin. When unbound it has a largely disordered structure with a weak α-helix near the C-terminus and no β-grasp fold. There is also far less sequence conservation of Pup between species, with large variation towards the N-terminus, however it does contain a C-terminal GGQ or GGE motif via which it forms an isopeptide bond with substrate lysines. (Striebel *et al.*, 2014). Pupylation requires the deamidation of this glutamine (Q) or glutamic acid (E) to a glutamate by Dop, after which PafA, a Pup ligase, utilises ATP to conjugate Pup to the substrate lysine (Striebel *et al.*, 2009) (Figure 6). Dop also acts as a depupylase, cleaving the isopeptide bond between Pup and the substrate via a mechanism that requires nucleotide binding (Imkamp *et al.*, 2010).

So far the only known role of Pup is as a marker for degradation by the 20S proteasome. Pup, conjugated to a target substrate, associates with the proteasomal ATPase Mpa via its central region, while its N-terminal region begins feeding the tagged protein into the proteasome. The Pup proteasomal system is non-essential and Actinobacteria can survive without it. However, given the multifaceted roles of ubiquitin, it is believed that Pup may play some other role in the bacterial cell (Striebel *et al.*, 2014).

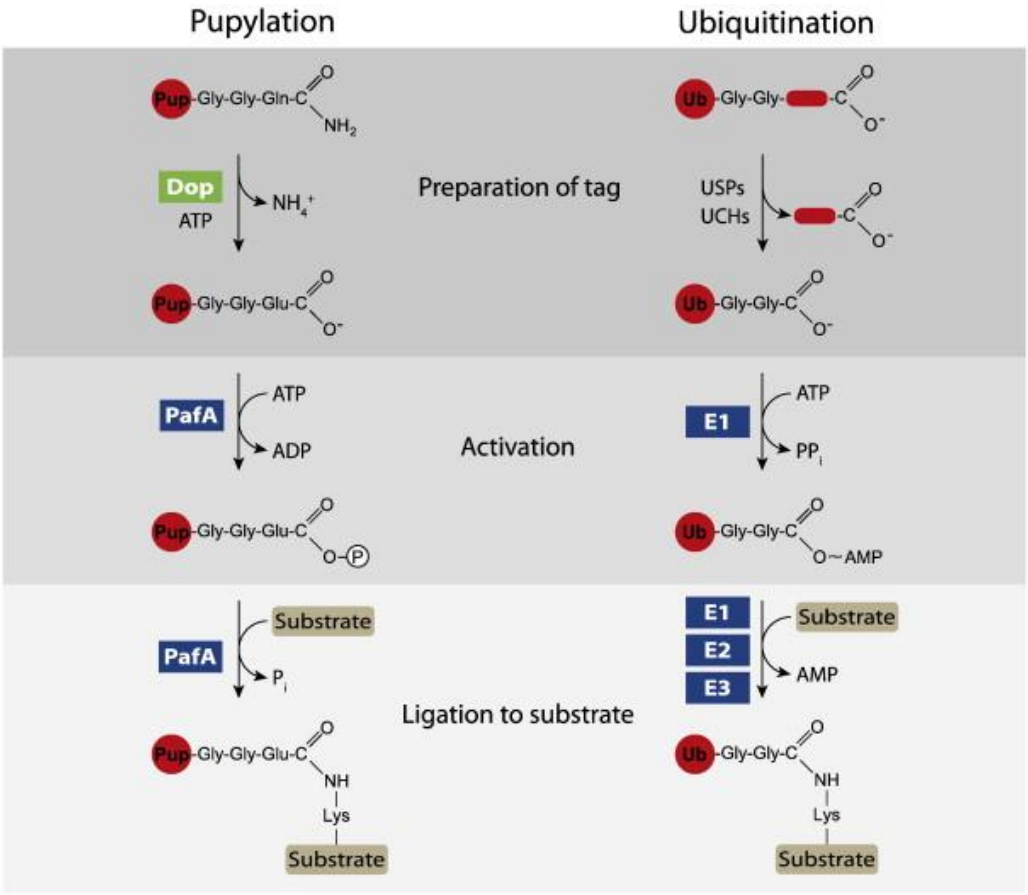


Figure 6 A comparison of pupylation and ubiquitylation, highlighting the similarities and differences between the two systems. Both Pup and ubiquitin require C-terminal processing prior to conjugation and both require ATP for the activation of the carboxyl group prior to ligation. However, whilst E1 adenylates ubiquitin, PafA phosphorylates Pup (although nucleophillic attack by the ε-group of the lysine remains the same). Additionally, ubiquitylation depends on a large hierarchy of E1, E2 and E3 enzymes, whilst pupylation requires only 2, Dop and PafA (Striebel *et al.*, 2014).

Many pathogenic bacteria use proteins, acquired by HGT, that interfere with the ubiquitin system, during infection of a host cell (Ashida *et al.*, 2014; Munro *et al.*, 2007). E3 ligases can be found in a range of gastrointestinal pathogens, for example, *Salmonella enterica* serovar Typhimurium SopA is a HECT-like E3, the target of which is unknown, but which has been shown to be essential for polymorphonuclear transepithelial migration of the bacterium into the host cell (Zhang *et al.*, 2006). LubX, a RING-type E3 of *Legionalla pneumophila*, ubiquitylates host cell kinase CLK1 and is essential for bacterial growth in macrophages (Kubori *et al.*, 2008). *Shigella flexneri* possesses a novel E3, IpaH0722, which poly-ubiquitylates TRAF2 (TNF receptor-associated factor), thus targeting it for degradation and inhibiting the NFκB inflammatory pathway (Ashida *et al.*, 2013). There are several documented cases of bacterial deubiquitylases contributing to virulence, for example YopJ of *Yersinia enterolitica* and TssM of *Burkholderia pseudomallei* both cleave ubiquitin from inflammatory response proteins, thus reducing the overall inflammatory response of the host cell (Ashida *et al.*, 2014; Haase *et al.*, 2005; Tan *et al.*, 2010). Some bacterial proteins also modify host cell SUMO. *Listeria monocytogenes* LLO causes the degradation of UBC9, a SUMO-ylating protein, which results in a general decrease of SUMO-ylated proteins in the cell. This, in turn, affects signalling pathways which rely on SUMO-ylation, such as the TGFβ (transforming growth factor β) pathway, a major player in host cell proliferation and response systems (Ribet *et al.*, 2010).

1.4 *B. fragilis*, outer membrane vesicles, and ubiquitin

B. fragilis produces a protein designated BfUbb which, unlike Pup, is a true homologue of eukaryotic ubiquitin and the result of an ancestral HGT (Patrick *et al.*, 2011). As seen above, many bacteria use HGT-acquired proteins to interfere with ubiquitin-dependent pathways in the host cell (Ashida *et al.*, 2014). This requires delivery of these proteins, usually via a secretion system. The only secretion system *B. fragilis* has that is capable of long-distance delivery of effector proteins to host cells is its OMV, which already have established roles in the delivery of other molecules (Shen *et al.*, 2012). Importantly, BfUbb is expressed with a periplasmic signal sequence and can be detected in concentrated supernatants of *B. fragilis* (Patrick *et al.*, 2011), suggesting that it can indeed be delivered to host cells via OMV.

The following chapters describe the methods and results of experiments designed to characterise both BfUbb and the OMV of *B. fragilis*. As BfUbb is able to form covalent bonds with E1 activating enzyme, assays were performed to determine whether BfUbb forms covalent bonds with other enzymes of the ubiquitin cycle, as well as other proteins within the eukaryotic cell. Constructs were created which would allow the purification of native (untagged) BfUbb and fluorescently tagged BfUbb for use in *in vitro* experiments, such as X-ray crystallography and ligand-binding assays. It has been hypothesised that BfUbb acts on the epithelial cells of the host gastro-intestinal tract, possibly to alter their response to bacterial stimulus. A set of constructs was therefore created for the expression of affinity tagged BfUbb and fluorescently tagged BfUbb in mammalian cells, thus allowing further studies to determine the effect of the presence of BfUbb on the cell, which proteins it forms covalent bonds with *in vivo* and its intracellular location. Clinical samples from patients with gastro-intestinal diseases or abscesses were screened for the presence of *B. fragilis* and the *ubb* gene, to determine whether any correlation exists between the expression of BfUbb by *B. fragilis* and the presence or absence of any of these diseases. OMV of Gram-negative

bacteria have a range of proposed functions, including delivery of proteins to other cells, eukaryotic or bacterial. The proteome of the OMV of *B. fragilis* was determined by LTQ-Orbitrap mass spectrometry with the intention of identifying any proteins that may elucidate the mechanism of biogenesis or function of the OMV, as well as proteins that have previously been identified in the concentrated supernatant of *B. fragilis*, such as BfUbb and a fibrinogen-binding protein. Finally, OMV of both WT *B. fraglis* and the Δ*ubb* mutant were tested for their ability to kill or inhibit growth of other gastro-intestinal bacteria, to determine whether BfUbb or the OMV themselves exhibit any antibacterial activity.

2. Materials and Methods

2.1 Bacterial culture

2.1.1 Bacterial strains

All bacterial strains used in this study are shown in Table 1.

Species	**Strain**	**Plasmid**	**Antibiotic**	**Source**
Bacteroides fragilis	NCTC 9343	pBF9343	Gentamicin (40µg/ml)	Patrick lab, Queens Belfast University
Bacteroides fragilis	NCTC 9343	pBF9343 pGB920	Tetracycline (5µg/ml)	(Patrick *et al.*, 2009)
Bacteroides fragilis	NCTC 9343 Δ*ubb*	pBF9343 pGB920	Erythromycin (10µg/ml)	(Patrick *et al.*, 2011)
Bacteroides fragilis	NCTC 9343 Δ*1705*	pBF9343 pGB920	Tetracycline (5µg/ml)	Blakely lab, University of Edinburgh
Salmonella enterica serovar Typhimurium	SL1344	None	None	Gallagher lab, University of Edinburgh
Esherichia coli	DH5α	pTRC99A:: 6xH-BfUbb	Ampicillin (100µg/ml)	(Patrick *et al.*, 2011)
Escherichia coli	JM109	None	None	French lab, University of Edinburgh
Escherichia coli	DH5α	None	None	New England Biolabs (NEB), UK
Escherichia coli	BL21 DE3	None	None	NEB, UK

Table 1 Bacterial strains used in this study, showing strain, mutations, plasmids, selection antibiotics, and source.

2.1.2 Culture conditions

B. fragilis solid cultures were grown on Brain-Heart Infusion (BD, USA) agar, supplemented with 5% cysteine, 10% sodium bicarbonate, 50 µg/ml haemin and 0.5 µg/ml menadione (BHI-S agar), and liquid cultures in either BHI-S solution or defined medium (Van Tassell & Wilkins, 1978), with the appropriate antibiotics (Table 1). All *B. fragilis* cultures were grown in a Minimacs anaerobic work station (Don Whitley Scientific, UK) at 37ºC in an anaerobic gas mix (10% Hydrogen, 10% Carbon dioxide, 80% Nitrogen).

Solid cultures of *E. coli* and *S. enterica* Typhimurium were grown on Luria-Bertani Broth (LB) agar, and liquid cultures were grown in LB, with shaking (200rpm), all at 37ºC and with appropriate antibiotics (Table 1).

2.2 DNA

2.2.1 Primers

All primers used in this study (Table 2) were synthesised by Eurofins MWG Operon (Germany), suspended in TE buffer (1M Tris-HCl, 0.5M EDTA, pH8.0) to a concentration of 100pg/µl and stored at -20ºC. Primers were diluted 1:10 in dH_2O before use (final concentration 10pg/µl).

Primer	Sequence 5' - 3'	Length	Tm (ºC)
Bfndei	GTGCATACATATGCAAGTTTTTATAAAAAACAG	33	66.5
Bfhindiii	AGCTTACAAGCTTCACCACCATCACCACCATATCGGC ATGCAAGTTTTTATAAAAAACAG	60	74.9
Bfx	CGAGCTGTACAAGATGCAAGTTTTTATAAAAAACAG	36	64.9
Brbamhi	TACTGTTGGATCCCTTAGTGTTAACAGTCGAC	32	66.9
Yfndei	ATACTTGCATATGGTGAGCAAGGGCGAGC	29	68.1
Yfhindiii	AGTAGCTAAGCTTATGGTGAGCAAGGGCGAG	31	69.5
Yrx	CTTGCATCTTGTACAGCTCGTCCATGC	27	66.5
Yrbamhi	GTCATCAGGATCCTTACTTGTACAGCTCG	29	66.7
6TRtetf	GTAAAGTGATTAACAGCGC	19	46.8
6TRtetr	TAAGATCTGAATTCCCGG	18	45.8

Table 2 Primers used in this study, showing sequence, length and melting temperature. Restriction sites are shown in red and homology to relevant gene is underlined

2.2.2 Plasmids

All plasmids used and generated in this study are shown in Table 3.

Plasmid	Size (bp)	Selection Antibiotic	Source
pET19b	5717	Ampicillin	Gallagher lab
pET19b 10xHBfUbb	5952	Ampicillin	This study
pET19b YFPBfUbb	6689	Ampicillin	This study
pcDNA6/TR	6662	Ampicillin/Blasticidin	Life Technologies
pcDNA4/TO	5078	Ampicillin/Zeocin	Life Technologies
pcDNA4/TO 10xHBfUbb	5340	Ampicillin/Zeocin	This study
pcDNA4/TO YFPBfUbb	6053	Ampicillin/Zeocin	This study
pcDNA4/TO YFP	5804	Ampicillin/Zeocin	This study

Table 3 Plasmids used and generated in this study, showing size and antibiotic used for selection of transformants. For pcDNA6/TR and pcDNA4/TO the second antibiotic was used for selection in Caco-2 cells.

2.3 Bacterial techniques

2.3.1 Preparation of competent cells

E. coli DH5α and BL21 DE3 competent cells were purchased from New England Biolabs (NEB, UK). 30ml *E. coli* JM109 cells were chilled on ice, pelleted at 8,000xg and resuspended in 13.5ml ice cold 100mM $MgCl_2$. Cells were pelleted again, resuspended in 6ml ice cold 100mM $CaCl_2$ and incubated on ice for 20 minutes. Cells were pelleted and resuspended in 1.5ml ice cold 100mM $CaCl_2$ (15% glycerol), pelleted for a final time and resuspended in 6ml $CaCl_2$ (15% glycerol). Competent cells were split into 50µl aliquots, flash frozen and stored at -80ºC.

2.3.2 Transformation of competent cells

Plasmid DNA, 2µl supercoiled plasmid or 10µl ligated product, was added to 50µl thawed competent cells and incubated on ice for 30 minutes. Cells were held at 42ºC for 90s, then returned to ice for 90s. 950µl pre-warmed LB was added to the cells, which were then incubated at 37ºC with shaking (200rpm) for 60 minutes. 50µl of either neat culture or a 10^{-1} dilution were plated onto LB amp agar.

2.4 DNA techniques

2.4.1 Plasmid purification

Plasmids were isolated from bacterial culture by Miniprep (Qiagen, Netherlands) for DNA modification and bacterial transformation experiments, and Maxiprep (Qiagen, Netherlands) for tissue transfection experiments.

For the Minipreps, 1.5-4.5ml bacterial culture were pelleted in microcentrifuge tubes at >16,000xg for 5 minutes. Pellets were resuspended in 250µl Buffer P1 (without LyseBlue reagent), then added 250µl Buffer P2 and mixed by inversion until the solutions became clear. Within 5 minutes 350µl of Buffer N3 was added and the solutions were mixed by inversion before being centrifuged at >16,000xg for 10 minutes. The supernatants were then applied to QIAprep spin columns and centrifuged at >16,000xg for 1 minute. The flow-through was discarded and 750µl Buffer PE applied to the top of the columns, followed by centrifugation at >16,000xg for 1 minute. The flow-through was discarded and the columns centrifuged again under the same conditions to remove residual buffer. The spin columns were then transferred into fresh microcentrifuge tubes and 30-50µl Buffer EB added to the top of the columns. The columns were incubated at room temperature for 1-3 minutes and

then centrifuged at 16,000xg for 1 minute. Collected DNA was stored at -20°C in Buffer EB until use.

For Maxipreps, 200-400ml overnight bacterial culture were centrifuged at 8,000xg for 15-30 minutes at 4°C and the pellets resuspended in 10ml Buffer P1 (no LyseBlue reagent used). 10ml Buffer P2 was added and the suspensions were mixed by swirling then incubated at room temperature for 5 minutes. After the lysis reaction, 10ml pre-chilled Buffer P3 was added and the lysate solutions were mixed by swirling then poured into a QIAfilter cartridge, where they were incubated at room temperature for >10 minutes. The lysate solutions were then filtered through the QIAfilter into 50ml falcon tubes. 2.5ml Buffer ER was added to the falcon tubes, the solutions mixed by inversion and then incubated on ice for 25-30 minutes. QIAGEN-tip 5000 columns were equilibrated with 10ml Buffer QBT. The chilled lysate solutions were then applied to the columns, followed by 2 x 30ml Buffer QC to wash the columns. The bound DNA was eluted using 15ml Buffer QN into depyrogenated 30ml centrifuge tubes. The solubilised DNA was precipitated by adding 10.5ml isopropanol and pelleted by centrifugation at 20,000xg for 30 minutes at 4°C. The supernatant was discarded and the pellet washed with 10ml endotoxin-free 70% ethanol, then centrifuged at 20,000xg for 15 minutes at 4°C. The supernatant was removed carefully so as not to disturb the pellet and the pellet allowed to air dry for up to 20 minutes before resuspension in <1ml Buffer TE. DNA was stored at -20°C in Buffer TE until use.

2.4.2 Polymerase chain reaction (PCR)

PCR reactions were performed using an Eppendorf Mastercycler (Eppendorf, UK) or an MJ Research DNA Engine PTC-200 Peltier thermocycler (Bio-Rad, USA). Mixtures typically contained: 2µl template DNA, 1µl forward primer, 1µl reverse primer, 10µl 10x GC buffer (NEB) 0.5µl Phusion polymerase (NEB, UK), 35.5µl dH_2O. A standard 30 cycle protocol was

used: 98ºC for 1 minute, followed by 30 cycles of 98ºC for 10s, annealing temperature for 20s, 72ºC for 30s, then 10 minutes at 72ºC. The annealing temperature was usually 2-6ºC lower than the lowest primer melting temperature (Table 2).

2.4.3 Agarose gel electrophoresis

DNA products were mixed with DNA gel loading buffer (NEB, UK) and loaded onto 1% (w/v) agarose gels, with 1kb or 100bp DNA markers (NEB, UK). Products were separated by electrophoresis at 80V for 30-60 minutes in 1x TAE buffer (40 mM Tris, 1 mM EDTA pH8, 19 mM Glacial acetic acid). Gels were stained in 1µg/ml ethidium bromide for 20 minutes and destained in dH_2O for 20 minutes, then imaged using a UVP VisiDoc- It Imaging System (UVP LLP, USA).

2.4.4 DNA purification

PCR products were cleaned using a QIAquick PCR Purification Kit (Qiagen, Netherlands). PCR reactions were transferred to microcentrifuge tubes and 5x starting volume Buffer PB (with pH indicator added as per kit instructions) added. Samples were vortexed then applied to QIAquick spin columns and centrifuged at >16,000xg for 1 minute. The flow-through was discarded and 750µl Buffer PE (with ethanol added as per kit instructions) was applied to the columns which was then centrifuged at >16,000xg for 1 minute. The flow-through was discarded and the columns centrifuged again under the same conditions to remove residual buffer. The columns were transferred into fresh microcentrifuge tubes. 20-50µl Buffer EB was applied to the columns which were incubated at room temperature for 1-5 minutes then centrifuged at >16,000xg for 1 minute. The cleaned PCR products were stored in Buffer EB at -20°C until use.

2.4.5 Gel extraction

Desired bands were extracted from agarose gels using a QIAquick Gel Extraction Kit (Qiagen, Netherlands). Bands were cut from the gel using a scalpel under a UV lamp and transferred to a tared microcentrifuge tube and weighed (should be <400mg). 3x gel weight of Buffer QG (assuming 100mg = 100µl) was added and the sample incubated at 50°C for ~10 minutes, with vortexing every 2-3 minutes. Once the gel slice had fully dissolved 1x gel weight isopropanol was added and the sample vortexed. The sample was then applied to a QIAquick spin column and centrifuged for 1 minute at >16,000xg. The flow-through was discarded and 500µl Buffer QG applied to the column, which was then centrifuged for 1 minute at >16,000xg. The flow-through was again discarded and 750µl Buffer PE (with ethanol, added per kit instructions) applied to the column and flushed through by centrifugation at >16,000xg for 1 minute. The flow-through was discarded and the column centrifuged again under the same conditions to remove residual buffer. The column was transferred to a fresh microcentrifuge tube. 20-50µl Buffer EB was applied to the column and incubated at room temperature for 1-5 minutes before centrifuging at >16,000xg for 1 minute. The purified DNA was stored in Buffer EB at -20°C until use.

2.4.6 DNA quantification

DNA products were quantified by applying 2µl DNA solution to a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific, USA) and performing analysis using NanoDrop 1000 software (Thermo Scientific, USA).

2.4.7 Restriction enzyme digestion

DNA products were digested with restriction enzymes in the appropriate conditions as recommended by the manufacturer (NEB, UK). Single digests typically contained 5-10µl purified DNA product, 1µl enzyme, 2µl buffer and 7-12µl dH_2O. Double digests typically contained 5-10µl 0.5µl enzyme 1, 0.5µl enzyme 2, 2µl buffer and 7-12µl dH_2O.

2.4.8 Dephosphorylation

Digested DNA was dephosphorylated to prevent religation. 1 µl of Antarctic Phosphatase (NEB, UK) and 5 µl Antarctic Phosphatase 1x buffer (NEB, UK) were added to the digest mixture and incubated at 37°C for 15 min. The enzyme was heat inactivated by incubation at 65°C for 5 min and the digested DNA cleaned with the QIAquick PCR Purification Kit (Qiagen, Netherlands).

2.4.9 Ligation

Digested plasmids and PCR products were mixed at a ratio of 1:3 and 1:6, with 0.5µl T4 DNA ligase (Promega, UK) and 1µl 10x reaction buffer, made up to 10µl with dH_2O. Mixtures were incubated at 16ºC overnight before cleaning (section 2.4.4) and transformation (section 2.3.2).

2.4.10 Single colony preparations

Transformant colonies were patched onto fresh LB amp agar and incubated at 37ºC overnight. Patches were picked and suspended in 50µl SCFS buffer (1.25% SDS, 2.5%

ficoll, 0.015% bromophenol blue, 10µg/ml RNase A, in 1x TAE buffer), incubated at room temperature for 15 minutes then centrifuged at 16,000xg for 30 minutes. The supernatant, containing the whole DNA of the cells, was analysed by agarose gel electrophoresis.

2.4.11 Genomic DNA purification

Genomic DNA was prepared from overnight *B. fragilis* cultures or clinical samples using a Wizard Genomic DNA Purification Kit (Promega, UK). 1ml samples were centrifuged at 16,000xg for 3 minutes and the pellets were washed twice in nuclease-free water before being resuspended in 600µl Nuclei Lysis Solution. Resuspended samples were incubated at 80°C for 5 minutes and then allowed to cool to room temperature. 3µl RNase Solution was added and the samples were vortexed and incubated at 37°C for 60 minutes. Samples were cooled to room temperature and 200µl Protein Precipitation Solution was added, the samples were vortexed and incubated on ice for 5 minutes. The samples were then centrifuged at 16,000xg for 3 minutes and the supernatant transferred into fresh microcentrifuge tubes containing 600µl isopropanol. The samples were mixed by inversion and then centrifuged at 16,000xg for 2 minutes. The supernatant was carefully removed and discarded and the pellet was washed with 600µl 70% ethanol before being centrifuged again at 16,000xg for 2 minutes. The wash solution was carefully removed and discarded and the pellet air dried for up to 30 minutes. The DNA was then rehydrated in 100µl Rehydration Solution overnight at 4°C. DNA was stored in Rehydration Solution at -20°C until use.

2.4.12 Quantitative PCR (qPCR)

Serial dilutions of *B. fragilis* genomic DNA and primers specific for the *gyrB* and *ubb* genes were used to generate calibration curves by Danny O'Connor (Patrick lab, Queen's University Belfast). Genomic DNA was isolated from 47 clinical samples or pure cultures of

B. fragilis (section 2.4.11) and qPCR mixtures prepared by the author using a Platinum SYBR Green qPCR SuperMix-UDG (Invitrogen, Thermo-Fisher Scientific, USA). qPCR mixtures were prepared in 384-well plates, with each well containing: 25µl SuperMix, 9µl nuclease-free water, 2.5µl 50mM $MgCl_2$, 1µl 1mg/ml BSA, 2µl genomic DNA, 0.2µl forward primer, 0.2µl reverse primer, 0.1µl ROX dye. qPCR analysis was performed by Dr. Derek Fairley (Belfast Health and Social Care Trust) and the number of copies of each gene in each sample was calculated from the calibration curves.

2.5 Protein techniques

2.5.1 Proteins

rBfUbb was purified from *E. coli* DH5α as described below, solid state bovine ubiquitin (Sigma Aldrich, USA) was suspended in phosphate buffered saline (PBS) (100mM NaH_2PO_4,18mM KH_2PO_4, 1.37M NaCl, 27mM KCl, pH7.5), and all ubiquitin-activating and ubiquitin-conjugating enzymes (CalTag Medsystems, UK) (Table 4) were diluted in E2 buffer (20mM Tris, 150mM NaCl, 2mM β-mercaptoethanol, 1% glycerol).

2.5.2 Expression

500ml cultures were grown to OD_{600nm} 0.4-0.6 and expression was induced with 1mM IPTG for 3h in *E. coli* DH5α pTRC99a::6xh-BfUbb, and with 400µM IPTG for 4h in *E. coli* BL21 DE3 pET19b::10xH-BfUbb.

2.5.3 Affinity purification

Induced 500ml culture was centrifuged at 8,000xg for 30 minutes and the pellet resuspended in 30ml loading buffer (26mM NaH_2PO_4, 0.5M NaCl, 20mM Imidazole, 1% PMSF) with protease inhibitors (Sigma Aldrich, USA). Cells were lysed using a 1.1 kW TS Benchtop Cell Disrupter (Constant Systems, UK) at 22kpsi and the lysate centrifuged at 50,000xg to pellet cell debris. The supernatant was filtered (0.45µm, Merck Millipore, Germany) and passed over a Ni-NTA agarose (Qiagen, Netherlands) column, then washed with loading buffer and eluted with elution buffer (26mM NaH_2PO_4, 0.5M NaCl, 0.5M Imidazole) in 1ml fractions (unless stated otherwise). Fractions were analysed by SDS-PAGE and coomassie staining and fractions containing the target protein were pooled.

2.5.4 Gel filtration

Pooled fractions from the affinity column were passed through a Superose 12 10/300 GL column (GE Healthcare, UK) and eluted with PBS (with 1mM EDTA and protease inhibitors) in 1ml fractions. Fractions were analysed by SDS-PAGE and coomassie staining and those containing pure protein were pooled and concentrated to 1-2ml in a Vivaspin 6, 5,000kDa (GE Healthcare, UK).

2.5.5 Quantification

Pure protein was quantified by applying 2µl homogenous sample to a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific, USA) and performing analysis using NanoDrop 1000 software (Thermo Scientific, USA).

2.5.6 Bradford assay

Bradford reagent (Sigma Aldrich, USA) was diluted 1:4 in assay buffer (3M Urea, 0.001% Triton X-100). Complex protein samples were diluted in assay buffer to a final volume of 100µl (in triplicate) then 900µl of diluted Bradford reagent was added to each sample (and a buffer control). Samples were incubated at room temperature for 15 minutes then the absorbance measured at 595nm. A standard curve was generated, using 0, 2, 4, 6, 8 and 10µg bovine serum albumin, and used to calculate the concentration of each sample.

2.5.7 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were mixed with SDS sample buffer (NEB, UK) with 100mM DTT and boiled (unless stated otherwise) and loaded onto Tris-tricine gels, consisting of a stacking gel and a resolving gel.

Stacking gel: 3.9ml dH_2O, 1.55ml gel buffer (3M Tris-Cl, 0.3% SDS, pH8.45), 0.8ml 30% bis-acrylamide, 100µl APS, 8µl TEMED
Resolving gel: 3.5ml dH_2O, 5ml gel buffer, 4.9ml bis-acrylamide, 1.6ml glycerol, 100µl APS, 8µl TEMED

Proteins were separated by electrophoresis at 120V for 60-90 minutes in tricine buffer (0.1M Tris, 0.1M Tricine, 0.1% (w/v) SDS).

2.5.8 Coomassie staining

Resolving gels were incubated in 50% dH_2O, 40% (v/v) methanol, 10% (v/v) glacial acetic acid, 0.125% (w/v) Coomassie Blue Reagent (Thermo Fisher Scientific, UK) for 20 minutes

and destained 70% dH_2O, 20% (v/v) methanol, 10% (v/v) glacial acetic acid for 1h. Coomassie stained gels were visualised with a Gel Doc XR imaging system (Bio-Rad, USA).

2.5.9 Silver staining

Resolving gels were stained using the Proteosilver Silver Stain Kit (Sigma Aldrich, USA) as per the manufacturer's instructions. After electrophoresis gels were incubated on a rocker in: Fixing solution (50% ethanol, 10% acetic acid) overnight, 30% ethanol for 10 minutes, Ultrapure water for 10 minutes, Sensitizer solution (1ml ProteoSilver Sensitizer, 99ml Ultrapure water) for 10 minutes, Ultrapure water for 2x 10 minutes, Silver solution (1ml ProteoSilver Silver Solution, 99ml Ultrapure water) for 10 minutes, Ultrapure water for 60 seconds, Developer solution (5ml ProteoSilver Developer 1, 0.1ml ProteoSilver Developer 2, 95ml water) until the bands started to appear. Once the bands of the gel were clearly visible ProteoSilver Stop Solution was added and the gel incubated for 15 minutes. The gel was then washed in Ultrapure water for 15 minutes before imaging. Silver stained gels were visualised with a Gel Doc XR imaging system (Bio-Rad, USA).

2.5.10 Western blotting

Proteins from resolving gels were transferred to a PVDF membrane (Life Technologies, Sigma Aldrich, USA) at 100V for 60 minutes in transfer buffer (25mM Tris, 192mM Glycine). Membranes were incubated in blocking buffer (5% (w/v) milk protein in PBS with 0.05% (w/v) Tween-20) for 1h then washed in PBS-T (PBS with 0.05% Tween-20). Membranes were then incubated in primary antibody diltued in blocking buffer (1:10,000 for anti-BfUbb and 1:5,000 for anti-1705) for 2h, washed in PBS-T, then incubated in secondary antibody diluted in blocking buffer (1:20,000 anti-rabbit). The membranes were washed in PBS-T and visualised using Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare,

UK) and X-ray films (GE Healthcare, UK) exposed to the chemiluminescent blots and developed using an X-ray processor (Medical Index GmbH).

2.5.11 Ubiquitin transfer assay

rBfUbb (His-tagged) was incubated at 37ºC for 1h in 4 different conditions:

1. 0.5mg/ml rBfUbb
2. 0.5mg/ml rBfUbb, 200mM DTT
3. 0.5mg/ml rBfUbb, 0.2µM E1 activating enzyme
4. 0.5mg/ml rBfUbb, 0.2µM E1 activating enzyme, 200mM DTT

All mixtures were prepared in E2 buffer. After the initial incubation, 10µl of each mixture was mixed with 10µl 0.5µM of each E2 conjugating enzyme (in E2 buffer) then incubated at 37ºC for a further 2h. 0.5mg/ml rBfUbb = 60µM rBfUbb, therefore the molar ratio of rBfUbb:E1:E2 is 60:0.2:0.25 (300:1:1.25). Samples were analysed by SDS-PAGE and Western blotting.

2.5.12 Pull down with His-tagged rBfUbb

Caco-2 cell lysate was incubated at 37ºC with gentle shaking for 3h in 4 different conditions:

1. 2mg/ml Caco-2 cell lysate
2. 2mg/ml Caco-2 cell lysate, 0.2mg rBfUbb
3. 2mg/ml Caco-2 cell lysate, 100mM DTT
4. 2mg/ml Caco-2 cell lysate, 0.2mg/ml rBfUbb, 100mM DTT

All mixtures were prepared in PBS. After incubation, samples were passed over a Ni-NTA agarose (Bio-Rad, USA) column, washed in PBS and eluted in a single fraction with elution buffer. Fractions were analysed by SDS-PAGE and coomassie staining or Western blotting.

2.6 Tissue culture techniques

2.6.1 Culture and passage

Human epithelial colorectal adenocarcinoma cell line Caco-2 cells were grown in Eagle's Minimum Essential Medium with Earle salts (Sigma Aldrich, USA), supplemented with 10% foetal bovine serum (Gibco, Thermo Fisher Scientific, UK), 1% non-essential amino acids (Gibco, Thermo Fisher Scientific, UK), 5mM L-glutamine (Gibco, Thermo Fisher Scientific, UK), 100U/ml penicillin and 10µg/ml streptomycin (as Pen-Strep, Gibco, Thermo Fisher Scientific, UK). Cultures were incubated at 37ºC in 5% CO_2.

For passage, culture medium was removed, cells were washed in PBS, and then incubated in 10ml 1x Trypsin (Thermo Fisher Scientific, UK) for 5 minutes. 1ml of suspended cells were then transferred into 24ml of fresh medium.

2.6.2 Transfection

24-well plates were seeded with 10^2 cells/ml in culture medium and grown for 48h. Plasmid DNA was mixed with Lipofectamine 2000 (Thermo Fisher Scientific, UK) in Opti-Mem (Thermo Fisher Scientific, UK) according to manufacturer's instructions. Culture medium was removed from the wells and cells were washed in Opti-Mem, then transfection mixtures were added. Cells were incubated at 37ºC for 18-24h then transfection mixtures were removed and replaced with culture media. Cells were incubated for another 24h then passaged into culture medium containing appropriate antibiotic (Table 3). After 1 week (replacing culture medium every 48h), dead cells were washed away with PBS and antibiotic resistant cells suspended in trypsin. 24 cells were picked under a light microscope and transferred to

individual wells containing culture medium with antibiotic. After ~2 weeks 8 of the 24 wells were used to establish cell lines containing the plasmid.

2.6.3 Preparation of cell lysate

Between 2 and 4 flasks (75cm^2) of Caco-2 cells were suspended in 10ml trypsin each. Cells were pooled and centrifuged at 5,000xg then resuspended in 10ml lysis buffer (PBS with 1mM EDTA, 0.2% Triton X-100, 1% PMSF, protease inhibitors). Cells were incubated on ice for 20 minutes (with intermittent inversions) then passed through a 25 gauge needle 4 times. Lysate was centrifuged at 16,000xg for 10 minute and the supernatant collected and used immediately.

2.7 Preparation of periplasmic extract and concentrated supernatants

2.7.1 Preparation of periplasmic extract

1ml of late-log phase/early stationary-phase *B. fragilis* cell culture were centrifuged at 16,000xg, resuspended in PBS, then centrifuged again. Pellets were then resuspended in 150µl ice cold 20% sucrose/10mM Tric-Cl (pH7.5) and vortexed. 5µl 0.5M EDTA (pH8) was added after which cells were incubated on ice for 10 minutes then centrifuged at 16,000xg for 5 minutes at 4ºC. Pellets were resuspended in 100µl ice cold dH_2O and incubated on ice for 10 minutes, then centrifuged at 16,000xg for 5 minutes. The supernatant was collected and stored at -20ºC.

2.7.2 Preparation of concentrated supernatant

400ml late log phase *B. fragilis* culture was centrifuged at 8,000xg and the supernatant filtered through a 0.45µm PVDF filter (Merck Millipore, Germany). Filtered supernatant was then concentrated to ~1-3ml in a Centricon Plus-70 Centrifugal Filter (Merck Millipore, Germany) according to the manufacturer's instructions.

2.8 Isolation of OMV

2.8.1 Tangential flow filtration

5L late log phase/early stationary phase WT *B. fragilis* culture was centrifuged at 8,000xg and the supernatant filtered (0.45µm) and concentrated to ~200ml using tangential flow filtration at a torque setting of 400 inch-lb.

2.8.2 Centrifugal filtration

The ~200ml concentrated supernatant was concentrated to ~1-3ml as described in section 2.7.2 and the concentration determined by Bradford assay.

2.8.3 Density centrifugation

A density gradient of Opti-Prep (Axis Shield, Sigma Aldrich, USA) was prepared in PBS. From bottom to top, Opti-Prep layers were: 1ml 30%, 1ml 25%, 2ml 20%, 2ml 15%, 2ml 10%. 3mg concentrated supernatant was loaded onto the top of the gradient, which was then centrifuged at 30,000xg for 3h. The tube content was then fractioned (0.5ml/fraction).

2.8.4 Alkaline phosphatase and β-galactosidase assays

For each preparation of OMV, the fractions of the Opti-prep density gradient were assayed for alkaline phosphatase activity (Mossner *et al.*, 1980) and β-galactosidase activity (Zhang & Bremer, 1995). The arbitrary units/ml of each enzyme were calculated and the ratio of alkaline phosphatase units to β-galatosidase units determined. The three fractions with the highest ratio were pooled.

2.8.5 Dynamic light scattering (DLS)

50µl of 0.5mg/ml WT *B. fragilis* concentrated supernatant was analysed by DLS, using a Zetasizer APS (Malvern Instruments, UK) according the manufacturer's instructions, to calculate an average diameter of the OMV. BHI-S and PBS controls were analysed for comparison.

2.9 Proteomic techniques

2.9.1 Sample preparation

1mg of the concentrated samples of OMV was precipitated using methanol and chloroform: to the sample was added 4x starting volume (SV) methanol, 2x SV chloroform and 2x dH_2O (with vortexing after each addition). The sample was then centrifuged at 16,000xg for 2 minutes, after which the top aqueous layer was removed and 3x SV methanol was added. The sample was vortexed and centrifuged at 16,000xg for 10 minutes. The supernatant was removed and discarded and the pellet was vacuum dried to remove any additional solvent (Le Bihan *et al.*, 2011). The dried pellet was resuspended in a mixture of 62.5µl 8M urea, 12.5µl 200mM DTT and 12.5µl 1M ammonium bicarbonate and incubated for 30 minutes at

room temperature. 12.5µl 500mM iodoactemide and 10µl 1mg/ml trypsin were added and the sample mixed thoroughly. The mixture was left shaking at room temperature overnight, after which 10µl were cleaned using Stagetips (Le Bihan *et al.*, 2011; Rappsilber *et al.*, 2007).

2.9.2 LTQ-Orbitrap mass spectrometry

The samples were loaded onto a 1200 binary HPLC system (Agilent, USA) for capillary HPLC, which delivered peptides to a hybrid LTQ-Orbitrap XL (Thermo-Fisher Scientific, UK) for mass spectrometry. These devices were controlled using Xcalibur 2.0.7 software (Thermo Scientific, UK) (Le Bihan *et al.*, 2010). HPLC-MS was performed by Dr Sarah Martin (Thierry lab, SynthSys, University of Edinburgh).

2.9.3 Analysis

Mascot (Matrix Science, USA) was used to compare experimental peptide hits with the pattern of predicted hits from the genome sequence of *B. fragilis* (NCBI) and produce a list of identified proteins. Progenesis (Nonlinear Dynamics, UK) was used for label-free quantitation and comparison of the OMV and PP samples (Le Bihan *et al.*, 2011). Further protein identification was performed using protein BLAST (NCBI).

2.10 OMV antibacterial assays

2.10.1 Bacterial culture and inoculation

Growth curves were prepared in triplicate in 5ml aliquots of LB without antibiotics. For the first experiment, aliquots were inoculated with 20µl overnight *S. enterica* Typhimurium

culture along with 500µl of either PBS, 1mg/ml WT concentrated supernatant or 1mg/ml Δ*ubb* concentrated supernatant. Cultures were grown at 37ºC with shaking, and at 0, 2, 4, 6 and 8 hours after inoculation 100µl of each culture was diluted to an appropriate concentration (10^{-1}-10^{-3}) in PBS and spread onto LB agar. 5ml aliquots of LB were also inoculated with just concentrated supernatants or PBS and grown for 8h at 37ºC with shaking, after which 100µl were spread onto LB agar. Plates were incubated at 37ºC for 48h, after which colonies were counted and Cfu/ml was calculated based on the dilution factor.

For the second experiment, aliquots were inoculated with 5µl *S. enterica* Typhimurium overnight culture then grown at 37ºC with shaking for 2h. 500µl of either PBS or 0.5mg/ml, 1mg/ml, or 2mg/ml of concentrated supernatant from WT, WT (gent), pGB920, or Δ*ubb B. fragilis* was then added to the cultures. The cultures were then grown for a further 8h, and at 2, 4, 6, 8 and 10h, 100µl were taken from each culture and serially diluted (10^{-1}-10^{-6}). 10µl of each dilution were spotted onto LB agar. Controls were set up the same as in experiment 1. Plates were grown at 37ºC for 48h, after which the number of colonies were counted from dilutions with 2-30 colonies and the Cfu/ml were calculated based the dilution factor.

2.10.2 T-test analysis

T-tests (Microsoft Excel 2010, USA) were used to compare 2 datasets at different time points. Results of the T-tests can be found in Appendix VIII.

3. Characterisation of *B. fragilis* ubiquitin (BfUbb)

3.1 Background

3.1.1 BfUbb

Patrick *et al* (2011) identified a gene (*ubb*, formerly designated BF3883) in NCTC9343, which shares 76% identity (151/199bp, e value of $5x10^{-29}$) with a ubiquitin-encoding gene from migratory grasshopper entomopoxvirus. The *ubb* gene encodes a 76aa protein (BfUbb) which shares 63% identity (48/76aa) with human ubiquitin (Ubc52) (Figure 7 A). The key lysines and arginines involved in interactions with E1 activating enzyme are conserved in BfUbb, however there are some significant differences between the primary structure of BfUbb and Ubc52. Firstly, BfUbb contains an N-terminal signal sequence, suggesting it is exported from the cytoplasm. Secondly, in place of the double glycine (GG) residues that are required for covalent interaction between ubiquitin and E1 activating/E2 conjugating enzymes, there is a single cysteine residue. Thirdly, BfUbb is missing 1 of the 7 lysines used in ubiquitin chain formation: K11, which is replaced with a tryptophan (W11) (Patrick *et al*., 2011). A comparison of the solved structure of Ubc52 (Vijaykumar *et al.*, 1987) and the predicted structure of BfUbb, created using PHYRE2 (Kelley & Sternberg, 2009), shows that BfUbb has a β-grasp fold with a free C-terminal tail, similar to ubiquitin. However, there are some differences in the β-sheets (note the yellow and orange sections of the 2 structures).

A

```
BfUbb MRFIKQVLLTITLCNIMLFALPSTVNAMQVFIKNRYGWTITLEVSPTDTVENVKQKIQDK
      :                            :.:.:.  : :::::: :.::.:::: :::::
Ubc52 M----------------------------QIFVKTLTGKTITLEVEPSDTIENVKAKIQDK
                                          10        20        30
               *     *     *        *        *
BfUbb EGFPPDKIRLIYGGKQMEDGRTLADYNVQKDSTILICIRDVDC
      ::.:::. :::..:::.:::::::.:::.::.::. . .:
Ubc52 EGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG
             40        50        60        70
```

B

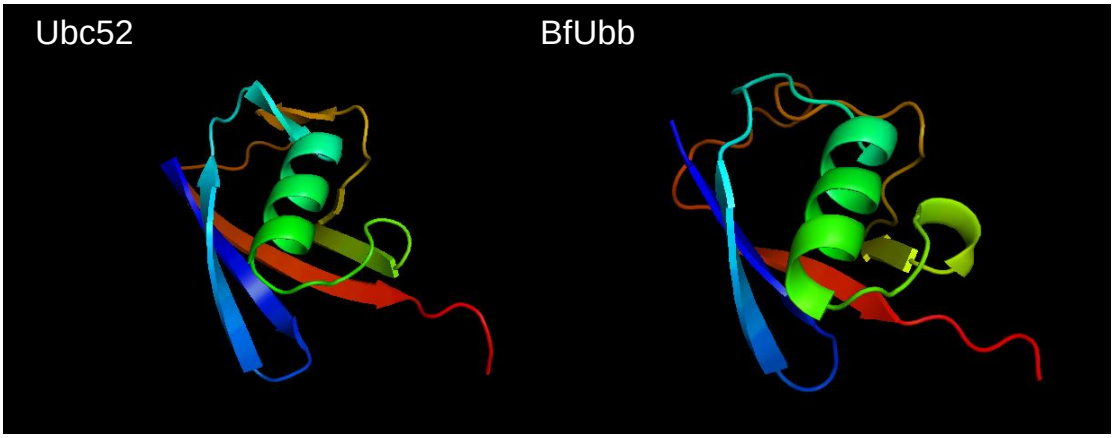


Figure 7 Comparisons of human ubiquitin Ubc52 and *B. fragilis* ubiquitin BfUbb. **(A)** A sequence comparison showing the additional periplasmic signal sequence at the N terminus of BfUbb, the lack of the C terminal double glycine residue, and the conservation of key arginine and lysine residues involved in interactions with E1 activating enzyme (bold, asterisks) (Patrick *et al.*, 2011). **(B)** The solved 3D structure of Ubc52 (Vijaykumar *et al.*, 1987) and the predicted 3D structure of BfUbb using PHYRE2 (Kelley & Sternberg, 2009). Structures are false coloured on a spectrum for comparison of regions.

The *ubb* gene in *B. fragilis* is believed to be the product of an interkingdom HGT and is located in an 11kb, low GC region of the NCTC9343 genome (Patrick *et al.*, 2011). It is also present in 7 other sequenced strains of *B. fragilis* (NCBI) and has been identified by PCR in several other strains isolated from clinical samples of various abscesses, however it has not yet been detected in any other Bacteroides species (Patrick *et al.*, 2011).

A deletion mutant of *ubb* (Δ*ubb*) was generated using a previously described method (Patrick *et al.*, 2009), replacing *ubb* with an *ermF* (erythromycin resistance) cassette. This mutation did not appear to affect growth or morphology of *B. fragilis* in BHI-S or DM. Reverse transcription PCR (RT-PCR) of *ubb* mRNA indicated that BfUbb is transribed in both the exponential and stationary phases of growth (Figure 8 a). A 6xHis fusion of *ubb* was expressed from pTRC99A in DH5α and purified. This recombinant (rBfUbb) was used to generate antibodies against BfUbb, which were then used to probe for BfUbb in whole cell extracts of NCTC9343 and the Δ*ubb* mutant by Western blotting (Figure 8 b). 3 bands were detected in the NCTC9343 strain: a ~12kDa band consistent with the size of the pre-processed holoprotein; a ~9kDa band, the strongest band, consistent with the size of the processed protein (8.6kDa); a smaller band which is believed to be a product of degradation. None of these bands were detected in the Δ*ubb* mutant. Cleavage of the signal sequence to produce a ~9kDa protein is consistent with BfUbb being exported to the periplasm. A smeared band the size of processed BfUbb could also be detected by Western blotting in the concentrated supernatants of NCTC9343 (Figure 8 c), suggesting that the protein is packaged into OMV and secreted (Patrick *et al.*, 2011).

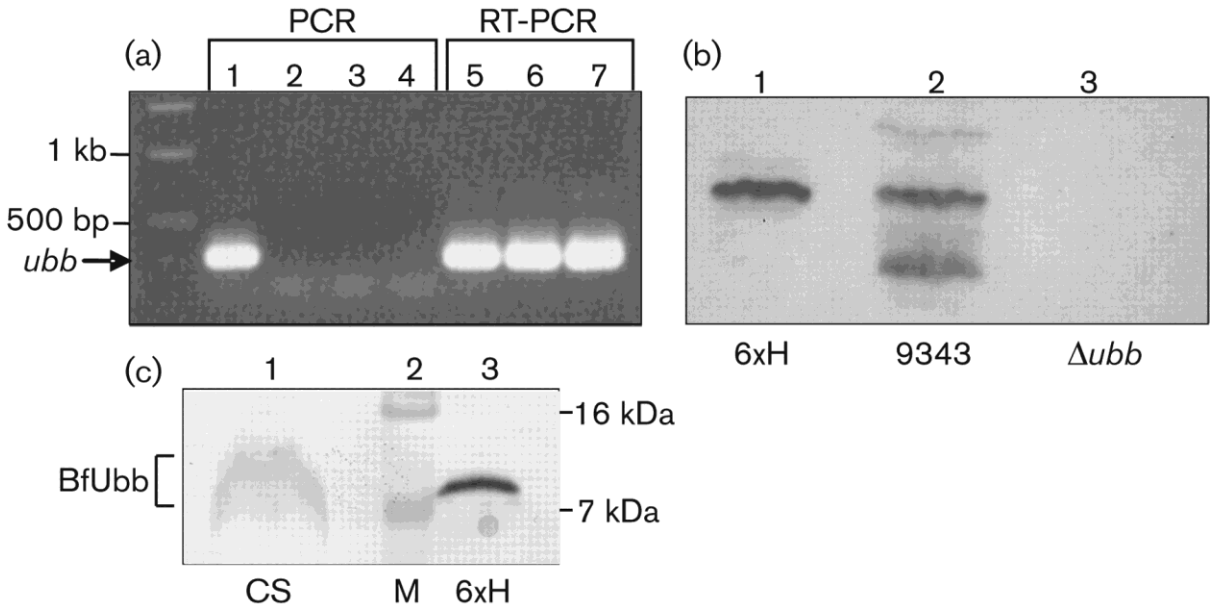


Figure 8 "Detecting expression of BfUbb. (a) Agarose gel showing transcription of ubb detected by RT-PCR. Lanes: 1, PCR using genomic DNA; 2–4, PCR using RNA from cultures at OD_{600} 0.4, 0.8 and 1.6 (indicates absence of DNA contamination); 5–7, RT-PCR using RNA from cultures at OD_{600} 0.4, 0.8 and 1.6. The position of ubb is indicated with an arrow. (b) Immunoblot using rabbit anti-BfUbb polyclonal antiserum. Lanes: 1, purified 6×His-BfUbb; 2, whole-cell extract of B. fragilis NCTC9343; 3, whole-cell extract of B. fragilis Δ ubb. The lowest band in lane 2 represents a degradation product of BfUbb. (c) Immunoblot of concentrated supernatant from a culture of NCTC9343 grown in DM (lane 1), compared with purified 6×His-BfUbb (lane 3). Molecular mass markers are shown in lane 2" (Patrick *et al.*, 2011).

Since the deletion of *ubb* did not appear to affect the viability of *B. fragilis* it was hypothesised that BfUbb may act as an effector in the host cells. The ubiquitylation cycle was recreated using HeLa extracts and a biotinylated lysozyme substrate, and visualised by Western blotting against the lysozyme. When the HeLa extract and lysozyme were mixed directly and incubated at 37ºC for 180 minutes, the increased molecular weight of the lysozyme indicates ubiquitylation of the substrate. However, if the HeLa extract was preincubated with BfUbb, ubiquitylation does not occur (Figure 9 a). Moreover, when mixed

together in non-reducing conditions and incubated at 37ºC for 60 minutes, BfUbb was found to covalently bind with E1 activating enzyme in the absence of ATP (Figure 9 b). It is possible, therefore, that BfUbb acts as a suicide substrate, binding and inactivating enzymes of the ubiquitylation cycle, by forming a disulphide bridge between the C-terminal cysteine and the active site of the enzyme (Patrick *et al.*, 2011).

Unlike the species of Actinobacteria that utilise Pup as a marker for protein degradation, *B. fragilis* does not have a 26S- or 20S-like proteasome system, nor does it encode any known E1 activating enzymes, E2 conjugating enzymes or E3 ligases (Patrick *et al.*, 2011). It is therefore unlikely that BfUbb is acting as a protein modifier within the bacterial cell. Many bacterial pathogens secrete virulence factors that interfere with the ubiquitin system of the host (Ashida *et al.*, 2014; Munro *et al.*, 2007); most of these are E3 ligases, deubiquitylases or other ubiquitin-binding proteins and usually have a specific target in a specific pathway. However, a bacterial homologue of ubiquitin itself, such as BfUbb, has the potential to interfere with any host cell process that utilises ubiquitin. In the case of *B. fragilis* this may be the TLR-activated inflammatory response; *B. fragilis* has been shown to interact with the inflammatory pathways by other mechanisms (Shen *et al.*, 2012) (section 1.1.4). Such activity could either contribute to inflammatory disease, such as that caused by ETBF, or protect the host against inflammatory disease, as observed with PSA by Shen *et al* (2012). Alternatively, BfUbb may act as a virulence factor in other diseases, or be a causative agent in autoimmune diseases, since it bears significant homology to a human protein (Patrick *et al.*, 2011).

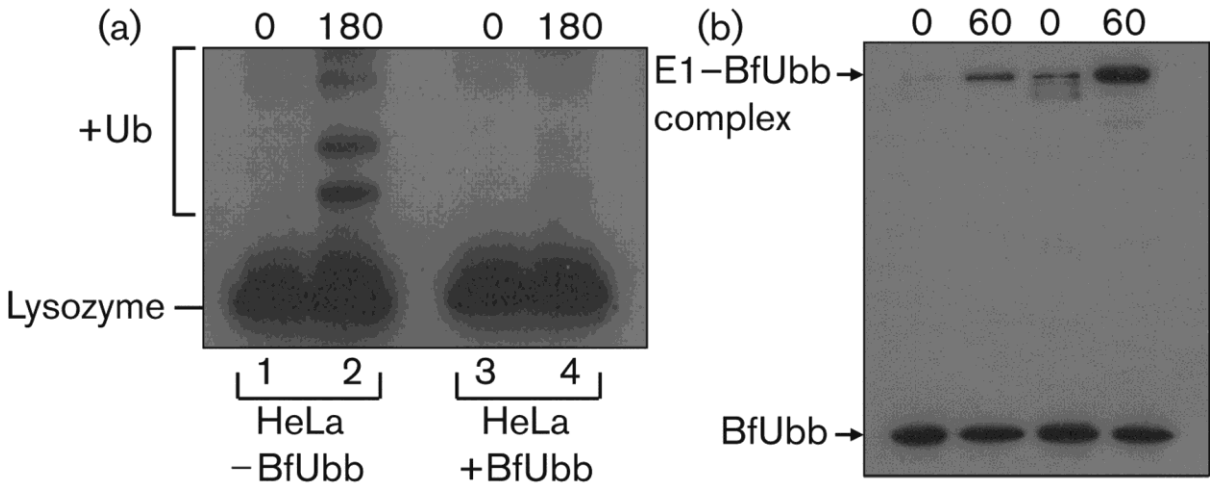


Figure 9 "In vitro activity of BfUbb. (a) Immunoblot of in vitro ubiquitination of lysozyme using HeLa cell extract in the absence (lanes 1 and 2) or presence (lanes 3 and 4) of BfUbb. Samples were taken at the start of the reaction (0) and at 180 min (180). Lane 2 shows the increase in molecular mass of the lysozyme substrate following covalent attachment of ubiquitin (+Ub). This covalent modification is inhibited by the addition of BfUbb (lane 4). (b) Covalent complexes between human E1 and BfUbb, under non-reducing conditions, were detected by immunoblotting using anti-bovine ubiquitin polyclonal serum. Two reactions using different concentrations of E1 are shown, with time points at 0 and 60 min. Note there was some antibody cross-reactivity with E1 at the higher concentration" (Patrick *et al.*, 2011).

3.1.2 Aims

There is evidence that BfUbb is capable of interfering the ubiquitylation cycle of eukaryotic cells (Patrick *et al.*, 2011). In order for this to occur *in vivo* the protein must be delivered to epithelial cells of the GI tract. BfUbb can be detected in the concentrated supernatant of WT cell culture, suggesting that it is present in OMV. *B. fragilis* OMV are known to be engulfed by DCs (Shen *et al.*, 2012) and may interact with other cells in the gastro-intestinal epithelium, thus providing a mechanism for delivery of BfUbb to the host.

BfUbb can bind E1 activating enzyme and inhibit ubiquitylation of lysozyme *in vitro* (Patrick *et al*., 2011), suggesting that it can act as a suicide substrate for ubiquitylation reactions. However, inhibiting all ubiquitylation in the host cell would lead to necrosis (Xu *et al.*, 2010), and it is more likely that *B. fragilis*, which exists as a commensal rather than a pathogen in the gut, would have retained BfUbb as a mechanism to affect a single target or pathway to benefit its own long term survival (and coincidentally the health of the host). Identifying binding partners for BfUbb, including specific E2 conjugating enzymes and E3 ligases, would aid significantly in determining its function.

On average mammalian cells contain ~10^8 molecules of ubiquitin (Haglund & Dikic, 2005; Komander & Rape, 2012). If BfUbb is utilising the host cell ubiquitin pathway it needs to be able to compete with endogenous ubiquitin, however it is unlikely that the OMV can deliver enough BfUbb for it to match the concentration of ubiquitin in the whole cell. Therefore, BfUbb needs to be highly concentrated at a specific location; the OMV may deliver its contents directly to a location or BfUbb may interact with a chaperone or similar protein. Determining the subcellular localisation of BfUbb may help determine its target and function. Lastly, the phenotypic effect resulting from the presence of BfUbb in the host cell may the biggest indicator of its function *in vivo*. Morphological changes would suggest an alteration of cell cycle or cell structure, cytokine changes would indicate induction or suppression of cell signalling for pathways such as inflammation, whilst necrosis would suggest that all ubiquitylation is being inhibited. Alternatively, various clinical isolates containing *B. fragilis* can be screened for the presence of the *ubb* gene, potentially revealing correlations between the presence or absence of BfUbb and 1 or more diseases.

The purpose of the following experiments was to characterise BfUbb, identify any target proteins/binding partners, determine its subcellular location within the host cell, and highlight any correlations between the presence of BfUbb and phenotypic changes in the host cells.

3.2 Results

3.2.1 Purification of rBfUbb

Expression of the his-tagged, recombinant Bfubb, rBfUbb, in 500ml *E. coli* DH5α pTRC99A-6xHBfUbb was induced with 1mM IPTG for 3 hours and the presence of the protein confirmed by SDS-PAGE and coomassie staining (Figure 10 A), after which cells were harvested and lysed. The lysate was filtered and passed over a nickel affinity (Ni) column, which was washed and then eluted with imidazole in a single fraction. The presence of rBfUbb in the elution fraction was confirmed by SDS-PAGE and coomassie staining (Figure 10 B). The whole fraction was then concentrated using a centrifugal filter to ~2-3mg/ml (total ~2ml, ~4-6mg). 0.5ml (~1-1.5mg) was passed through a Sepharose gel filtration column and eluted in 1ml fractions. Pure rBfUbb-containing fractions were identified by SDS-PAGE and coomassie staining (Figure 10 C) and pooled, then concentrated to a final concentration of ~0.5-1mg/ml (total ~1ml, ~0.5-1mg).

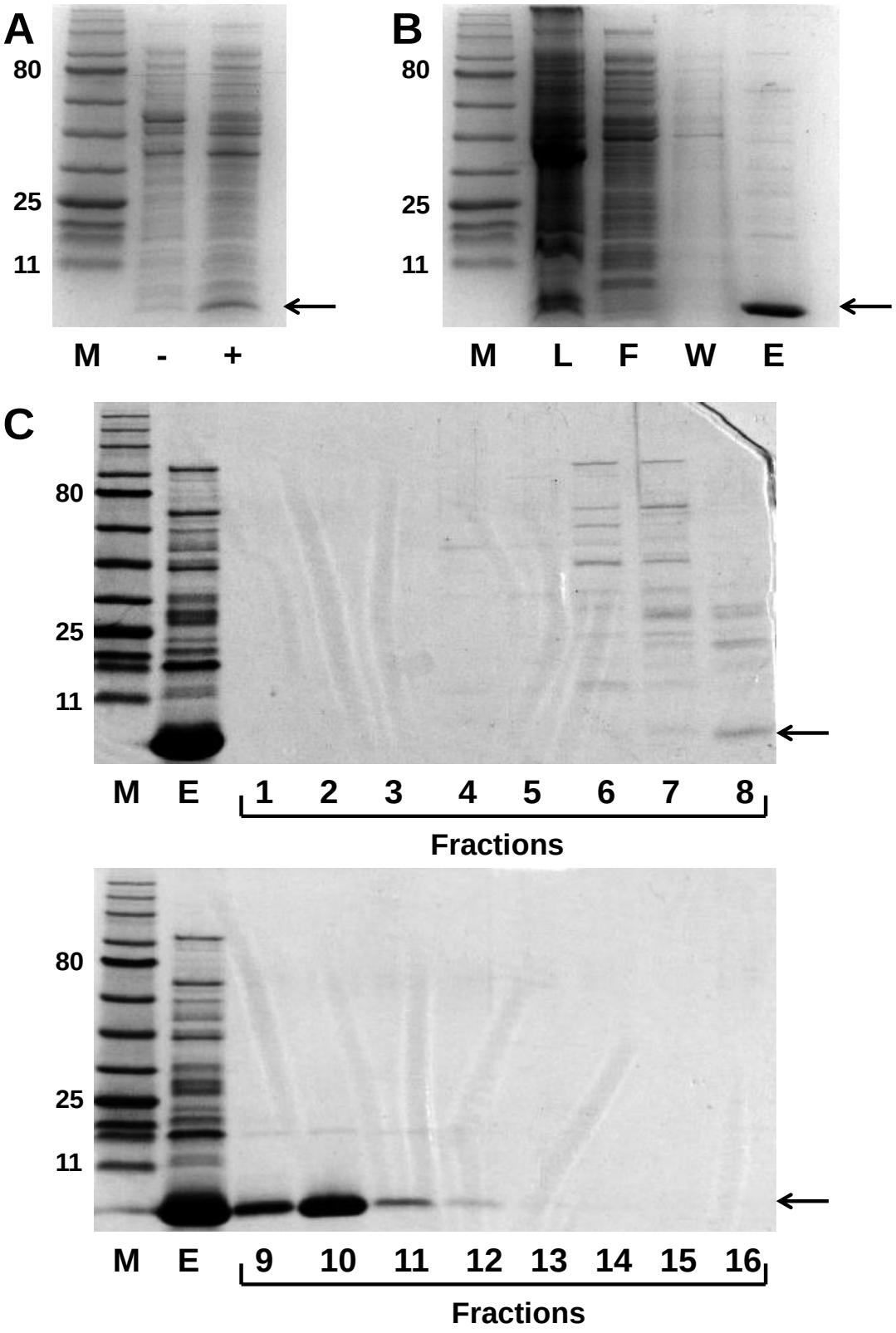

A
80
25
11
M
-
+
B
80
25
11
M
L
F
W
E
C
80
25
11
M
E
1
2
3
4
5
6
7
8
Fractions
80
25
11
M
E
9
10
11
12
13
14
15
16
Fractions

Figure 10 Coomassie stained gels showing the expression and purification of rBfUbb from *E. coli* DH5α pTRC99A-6xHBfUbb. **(A)** Whole cell extract of culture without (-) and with (+) 1mM IPTG. **(B)** Samples from the Ni affinity column: lysate prior to loading (L), the flow-through during loading of the column (F), the wash-off (W) and the eluted fraction (E). **(C)** The samples from the gel filtration column: eluted fraction from the Ni column (E), gel filtration fractions 1-8 and 9-16. The expected location of rBfUbb (9.6kDa) is indicated by an arrow on each gel. A second band, ~17kDa, can be seen in most lanes, which is believed to be a dimeric form of rBfUbb.

3.2.2 BfUbb forms multimers *in vitro*

When kept at 37ºC for >1 hour (or >6 months at -20ºC) in the absence of glycerol or a reducing agent, rBfUbb forms multimeric structures. A dimer is visible by coomassie staining (Figure 11 A) and up to a tetramer is visible by Western blotting (Figure 11 B). This effect is reproducible between different preparations of rBfUbb. The tetrameric and trimeric structures can be broken down by adding DTT to the sample (100mM in 0.5mg/ml BfUbb) or by boiling at 90ºC for 10 minutes, whilst the dimeric structure seems very stable (Figure 11 B). DTT-sensitivity suggests that the intermolecular bonds between BfUbb residues in these multimers are disulphide bridges between cysteine resides. BfUbb contains 2 cysteine residues (C70 and C76), both in the C-terminal tail (Figure 7 A), which would allow formation of a unique chain (Figure 12). Interestingly, when incubated with human E1 activating enzyme (Uba1), pentamers and hexamers have also been observed by Western blotting (Figure 11 C). The E1 enzyme is not involved in ubiquitin chain formation so there is no clear mechanism for multimerisation involving this protein. BfUbb multimers have never been observed in samples extracted from *B. fragilis* culture (Figure 8, Figure 26), so the biological relevance of this phenomenon is therefore unclear.

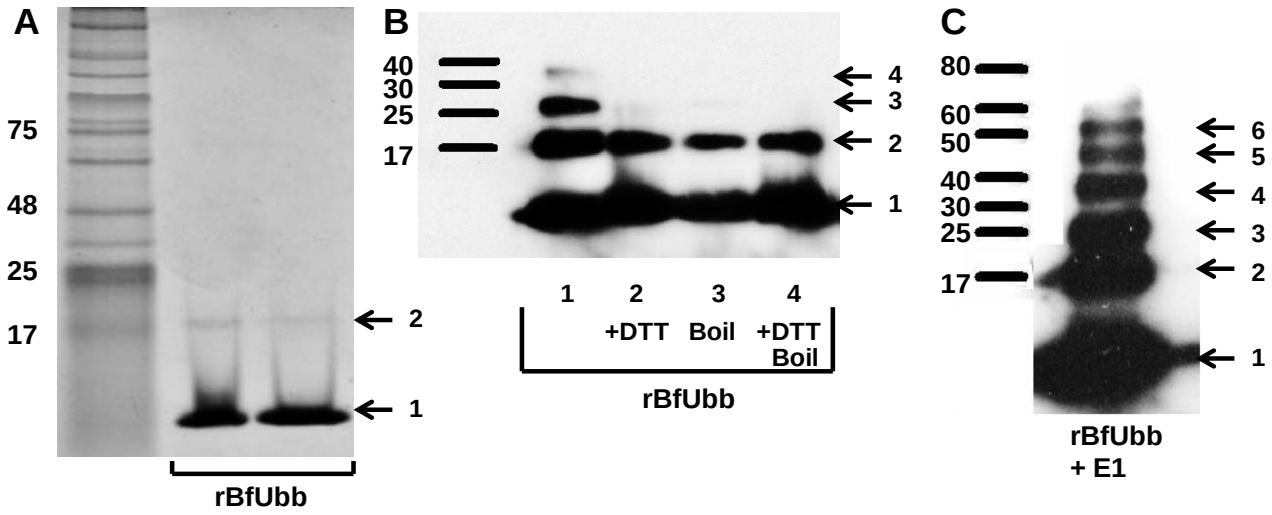


Figure 11 Multimerisation of rBfUbb. Arrows indicate the number of rBfUbb molecules in the complex represented by each band. **(A)** Coomassie stain of 2 independently prepared samples of rBfUbb, 1mg/ml, stored at -20ºC for >6 months. **(B)** Western blot of 0.5mg/ml rBfUbb incubated at 37ºC for 2 hours. Lanes: 1, rBfUbb only; 2, + 100mM DTT; 3, boiled at 90ºC post-incubation; 4, + DTT and boiled post-incubation using rabbit anti-BfUbb primary antibodies and HRP-conjugated anti-rabbit secondary antibodies. X-ray film was exposed to chemiluminescent blot for 30s. **(C)** Western blot of 0.5mg/ml rBfUbb incubated with 1mM E1 at 37ºC for 2 hours, exposure of 1 minute.

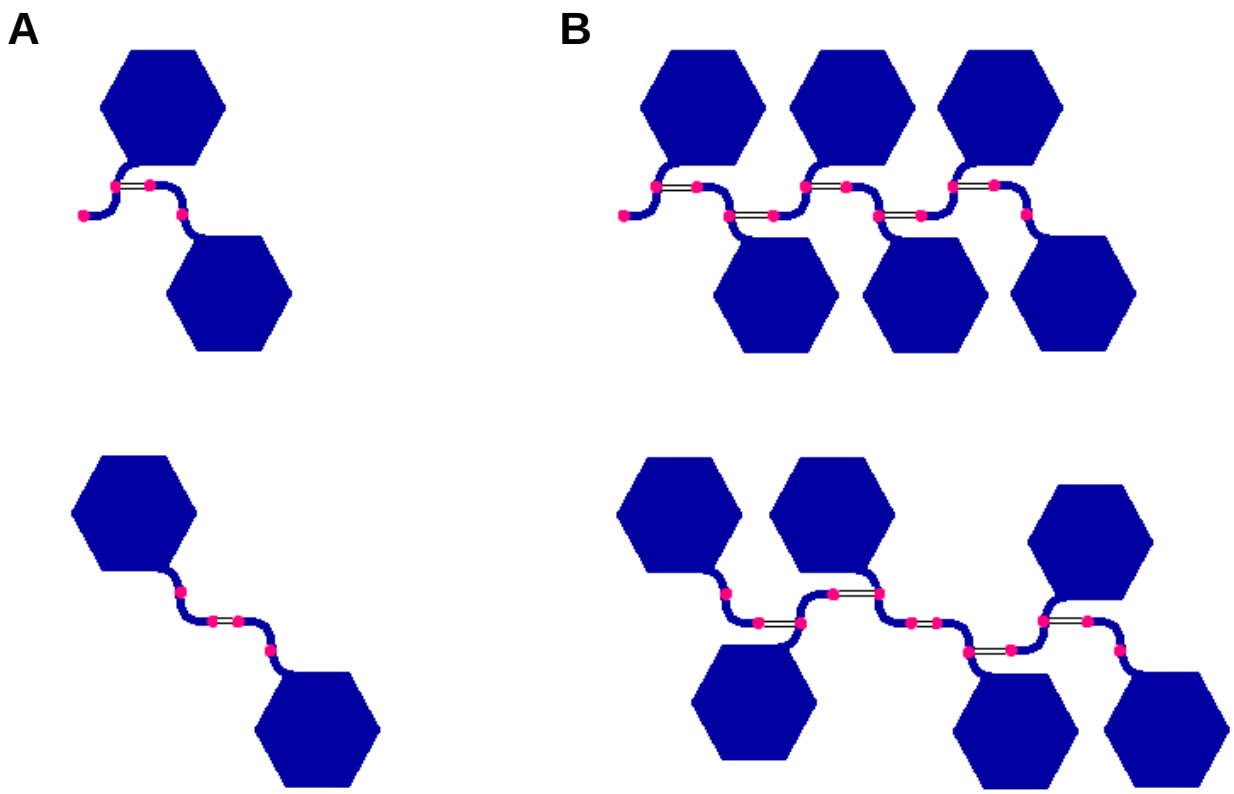
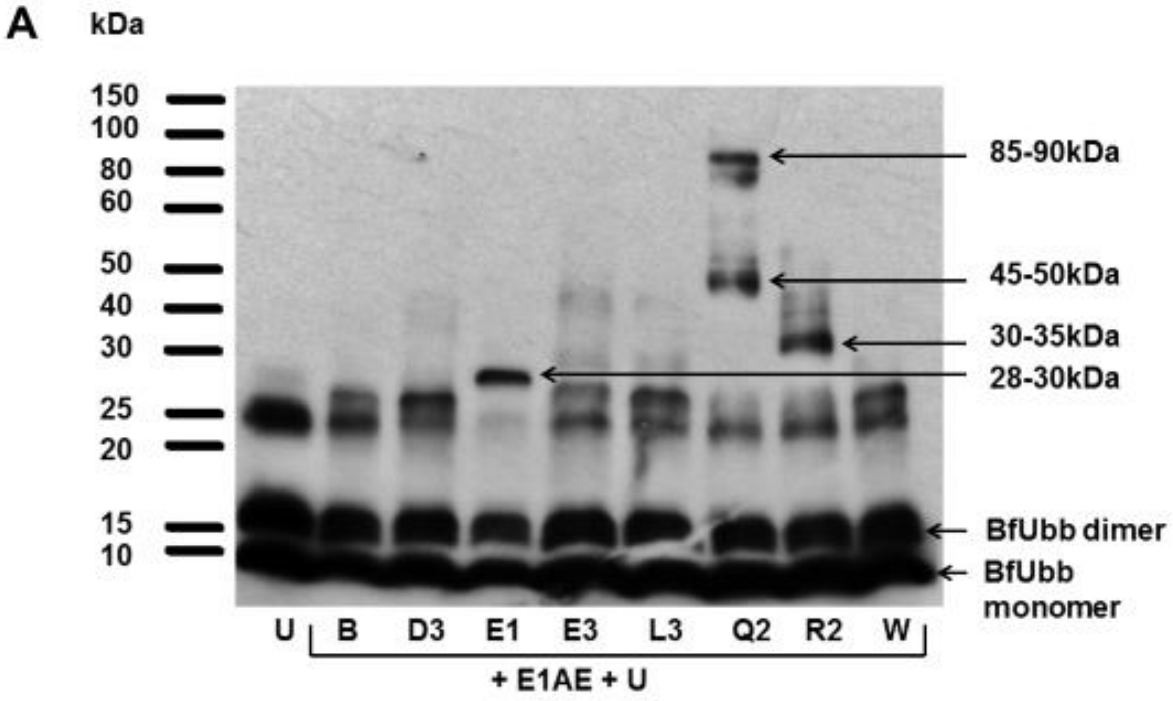
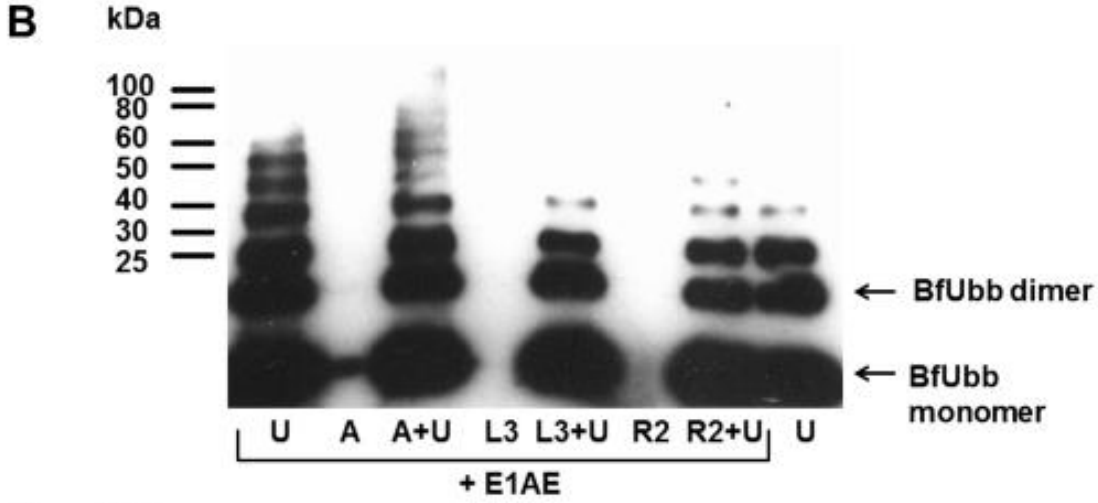


Figure 12 Possible structures of BfUbb dimers (**A**) and hexamers (**B**). A C70-C76 linkage (upper figures) would allow an elongating chain, similar to, but distinct from, those formed via lysine residues. Alternatively, a C76-C76 (or end-to-end) linkage would form a unique complex unseen in eukaryotic ubiquitin. The lack of additional cysteine residues prevents branched chain formation. Note that C70-C70 linkages may also occur and that any or all combinations may be occurring *in vitro*.

3.2.3 BfUbb interacts with E2 conjugating enzymes *in vitro*

To determine whether BfUbb can form a covalent bond with E2 conjugating enzymes, as it does with E1 activating enzyme, rBfUbb was incubated with E1 in the absence of ATP at 37ºC for 1 hour, after which the sample was mixed with 8 randomly chosen human E2 enzymes and incubated for another 2 hours at 37ºC. Final mixtures were analysed SDS-PAGE (without boiling the samples in DTT) and Western blotting (Figure 13 A). Visible in most lanes are bands consistent with the expected size of covalent complexes formed between rBfUbb and the various E2 enzymes (arrows), as well as the 3 bands representing the monomeric, dimeric and trimeric forms of rBfUbb. For example, a ~28-30kDa band in the lane containing Ube2E1 (lane "E1") is approximately the size of the enzyme (21.4kDa) combined with rBfUbb (9.6kDa). Similarly, a covalent complex between Ube2R2 (27.2kDa) and rBfUbb would be ~35.8kDa and a band of roughly this size can be seen in the Ube2R2 sample (lane "R2"). The most striking example is the covalent complexes formed with Ube2Q2 (42.8kDa): a band of ~45-50kDa is consistent with Ube2Q2 combined with rBfUbb (lane "Q2"), however a larger band of ~90kDa can also be seen which might indicate that 2 enzyme molecules are binding the same rBfUbb molecule. The expected molecular weight of that complex would be 94.2kDa. Other reactions, e.g. Ube2E3 and Ube2L3 (Figure 13 A, lanes "E3+U" and "A+U"), also contain multiple bands larger than the rBfUbb trimer, but smaller than a complex containing 2 copies of the enzyme, suggesting either that rBfUbb is binding the E2 enzymes in multimeric form or that E2 enzymes are binding multiple monomers. The most striking example of this is Ube2K (Figure 13 C). Lane "K+U" contains a band of ~50kDa, which is consistent with a complex formed between Ube2K (22.4kDa) and 3 rBfUbb residues (again, whether this is trimeric rBfUbb or 3 monomers is unknown). No bands can be detected in any samples containing only an E2 enzyme (Figure 13 B, C), indicating that at least these 6 enzymes do not cross react with the anti-BfUbb antibody.


A
kDa
150
100
80
60
50
40
30
25
20
15
10
85-90kDa
45-50kDa
30-35kDa
28-30kDa
BfUbb dimer
BfUbb monomer
U B D3 E1 E3 L3 Q2 R2 W
+ E1AE + U


B
kDa
100
80
60
50
40
30
25
BfUbb dimer
BfUbb monomer
U A A+U L3 L3+U R2 R2+U U
+ E1AE

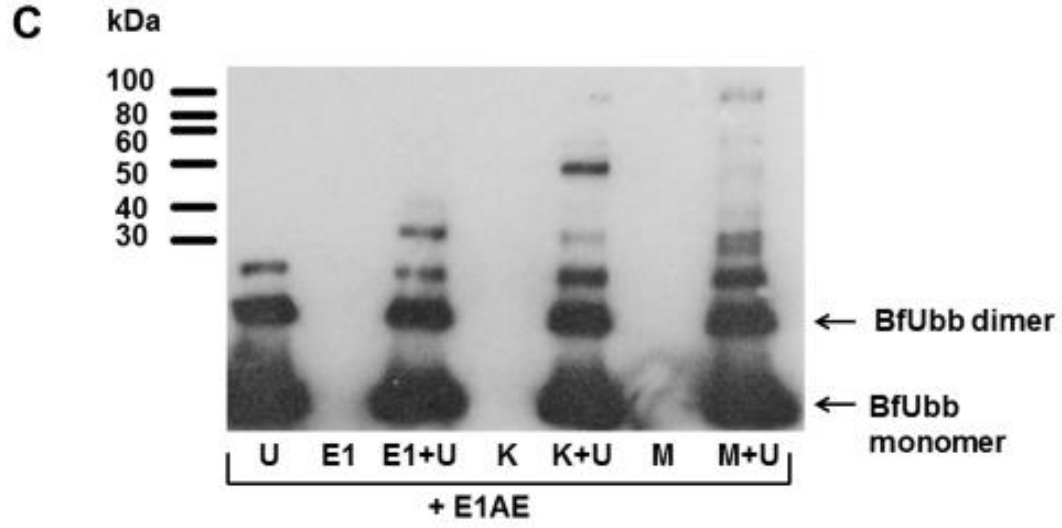

C
kDa
100
80
60
50
40
30
BfUbb dimer
BfUbb monomer
U E1 E1+U K K+U M M+U
+ E1AE

Figure 13 Western blots showing mixtures of BfUbb (U) incubated with E1 activating enzyme (E1AE) and randomly selected E2 enzymes, separated by SDS-PAGE, reacted with anti-rBfUbb polyclonal antisera, visualised with anti-rabbit-HRP antibody. rBfUbb was incubated with E1 activating enzyme at 37ºC for 1h then mixed with each E2 enzyme and incubated at 37ºC for 2h in non-reducing conditions (alternatively, rBfUbb was incubated for 3h on its own, or single E2 enzymes were incubated for 2h in the absence of E1 or rBfUbb). Final concentrations were: 0.25mg/ml (30µM) rBfUbb, 0.1µM E1, 0.25µM E2. 10µl of each sample was mixed with 5µl loading buffer and analysed by SDS-PAGE and Western blotting using rabbit anti-BfUbb primary antibody and anti-rabbit secondary antibody. (**A**) 3 bands representing the monomeric, dimeric and trimeric forms of rBfUbb can be seen in most lanes. Each lane also contains at least 1 other band which corresponds to the expected size of a complex containing an E2 enzyme and at least 1 copy of rBfUbb. Some of these bands are indicated by arrows. (**B**) and (**C**) The 3 rBfUbb bands can be seen in all lanes containing rBfUbb. There are no bands in any lanes containing only an E2 enzyme and the E1 activating enzyme, indicating that these enzymes are not cross-reactive with the anti-BfUbb antibody. In these blots, bands representing E2 enzymes bound to both monomers and multimers (or multiple monomers) of rBfUbb are visible.

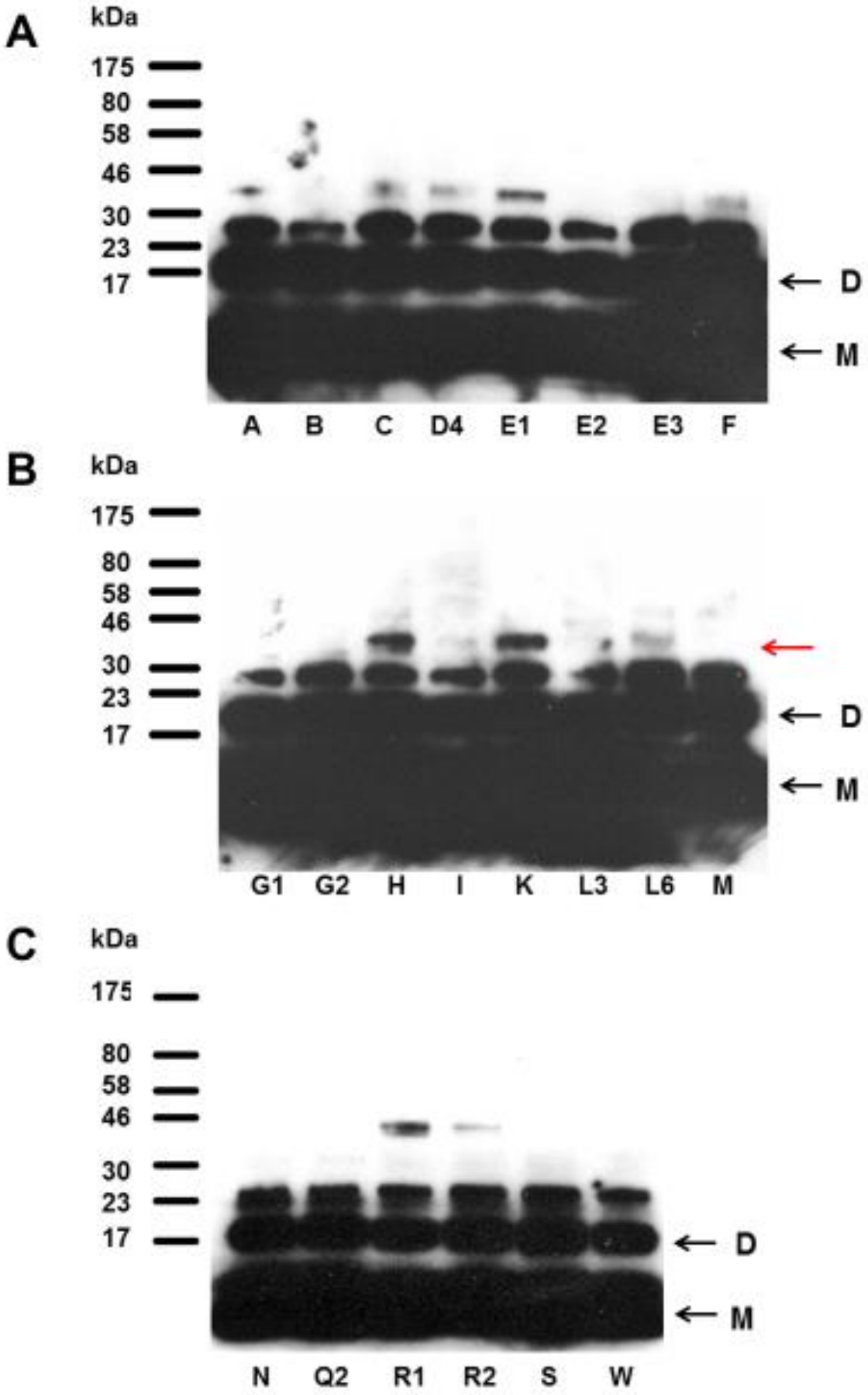
To further investigate the nature of these interactions, rBfUbb was incubated with 22 E2 enzymes (Table 4) in 4 conditions: 1, rBfUbb was incubated directly with each E2 enzyme; 2, rBfUbb was pre-incubated with E1 activating enzyme then incubated with each E2 enzyme; 3, same as condition 1 but in the presence of reducing agent DTT (dithiothreitol); 4, same as condition 2 but in the presence of DTT. Each sample was analysed by SDS-PAGE and Western blotting using the rabbit anti-BfUbb antibody.

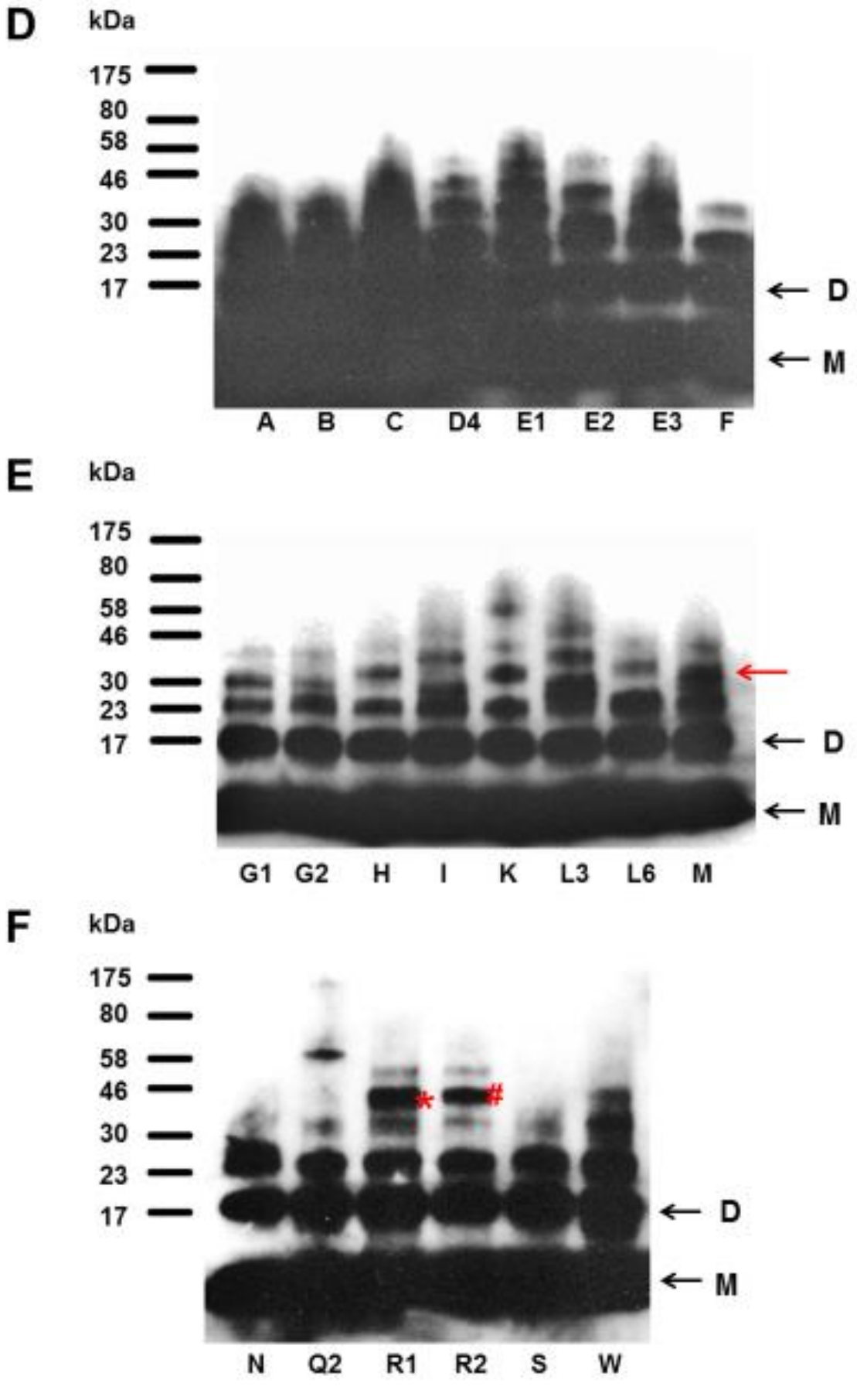
Name	Alternative Name	Size (kDa)	Associated Function/Pathway
UBA1		118	Initiating Ubb transfer
Ube2A	(Rad6A)	17.1	Post-replicative DNA repair
Ube2B		17.3	Post-replicative DNA repair
Ube2C	UbcH10	19.7	Destruction of mitotic cyclins
Ube2D1	Ubc5A	16.6	Cell cycle regulation
Ube2D2	UbcH5B	16.7	Cell cycle regulation
Ube2D3	UbcH5C	16.7	Cell cycle regulation
Ube2D4	UbcH5D	16.6	Unknown
Ube2E1	UbcH6	21.4	Unique N-terminal extension
Ube2E2	UbcH8	22.3	Mediate degradation of misfolded proteins
Ube2E3	UbcH9	22.9	Mediate degradation of misfolded proteins
Ube2F		21.1	Unknown
Ube2G1		19.5	Degradation of muscle-specific proteins
Ube2G2		18.6	Protein degradation
Ube2H	UbcH2	20.7	Histones and cytoskeletal proteins
Ube2I		18.0	Transcription
Ube2K	E2-25K	22.4	Apoptosis suppression
Ube2L3		17.9	p53 and NFkB degradation
Ube2L6		17.8	p53 and NFkB degradation
Ube2M	UbcH12	20.9	Cell cycle regulation, mediates cytotoxicity
Ube2N	Ubc13, Uev1a	17.1	Post replicative DNA repair
Ube2Q2		42.8	Cell cycle regulation
Ube2R1	Ubc3, Cdc34	26.7	Cell cycle regulation
Ube2R2	Ubc3B	27.2	Cell cycle regulation, cell-cell adhesion
Ube2S		23.8	Cell cycle regulation
Ube2T		22.5	DNA damage repair
Ube2W		18.7	N-terminal monoubiquitylation activity

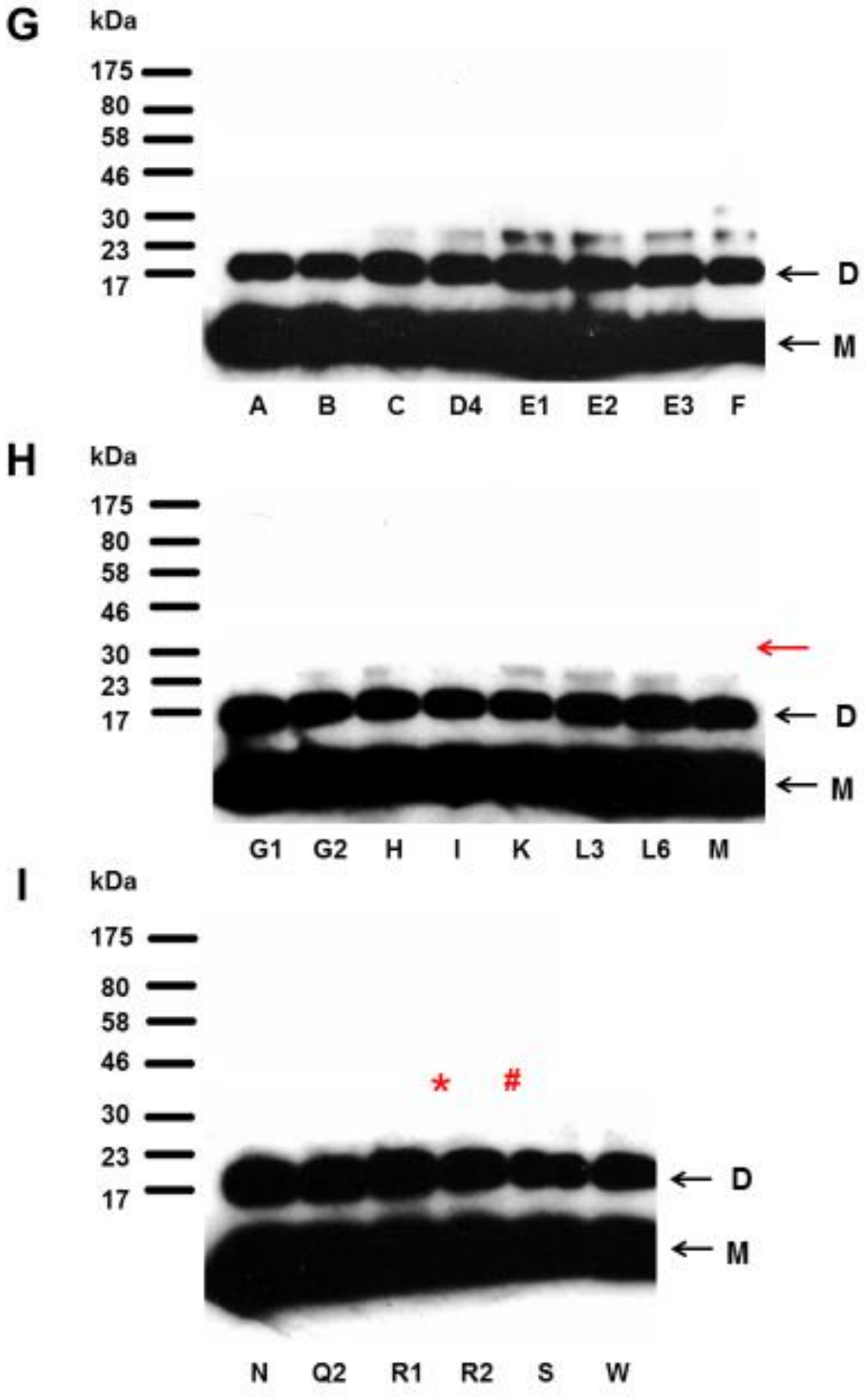
Table 4 All E1 activating and E2conjugating enzymes used in this study, showing all known nomenclature, molecular weight and the functions or pathways with which they are associated (if known).

rBfUbb is able to form covalent complexes with some E2-conjugating enzymes when not pre-incubated with E1-activating enzyme (Figure 14 A, B, C). For example, the Ube2H (20.7kDa) and Ube2K (22.4kDa) reactions both contain a band of approximately 35-40kDa, whilst the Ube2R1 (26.7kDa) and Ube2R2 (27.2kDa) both contain a band of approximately 40-45kDa. This is consistent with each of these enzymes binding 2 copies of rBfUbb, however these blots do not allow for differentiation between E2 enzymes binding dimeric rBfUbb or 2 copies of monomeric rBfUbb binding a single E2 enzyme. When pre-incubated with E1-activating enzyme, rBfUbb forms covalent complexes with most E2 enzymes in both monomeric and multimeric form, as indicated by multiple bands above 28.8kDa (the molecular weight of trimeric rBfUbb) (Figure 14 D, E, F). These interactions are DTT sensitive; there is no indication of interactions between rBfUbb and any of the E2 enzymes without pre-incubation with E1 activating enzyme when DTT is present (Figure 14 G, H, I). There are also significantly fewer detectable interactions between rBfUbb and the E2 enzymes when pre-incubated with E1 activating enzyme in the presence of DTT (Figure 14 J, K, L), however, some bands still appear under these conditions. A ~30kDa band can be seen in the Ube2M sample (red arrow) in conditions 2 and 4, consistent with Ube2M (20.9kDa) combined with 1 molecule of rBfUbb. This band is absent in conditions 1 and 3, suggesting that binding of rBfUbb to Ube2M is dependent on E1 activating enzyme. A ~40-45kDa band can be seen in both the Ube2R1 (red asterisk) and Ube2R2 (red hash) samples, appearing faintly in conditions 1 and 4 and strongly in condition 2 (but not appearing in condition 3). This suggests that the interaction between rBfUbb and these enzymes is not completely dependent on E1 activating enzyme, but that E1 is able to partially rescue the DTT sensitivity of these complexes. Pre-incubation of E1 also enables

the formation of the Ube2M-rBfUbb complex in the presence of DTT. These results indicate that a) rBfUbb most likely forms disulphide bonds with E2 enzymes via 1 of its cysteines, and b) the E1 enzyme is involved in the transfer of rBfUbb to at least some E2 enzymes. This contradicts the previous hypothesis, that BfUbb acts as a suicide substrate, permanently binding the E1 enzyme to inhibit ubiquitylation.


A
kDa
175
80
58
46
30
23
17
D
M
A
B
C
D4
E1
E2
E3
F
B
kDa
175
80
58
46
30
23
17
D
M
G1
G2
H
I
K
L3
L6
M
C
kDa
175
80
58
46
30
23
17
D
M
N
Q2
R1
R2
S
W


D
kDa
175
80
58
46
30
23
17
D
M
A B C D4 E1 E2 E3 F
E
kDa
175
80
58
46
30
23
17
D
M
G1 G2 H I K L3 L6 M
F
kDa
175
80
58
46
30
23
17
*
#
D
M
N Q2 R1 R2 S W


G
kDa
175
80
58
46
30
23
17
D
M
A
B
C
D4
E1
E2
E3
F
H
kDa
175
80
58
46
30
23
17
D
M
G1
G2
H
I
K
L3
L6
M
I
kDa
175
80
58
46
30
23
17
*
#
D
M
N
Q2
R1
R2
S
W

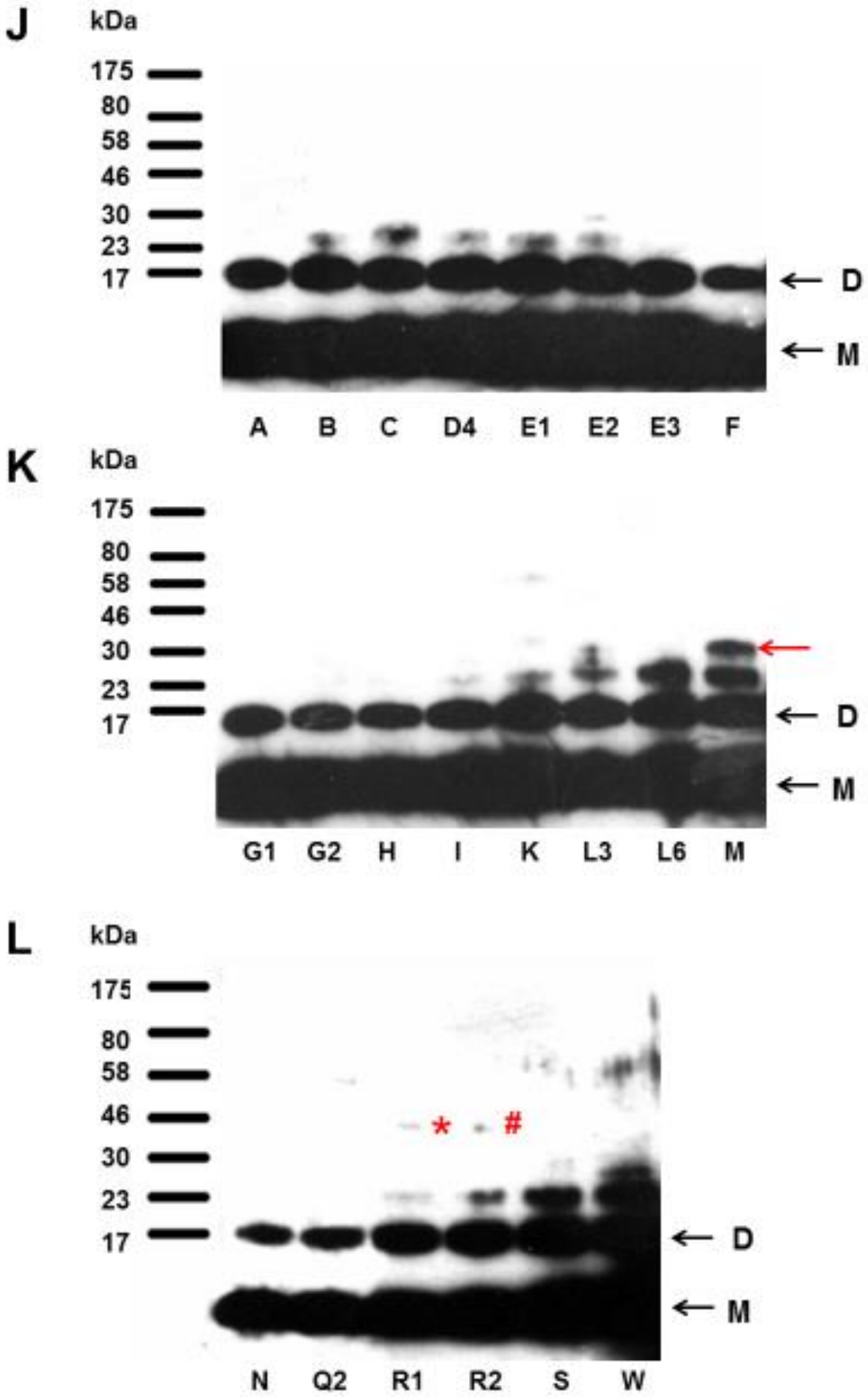
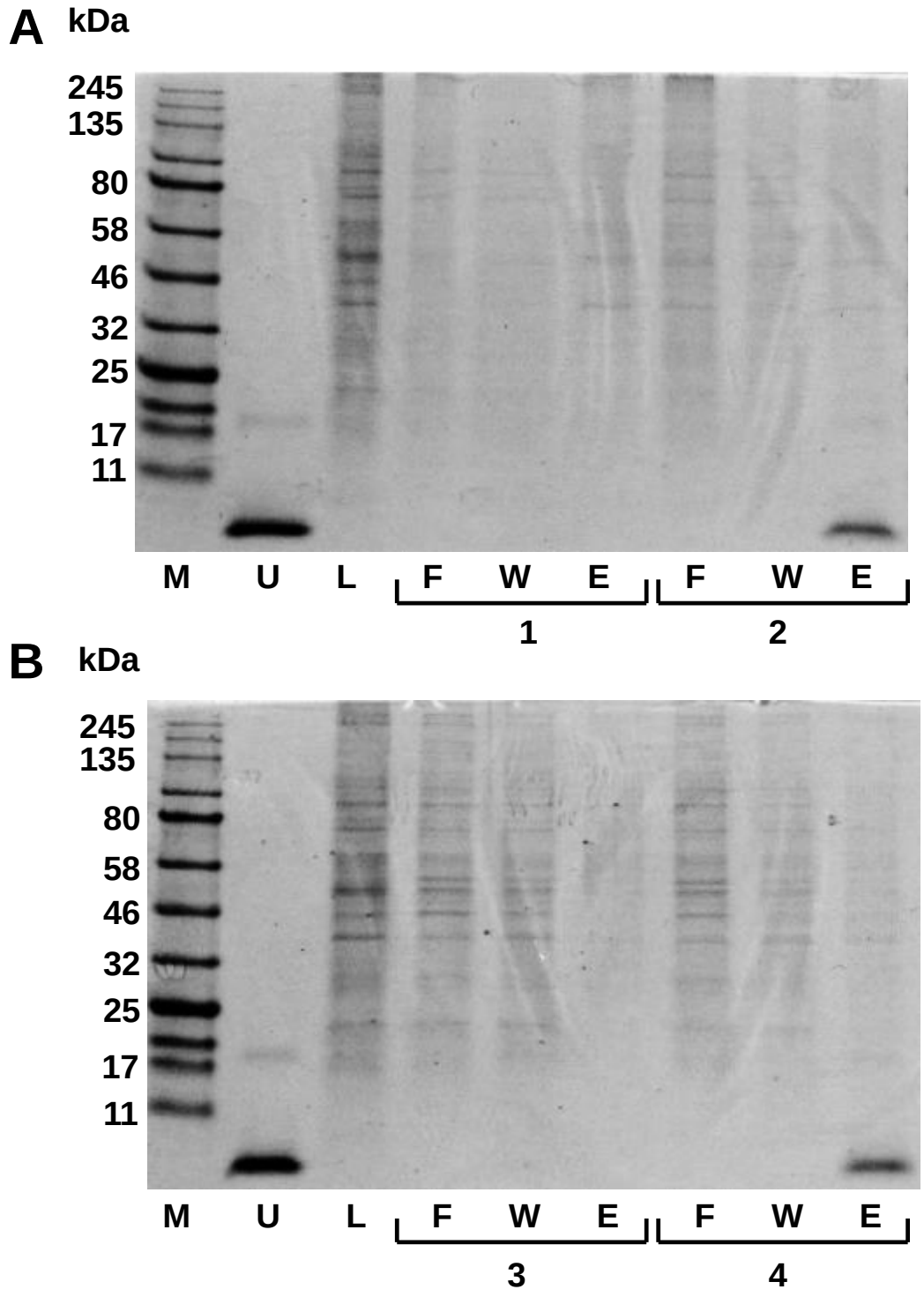

J
kDa
175
80
58
46
30
23
17
D
M
A
B
C
D4
E1
E2
E3
F
K
kDa
175
80
58
46
30
23
17
D
M
G1
G2
H
I
K
L3
L6
M
L
kDa
175
80
58
46
30
23
17
*
#
D
M
N
Q2
R1
R2
S
W

Figure 14 Western blots of mixtures containing rBfUbb and 22 E2 conjugating enzymes in 4 different conditions, reacted with rabbit polyclonal antisera raised against rBfUbb, and visualised with anti-rabbit-HRP antibody. Condition 1 (**A, B, C**): rBfUbb was incubated at 37ºC for 1h then mixed with each E2 enzyme and incubated at 37ºC for 3h in non-reducing conditions. Condition 2 (**D, E, F**): rBfUbb was mixed with E1 activating enzyme at 37ºC for 1h then mixed with each E2 enzyme and incubated at 37ºC for 3h in non-reducing conditions. Condition 3 (**G, H, I**): same mixtures as condition 1 but in the presence of DTT. Condition 4 (**J, K, L**): same mixtures as condition 2 but in the presence of DTT. Final concentrations were: 0.25mg/ml (30µM) BfUbb, 0.1µM E1, 0.25µM E2, 100mM DTT. 10µl of each sample was mixed with 5µl loading buffer and analysed by SDS-PAGE and Western blotting using rabbit anti-BfUbb primary antibody and anti-rabbit secondary antibody. The bonds formed between rBfUbb and E2 enzymes are DTT-sensitive and partially E1-dependent. Some E2 enzymes form covalent complexes with rBfUbb in the absence of E1 (**A, B, C**) but not if DTT is present (**G, H, I**). When pre-incubated with E1-activating enzyme rBfUbb forms complexes with most E2s (**D, E, F**). Again, these interactions are DTT sensitive (**J, K, L**), however some bands can still be detected, specifically in Ube2M (**arrow**), Ube2R1 (**asterisk**) and Ube2R2 (**hash**). There is a large amount of background noise, believed to be caused by cross-reactivity with the polyclonal antisera, on several of these blots, which often obscures the bands representing the monomeric (M) and dimeric (D) forms of BfUbb.

3.2.4 No binding partners of BfUbb could be identified *in vitro*

To identify possible binding partners of BfUbb in the host cell, rBfUbb was incubated with Caco-2 cell extracts, in the presence or absence of DTT. The mixtures were then passed over a Ni column and the eluted fractions were analysed by non-reducing SDS-PAGE and Coomassie staining (Figure 15 A and B). There were no bands exclusive to those samples incubated with rBfUbb that would be indicative of rBfUbb complexed with a binding partner. The samples were also analysed via Western blotting using anti-BfUbb antibody and no bands larger than the BfUbb monomer, dimer and trimer could be detected (Figure 15 C and

D) The presence of DTT did not significantly impact the results. So far, the only known binding partners for BfUbb *in vitro* are the E1 activating and E2 conjugating enzymes.



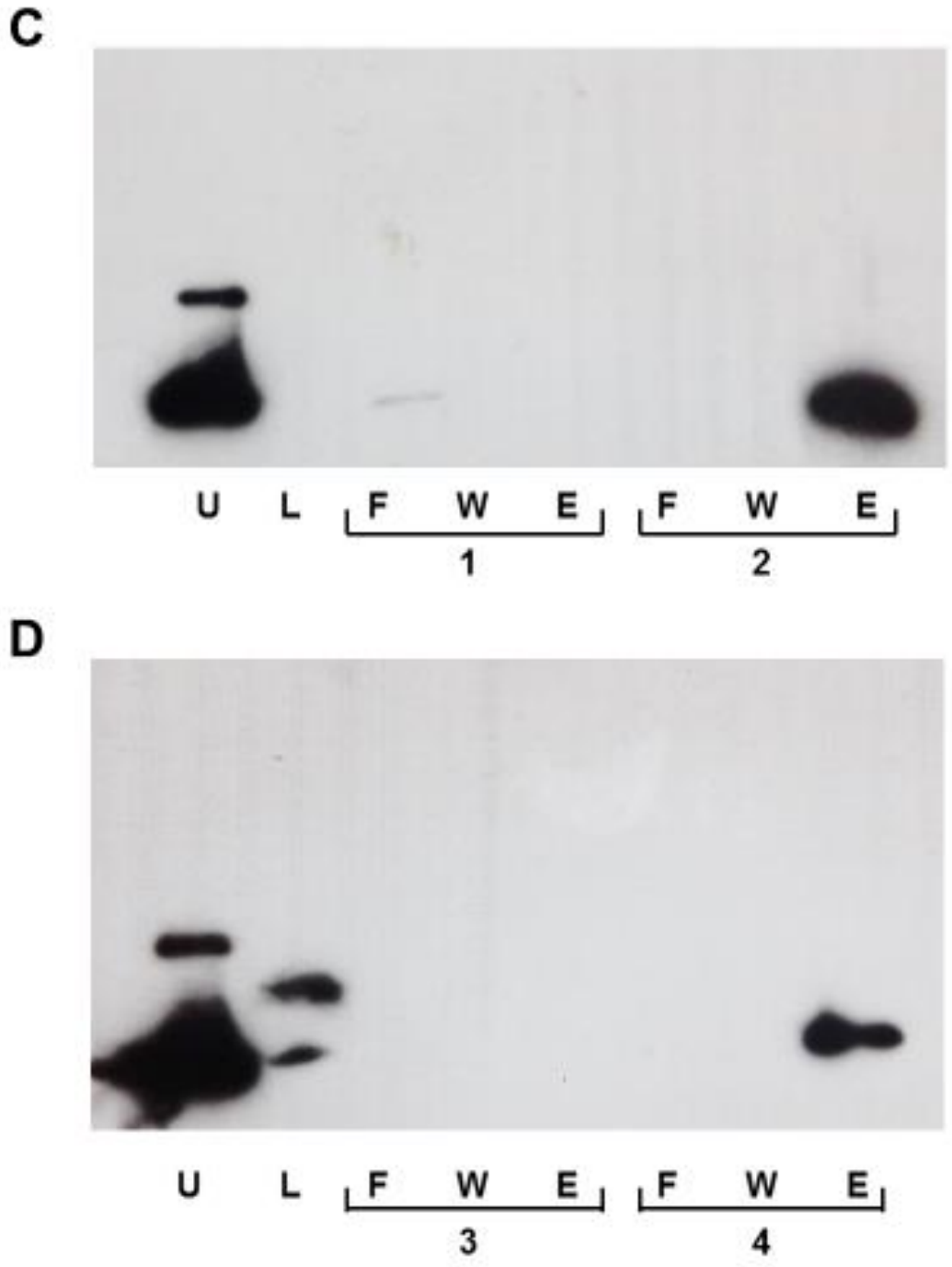


Figure 15 SDS-PAGE gels imaged with Coomassie (**A** and **B**) and Western blotting (**C** and **D**), showing Ni column fractions of Caco-2 cell extract incubated in 4 conditions. Condition 1: 10mg Caco-2 cell lysate was incubated at 37ºC for 3h with gentle shaking. Condition 2: 10mg Caco-2 cell lysate was incubated with 1mg rBfUbb at 37ºC for 3h with gentle shaking. Condition 3: Same as 1 but with 100mM DTT. Condition 4: Same as 2 but with 100mM DTT. The final concentrations were 2mg/ml Caco-2 cell lysate, 0.2mg/ml rBfUbb, 100mM DTT. All samples were mixed 2:1 with loading buffer

(but were not boiled in DTT) and analysed by SDS-PAGE and Coomassie staining or Western blotting. (**A**) rBfUbb (U), Caco-2 cell lysate (L) and the flowthrough (F), wash (W) and elution (E) fractions of the samples from conditions 1 and 2. Aside from the rBfUbb band, there were no bands present in the elution fraction of condition 2 that were not present in that of condition 1. (**B**) U, L and the F, W and E fractions of conditions 3 and 4. The presence of DTT did not significantly impact the samples. (**C**) and (**D**) showed the same results as (**A**) and (**B**), respectively. Western blotting was performed using the anti-BfUbb antibody, the exposure time was 2 minutes. The only band detectable in the eluted fractions for conditions 2 and 4 was BfUbb itself. No larger bands were evident. The apparent banding in the "L" lane in (**D**) was most likely due to accidental exposure of the film.

3.2.5 Generation of a recombinant BfUbb protein with a cleavable affinity tag

Native (untagged) and fluorescently tagged forms of BfUbb would allow further characterisation of the protein. For example, native BfUbb can be used for crystallisation and X-ray crystallography, and fluorescently tagged BfUbb can be used in ligand-binding assays. Expression plasmid pET19b (Novagen), which encodes an N-terminal 10xHis affinity tag followed by an enterokinase cleavage sequence, was chosen and an analytical digest of this plasmid was performed to confirm its identity (Figure 16 A). To create the cleavable affinity tagged recombinant the *ubb* gene was amplified from *B. fragilis* genomic DNA using primers Bfndei and Brbamhi (PCR product designated Pbfubb) (Figure 16 B). To create the fluorescent protein fusion, the *ubb* gene was amplified from *B. fragilis* genomic DNA using primers Bfx and Brbamhi (PCR product designated Pbfubbx) and the yellow fluorescent protein (YFP) gene, *yfp*, was amplified from pDL3196 using primers Yfndei and Yrx (PCR product designated PYx) (Figure 16 B). A crossover PCR was performed using PYx, Pbfubbx and primers Yndei and Bbamhi (PCR product designated PYbfubb) (Figure 16 C). pET19b, Pbfubb and PYbfubb were sequentially digested with BamHI then NdeI. Neither enzyme caused unexpected digestion of the plasmid or PCR products (Figure 16 D and E).

The digested products were then mixed (pET19b with Pbfubb and pET19b with PYbfubb) and ligated to produce pET19b-10xH-BfUbb and pET19b-10xH-YFP-BfUbb. *E. coli* DH5α was transformed with both ligations (DH5α was also transformed with the empty vector); transformed cells were then grown on LB ampicillin plates at 37ºC. 8 individual colonies of both pET19b-10xH-BfUbb and pET19b-10xH-YFP-BfUbb transformed cells were picked, transferred onto fresh LB agar containing ampicillin and incubated at 37ºC for a further 24h. DNA was isolated from each colony by single colony preparation and analysed by agarose gel electrophoresis and staining with ethidium bromide (Figure 16 F and G). The majority of selected colonies contained the ligated products.

Three successful transformants containing each construct were grown in LB ampicillin and the plasmid DNA purified by miniprep (Qiagen). The plasmids were digested with NdeI and BamHI and analysed by agarose gel electrophoresis and ethidium bromide staining (Figure 16 H and I). The change in banding pattern after digestion confirms the presence of inserts the size of Pbfubb and PYbfubb in the respective plasmids. The inserts in pET19b-10xH-BfUbb and pET19b-10xH-YFP-BfUbb were sequenced (DNA Sequencing and Services, MRC, Dundee) and the results (Appendix II) indicate that the DNA sequence of both inserts is correct. *E. coli* BL21 DE3 was then transformed with both recombinant plasmids.

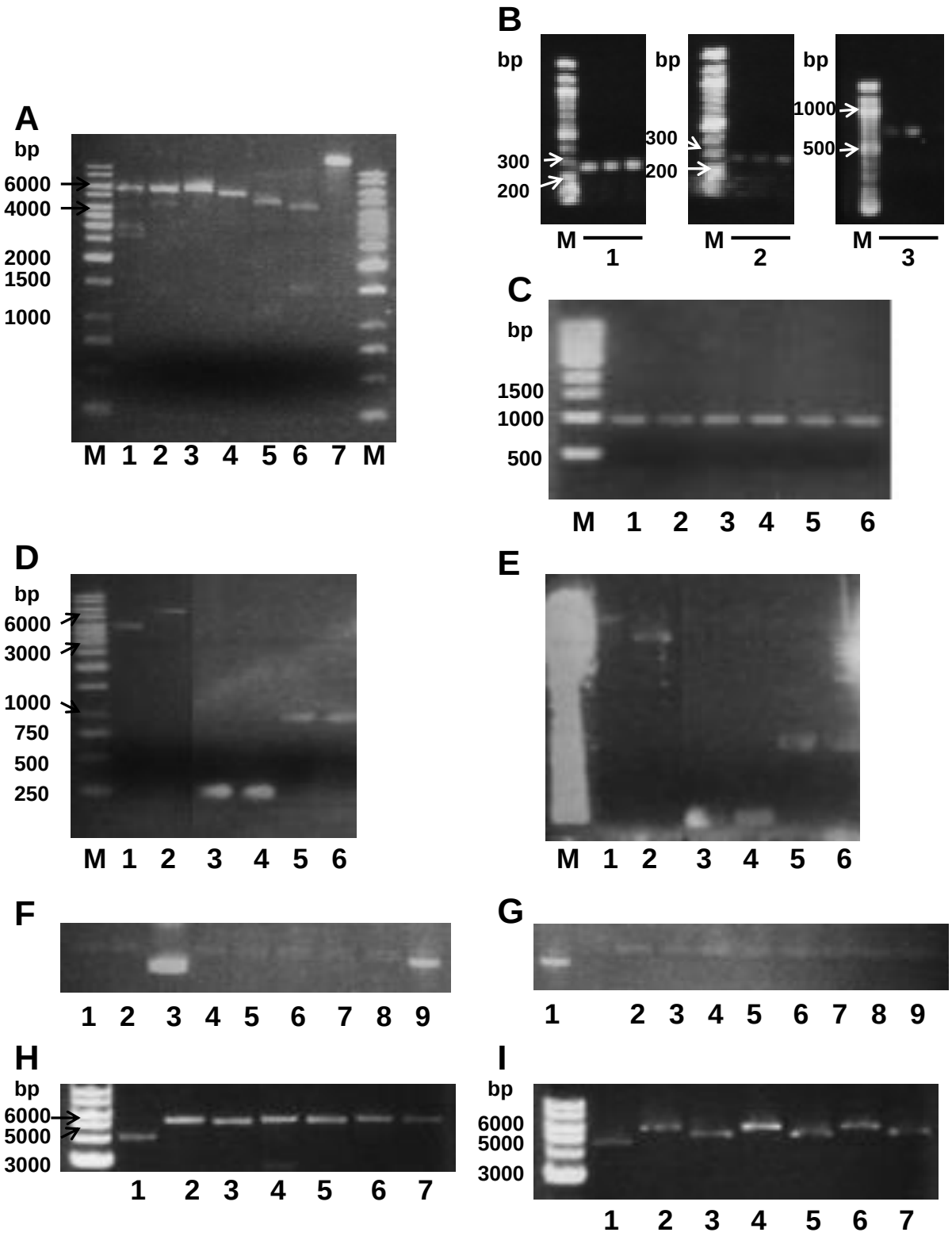

A
bp
6000
4000
2000
1500
1000
M 1 2 3 4 5 6 7 M
B
bp
300
200
M
1
bp
300
200
M
2
bp
1000
500
M
3
C
bp
1500
1000
500
M 1 2 3 4 5 6
D
bp
6000
3000
1000
750
500
250
M 1 2 3 4 5 6
E
M 1 2 3 4 5 6
F
1 2 3 4 5 6 7 8 9
G
1 2 3 4 5 6 7 8 9
H
bp
6000
5000
3000
1 2 3 4 5 6 7
I
bp
6000
5000
3000
1 2 3 4 5 6 7

Figure 16 DNA agarose gels showing the construction of 2 plasmids, 1 encoding a form of BfUbb with a cleavable affinity tag (pET19b-10xH-BfUbb) and 1 encoding an N-terminal YFP-fusion of BfUbb (pET19b-10xH-YFP-BfUbb), visualised with ethidium bromide under UV excitation. (**A**) Analytical digest of pET19b. Lanes: pET19b digested with: 1) BamHI 2) EcoRI 3) ApaI 4) BamHI+EcoRI 5) BamHI+ApaI 6) EcoRI+ApaI 7) Undigested. pET19b is 5717bp - a band of this size was seen in lanes 1-3 (there is some *activity in BamHI due to the buffer used). The double digests should give bands as follows: BamHI+EcoRI - 5396, 321; BamHI+ApaI - 4600, 1117; EcoRI+ApaI - 4279, 1438. Although the smaller bands were difficult to see, the banding patterns in lanes 4-6 matched those expected. (**B**) Initial PCR products: 1) Pbfubb, 255bp 2) Pbfubbx, 258bp 3) PYx, 734bp. All bands matched expected sizes. (**C**) Crossover PCR product. Lanes 1-6) PYbfubb, 992bp. (**D**) BamHI digest of plasmid and PCR products. Lanes: 1) Undigested pET19b, 2) pET19b+BamHI, 3) Undigested Pbfubb, 4) Pbfubb +BamHI, 5) Undigested PYbfubb, 6) PYbfubb +BamHI. There was no obvious degradation of the plasmid or PCR products. (**E**) NdeI digest of BamHI digests from (**D**). Lanes: 1) Double digested pET19b, 2) Undigested pET19b, 3) Double digested Pbfubb, 4) Undigested Pbfubb, 5) Double digested PYbfubb, 6) Undigested PYbfubb. Again, there was no obvious degradation of any bands. (**F**) Plasmids seen in the single colony preparations of pET19b-10xH-BfUbb transformed cells. Lanes: 1-8) Single colony preparation s from 8 different colonies, 9) Undigested pET19b. Lanes 1-2 and 4-8 contain plasmids larger than pET19b, suggesting that they carry the insert. These bands were faint because pET19b is a low copy number plasmid. (**G**) Plasmids seen in the single colony preparations of pET19b-10xH-YFP-BfUbb transformed cells. Lanes: 1) Undigested pET19b, 2-9) Single colony preparations from 8 different colonies. Lanes 2-9 all contain a plasmid larger than pET19b, suggesting that they carry the insert. (**H**) Digests of miniprepped plasmids from cells transformed with pET19b-10xH-BfUbb. Lanes: 1) Undigested pET19b-10xH-BfUbb, 2,4,6) pET19b-10xH-BfUbb +BamHI, 3,5,7) pET19b-10xH-BfUbb +BamHI+NdeI. The excised bands were not visible, however the reduction in size in lanes 3, 5 and 7 is consistent with excision of a ~200-300bp band. (**I**) Digests of miniprepped plasmids from cells transformed with pET19b-10xH-YFP-BfUbb. Lanes: 1) Undigested pET19b-10xH-YFP-BfUbb, 2,4,6) pET19b-10xH-YFP-BfUbb +BamHI, 3,5,7) pET19b-10xH-YFP-BfUbb +BamHI+NdeI. The excised bands were not visible, however the reduction in size in lanes 3, 5 and 7 is consistent with excision of a ~1000bp band.

Expression of 10xH-BfUbb was induced in 500ml *E. coli* BL21 DE3 pET19b-10xH-BfUbb with 400µM IPTG for 4h, after which the cells were harvested and lysed. The lysate was clarified and filtered, then passed over a Ni column, washed and eluted in 8 fractions using imidazole. Presence of 10xH-BfUbb was confirmed by SDS-PAGE and Coomassie staining (Figure 17 A and B). The 10xH-BfUbb-containing fractions were pooled and concentrated to ~3-4mg/ml (total ~2ml, 6-8mg), then 0.5ml (~1.5-2mg) were loaded onto a Sepharose gel filtration column and eluted in 1ml fractions. Fractions were analysed by SDS-PAGE and Coomassie staining (Figure 17 C). 10xH-BfUbb should have a molecular weight of 11.6kDa and a band of this size can be seen in lanes containing the protein, however there is also a smaller band (~10kDa) which was not removed by gel filtration. The origin of this band is unknown. Multiple attempts were made to cleave the His-tag with enterokinase, however none resulted in the isolation of native ubiquitin. Cleavage of the post-gel filtration fraction was unsuccessful (Figure 17 D), whereas cleavage of the pooled elution fraction from the Ni column resulted in complete degradation of the sample, suggesting that there is a protease present in the elution fraction (Figure 17 E). Expression of 10xH-YFP-BfUbb has also been confirmed (Appendix III) however the protein has not yet been purified.

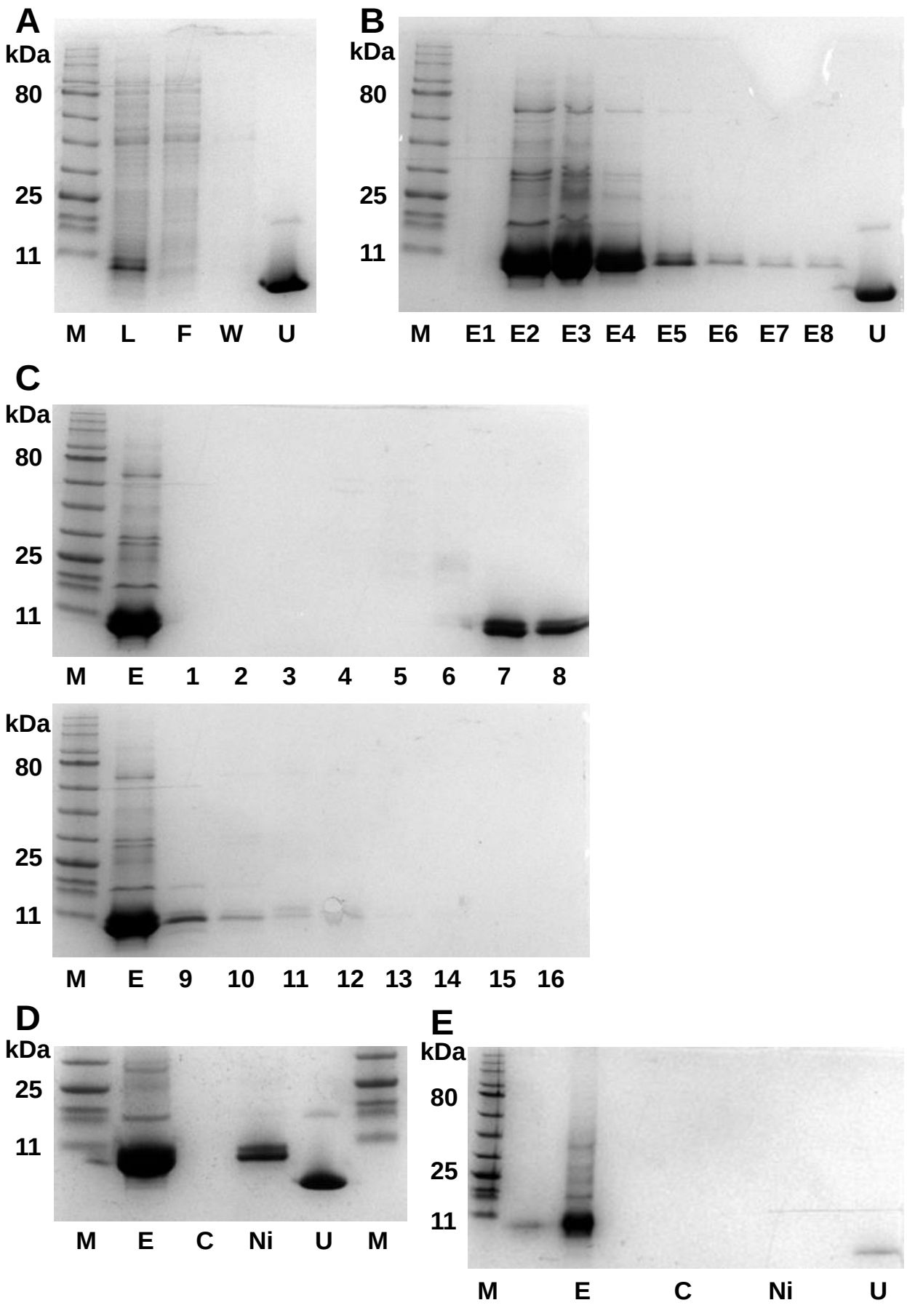

A
kDa
80
25
11
M L F W U
B
kDa
80
25
11
M E1 E2 E3 E4 E5 E6 E7 E8 U
C
kDa
80
25
11
M E 1 2 3 4 5 6 7 8
kDa
80
25
11
M E 9 10 11 12 13 14 15 16
D
kDa
25
11
M E C Ni U M
E
kDa
80
25
11
M E C Ni U

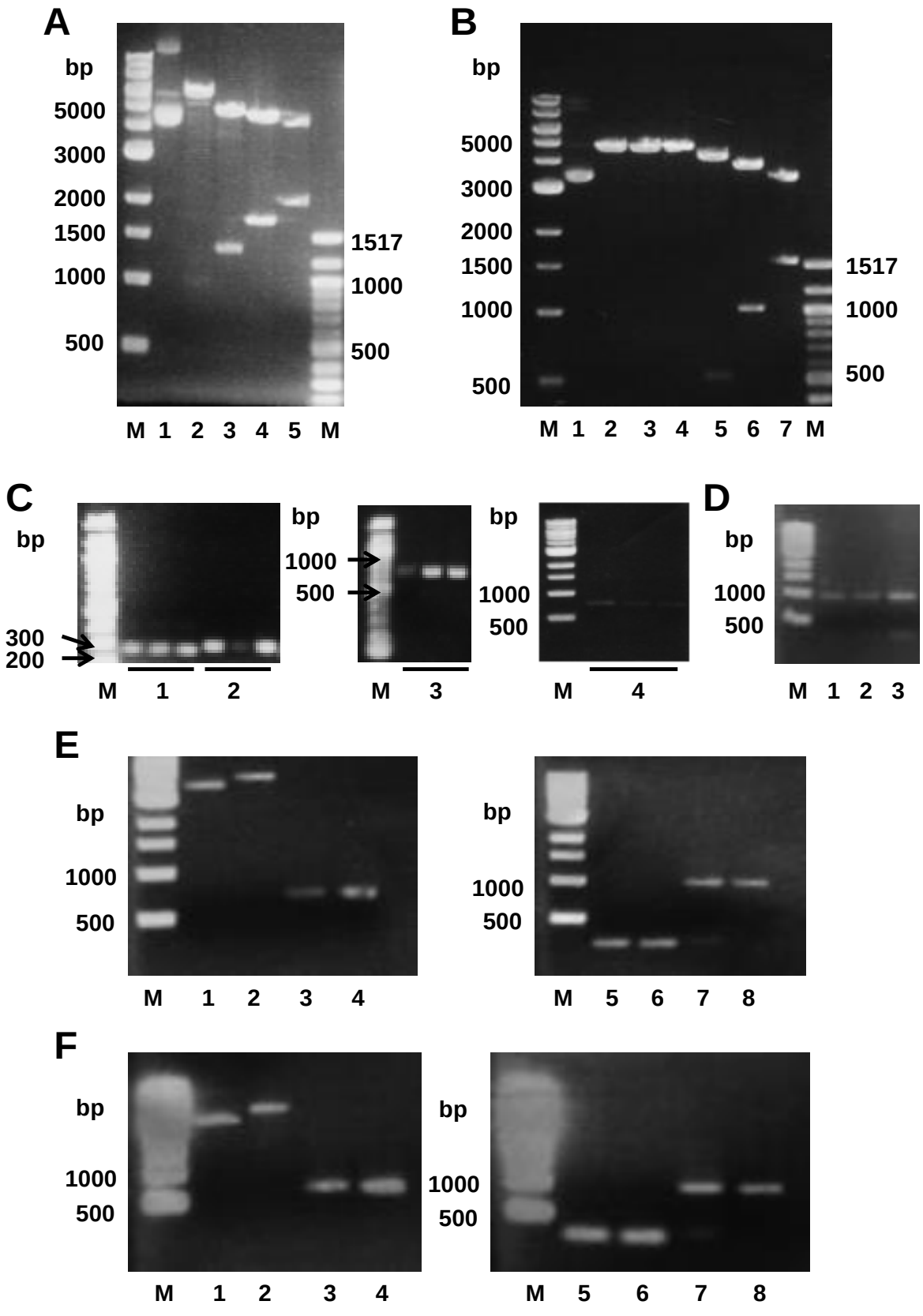
Figure 17 Gels demonstrating the expression of 10xH-BfUbb in bacterial culture samples, the purification of this by Ni-NTA affinity, and the result of incubation of 10xH-BfUbb with enterokinase, visualised by Coomasie staining. (**A**) and (**B**) Cell lysate (L), which shows overexpression of 10xH-BfUbb, prior to loading was loaded onto a Ni column. Subsequent lanes show: flowthrough from the column (F); the wash-off from the column (W); rBfUbb (U); elution fractions (E1-8); rBfUbb. (**C**) Eluted fractions were pooled (E), 0.5ml loaded onto a gel filtration column and eluted in 1ml fractions (1-16). Fractions 7 and 8 were pooled. (**D**) 0.5mg 10xH-BfUbb was incubated with 5µg enterokinase in 10ml enterokinase buffer at 37ºC for 16h. The mixture was then passed over a Ni column, washed and eluted in a single fraction. Lanes: pooled eluted fractions (E); the flowthrough and wash fraction, i.e. cleaved BfUbb (C); the elution fraction, i.e. uncleaved 10xHBfUbb and remaining his-tags (Ni); rBfUbb (U). The enterokinase does not appear to have cleaved any of the protein. (**E**) 0.5mg/ml 10xH-BfUbb was incubated with 10µg enterokinase in 10ml enterokinase buffer containing Ni-Agarose at 37ºC with gentle shaking for 16h. The sample was then centrifuged to pellet the Ni-Agarose. The supernatant was collected and the Ni-Agarose pellet was incubated in imidazole-containing buffer at room temperature to release any bound protein. Lanes: pooled elution fraction (E); supernatant from the cleavage mixture (C); fraction eluted from the Ni-Agarose (Ni); rBfUbb (U). No protein was detected in any of the enterokinase-incubated samples, suggesting that the protein may have been completely degraded.

3.2.6 Generation of BfUbb constructs for expression in mammalian cells

To determine whether BfUbb has an effect on, or localises to a specific area of, a host epithelial cell, it was necessary to create constructs allowing the expression of an affinity tagged form of BfUbb, and a fluorescent protein fusion of BfUbb, in mammalian cells. The T-REx mammalian expression system, consisting of the regulator plasmid pcDNA6/TR and the expression plasmid pcDNA4/TO, was used as it allows high levels of inducible expression (Life Technologies, 2011). Analytical digests of both plasmids (which were not sourced

directly from the manufacturer) were performed to confirm their identity (Figure 18 A and B). To create a His-tagged construct the *ubb* gene was amplified from *B. fragilis* genomic DNA using primers Bfhindiii and Brbamhi (PCR product designated 4bfubb) (Figure 18 C). To create the fluorescent protein fusion, the *ubb* gene was amplified from *B. fragilis* genomic DNA using primers Bfx and Brbamhi (PCR product designated 4bfubbx) and the *yfp* gene was amplified from pDL3196 using Cfhindiii and Crx (PCR product designated 4Yx) (Figure 18 C). Crossover PCR was performed using 4Yx, 4bfubbx and primers Cfhindiii and Brbamhi (PCR product designated 4Ybfubb) (Figure 18 D). To create a YFP control construct the *yfp* gene was amplified from pDL3196 using Cfhindiii and Crbamhi (PCR product designated 4Y) (Figure 18 C). pcDNA4/TO and PCR products 4bfubb, 4Ybfubb and 4Y were digested first with BamHI then with HindIII (Figure 18 E and F). Digested products were mixed (pcDNA4/TO with each of the PCR products) and ligated. *E. coli* JM109 cells were transformed with each ligated product (and the empty vector) and cells were grown on LB ampicillin plates at 37ºC. The transformation efficiency of the ligations was ~2-6 transformants/µl ligation.

Patches of each ligation (8 each of pcDNA4/TO-6xHBfUbb and pcDNA4/TO-YFP-BfUbb, and 6 of pcDNA4/TO-YFP) were made onto fresh LB agar containing ampicillin. DNA was extracted by single colony preparations and analysed by agarose gel electrophoresis and stained with ethidium bromide (Figure 18 G, H and I). Three colonies from each transformation, believed to contain the correct version of the relevant plasmid, were grown in LB ampicillin. The plasmids were purified by miniprep, digested with BamHI and HindIII, and analysed by agarose gel electrophoresis and ethidium bromide staining (Figure 18 J, K and L). Bands the size of 4bfubb, 4Ybfubb and 4Y were excised from pcDNA4/TO-6xHBfUbb, pcDNA4/TO-YFP-BfUbb and pcDNA4/TO-YFP, respectively. Each plasmid was sequenced (Appendix IV) and confirmed to contain the expected inserts. Large (400-600ng), endotoxin-free batches of plasmids pcDNA6/TR, pcDNA4/TO, pcDNA4/TO-6xHBfUbb, pcDNA4/TO-YFP-6xHBfUbb and pcDNA4/TO-YFP were purified by maxiprep (Qiagen).


A
bp
5000
3000
2000
1500
1000
500
1517
1000
500
M 1 2 3 4 5 M
B
bp
5000
3000
2000
1500
1000
500
1517
1000
500
M 1 2 3 4 5 6 7 M
C
bp
300
200
M 1 2
bp
1000
500
M 3
bp
1000
500
M 4
D
bp
1000
500
M 1 2 3
E
bp
1000
500
M 1 2 3 4
bp
1000
500
M 5 6 7 8
F
bp
1000
500
M 1 2 3 4
bp
1000
500
M 5 6 7 8

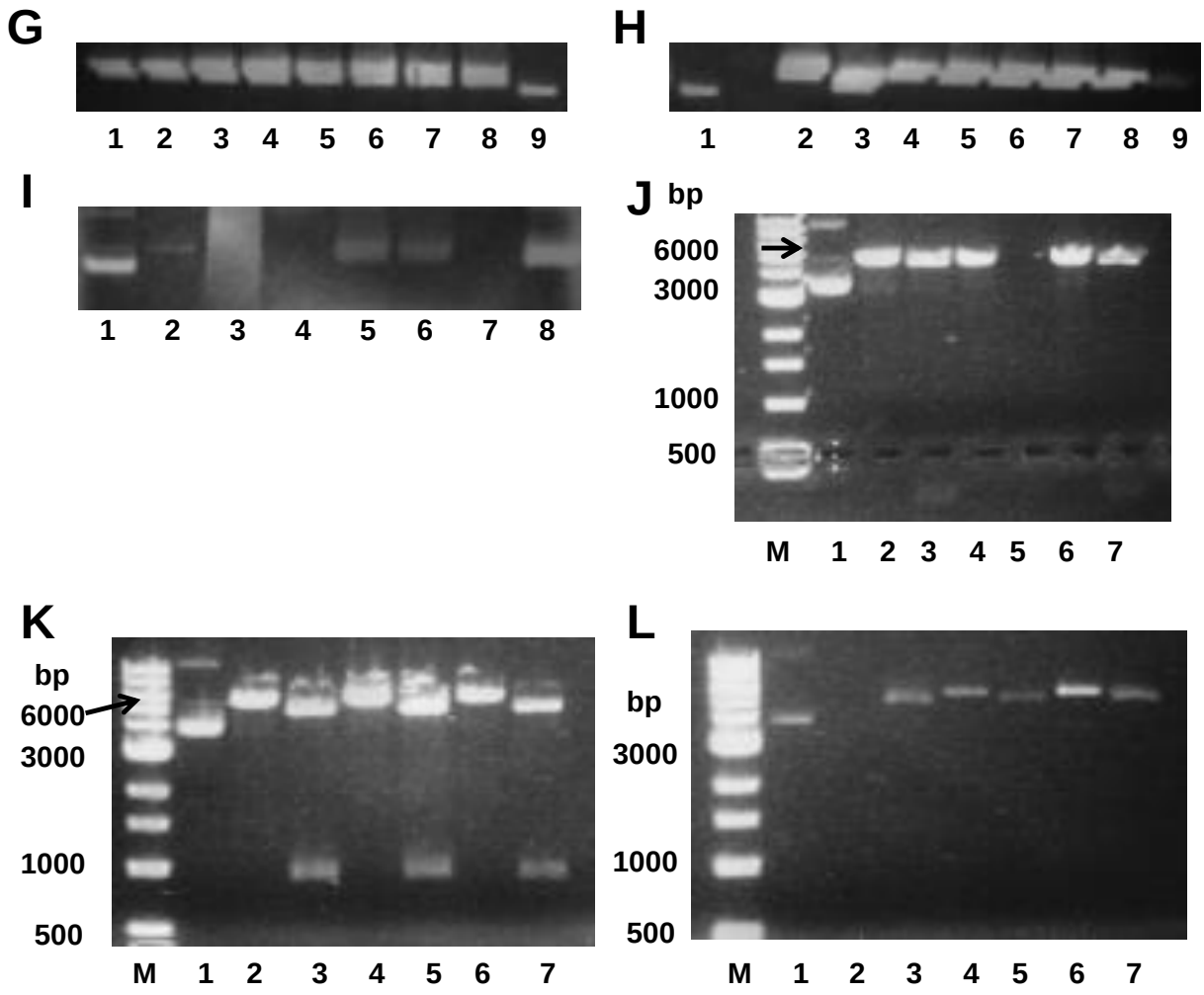


Figure 18 DNA agarose gels showing the intermediate and final products of the construction of 2 inserts for a mammalian 2-plasmid expression system, 1 encoding a 6xH-BfUbb (pcDNA4/TO-6xHBfUbb) and 1 encoding an N-terminal YFP-fusion of BfUbb (pcDNA4/TO-YFP-BfUbb), visualised using ethidium bromide excited by UV light. (**A**) Analytical digest of pcDNA6/TR. Lanes: 1) Undigested pcDNA 6/TR, 2) pcDNA6/TR digested with EcoRI, 3) BamHI, 4) NdeI, 5) PvuI. pcDNA6/TR is 6662bp. The digests should give bands as follows: EcoRI - 6662bp; BamHI - 5324, 1338; NdeI - 4883, 1779; PvuI - 4570, 2092. The banding pattern in lanes 2-5 matches that expected. (**B**) Analytical digest of pcDNA4/TO. Lanes: 1) Undigested pcDNA4/TO, 2) pcDNA4/TO digested with XhoI, 3) EagI, 4) PvuI, 5) XhoI+EagI, 6) XhoI+PvuI, 7) EagI+PvuI. pcDNA4/TO is 5078bp, bands of this size can be seen in lanes 2-4. The double digests should give bands as follows: XhoI+EcoRI - 4543, 535; XhoI+PvuI - 4041, 1037; EcoRI+PvuI - 3506, 1572. The banding pattern in lanes 5-7 matches that expected. (**C**) Initial PCR products: 1) 4bfubbx, 258bp, 2) 4bfubbx, 282bp, 3) 4Yx,

737bp, 4) 4Y, 746. All bands match expected sizes. (**D**) Crossover PCR product. Lanes 1-3) 4Ybfubb, 995bp. (**E**) BamHI digest of pcDNA4/TO and PCR products. Lanes: 1) Undigested pcDNA4/TO, 2) pcDNA4/TO+BamHI, 3) Undigested 4Y, 4) 4Y +BamHI, 5) Undigested 4bfubb, 6) 4bfubb +BamHI, 7) Undigested 4Ybfubb, 8) 4Ybfubb +BamHI. There is no obvious degradation of the plasmid or PCR products. (**E**) NdeI digest of BamHI digests from (**F**). Lanes: 1) Undigested pcDNA4/TO, 2) Double digested pcDNA4/TO, 3) Undigested 4Y, 4) Double digested 4Y, 5) Undigested 4bfubb, 6) Double digested 4bfubb, 7) Undigested 4Ybfubb, 8) Double digested 4Ybfubb. Again, there is no obvious degradation of any bands. (**G**) Plasmids seen in the single colony preparations of pcDNA4/TO - 6xHBfUbb transformed cells. Lanes: 1-8) Single colony preparations from 8 different colonies, 9) Undigested pcDNA4/TO. Lanes 1-8 all contain plasmids larger than pcDNA4/TO, suggesting that they carry the insert. The double banding is a gel artefact. (**H**) Plasmids seen in the single colony preparations of pcDNA4/TO -YFP-BfUbb transformed cells. Lanes: 1) Undigested pcDNA4/TO, 2-9) Single colony preparations from 8 different colonies. Lanes 2 and 4-9 all contain a plasmid larger than pcDNA4/TO, suggesting that they carry the insert. The double banding is a gel artefact. (**I**) Plasmids seen in the single colony preparations of pcDNA4/TO -YFP transformed cells. Lanes: 1) Undigested pcDNA4/TO, 2) Linearised pcDNA4/TO, 3-8) Single colony preparations from 6 different colonies. Lanes 5, 6 and 8 all contain a plasmid larger than pcDNA 4/TO, suggesting that they carry the insert. (**J**) Digests of miniprepped plasmids from cells transformed with pcDNA4/TO-6xHBfUbb. Lanes: 1) Undigested pcDNA4/TO-6xHBfUbb, 2,4,6) pcDNA4/TO-6xHBfUbb +BamHI, 3,5,7) pcDNA4/TO-6xHBfUbb +BamHI+HindIII. Excised bands of ~200-300bp are just visible at the bottom of lanes 3 and 7. The lack of DNA in lane 5 is most likely due to experimental error. (**K**) Digests of miniprepped plasmids from cells transformed with pcDNA4/TO-YFP-BfUbb. Lanes: 1) Undigested pcDNA4/TO-YFP-BfUbb, 2,4,6) pcDNA4/TO-YFP-BfUbb +BamHI, 3,5,7) pcDNA4/TO-YFP-BfUbb +BamHI+HindIII. Excised bands of ~1000bp can be clearly seen in lanes 3, 5 and 7. (**L**) Digests of miniprepped plasmids from cells transformed with pcDNA4/TO-YFP. Lanes: 1) Undigested pcDNA4/TO-YFP, 2,4,6) pcDNA4/TO-YFP +BamHI, 3,5,7) pcDNA4/TO-YFP +BamHI+HindIII. The reduction in size in lanes 3, 5 and 7 is consistent with excision of a ~700-800bp band.

The epithelial cell line Caco-2 (ATCC HTB-37) was transfected using various combinations of pcDNA6/TR and Lipofectamine 2000 (Life Technologies) concentrations, diluted in Opti-Mem reduced serum medium (Life Technologies). 24-48h post-transfection, cells were moved into 5µg/ml blasticidin to select for stable transfectants. 10 foci (R1-10) were selected from a range of successful conditions, grown to confluence and aliquots stored in liquid nitrogen. The pcDNA4/TO-YFP plasmid would be used as a positive control to determine basal and inducible levels of expression from the recombinant plasmids (and therefore the level of expression of TetR from pcDNA6/TR). Attempts were made to transfect cell lines R1, 2, 3 and 4 with pcDNA4/TO-YFP, pcDNA4/TO-6xHBfUbb, pcDNA4/TO-YFP-BfUbb and pcDNA4/TO (empty vector). None of these transfections were successful due to a large amount of cell death (70-80%), within 24h of transfection, prior to movement into 1mg/ml zeocin. This level of toxicity was not observed in cells incubated with just the plasmids or Lipofectamine 2000 for the same amount of time, suggesting that there is no contamination of either the DNA or the Lipofectamine with toxic material, such as endotoxin. It is possible that basal expression of BfUbb from the inserts was enough to kill the cells, however toxicity was also observed in cells transfected with the YFP and empty vector controls. The exact cause of cell death is unknown.

3.2.7 The presence of *the* ubb gene does not appear to correlate with any particular disease

To determine how frequently the *ubb* gene occurs in clinical strains of *B. fragilis*, and whether carriage of the *ubb* gene correlates with any particular diseases, real time quantitative PCR (qPCR) was performed on a range of clinical samples from patients with *B. fragilis*-associated diseases. Genomic DNA was extracted either from bacterial cultures isolated from clinical samples or directly from the samples themselves, depending on the nature of the sample. The presence and relative quantity of *B. fragilis* in the samples was determined using a primer set specific for the *gyrB* gene of human-associated *B. fragilis* strains (Lee & Lee, 2010). Mixtures were prepared in triplicate and qPCR was performed by a lab associated with the Belfast Health and Social Care Trust, then the mean number of copies of each gene per sample was calculated using previously generated calibration curves (not shown). 54 analysed samples were found to encode *gyrB*, indicating that these samples contain *B. fragilis* (Table 5). Of these 54 samples, 13 are from patients with abscesses or bacteraemia, 14 are from patients with colonic diseases, and 13 are of unknown origin. The *ubb* gene was detected in 8 samples. Of these 8 samples, 4 are from a single patient with Diverticular Disease (3496ii and 3496vii from the ascending colon, 3497iii and 3497vi from the rectum), 1 is from an ischiorecal abscess, 1 is from a perianal abscess, 1 is from a patient with irritable bowel syndrome and 1 is from a patient who had undergone a colostomy. From these data there is no obvious association between the presence of the *ubb* gene and any particular disease.

Identity	**Infection**	**Original sample**	***gyrB***	***ubb***
LS11	Abdominal Abscess	Pus/blood	485446	0
LS14	Perianal Abscess	Pus/blood	12936708	0
LS16	Ischiorecal Abscess	Culture	14278	163921
LS18	Bartholins Abscess	Culture	2068	0
LS27	Neoplasm Abscess	Pus/blood	276373	0
LS52	Perianal Abscess	Culture	16652970	0
LS54	Perianal Abscess	Pus/blood	1257490	0
LS66	Abdominal Abscess	Pus/blood	11365376	0
LS67	Perianal Abscess	Culture	44228	33452
LS75	Perianal Abscess	Culture	6442	0
LS80	Perianal Abscess	Culture	4531898	0
LS84	Diverticular Abscess	Pus/blood	16.1	0
LS98	Bacteraemia	Culture	6903352	0
3496 ii	Diverticular Disease	Culture	8063952	1558257
3496 vii	Diverticular Disease	Culture	81816	201778
3497 iii	Diverticular Disease	Culture	2264	1011
3497 vi	Diverticular Disease	Culture	4835040	428242
3498 vi	Irritable Bowel Syndrome	Culture	56.15	0
3498 ix	Irritable Bowel Syndrome	Culture	1257490	0
3498 x	Irritable Bowel Syndrome	Culture	858216	53.6
3499 vi	Irritable Bowel Syndrome	Culture	898008	0
3499 x	Irritable Bowel Syndrome	Culture	56201	0

3528 vii	Quiescent Ulcerative Colitis (pancolitis)	Culture	8222122	0
3529 iii	Quiescent Ulcerative Colitis (pancolitis)	Culture	3683801	0
3529 iv	Quiescent Ulcerative Colitis (pancolitis)	Culture	1171042	0
3532 x	Cancer of the Ascending Colon	Culture	1009013	0
3533 v	Cancer of the Ascending Colon	Culture	1772312	0
3533 viii	Cancer of the Ascending Colon	Culture	455009	0
3534 ii	Quiescent Ulcerative Colitis	Culture	10	0
3535 v	Quiescent Ulcerative Colitis (extensive)	Culture	4893	0
3536 i	Active Ulcerative Colitis (pancolitis)	Culture	3050	0
3536 ii	Active Ulcerative Colitis (pancolitis)	Culture	203	0
3536 vii	Active Ulcerative Colitis (pancolitis)	Culture	3254	0
3536 ix	Active Ulcerative Colitis (pancolitis)	Culture	628	0
3536 x	Active Ulcerative Colitis (pancolitis)	Culture	897	0
3537 ii	Active Ulcerative Colitis (pancolitis)	Culture	4055	0
LS15	Colostomy	Culture	2124596	15
EOC	Unknown	Culture	223203	0
VH	Unknown	Culture	369859	0
RS	Unknown	Culture	13.34	0
AI	Unknown	Culture	13020743	0
CN	Unknown	Culture	7.89	0
BE1	Unknown	Culture	12125616	0
BE3	Unknown	Culture	15710272	0

GNAB85	Unknown	Culture	3707731	0
GNAB92	Unknown	Culture	1459421	0
JC6	Unknown	Culture	7223438	0
JC17	Unknown	Culture	2679.8	0
JC19	Unknown	Culture	176.6	0
NCTC 9344	Unknown	Culture	915622	0
NCTC 10583	Unknown	Culture	21.4	0
NCTC 10584	Unknown	Culture	11892353	0
ATCC 23745	Unknown	Culture	4502649	0
ATCC 29765	Unknown	Culture	17998473	0

Table 5 A table showing the samples in which *gyrB* was detected by qPCR and whether *ubb* was also detected. Columns, from left to right: the identity of the sample; the disease with which it is associated; whether genomic DNA was extracted from the biological sample or from a derived bacterial culture; the number of copies of *gyrB*; the number of copies of *ubb*. Positive are shown in green (>0 copies), and negative results (0 copies) are shown in red. Full details of the samples (excepting patient information) can be found in Appendix V.

3.3 Discussion

3.3.1 Protein purification

rBfUbb had previously been successfully expressed in *E. coli* DH5α pTrc99A-6xHBfUbb, and purified in a previous study (Patrick *et al.*, 2011), therefore no difficulties were expected. The protocol described here gave a consistent yield of pure rBfUbb between 6 and 10mg/L initial culture. The recombinant protein is the expected molecular weight (9.6kDa) and can be successfully isolated by gel filtration. A second band of ~17kDa can be seen in most lanes containing rBfUbb which is most likely the dimeric form of ubiquitin forming post-filtration, since the gel filtration buffer did not contain a reducing agent.

3.3.2 BfUbb multimers

Dimeric rBfUbb was first observed in samples of rBfUbb that had been stored at -20ºC for >6 months (Figure 11 A), and when the purified protein was incubated at 37ºC in the absence of a reducing agent multimers formed within 3 hours, detectable by Western blotting (Figure 11 B). Whilst only dimers, trimers and tetramers were detected when rBfUbb was incubated alone, pentamers and hexamers could be detected when it was incubated with E1 activating enzyme for the same amount of time (Figure 11 C).

Given that a) lysine-linked ubiquitin chain formation is ATP-dependent (Wickliffe *et al.*, 2011), b) rBfUbb lacks the C-terminal double glycine involved in chain formation (Figure 7 A), and c) these interactions are DTT-sensitive (Figure 11 B), it appears that the bonds between rBfUbb molecules are disulphide bridges formed between cysteine residues, rather than lysine-C-terminal bonds. There are only 2 cysteine residues in rBfUbb, C70 and C76, both in the C-terminal tail, meaning that there are a limited number of chain conformations that can be formed (Figure 12 B). These will be distinct from any previously documented

ubiquitin chains, however, as eukaryotic ubiquitin has no cysteine residues and none of its lysines are located in the C-terminal tail.

The mechanism by which E1 activating enzyme promotes multimerisation of rBfUbb is unclear, as this enzyme has no known role in polymerisation of ubiquitin. However, rBfUbb is known to associate with E1 enzyme in the absence of DTT, and it may be that the enzyme acts as a scaffold to bring together individual rBfUbb molecules, thus promoting formation of disulphide bridges.

The ability of BfUbb to form unique chain structures could have implications if this occurs in the host cell. The chains may associate with a UBD to affect cellular pathways, or they may interfere with existing chains, preventing them from being recognised by their relevant UBD. It should be noted, however, that, since the formation of these multimers is inhibited by DTT and other reducing agents, it is unlikely that they will be able to form in the reducing conditions of the host cell cytoplasm. BfUbb would either need to be compartmentalised in a non-reducing environment, such as the lysosome, or assisted by other proteins in order to form multimers. It is also possible that the formation of multimers is an artefact of rBfUbb existing at unnaturally high concentrations *in vitro*, thus promoting the formation of disulphide bridges. BfUbb would likely never reach such concentrations in the host cell, or even the bacterial cell; in fact, no BfUbb multimers have ever been identified in whole cell extract or concentrated supernatants of *B. fragilis*. Therefore, the biological significance, if any, of these multimers has yet to be determined.

3.3.3 BfUbb and E2 conjugating enzymes

The enzymes of the ubiquitin cycle - the E1 activating enzyme, the E2 conjugating enzymes, and the E3 ligases - demonstrate increasing specificity (Komander & Rape, 2012; Wickliffe *et al.*, 2011). rBfUbb has already been shown to form a covalent bond with the E1 enzyme Uba1 and to inhibit ubiquitylation of a lysozyme substrate, possibly by binding the acitve site cysteine of the E1 enzyme (Patrick *et al.*, 2011). However, in order to do this BfUbb would have to reach a high enough concentration in the cell to out-compete eukaryotic ubiquitin. Additionally, cell-wide inhibition of ubiquitylation would result in necrosis (Xu *et al.*, 2010), which would be disadvantageous to the host epithelium and, consequently, the commensal *B. fragilis*. It is therefore more likely that BfUbb acts on a specific pathway or protein, possibly other ubiquitin cycle enzymes. There are hundreds of target-specific E3 ligases, meaning screening them for interactions with rBfUbb via ubiquitin transfer assay would be impractical. Instead, SDS-PAGE and Western blotting was used to determine whether rBfUbb has a preference for 1 or more E2 conjugating enzymes in an effort to identify a potential target pathway. This method determined that rBfUbb was able to interact directly with several E2 enzymes in non-reducing conditions and in the absence of ATP and the E1 enzyme (Figure 14 A, B and C). These interactions are DTT sensitive (Figure 14 G, H and I), and therefore the bonds are most likely disulphide bridges between cysteine residues, meaning rBfUbb may be binding the active site cysteine of an E2 enzyme via rBfUbb C70 or C76. Interestingly, the interactions between rBfUbb and most E2 enzymes become more common if rBfUbb is pre-incubated with the E1 enzyme (Figure 13 A, B and C; Figure 14 D, E and F), suggesting that the interaction between enzymes promotes the forming of a bond between BfUbb and an E2 enzyme. This is unexpected since BfUbb appears to covalently bind the E1 activating enzyme in a manner that should not allow nucleophillic attack by the E2 active site cysteine. The mechanism behind this apparent transferral is therefore a mystery.

Some rBfUbb-E2 enzyme interactions can form in the presence of DTT, if rBfUbb is pre-incubated with the E1 enzyme (Figure 14 J, K and L). As well as confirming that the E1 enzyme promotes rBfUbb-E2 enzyme bonds, this suggests that rBfUbb may have a preference for certain E2 enzymes, most notably Ube2M, Ube2R1 and Ube2R2 (arrow, asterisk and hash, respectively). Ube2R1 and Ube2R2 are also 2 of the enzymes with which rBfUbb is able to interact in the absence of the E1 enzyme. Ube2M is a Nedd8-conjugating enzyme, known to be involved in the S-phase DNA damage repair system which is required for cell cycle progression at this phase. Specifically, Ube2M is required for neddylation of certain cullin-RING ligases (cullins, a class of E3 ligases), mostly those involved in the DNA repair- associated ubiquitylation. Neddylation is essential for cullin activity (Cukras *et al.*, 2014; Huang *et al.*, 2011a). The ability of the ubiquitin E1 activating enzyme to interact with Nedd8 E2 conjugating enzymes is not unprecedented as crosstalk between the 2 pathways has been documented during cellular stress and depletion of ubiquitin (Leidecker *et al.*, 2012). Ube2R1 (Cdc34) interacts exclusively with the cullin class of E3 ligases to promote degradation of cell cycle regulators via K48-linked ubiquitylation, thus allowing progression from G1 to S phase of the cell cycle (Ceccarelli *et al.*, 2011; Harper & King, 2011). Ube2R2, which bears significant homology to Ube2R1, interacts with β-TrCP (Semplici *et al.*, 2002), which associates with the SCF (Skp, cullin, F-box complex) to ubiquitylate cell cycle regulators for degradation, thus allowing progression from G1 to S and from G2 to M (Lau *et al.*, 2012). Common to all 3 of these E2 enzymes is their role in cell cycle progression. Uncontrolled or rapid mitotic progression is associated with cancerous cells, and proteins associated with these 3 E2 enzymes are known oncogenes, e.g. p27, a degradation target in the Ube2R1 pathway, is a cell cycle inhibitor (Ceccarelli *et al.*, 2011; Harper & King, 2011), and β-TrCP is over expressed in several forms of cancer, including colonic cancer (Lau *et al.*, 2012). Inhibition of Ube2M results in extensive DNA damage, which delays progression of the cell cycle at S phase (Cukras *et al.*, 2014), inhibition of Ube2R1 halts progression from G1 to S (Ceccarelli *et al.*, 2011), and inhibition of Ube2R2 activity would most likely inhibit β-TrCP/SCF-dependent cell cycle progression. In addition, Ube2M is the E2 enzyme

responsible for neddylation (and thus activation) of the cullin involved in the SCF (Chen *et al.*, 2000), therefore inhibition of Ube2M would result in an extra level of cell cycle progression inhibition. If BfUbb can interact with and inhibit any or all of these 3 E2 enzymes, or their associated E3 ligases, it may act as an inhibitor of uncontrolled cell cycle progression and consequently protect against tumorigenesis. It should be noted that inhibition of Ube2R2 could also promote tumorigenesis as it is known to be involved in β-catenin degradation (Semplici *et al.*, 2002) and accumulation of β-catenin in the cytoplasm is associated with tumour cells (Lau *et al.*, 2012).

Preference of rBfUbb for these, or any other, E2 enzymes could be confirmed by other biochemical techniques. SPR was performed, using a rBfUbb/ubiquitin coated chip and the E1 and E2 enzymes as analyte, however a high enough concentration of E2 enzyme could not be reached to generate a conclusive result (Appendix I). A repeat of this assay at higher concentrations could determine the relative strength of the interactions between rBfUbb/ubiquitin and the E1/E2 enzymes. Thermal denaturation fluorescence was also used to study the stability of ubiquitin/BfUbb-E2 interactions, however both ubiquitin and BfUbb demonstrate high thermostability and melt curves could not be generated (data not shown).

The fact that covalent bonds between BfUbb and the E2 enzymes are DTT sensitive suggests, as with the formation of multimers, that they are less likely to form in the reducing environment of the host cell cytoplasm without assistance. However, the ability of the E1 enzyme to promote rBfUbb-E2 enzyme complexes indicates that rBfUbb may interact with the host cell ubiquitin pathway in much the same way as eukaryotic ubiquitin. rBfUbb was seen binding E2 enzymes in both monomeric and multimeric form (Figure 13 and Figure 14). E2 enzymes do not usually bind ubiquitin chains and therefore the biological significance of this finding is unknown.

3.3.4 Binding partners of BfUbb

To identify host cell proteins that may act as binding partners for rBfUbb, the purified protein was incubated with Caco-2 whole cell extract and the results analysed by SDS-PAGE and coomassie staining (Figure 15). Even in the absence of DTT, no bands could be detected in the elution fraction of samples containing rBfUbb that were not also found in that of samples without rBfUbb. This was also observed by Western blotting (Figure 15), however a mass spectrometric analysis of both samples could confirm that they contain the same proteins. It is likely that, since ubiquitin, and presumably BfUbb, has a low affinity for all binding partners (Grabbe & Dikic, 2009), a short incubation period at these concentrations may not be sufficient to form detectable interactions. We know that rBfUbb is at least able to form complexes with enzymes of the ubiquitin cycle (Figure 13 and Figure 14), however these are at much lower concentrations in the cell extract and rBfUbb may be out competed by eukaryotic ubiquitin for the binding sites. For BfUbb to bind an E2 enzyme it would therefore need to be highly concentrated in a specific location where the E2 enzyme operates. This requirement means that even if a binding partner were identified by the method described here it would not necessarily mean that the 2 proteins would interact *in vivo*.

3.3.5 10xH-BfUbb and 10xH-YFP-BfUbb: production and uses

A construct was generated for the expression of BfUbb with an N-terminal, cleavable 10xHis tag. Initially, the results suggested the construct had been made successfully: the recombinant plasmid was found to contain an insert of the correct size (Figure 16 H), the sequence of the insert matched that expected (Appendix II), and the expression profile, and the Ni and gel filtration column elution fractions, contained a band of ~11kDa (the expected molecular weight of 10xH-BfUbb is 11.6kDa) (Figure 17 A-D). However, the gel filtration column elution fractions (and earlier samples) also contain a second band, ~10kDa. This is

unlikely to be a premature cleavage product as it was isolated in large amounts from the Ni column and therefore probably contains a his-tag. In addition, it is larger than rBfUbb and therefore too large to be native BfUbb. To resolve this, the recombinant plasmid should be resequenced to confirm that the original sequences of the his-tag, cleavage site and BfUbb are intact. The protein should then be expressed and purified again in the presence of enterokinase inhibitors. If the problem is recurring, mass spectrometric sequencing could reveal the exact nature of the smaller protein. Failing this, an alternative version of the cleavable his-tag could be used.

Once cleaved from the 10xHis-tag, a purified, native form of BfUbb would be useful in further characterisation of the protein. A predicted structure of BfUbb was produced using PHYRE2 (Figure 7 B), however in order to compare the structures of eukaryotic ubiquitin and BfUbb, the 3D structure of the latter will need to be solved, e.g. by X-ray crystallography. Since BfUbb is only 76aa long, a 6xHis tag will make a significant difference to the solubility and structure of the protein, therefore it is better to use an un-tagged version.

A second construct was generated for the expression of an N-terminal YFP-fusion of BfUbb, also with a 10xHis-tag at the extreme N-terminus. Again, the results suggest a successful construct: the insert is the correct size (Figure 16 I), the sequence of the insert matches that expected (Appendix II), and the transformed strain expresses a protein of approximately the correct size (38kDa) (Appendix III). This protein has not yet been purified so it is not known whether there will be similar problems with additional bands or cleavage of the his-tag.

A fluorescent fusion of BfUbb can be used in *in vitro* ligand-binding assays, thus allowing determination of the affinity of BfUbb for various E2 enzymes and other ubiquitin-binding proteins. Competition assays, specifically, could determine whether BfUbb and eukaryotic ubiquitin compete for the same binding sites and if so which has the higher affinity. Although

not a particularly large protein (238aa, 24.6kDa), YFP is >3 times the size of BfUbb and may affect binding or function of the smaller protein. This is something that should be taken into consideration when using this fusion protein.

3.3.6 Expression of 6xHBfUbb and YFP-BfUbb in mammalian cells

Constructs were generated for the expression of 6xHBfUbb, YFP-BfUbb and a YFP control in mammalian cells, using the T-REx 2-plasmid system. pcDNA6/TR expresses the tetracycline-associated repressor TetR, which binds the $TetO_2$ region of pcDNA4/TO, preventing transcription initiation. When present, tetracycline binds TetR, causing a conformational change which prevents the repressor from binding TetO, therefore inducing expression (Yao *et al.*, 1998). There is some basal expression from pcDNA4/TO even when pcDNA6/TR is present; the level of basal expression depends on the level of TetR expression. This can be assessed using a recombinant pcDNA4/TO containing a reporter gene (Life Technologies, 2011), in this case *yfp*. The results suggest that the constructs were generated successfully: the inserts are all of the correct size (Figure 18 J, K and L), and the sequences match those expected (Appendix IV). Caco-2 cells (immortal colonic epithelial cells) were chosen as the model for the target cell. Although there is no clear evidence of *B. fragilis* OMV being endocytosed by gastro-intestinal epithelial cells, OMV are known to be taken up by DCs in tissue culture (Shen *et al.*, 2012), and *B. fragilis* has been seen invading epithelial cells of patients with Crohn's disease (Swidsinski *et al.*, 2005). It is not unreasonable to suggest that OMV of *B. fragilis* deliver their contents to these cells. Caco-2 cells were transfected with pcDNA6/TR and stable blasticidin-resistant cell lines were established. Several of Caco-2 pcDNA6/TR cell lines were then transfected for a second time either with pcDNA4/TO-YFP, pcDNA4/TO-6xHBfUbb, pcDNA4/TO-YFP-BfUbb or pcDNA4/TO (empty vector). So far no attempts at establishing a stable cell line containing both pcDNA6/TR and 1 of the pcDNA4/TO plasmids have been successful since the

transfected cells undergo necrosis in the 24h recovery period prior to the addition of zeocin to the culture medium. This is in spite of the fact that this procedure does not differ significantly from that of pcDNA6/TR transfection. Cells were also incubated with either Lipofectamine 2000 or plasmid DNA in conditions otherwise identical to the transfection conditions, to test for the presence of cytotoxic contaminating factors, such as endotoxin, in the reagent or the DNA. However, no necrosis was observed in these cells prior to the addition of zeocin, suggesting that only actual transfectants experience toxicity. Although there will be some basal expression of the inserts, BfUbb is unlikely to be the causative agent as necrosis also occurs in cells transfected with the empty vector. It may be that pcDNA4/TO, and recombinant plasmids, are recombining with the Caco-2 genome in a way that inhibits proper cell function. Linearising the plasmids prior to transfection may mitigate the problem. Alternatively, success may be achieved with a different cell line.

Expression of 6xHBfUbb in Caco-2 cells is an important experiment for several reasons. Firstly, it will determine whether the presence of BfUbb in a eukaryotic cell would result in cell death, a distinct possibility if BfUbb is able to inhibit all ubiquitylation. Secondly, if the cells survive expression of the protein, they can be observed for phenotypic changes, such as changes in morphology (via microscopy), cytokine expression (via cytokine assay) or gene expression (via RNA microarray). Thirdly, the 6xHis-tag can be used to pull down BfUbb and any binding partners, followed by SDS-PAGE and mass spectrometry for identification. This expression would not be representative of delivery of BfUbb via OMV as the concentration of the protein in the host cell would be higher than could be achievable *in vivo*, however it may help to determine the function and potential effects of the presence of BfUbb in the epithelial cells. Expression of YFP-BfUbb in Caco-2 cells would allow tracking of BfUbb within the cells (via fluorescence microscopy), i.e. where, if anywhere, it localises to. Again, this would not be representative of delivery of BfUbb via OMV as the protein may be chaperoned to a particular area of the host cell by a hitherto unidentified bacterial protein.

3.3.7 Characterisation of BfUbb by qPCR

Multiple clinical samples in which *B. fragilis* could be identified were screened for the presence of the *ubb* gene using qPCR (Table 5). Of the 54 samples in which the *B. fragilis* identifier *gyrB* was detected with confidence, 8 also contained the *ubb* gene. These samples came from patients with Diverticular Disease, an ischiorecal abscess, a perianal abscess, IBS and one who had undergone a colostomy. From this data no correlation can be made between the presence or absence of *ubb* and any particular disease (since all 4 Diverticular Disease samples came from the same patient). A larger scale version of this study has been performed (unpublished data) but still no correlation was observed. qPCR can be a powerful tool in the characterisation of a gene, and its application to a large number of clinical *B. fragilis* samples could confirm or deny the role of BfUbb in colonic diseases or abscess formation. There are some drawbacks to this approach, however, most notably that there are very few clinical samples (especially biopsy samples) from people without colonic diseases, making it difficult to compare healthy and diseased samples, and that whilst it can suggest a correlation it cannot determine causality, or mechanism of action.

3.3.8 Future work and considerations

There is much work still to do in the characterisation of BfUbb. Once 10xH-BfUbb has been successfully purified, the tag must be cleaved and the tag and enterokinase removed from the sample. The crystallisation conditions of eukaryotic ubiquitin should be used as a starting point for the crystallisation of native BfUbb, and the use of high-throughput methods could significantly speed up the process. Solving the structure of BfUbb would allow a direct comparison between it and eukaryotic ubiquitin, perhaps highlighting significant differences or similarities in their ability to interact with other proteins. Isolating the BfUbb multimers and BfUbb-E2 complexes of interest and solving their structure would also be of use in

characterising these interactions and their potential biological significance. Biochemical assays, such as SPR, using BfUbb/ubiquitin-coated chips, and ligand-binding assays, using purified YFP-BfUbb and fluorescently tagged ubiquitin, could determine the relative affinities of BfUbb and eukaryotic ubiquitin for various E2 conjugating enzymes. Such assays may confirm whether BfUbb has particular affinity for Ube2M, Ube2R1 and Ube2R2. Other ubiquitin-binding proteins of interest, such as E3 ligases or suspected targets from a pull-down experiment, could also be screened by these methods.

Binding partners of BfUbb could be identified by expressing 6xHBfUbb in Caco-2 cells (from pcDNA4/TO-6xHBfUbb) and, with or without cross-linking, passing the cell lysate over a Ni column to pull down BfUbb and any bound/cross-linked proteins. These proteins could then be identified by SDS-PAGE and/or mass spectrometry. Expression of BfUbb in Caco-2 cells may cause cell death or halt cell cycle progression, however it may also allow determination of the function of BfUbb, if any, in the host cell through the observation of phenotypic changes, e.g. cell morphology, cytokine or protein expression profiles. Expression of YFP-BfUbb (from pcDNA4/TO-YFP-BfUbb) in Caco-2 cells can be combined with fluorescence microscopy to observe the subcellular localisation of BfUbb, using the YFP expressed from pcDNA4/TO-YFP as a control. Expression of BfUbb recombinants in the Caco-2 cells requires the successful transfection of a Caco-2 pcDNA6/TR cell line with the recombinant plasmids. Transfection may be achievable if the plasmids are first linearised by restriction enzyme digest, however it this does not help a different colonic epithelial cell line, such as FHC, could be used. Finally, the continued use of qPCR as a tool for the screening of clinical samples for the presence of *B. fragilis* and the *ubb* gene could establish a correlation, or lack thereof, between BfUbb and a colonic disease or abscess formation.

4. The proteome of the outer membrane vesicles of *B. fragilis*

4.1 Background

4.1.1 *B. fragilis* OMV

OMV are produced by commensal, clinical and environmental strains of *B. fragilis* (Domingues *et al.*, 1997) and can be observed by electron microscopy budding from the cell surface (Lutton *et al.*, 1991; Patrick *et al.*, 1996) (Figure 19). These vesicles are usually between 50 and 100nm in diameter and are produced mostly by cells expressing the MC (Patrick *et al.*, 1996; Shen *et al.*, 2012). LC cells produce significantly fewer OMV and these appear to become entangled in the complex polysaccharide as they are released from the cell. Antibodies raised against specific *B. fragilis* CPS molecules are cross reactive with both the cells and the OMV of a particular culture, suggesting that the OMV carry the same epitopes as the cells that produce them. The OMV also have the same haemagglutinating and enzymatic activity as *B. fragilis* whole cells and therefore have the potential to adhere to and enymatically attack host cells (Patrick *et al.*, 1996).

The cephalosporinase CepA is associated with the surface of the OMV produced by several *Bacteroides* species, including *B. fragilis* and *B. thetaiotaomicron*. OMV derived from these species can protect other bacteria from the effects of cefotaxime when pre-incubated with the antibiotic. The OMV of *B. thetaiotaomicron* are also able to protect against the effects of β-lactams and it is therefore likely that the OMV of *B. fragilis* can do the same (Stentz *et al.*, 2015). This suggests that these OMV are able to present proteins on their surface that degrade antimicrobials in the extracellular medium before the antibiotics reach the bacterial cells.

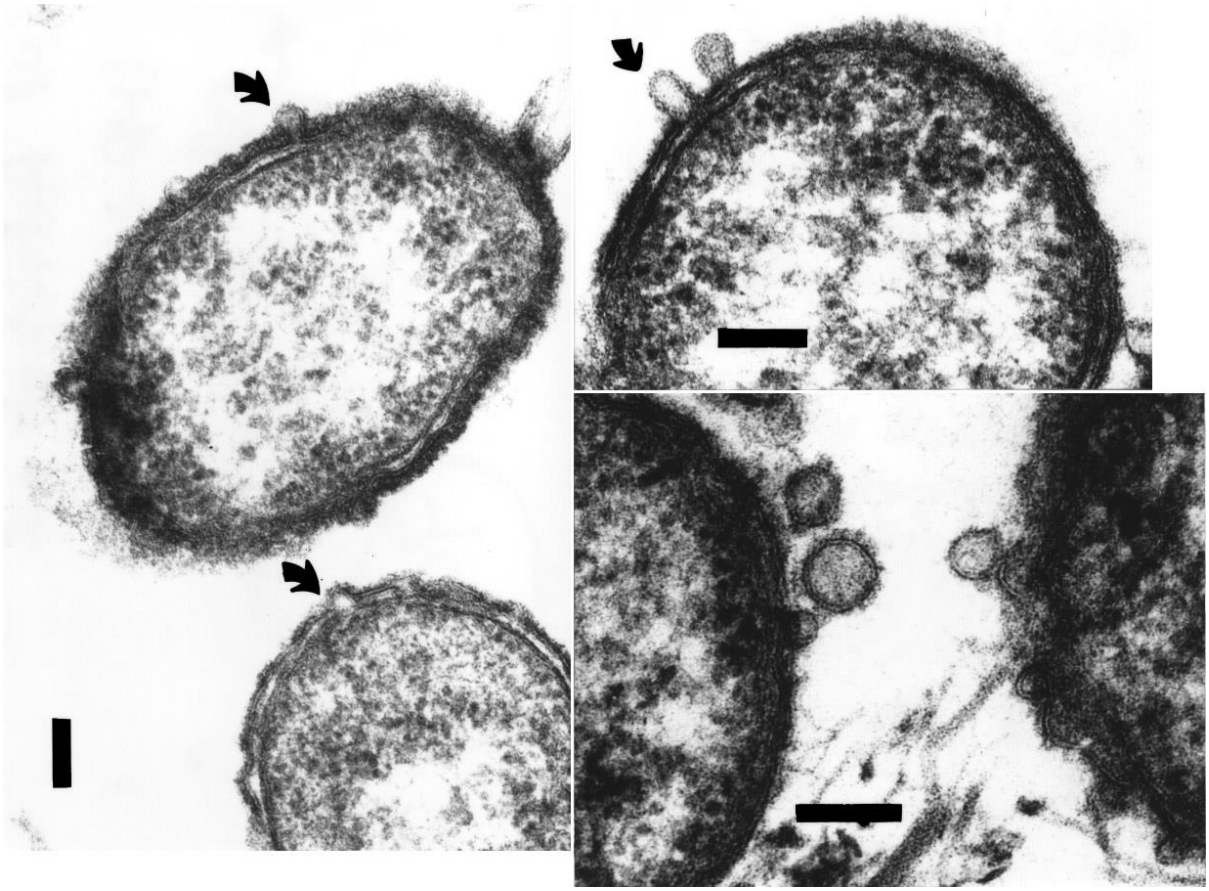


Figure 19 Electron micrographs of thin sections of *B. fragilis* cells, showing OMV budding from the cell surface (**arrows**) (bar:0.1µm) (Patrick *et al.*, 1996).

As mentioned previously (section 1.1.4), *B. fragilis* PSA is necessary and sufficient to prevent experimentally-induced colitis in GF mice (Mazmanian *et al.*, 2008). Further research indicated that PSA is delivered to host cells on the surface of the OMV. When administered with 2-4-6-trinitrobenzene sulfonic acid (TNBS), mice experience artificial inflammation and associated colitis and weight loss. WT OMV were able to prevent weight loss in TNBS mice when administered orally and were able to prevent colitis when administered rectally. Additionally, when applied to a co-culture of DCs and $CD4^+$ T-cells, WT OMV were actively internalised by DCs and induced expression of the anti-inflammatory cytokine IL-10 and the anti-inflammatory protein Foxp3 in T-cells. ΔPSA OMV were also

taken up by DCs but were not able to prevent colitis, weight loss, or induce expression of IL-10 and Foxp3. These results suggest that *B. fragilis* OMV are important in proper immune development and protection from inflammation (Shen *et al.*, 2012).

B. fragilis does not encode any of the major secretion systems (TI - TV), and therefore OMV constitute an important mechanism by which this species interacts with its host and its environment.

4.1.2 Proteomic studies of OMV

Recently, there has been a notable increase in the number of proteomics studies focused on extracellular vesicles, both prokaryotic and eukaryotic (Choi *et al.*, 2015). Proteomics is a powerful and important tool in biological studies as it can provide a large amount of information about an organism or subcellular fraction much faster and cheaper than classical biochemical and genetic techniques. Knowledge of the proteome of the OMV from any bacterial species can help to determine OMV function and mechanism of biogenesis, as well as support their use in biotechnological applications (Kulp & Kuehn, 2010; Lee *et al.*, 2008). A variety of methods have been used in OMV proteomics studies (Figure 20, Figure 21), however the two most important steps are the efficient isolation of pure vesicles and the fractionation of proteins or peptides prior to identification (Choi *et al.*, 2015; Lee *et al.*, 2008).

There are 3 types of OMV that can be isolated from bacterial cell culture. The first are detergent-extracted OMV (DOMV) - the detergent strips endotoxin from the OMV meaning they can be safely used as vaccines, however lipoproteins are also be stripped and DOMV tend to aggregate. Native OMV (NOMV) are extracted using a chelating agent which causes less damage than detergent, and spontaneous OMV (SOMV) are released naturally by the bacterial cells (van de Waterbeemd *et al.*, 2013). Mechanically and chemically induced OMV

are not considered representative of naturally formed OMV (Kulp & Kuehn, 2010). Some early OMV proteomics studies used DOMV and therefore the data produced by these studies may not be reliable (Lee *et al.*, 2008). Most studies now use SOMV concentrated from the supernatant of the cell culture.

The first step in isolating OMV is differential centrifugation to remove whole cells (Figure 20), at around 6,000xg, after which the supernatant is usually filtered (0.45µm or 0.22µm) to remove any remaining cells. The OMV are then isolated from the supernatant via a variety of methods, the most common of which is ultracentrifugation at 80,000-100,000xg (Altindis *et al.*, 2014; Berlanda Scorza *et al.*, 2012; Haurat *et al.*, 2011; Kreimer *et al.*, 2015; Kulp & Kuehn, 2010; Lee *et al.*, 2008) (Figure 20). Others have used density centrifugation on an Opti-Prep gradient (Bauman & Kuehn, 2006; Jang *et al.*, 2014; Kreimer *et al.*, 2015), tangential flow filtration (TFF) (Berlanda Scorza *et al.*, 2012), centrifugal filtration (Bauman & Kuehn, 2006), precipitation with ammonium sulphate (Bauman & Kuehn, 2006; Mullaney *et al.*, 2009), immunoaffinity pull-down (Kreimer *et al.*, 2015) or gel filtration chromatography (Post *et al.*, 2005) to isolate and concentrate OMV (Figure 20). Lee *et al* (2008) assert that ultracentrifugation is not sufficient to separate OMV from other membrane debris and large protein aggregates. Some studies used serial centrifugation at lower speeds (20,000-40,000xg) to pellet larger debris prior to ultracentrifugation at high speed (150,000xg) and other studies used ultracentrifugation in combination with a second technique (Bauman & Kuehn, 2006; Jang *et al.*, 2014; Lee *et al.*, 2007). Gel filtration chromatography relies on size exclusion to isolate vesicles and can be effective, however a prerequisite is that the majority of OMV are approximately the same size (Post *et al.*, 2005). Filtration methods, such as TFF and the use of centrifugal filters risk the loss of sample as OMV can adhere to filters. Low protein binding material, such as PVDF, is recommended when using filters. Precipitation with ammonium sulphate produces a larger number of vesicles than some other methods, but must be combined with another method, such as density centrifugation, to ensure purity (Bauman & Kuehn, 2006). Several authors agree that density centrifugation, ideally with

Opti-Prep, is an important step in the isolation of pure OMV (Bauman & Kuehn, 2006; Kalra *et al.*, 2013; Kreimer *et al.*, 2015; Lee *et al.*, 2008). The proteomic contents of extracellular vesicles as identified by mass spectrometry (MS) depend heavily on the isolation method used (Choi *et al.*, 2015). Currently, different labs use a wide range of different methods for vesicle isolation (Figure 20). There is therefore an urgent need for a standardised procedure for vesicle isolation as proteomes of OMV prepared in different ways cannot reliably be compared.

There are a range of proteomic techniques available in the study of OMV (Figure 21). Fractionation of OMV proteins and peptides can be performed before or after enzymatic digestion (or both before and after). Gel-based techniques use 1D or 2D gel electrophoresis to separate proteins. Individual bands or gel segments are then either excised and digested or digested in-gel prior to ionisation and MS analysis. This results in a low complexity mix of proteins meaning peptides can be more easily identified and also allows a more direct comparison of multiple proteomes. Fractionation of proteins prior to digestion has the added advantage of reducing interference of lipids and high abundance proteins (Kreimer *et al.*, 2015). Additional separation can be achieved by performing liquid chromatography (LC) after digestion (Aguilera *et al.*, 2014; Berlanda Scorza *et al.*, 2012; Choi *et al.*, 2011; Lee *et al.*, 2008). One of the greatest drawbacks of gel-based approaches is that excision and analysis of individual bands is expensive and time consuming which means that often only certain bands are analysed. This leads to an incomplete proteome which is biased towards abundant proteins (those which constitute the strongest bands). Many early studies using this approach identified a total of less than 50 proteins (Choi *et al.*, 2015; Kreimer *et al.*, 2015; Lee *et al.*, 2008). 2D gels are especially unsuitable for OMV studies as vesicles are enriched in hydrophobic membrane proteins, which are difficult to separate on 2D gels (Lee *et al.*, 2008).

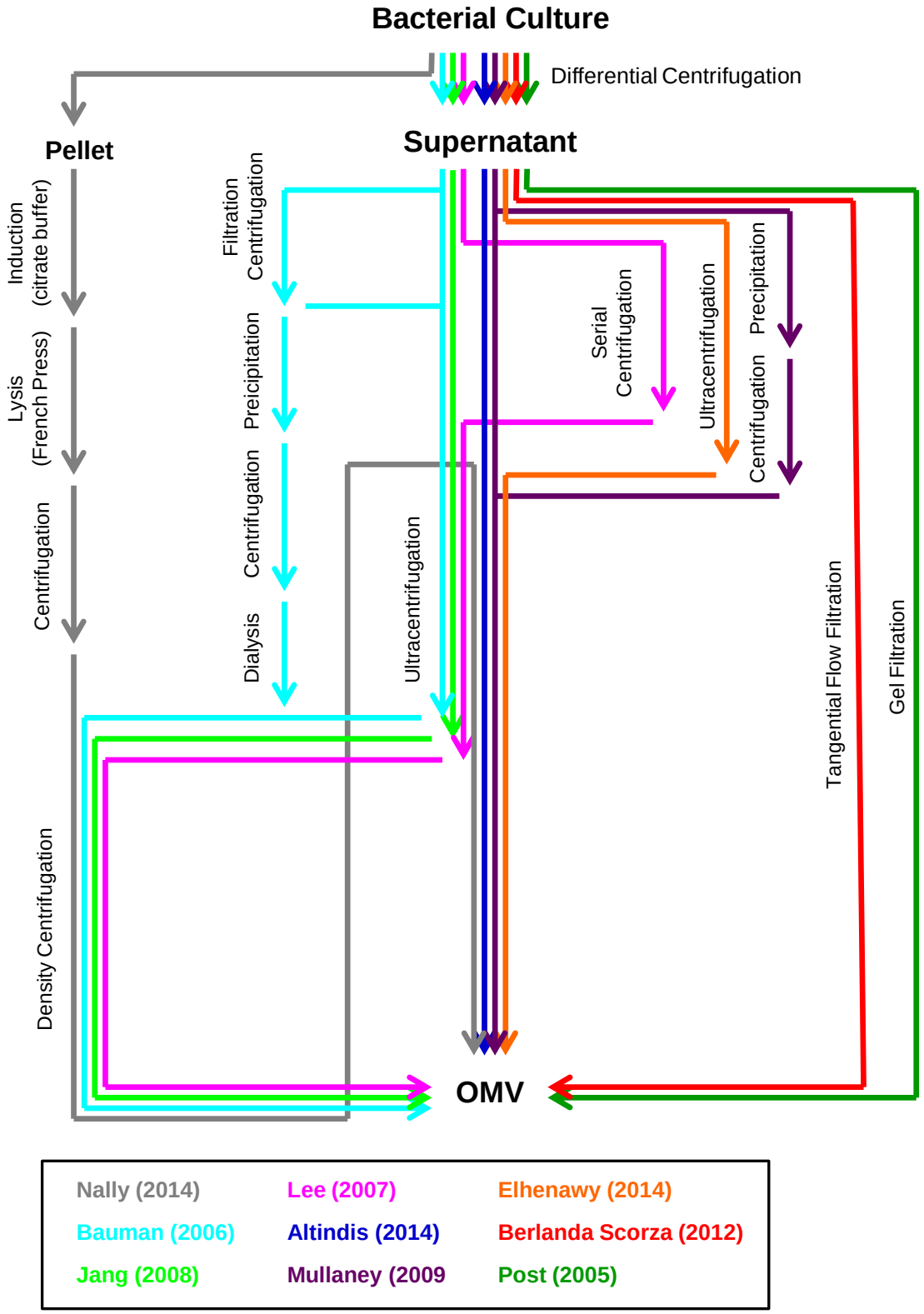

Bacterial Culture
Differential Centrifugation
Pellet
Supernatant
Induction (citrate buffer)
Lysis (French Press)
Centrifugation
Density Centrifugation
Filtration Centrifugation
Preicipitation
Centrifugation
Dialysis
Ultracentrifugation
Serial Centrifugation
Ultracentrifugation
Precipitation
Centrifugation
Tangential Flow Filtration
Gel Filtration
OMV
Nally (2014)
Lee (2007)
Elhenawy (2014)
Bauman (2006)
Altindis (2014)
Berlanda Scorza (2012)
Jang (2008)
Mullaney (2009
Post (2005)

Figure 20 A flow chart showing the different combinations of techniques used to isolate OMV in 9 independent proteomics studies. Each colour of arrow corresponds to a different study as indicated at the bottom of the figure. Bauman (2006) and Mullaney (2014) both used 2 different isolation methods. Aside from differential centrifugation to remove whole cells, the most commonly used techniques are ultracentrifugation and density centrifugation. (Altindis *et al.*, 2014; Bauman & Kuehn, 2006; Berlanda Scorza *et al.*, 2012; Elhenawy *et al.*, 2014; Jang *et al.*, 2014; Lee *et al.*, 2007; Lee *et al.*, 2008; Mullaney *et al.*, 2009; Nally *et al.*, 2005; Post *et al.*, 2005)

van de Waterbeemd *et al* (2013) state that gel-free approaches are more compatible with the hydrophobic membrane proteins of OMV. The most common gel-free approach is "bottom-up" proteomics in which the entire sample is digested, fractionated by LC or isoelectric focusing (IEF) and analysed by high-sensitivity MS (such as LTQ-Orbitrap) (Jang *et al.*, 2014; Kreimer *et al.*, 2015). Using this high-throughput method, thousands of proteins can be identified in a matter of hours, however the signal of small, low abundance proteins can be lost in such a complex mixture. High-sensitivity methods can adjust for this, for example ultra-low flow (ULF) fractionation reduces the dilution effects of LC and produces smaller droplets which improves ionisation and increases the likelihood of detecting low abundance proteins (Kreimer *et al.*, 2015). Tagging methods, such as phospho-tag (PTAG) can also reduce the effects of large or abundant proteins. PTAG modifies the internal peptides of proteins which are then removed by TiO_2 affinity post-cleavage. This leaves only the N-terminal peptide of each protein, meaning all proteins are equally represented (van de Waterbeemd *et al.*, 2013). Multiple samples can be more easily compared with quantitative proteomics. In iTRAQ (isobaric tag for absolute quantitation) proteomics, the different samples are labelled with isobaric tags which have the same mass but a different mass-to-charge ratio. The samples are then pooled and analysed together and the relative quantities of each peptide in the different samples is determined by measuring the relative intensities of each tag. Relative quantities of proteins can also be measured using spectral counts. In this

case, 2 samples are analysed separately and the derived spectral count for each protein is compared between the 2 samples. This method is cheaper and easier that those described previously as it is label-free, however it is considered to be the least accurate method of protein quantitation (Kreimer *et al.*, 2015). Spectral counting is less likely to detect small fold changes in protein abundance (Li *et al.*, 2012) and there is a greater risk of proteins going undetected due to the random-sampling effects of high-throughput MS (Cooper *et al.*, 2010).

Once the raw peptide sequences have been determined they are used to search databases of the genome/proteome of the organism of interest. Identified proteins are considered "true" hits if they are identified using multiple peptides and have a p or e value below a pre-determined threshold (usually 0.05). When the whole sample is analysed, around 60-340 proteins are typically identified in OMV (although some studies have identified more) (Aguilera *et al.*, 2014; Altindis *et al.*, 2014; Berlanda Scorza *et al.*, 2012; Choi *et al.*, 2011; Elhenawy *et al.*, 2014; Jang *et al.*, 2014; Lee *et al.*, 2007; Mullaney *et al.*, 2009; van de Waterbeemd *et al.*, 2013).

Characterisation of OMV extends beyond MS proteomics as the identification of proteins alone does not provide a complete understanding of their role in the OMV. Post-translational modifications (PTMs) play an important role in protein function and protein-protein interactions. Identification of PTMs requires a "top-down" approach, starting with the intact protein. The characterisation of other components of OMV, e.g. identification of lipids by lipidomics, can also provide information about the origin and function of OMV (Kreimer *et al.*, 2015).

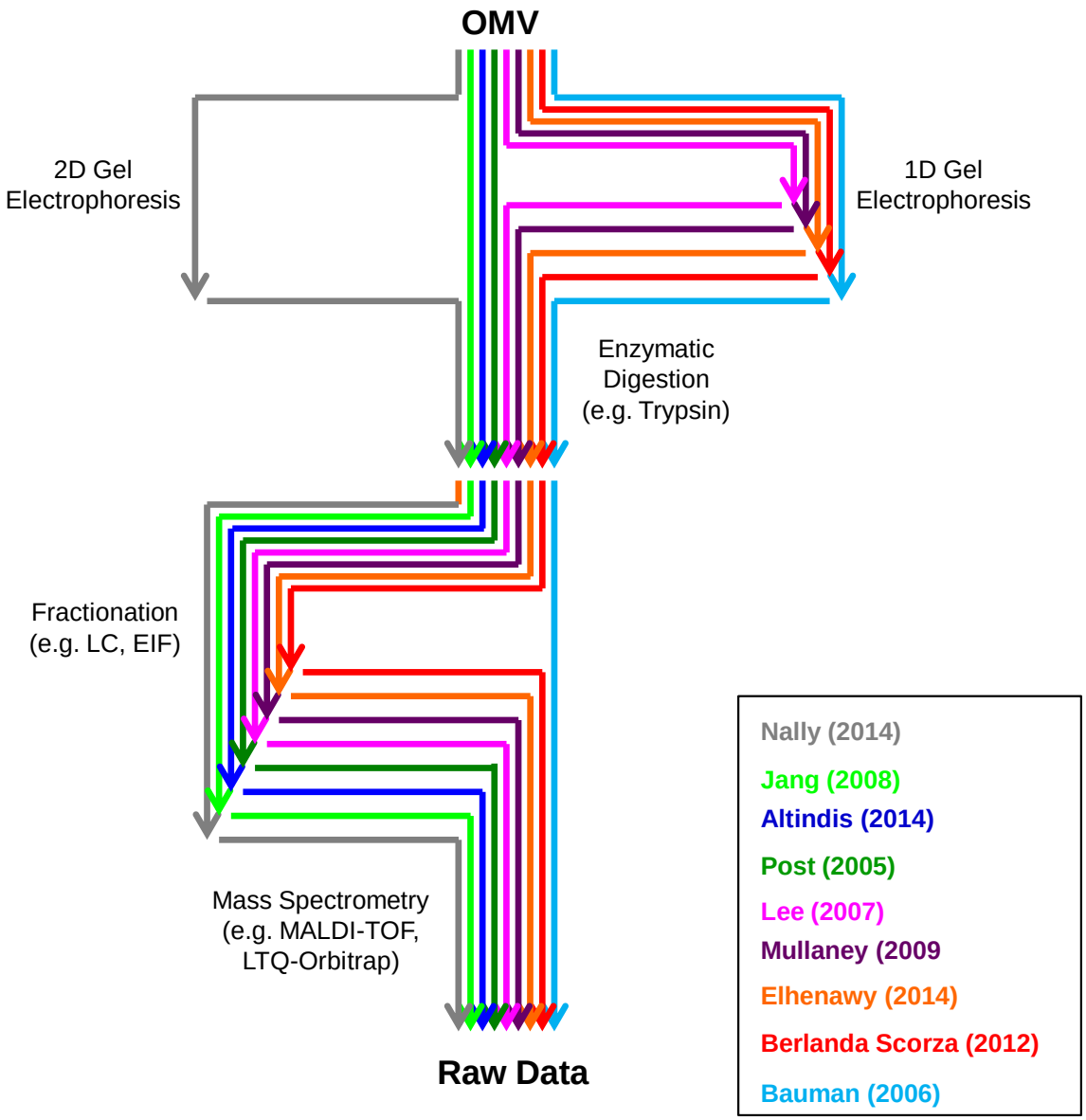


Figure 21 A flow chart showing the different combinations of proteomic techniques used in 9 independent OMV proteomics studies. Each coloured arrow corresponds to a different study as indicated at the bottom of the figure. (Altindis *et al.*, 2014; Bauman & Kuehn, 2006; Berlanda Scorza *et al.*, 2012; Elhenawy *et al.*, 2014; Jang *et al.*, 2014; Lee *et al.*, 2007; Mullaney *et al.*, 2009; Nally *et al.*, 2005; Post *et al.*, 2005)

Only one comprehensive proteomic study of *B. fragilis* OMV has been published so far (Elhenawy *et al.*, 2014). OMV were isolated by differential centrifugation to remove cells followed by 2 rounds of ultracentrifugation at 100,000xg for 2h. The sample was then fractionated by 1D SDS-PAGE and individual bands were excised, trypsin digested and analysed by LC-MS (quadrupole orthogonal acceleration time-of-flight, Q-TOF). A sample of isolated OM was analysed in parallel with the OMV. Proteins were identified using the Mascot search engine and the NCBI non-redundant database. This study identified 115 OMV proteins and 102 OM proteins; 69 proteins were shared between the two samples. OMV-exclusive proteins were found to be more acidic than OM proteins, possibly indicating that acidic proteins are preferentially packaged into the OMV. Several potential proteases were identified by MS. OMV proteins were separated by SDS-PAGE and bands corresponding to these proteases were assayed for proteolytic activity. This confirmed the presence of at least 11 proteases in the OMV. Eight sugar hydrolases were found to be exclusive to the OMV and whole OMV possess fucosidase activity. The OMV were also found to be enriched in SusD-like starch-binding proteins. The enrichment of certain proteins in the OMV over the OM supports the active biogenesis theory of OMV in *B. fragilis*. The enrichment of sugar hydrolases in particular suggests that OMV may play a role in sugar degradation in the gut (Elhenawy *et al.*, 2014).

4.1.3 Aims

The aim of this study was to characterise the OMV of *B. fragilis*. DLS was used to determine the size of OMV in concentrated supernatants of *B. fragilis* culture. In addition, mass spectrometry was used to identify the proteins present in the OMV and the periplasm of *B. fragilis* NCTC 9343, and the two samples were compared. OMV were isolated using a combination of the techniques described above. Whole cells were removed by differential centrifugation and the supernatant was concentrated, first by TFF and then by centrifugal filtration. Concentrated supernatants were then applied to an Opti-Prep density gradient and centrifuged. OMV and periplasmic proteomes were determined using a gel-free "bottom-up" approach; whole samples were trypsin digested and cleaned then fractionated by HPLC and analysed by LTQ-Orbitrap MS. The raw peptide data was used to search the NCBI non-redundant database of all available *B. fragilis* strains via Progenesis LC-MS software. The results were further analysed to determine "true" hits, the most abundant proteins in the OMV, and the proteins most enriched in the two samples. Proteins of interest, including previously identified proteins and novel proteins, were identified using BLASTP.

4.2 Results

4.2.1 Dynamic light scattering of *B. fragilis* OMV

DLS was used to determine the average size of WT OMV. Concentrated supernatant of WT *B. fragilis* (grown in BHI-S with 20µg/ml gentamicin) was diluted to 0.5mg/ml in PBS. 50µl of this dilution was analysed by DLS (Zetasizer APS, Malvern Instruments, UK). Each sample was measured 3 times and the average size calculated from the first 2, since the third reading was of low quality (not shown). A single peak in the WT concentrated supernatant sample represents a monodisperse solution of a globular species approximately 250nm in diameter (Figure 22 A). There was little consistency between readings (Appendix VI), possibly due to the effects of repeated asipration of the sample on the OMV. A control sample of BHI-S shows multiple peaks of low intensity, representing a polydisperse solution (Figure 22 B), most likely of the large molecules and possible precipitates formed in the medium. None of these peaks appear in the concentrated supernatant sample, suggesting that they have been successfully removed during the concentration and washing procedure. Additionally, the 250nm peak of the supposed OMV does not appear in the BHI-S control, suggesting that it was not an artefact of the medium.

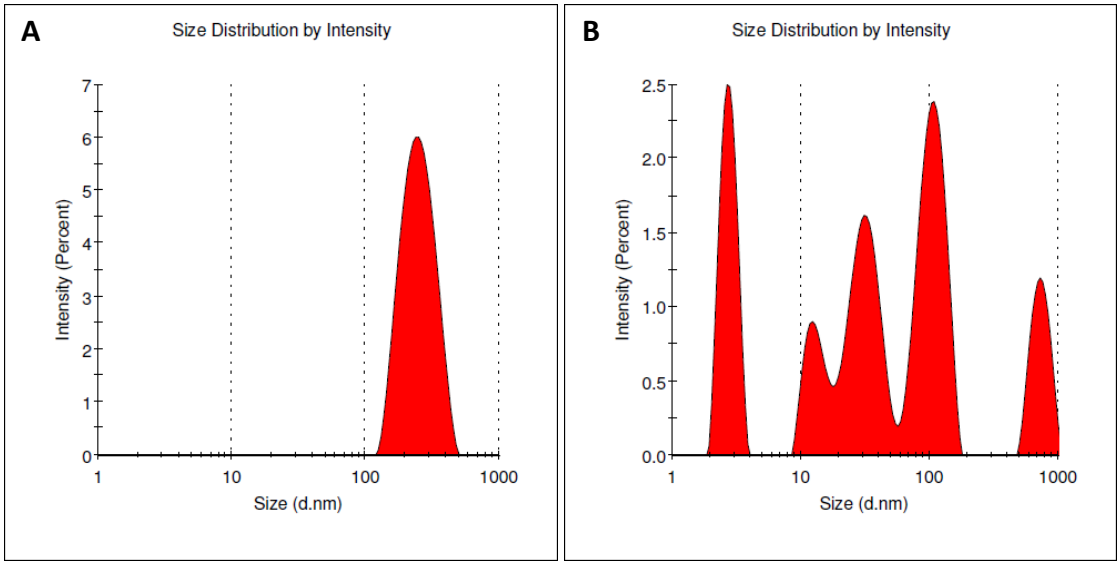


Figure 22 Dynamic light scattering of WT *B. fragilis* concentrated supernatant grown in BHI-S **(A)** and a BHI-S control **(B)**. **(A)** shows a monodisperse solution with a single peak ranging from 150-500nm with an average of approximately 200nm. **(B)** shows a polydisperse solution with at least 5 peaks of varying sizes. These most likely represent large molecules or precipitates in the BHI-S. None of the peaks were the same size as that seen in **(A)**.

4.2.2 Proteome of the OMV of *B. fragilis*

The complete proteome of both the OMV and periplasm of WT *B. fragilis* was determined by gel-free mass spectrometry. OMV were isolated from 3 separate WT *B. fragilis* cultures. The supernatant of 5L WT *B. fragilis* culture was concentrated to ~200ml using TFF, then to ~3ml by centrifugal filtration, then 3mg of total protein were loaded onto an Opti-Prep gradient and ultracentrifuged to separate OMV from contaminants. The Opti-Prep gradient was unloaded in 1ml fractions and each fraction tested for the level of β-galactosidase (cytoplasmic) and alkaline phosphatase (periplasmic) activity (Table 6). The 3 fractions (those with the highest alkaline phosphatase/β-galactosidase ratios) in each sample were pooled. The pooled samples had an average concentration of ~1.77mg/ml, as determined by Bradford assay.

Periplasmic extract was also prepared from each bacterial culture and found to have an average of concentration of 3.83mg/ml. 1mg of WT OMV and 1mg of WT periplasmic extract (prepared in triplicate) were used for proteomic analysis. The total protein of each sample was precipitated using methanol and chloroform, then digested with trypsin and cleaned. Samples were analysed separately by HPLC-MS (LTQ-Orbitrap) and proteins identified using Progenesis LC-MS software.

Sample A				Sample B				Sample C			
Fraction	AP	β-G	Ratio	Fraction	AP	β-G	Ratio	Fraction	AP	β-G	Ratio
A1	0.04	269.10	0.0001	B1	0.05	137.91	0.0004	C1	0.11	175.93	0.0006
A2	0.02	218.52	0.0001	B2	0.03	171.72	0.0002	C2	0.11	165.81	0.0007
A3	0.05	155.71	0.0003	B3	0.02	124.34	0.0002	C3	0.14	124.62	0.0011
A4	0.11	124.58	0.0009	B4	0.06	77.33	0.0008	C4	0.12	82.99	0.0015
A5	0.12	77.91	0.0016	B5	0.09	80.13	0.0011	C5	0.16	63.93	0.0025
A6	0.22	64.20	0.0035	B6	0.15	65.26	0.0022	C6	0.21	68.52	0.0030
A7*	0.31	41.30	0.0075	B7	0.19	60.48	0.0032	C7*	0.24	64.66	0.0037
A8*	0.27	34.72	0.0078	B8	0.26	55.56	0.0048	C8*	0.27	32.47	0.0084
A9*	0.25	29.74	0.0084	B9*	0.35	34.50	0.0101	C9*	0.25	47.84	0.0053
A10	0.18	45.19	0.0041	B10*	0.33	42.78	0.0077	C10	0.19	76.25	0.0025
A11	0.21	83.33	0.0025	B11*	0.28	52.33	0.0054	C11	0.15	64.97	0.0024
A12	0.23	63.13	0.0036	B12	0.23	54.11	0.0042	C12	0.15	47.62	0.0031

Table 6 The results of the alkaline phosphatase and β-galactosidase assays. Shown are the calculated units/ml of both alkaline phosphatase (AP) and β-galactosidase (β-G) in each fraction and the ratio of AP/β-G. The fractions which were pooled for analysis are indicated by *.

A total of 1,009 proteins were identified by MS between the OMV and periplasmic extract samples. Of these, 783 were identified using 2 or more peptides. Label-free quantitation was performed using peptide spectra (Le Bihan *et al.*, 2011) and used to compare the OMV and PP samples. To be considered enriched in a particular sample, a protein must have been identified using 2 or more peptides and be more than 2-fold more abundant in that sample with an associated ANOVA p value of <0.05. By these standards 88 proteins are considered

to be significantly more enriched in OMV than in the PP; these proteins will be referred to as OMV proteins from here (all of these proteins were also identified in the PP). Meanwhile 84 proteins were considered to be more enriched in the PP than in the OMV (Figure 23).

Figure 23 Venn diagram showing the proportion of proteins (those which were identified using 2 or more peptides) which were significantly enriched in the OMV or periplasm by abundance. Proteins are considered to be significantly enriched if the ratio of the mean abundance of the protein in the OMV ($\bar{x}$ OMV) over the mean abundance of the protein in the periplasm ($\bar{x}$ PP) is >2 (or vice versa) with an associated ANOVA p value of <0.05.

Only 29% of OMV proteins could be identified as outer membrane and periplasmic proteins (Figure 24 A), however several proteins of unknown location are known to contain a signal sequence and are likely to originate from these subcellular fractions. 21% of enriched OMV proteins were of cytoplasmic origin. This is an unexpectedly high number and may be indicative of contamination during the isolation of the OMV. The presence of OM, IM and cytoplasmic proteins in the PP sample (Appendix VII) suggests that significant contamination occurred during the periplasmic extraction procedure.

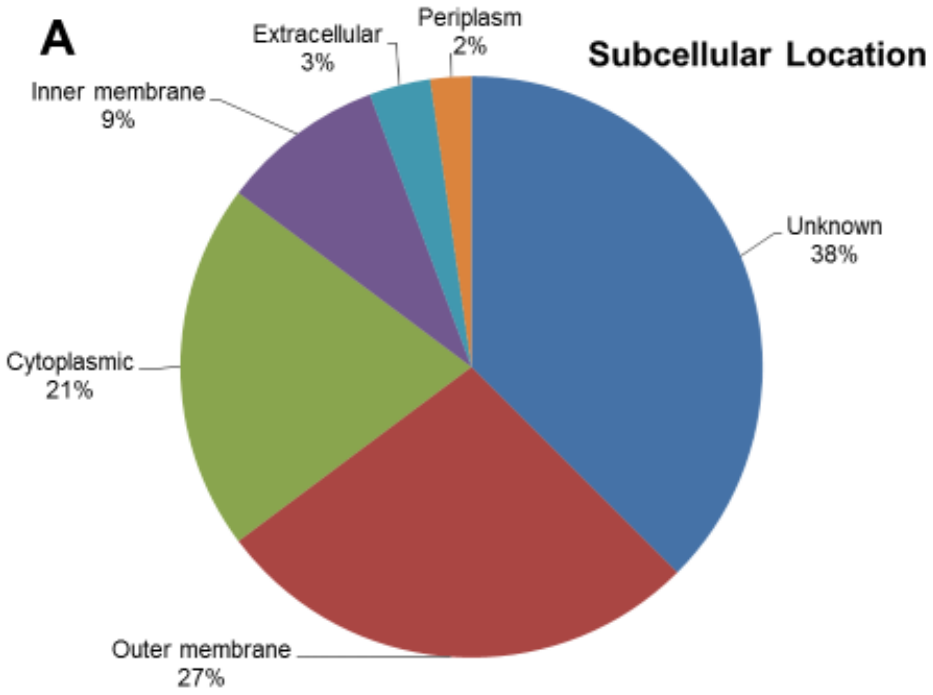
The identity of proteins in the OMV were predicted using Mascot (Matrix Sciences, USA) or BLASTP (NCBI), based on a protein sequence comparison against NCBI non-redundant databases. The OMV proteins were of a wide range of predicted functions (Figure 24 B), the most common of which were metabolic and biosynthetic enzymes. 42% of OMV proteins remain uncharacterised.

There was no enrichment of any particular function or subcellular location amongst the most abundant of the 783 proteins identified in the OMV (Table 7). The most commonly seen in these 20 proteins were metabolic or biosynthesis enzymes of cytoplasmic origin and outer membrane nutrient acquisition proteins. Of the 20 proteins which show the largest fold increase in abundance over the PP (Table 8), more than half are uncharacterised and of unknown subcellular origin, whilst the remaining proteins consist mainly of outer membrane and extracellular proteins. There are a large number of nutrient binding and acquisition proteins in the OMV (Figure 24), however due to the lack of characterisation it is unclear which functional group is actually the most enriched. For a full list of proteins enriched in the OMV see Appendix VII.

There are 226 proteins identified in this study which cannot be considered "true" hits as they were identified using only 1 peptide. There are 84 which are significantly less abundant in the OMV than in the PP and a further 611 proteins which are not significantly enriched in the OMV. There are some proteins of interest in this last group, including a previously documented fibrinogen-binding protein (section 4.2.6). It is probable that other proteins thought to exist in the OMV can be found in this group.

Accession Number	Predicted Identity	Functional Group	Subcellular Origin
YP_213166.1	Xylanase	Nutrient binding	Extracellular
YP_097602.1	Hypothetical protein	Unknown	Unknown
YP_101476.1	Elongation factor Tu	DNA/RNA	Cytoplasm
YP_099239.1	Starch-binding protein	Nutrient binding	Outer membrane
YP_101460.1	Elongation factor G	DNA/RNA	Cytoplasm
YP_099848.1	tRNA synthetase subunit	DNA/RNA	Cytoplasm
YP_100577.1	Fructose-bisphosphate aldolase	Metabolism	Cytoplasm
YP_100610.1	Pyruvate-flavodoxin oxidoreductase	Metabolism	Cytoplasm
YP_098358.1	Hypothetical protein	Unknown	Unknown
YP_098251.1	Glyceraldehyde-phosphate dehydrogenase	Metabolism	Cytoplasm
YP_098327.1	30S ribosomal protein S1	DNA/RNA	Cytoplasm
YP_101034.1	Ketol-acid reductoisomerase	Biosynthesis	Cytoplasm
YP_098961.1	Hypothetical protein	Unknown	Unknown
YP_101717.1	Phophoenolpyruvate carboxykinase	Metabolism	Periplasm
YP_213038.1	Xylanase	Nutrient binding	Outer membrane
ZP_07811280.1	Two component system response regulator	Signal transduction	Outer membrane
YP_100507.1	Lipoprotein	Unknown	Outer membrane
YP_099179.1	OmpA family protein	Unknown	Outer membrane
YP_101286.1	Elongation factor Ts	DNA/RNA	Unknown
YP_097970.1	Ribosome recycling factor	DNA/RNA	Unknown

Table 7 The 20 most abundant proteins amongst the 783 identified using 2 or more peptides in the OMV sample. Included is the accession number for each protein (*B. fragilis* NCTC 9343), the predicted identity of the protein according to Mascot or a BLASTP search against all NCBI protein databases, and the functional group and predicted subcellular origin of each protein.



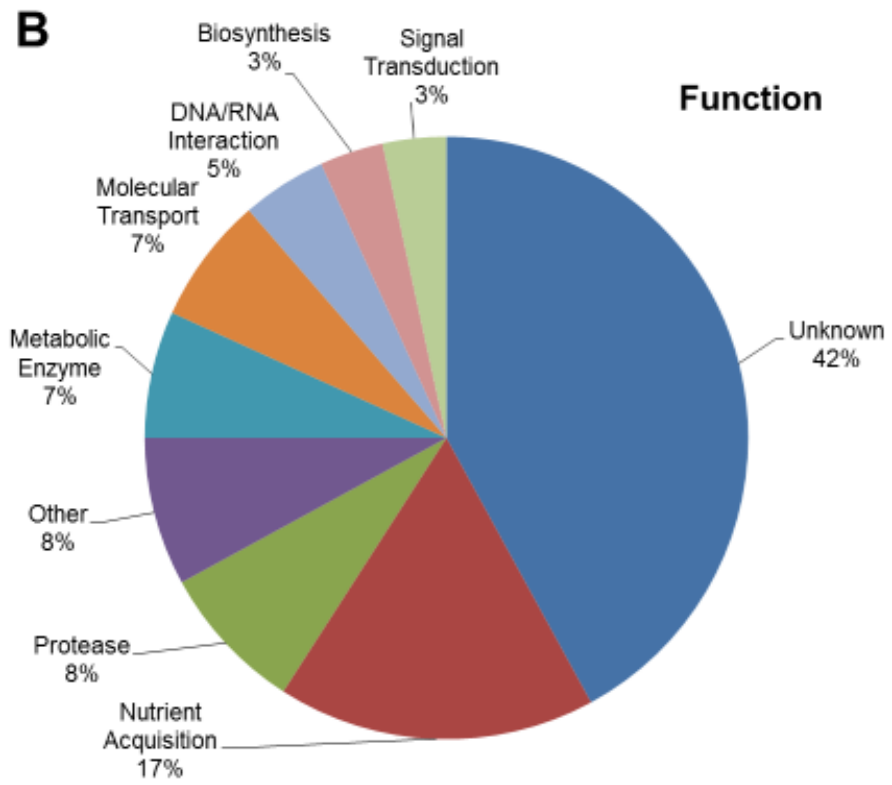


Figure 24 Pie charts showing the subcellular origin **(A)** and predicted function **(B)** of the 88 proteins enriched in the OMV.

Accession No.	Predicted Identity	Functional Group	Subcellular Origin
YP_210059.1	TonB-dependent receptor	Signal Transduction	Outer Membrane
YP_210687.1	Putative lipoprotein	Unknown	Unknown
YP_210300.1	SusC/RagA	Nutrient	Outer Membrane
YP_210172.1	Hypothetical protein	Unknown	Extracellular
YP_210578.1	Hypothetical protein	Unknown	Unknown
YP_211023.1	Hypothetical protein	Unknown	Unknown
YP_210845.1	Hypothetical protein	Unknown	Unknown
YP_212392.1	Hypothetical protein	Unknown	Unknown
YP_213166.1	Xylanase	Nutrient	Extracellular
YP_213613.1	OM protein	Unknown	Outer Membrane
YP_213037.1	Hypothetical protein	Unknown	Unknown
YP_211269.1	ATP/GTP-binding subunit	Unknown	Unknown
YP_212656.1	Lipoprotein, LigD superfamily	DNA/RNA binding	Unknown
YP_211022.1	Hypothetical protein	Unknown	Unknown
YP_212584.1	OM protein	Unknown	Outer Membrane
YP_211011.1	Alpha-2 macroglobulin	Protease inhibitor	Outer Membrane
YP_213040.1	Hypothetical protein	Unknown	Unknown
YP_211278.1	C10-like peptidase	Protease	Unknown
YP_213780.1	TonB-dependent receptor	Signal transduction	Outer Membrane

Table 8 The 20 OMV proteins with the largest fold increase in abundance over the periplasm, i.e. the ratio of the mean abundance of the protein in the OMV ($\overline{x}$ OMV) over the mean abundance of the protein in the periplasm ($\overline{x}$ PP). Included is the accession number for each protein (*B. fragilis* NCTC 9343), the predicted identity of the protein according to Mascot or a BLASTP search against all NCBI protein databases, and the functional group and the predicted subcellular origin of each protein.

4.2.3 Potential proteases in the OMV of *B. fragilis*

This study identified 7 putative proteases, by a BLASTP search against the NCBI non-redundant database, which are enriched in the OMV (Table 9). Each of these proteases contains an N-terminal signal sequence, indicating that they are exported from the cytoplasm, and have a range of potential functions, including virulence, nutrient acquisition, cell wall turnover and degradation of misfolded proteins. However, these functions are speculative and would need to be confirmed by biochemical techniques. An additional protease was detected in the remaining 611 proteins, however it is not considered enriched in the OMV as it has an associated p value of >0.05.

Accession No.	Size (kDa)	Predicted Identity	Potential Function	Signal Sequence?
YP_211278.1	56.55	Related to Peptidase C10	Virulence	Yes
YP_209856.1	45.29	Related to Peptidase C10	Virulence	Yes
YP_212625.1	81.39	Endopeptidase, conserved in Bacteroidetes	Unknown	Yes
YP_211064.1	61.37	Related to Peptidase C69	Various	Yes
YP_213364.1	48.94	Serine protease	Nutrient acquisition	Yes
YP_210628.1	106.42	Related to Peptidase M16	Unknown	Yes
YP_213342.1	123.39	Related to Peptiase S41	Degradation of incorrectly folded proteins	Yes

Table 9 Potential proteases enriched in the OMV sample. Included is the accession number in NCTC 9343, the predicted molecular weight of the protein (as determined by Artemis), the predicted identity (as determined by a BLASTP search against all available databases), the predicted function of these proteases based on their function in other bacteria, and whether they contain a signal sequence for export from the cytoplasm.

4.2.4 A *B. fragilis* α-2 macroglobulin is present in the OMV

One of the OMV proteins with a large fold increase in abundance over the periplasm, BF1415 (Accession no. YP_211011.1) is a putative α-2 macroglobulin (α2M), identified using a BLASTP search against all available databases. α2ms are large protease traps, found in both prokaryotes and eukaryotes (Budd *et al.*, 2004). An α2m in the OMV of *B. fragilis* may sequester, inhibit and transport proteases either of bacterial or host origin. BF1415 shares 75% identity (across 1734aa, e value 0.0) with an α2M from *B. thetaiotaomicron*. The gene encoding this α2m is directly adjacent to a gene encoding an OM protein, BF1416. The active domains of this α2M are predicted to exist on the extracellular side the OM, according to 3 independent transmembrane prediction servers; CBS TMHMM v.2.0 (Krogh *et al.*, 2001), HMMTOP v2.0 (Tusnady & Simon, 1998) and SACS MEMSAT v1.8 (Jones *et al.*, 1994).

4.2.5 A known fibrinogen-binding protein was detected in the OMV of *B. fragilis*

BF1705 of NCTC 9343 (accession number YP_211343.1) is a previously characterised, 54kDa, outer membrane fibrinogen-binding protein. A deletion mutant, *B. fragilis* Δ*1705* was generated using the I-SceI-expressing plasmid pGB920 (Houston *et al.*, 2010). BF1705 was identified amongst the 783 proteins identified in the OMV, however it is not significantly enriched in the OMV (p value of 0.178). To determine whether this is a true hit, 10µg of concentrated supernatant of WT, pGB920 and Δ*1705 B. fragilis* were analysed by SDS-PAGE followed by Western blotting with rabbit anti-BF1705 antibodies (secondary staining with HRP-conjugated anti-rabbit antibodies). The Western blot shows that a 54kDa band present in both the WT and pGB920 concentrated supernatant is absent in the Δ*1705* concentrated supernatant (Figure 25 C). Gels stained with Coomassie blue and Silver stain show that neither *B. fragilis* pGB920 nor the Δ*1705* mutation have a significantly different pattern of protein expression to the WT (Figure 25 A and B).

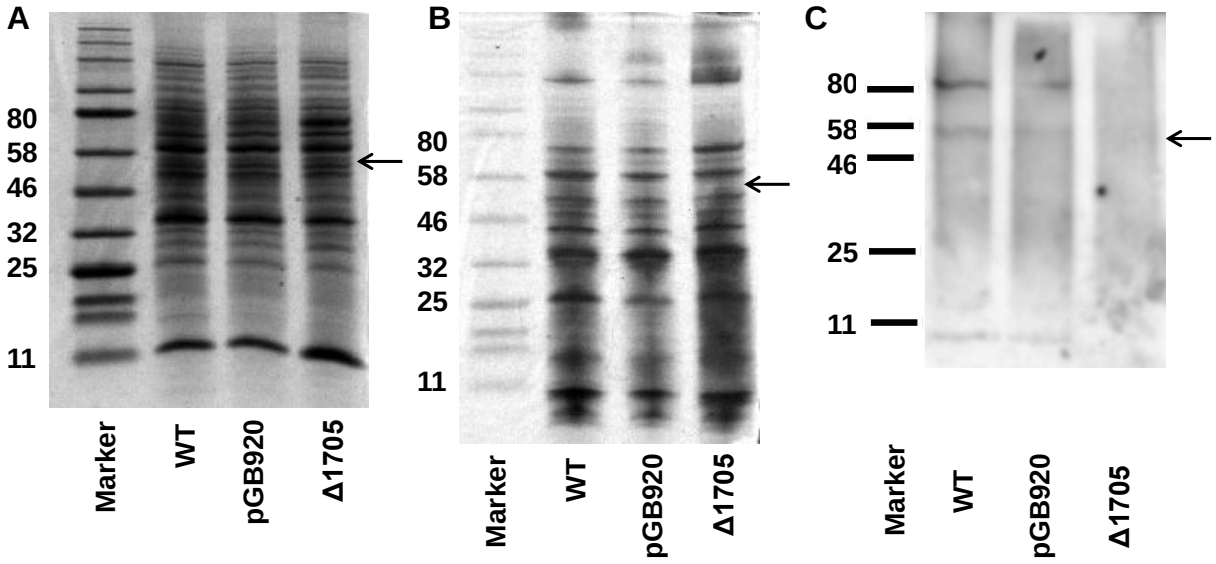


Figure 25 Comparison of WT, pGB920 and Δ1705 concentrated supernatant. Wells were loaded with 10µg total protein and stained by Coomassie **(A)**, Silver stain **(B)** and Western blotting **(C)** using rabbit anti-1705 primary antibodies and HRP-conjugated anti-rabbit secondary antibodies. Arrows indicate the expected location of fibrinogen-binding protein (54kDa).

4.2.6 BfUbb in the OMV of *B. fragilis*

BfUbb is an 8.6kDa, prokaryotic homologue of ubiquitin, which has a periplasmic signal sequence and is believed to be packaged into OMV of *B. fragilis* (Patrick *et al.*, 2011). However, this protein was not identified amongst the 783 proteins (those identified using ≥2 peptides). 10µg of concentrated supernatant of WT, pGB920 and Δ*ubb* were analysed by SDS-PAGE followed by either Coomassie staining, Silver Staining, or Western blotting with rabbit anti-BfUbb antibodies (secondary staining with HRP-conjugated anti-rabbit antibodies). There were no significant differences in protein laddering patterns between the WT, pGB920 and Δ*ubb* strains in the Coomassie- and Silver-stained gels (Figure 26 A and B). A ~8-9kDa band could be detected by Western blotting in WT and pGB920 concentrated supernatant but not in Δ*ubb* concentrated supernatant (Figure 26 C). BfUbb may not have been detected by LTQ-Orbitrap LC-MS due to its relatively small size.

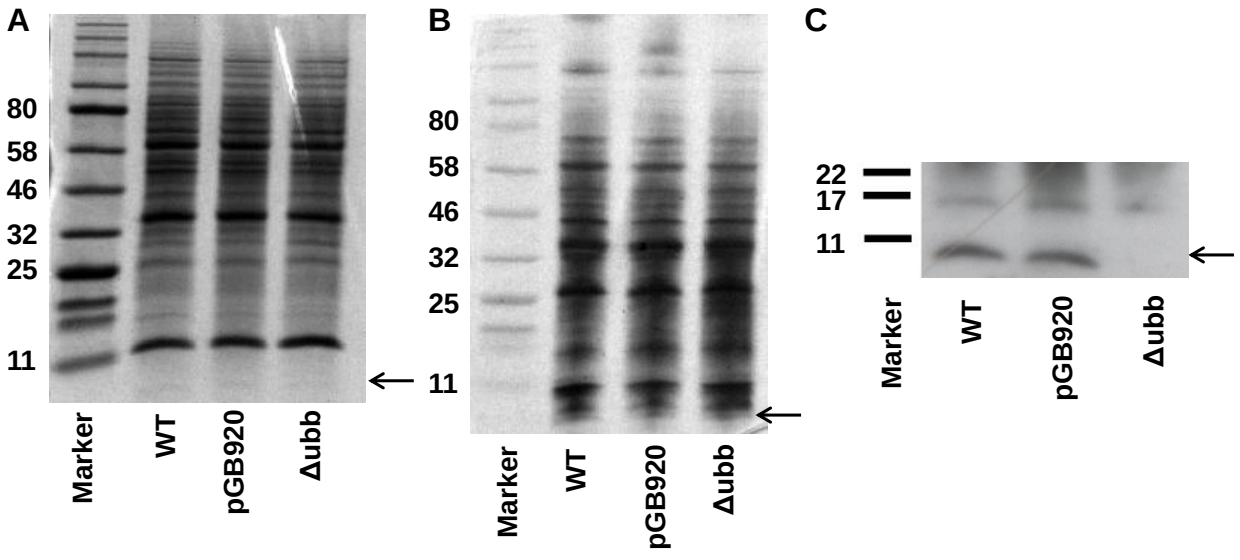


Figure 26 Comparison of WT, pGB920 and Δubb concentrated supernatant. Wells were loaded with 10µg total protein and stained by Coomassie **(A)**, Silver stain **(B)** and Western blotting **(C)** using rabbit anti-BfUbb primary antibodies and HRP-conjugated anti-rabbit secondary antibodies. Arrows indicate the expected location of BfUbb (8.6kDa).

4.3 Discussion

4.3.1 OMV isolation procedure and LTQ-Oribtrap LC-MS

The OMV isolation method used in this study was a combination of techniques established in previous OMV proteomics studies, closely based on method 1 from Bauhman and Kuehn (2006). Whole cells were removed from 5L of *B. fragilis* culture by differential centrifugation (6,000xg) and the supernatant was filtered (0.45µm), concentrated to <500ml by TFF, filtered a second time (0.45µm) and then concentrated by centrifugal filtration to <3ml. The concentration of the concentrated supernatant was ~14.7mg/ml, with a yield of ~5.38mg/L initial culture. Approximately 10mg of total protein were loaded onto an Opti-Prep density gradient (15-40%) and centrifuged (100,000xg). OMV-containing fractions were identified as those with the highest ratio of alkaline phosphatase activity (periplasmic) to β-galactosidase activity (cytoplasmic) and pooled. The final concentration of OMV was ~0.77mg/ml, a yield of 1.24mg/L.

This isolation method should have ensured a high concentration and purity of OMV. Starting with 5L and concentrating the supernatant to <3ml prior to density centrifugation resulted in a high concentration of protein, however, of the ~10mg of protein loaded onto the Opti-Prep gradient, only ~2.31mg was recovered in the OMV-containing fractions. This is a loss of ~77% of total protein, which reduced the overall yield of pure OMV. The concentration of the sample at each stage of the isolation procedure is not known for this study, thus it is unclear whether protein was lost during the concentration of the supernatant. PVDF filters were used to prevent adherence during centrifugal filtration, but not during TFF. The high pressure exerted on the sample during TFF may also have caused lysis of some OMV. Some researchers have added protease inhibitors to the supernatant after filtration to prevent protein degradation (Kolodziejek *et al.*, 2013). Although samples were kept at 4ºC throughout OMV isolation in this study, the whole procedure took up to 48h. It is possible

that proteins were degraded during this time and that the addition of protease inhibitors would have increased overall yield. This may also account for the loss of small or low abundance proteins which will undergo complete degradation faster than larger and more abundant proteins (section 4.3.7). Opti-Prep density centrifugation is considered essential to the isolation of pure OMV (Lee *et al.*, 2008), therefore its use here should ensure a certain level of purity. The alkaline phosphatase and β-galactosidase assays were used to determine which Opti-Prep fractions contain the highest proportion of OMV, on the assumption that OMV are enriched in periplasmic proteins and that cytoplasmic proteins are contaminants. Although both alkaline phosphatase and β-galactosidase peptides were detected in the OMV neither were considered true hits and many other cytoplasmic proteins were identified in the OMV. This suggests that this approach is not sufficient to distinguish OMV from cytoplasmic contaminants.

OMV and PP samples were analysed using a gel-free approach, as is appropriate when dealing with a large number of hydrophobic membrane proteins (van de Waterbeemd *et al.*, 2013). This approach also means that proteins are not excluded by the picking and choosing of particular bands (Choi *et al.*, 2015). The disadvantage of this approach is that digestion of a whole sample without pre-fractionation leads to a very complex peptide mixture. Although peptides were fractionated by HPLC prior to analysis there is still a risk that the signal of small or low abundance proteins is drowned out by that of large or high abundance proteins (Kreimer *et al.*, 2015). Small proteins known to be exported to the periplasm could not be detected in this study (sections 4.2.5 and 4.3.7), suggesting that this effect is occurring. This could be countered by using high-sensitivity approaches such as ULF LC to further separate peptides prior to analysis (Kreimer *et al.*, 2015). Alternatively, quantitative proteomics could determine the absolute abundance of proteins, e.g. by iTRAQ. Spectral counting was used for label-free quantitation in this study, however this is known to result in some proteins with low ion abundance being overlooked (Cooper *et al.*, 2010) and other quantitative methods may be more representative.

4.3.2 Dynamic light scattering of OMV

DLS is cheaper, easier and faster than electron microscopy, which is the standard technique used to determine the average diameter of OMV. Some proteomics studies have used DLS in parallel with TEM and the two methods seem to produce consistent measurements (Choi *et al.*, 2011; Jang *et al.*, 2014). In this study DLS was used to analyse concentrated supernatant of WT *B. fragilis*. A single peak was detected, suggesting a monodisperse solution of globular molecules with an average diameter of ~200nm (Figure 22 A). This peak is considered to be of cellular origin as it does not appear in a BHI-S only control (Figure 22 B), however it indicates a diameter >100nm greater than that of *B. fragilis* OMV as determined by TEM in previous studies of *B. fragilis* OMV (Figure 19) (Patrick *et al.*, 1996). Additionally, there was little consistency between the repeated measurements (Appendix VI). It is possible that repetitive aspiration by the needle of the Zetasizer APS causes turbulence which strips the OMV of their CPS, causing them to aggregate into clumps with a larger average diameter than individual OMV. Additional measurements of pure OMV or concentrated supernatant are required to determine whether the Zetasizer APS can accurately measure OMV diameter. The single peak in the concentrated supernatant sample suggests that the large molecules present in BHI-S are effectively removed during the concentration by centrifugal filtration.

4.3.3 The contents of *B. fragilis* OMV

This study identified 783 OMV proteins using 2 or more peptide, 88 of which are enriched in the OMV (Figure 23). This is consistent with previous studies of OMV proteomes, which have identified around 60-340 proteins (discounting studies where only certain bands on a 1D gel were selected). All 88 proteins were also detected in the PP. Some of the proteins most enriched in the OMV are nutrient-binding proteins and proteases as well as various OM proteins (Table 8, Figure 24). The enrichment of OM proteins in the OMV sample is to be expected.

A significant proportion (21%) of the 88 OMV proteins are of cytoplasmic origin. Similar numbers of cytoplasmic proteins have been identified in other studies, despite the fact that most proteins are expected to be of periplasmic or OM origin (Aguilera *et al.*, 2014; Altindis *et al.*, 2014; Choi *et al.*, 2011; Elhenawy *et al.*, 2014). Some authors claim that if proper care is taken in isolating OMV from log-phase cells there should be no IM or cytoplasmic proteins in the sample (Berlanda Scorza *et al.*, 2012; Kulp & Kuehn, 2010; Scorza *et al.*, 2008). The OMV in this study were isolated from late log-phase/early stationary-phase cultures and may have been contaminated with cytoplasmic proteins from lysed cells. Some studies have described the formation of larger "complex OMV" which incorporate both the inner and outer membrane, as well as sections of the periplasm and cytoplasm. This model of OMV biogenesis accounts for the inclusion of proteins from all subcellular compartments, however it is believed to be damaging to the bacterial cell (Perez-Cruz *et al.*, 2013) and these complex OMV may therefore be too rare to account for the high numbers of cytoplasmic proteins seen in this study. Moreover, these OMV have never been observed by TEM in *B. fragilis* culture (Lutton *et al.*, 1991; Patrick *et al.*, 1996; Pumbwe *et al.*, 2007). Many of the cytoplasmic proteins identified in the OMV are metabolic "house-keeping" enzymes, some of which are thought to have alternative functions in other subcellular compartments. For example, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is known to accumulate of

the cell surface of various bacterial species, including *Streptococcus* spp, *Staphylococcus* spp, *Neisseria* spp, enteropathogenic *E. coli* (EPEC) and *Mycobacterium avium* (Espinosa-Urgel & Kolter, 1998; Grifantini et al., 2002; Lottenberg et al., 1992; Modun et al., 2000; Pancholi & Fischetti, 1992; Pancholi & Chhatwal, 2003). Surface associated GAPDH is known to have ADP-ribosylating activity (Pancholi & Chhatwal, 2003), a feature of many enteric toxins, and in EPEC it is required for initiation of signalling between the bacterium and the host cell (Espinosa-Urgel & Kolter, 1998). One of the most abundant proteins identified in *B. fragilis* OMV in this study was a GAPDH (Table 7). This protein, presented on the surface of OMV, may allow *B. fragilis* to interact with its host. Other cytoplasmic enzymes have been found on the surface of bacteria, despite lacking signal sequences, including phosphoglycerate mutase, triosephosphate isomerase, aldolase, pyruvate kinase, alcohol dehydrogenase and succinyl-CoA synthetase (Pancholi & Chhatwal, 2003). All of these enzymes can be found amongst the total 783 proteins identified in the OMV of *B. fragilis* in this study (Appendix VII).

Whilst the presence of some cytoplasmic proteins can be accounted for, others are more likely to be contaminants, for example, 5% of the enriched OMV proteome consists of DNA- and RNA-associated proteins. OMV are known to contain DNA (Kulkarni & Jagannadham, 2014; Kulp & Kuehn, 2010), and therefore DNA-binding proteins might be expected, however proteins such as DNA polymerase and 50S ribosomal proteins, both of which were detected in the OMV (Appendix VII), likely should not. Cytoplasmic contaminants present a similar problem to large, abundant proteins in that they can be over represented in the digested sample, suppressing the ionisation and detection of small and low abundance proteins (Kreimer *et al.*, 2015).

17% of OMV proteins are nutrient-binding and acquisition proteins (Figure 24 Pie charts showing the subcellular origin (A) and predicted function (B) of the 88 proteins enriched in **the OMV.** B), such as SusC, SusD and xylanase. These also represent some of the most abundant proteins in the OMV (Table 7) (Appendix VII). Nutrient acquisition is believed to be a major function of OMV (Kulkarni & Jagannadham, 2014; Kulp & Kuehn, 2010) and a high number of these proteins is consistent with the known roles of *Bacteroides* spp in the uptake and degradation of complex intestinal sugars (Wexler, 2007).

The majority of enriched *B. fragilis* OMV proteins identified by Mascot were designated "hypothetical proteins". A BLASTP search of each hypothetical protein against the non-redundant NCBI database identified many of the hits, however 42% of OMV proteins remain uncharacterised (Figure 24 B). The lack of characterisation for a large number of entries in protein databases means that proteomic methods cannot be used to fully characterise any biological sample. In this case, the mechanism of biogenesis and all the functions of *B. fragilis* OMV are still to be determined.

4.3.4 *B. fragilis* proteases in the OMV

This study identified 7 proteases enriched in the OMV of *B. fragilis* (Table 9), whereas another study identified 11 potential proteases (Elhenawy *et al.*, 2014). The proteases all contain signal sequences, range in size from 58.2-131.4kDa and may have a range of functions. Two of these proteases are closely related to C10 peptidases, which act as virulence factors in *Streptococcus pyogenes*, *Prevotella intermedia* and *P. gingivalis*. *B. fragilis* NCTC 9343 only encodes 2 C10 genes, BF0116 and BF1640, both of which are predicted to contain lipoprotein (OM anchor) sequences (Thornton *et al.*, 2010). Both of these proteins are significantly more abundant in the OMV than in the PP (Appendix VII). BF0116 is co transcribed with 2 staphostatin-like inhibitors; BF0115 is exported to the PP, BF0117 is exported to the OM (Thornton *et al.*, 2010). Neither of these were detected in the

OMV or PP samples, possibly due to their small size (<100aa). The function of these C10 peptidases is unknown. Although such proteases are known to act as toxins in other species, *B. fragilis* NCTC 9343 is a non-enterotoxigenic strain and it is therefore unlikely that these C10 peptidases have any pathogenic function. Since NCTC 9343 does not express Bft, it is still unknown whether this toxin is packaged into the OMV of ETBF strains.

4.3.5 *B. fragilis* α-2 macroglobulin in the OMV

α-2 Macroglobulins (α2ms) are large (~1,500aa) protease inhibitors, found in both eukaryotes and prokaryotes, which form molecular traps for proteases. Usually acting as homotetramers, α2ms contain a small (30-40aa), flexible, solvent-exposed bait region containing 1 or more proteolytic cleavage sites. When a protease cleaves the bait region the α2m undergoes a conformational change; an exposed glutamyl links with a lysine in the target protease or in another part of the α2m, trapping the protease (Armstrong & Quigley, 1999; Budd *et al.*, 2004; Kantyka *et al.*, 2010; Robert-Genthon *et al.*, 2013). Trapped proteases can no longer cleave macromolecular substrates, however they can cleave smaller substrates. They are also protected from other protease inhibitors and from being proteolytically digested themselves (Armstrong & Quigley, 1999; Kantyka *et al.*, 2010). Different α2ms are able to trap a wide range of proteases due to significant variation in the bait region and active domains, even between closely related species (Armstrong & Quigley, 1999; Budd *et al.*, 2004; Kantyka *et al.*, 2010; Robert-Genthon *et al.*, 2013). In eukaryotes, α2ms form part of the innate immune system, clearing proteases from tissue fluids post-infection. The conformational change in the α2m results in the exposure of a C-terminal motif that binds blood cell surface protein CD19, resulting in endocytosis and digestion of the α2m and the trapped protease (Armstrong & Quigley, 1999; Budd *et al.*, 2004). α2ms are also found in 14 of the 29 bacterial phyla (Kantyka *et al.*, 2010). It is believed that an inter-kingdom HGT occurred between eukaryotes and prokaryotes, however it is unclear in which

kingdom α2ms originated. The primary function of bacterial α2ms, which typically exist in the PP of Gram-negative bacteria, is believed to be the trapping of invading proteases during a breach of the OM (Budd *et al.*, 2004; Kantyka *et al.*, 2010; Robert-Genthon *et al.*, 2013). This study identified a homologue of a bacterial α2m in the OM portion of the OMV of *B. fragilis.* This protein showed significant homology with an α2m in *B. thetaiotaomicron* by BLASTP, with an e value of 0.0 and 75% identity across 1734aa (with a total length of 1864aa), and is considered enriched in the OMV by the parameters of this MS study. α2ms in Gram-negative bacteria are usually encoded near genes for PG synthesis or IM proteins and are usually anchored in the IM, facing the PP (Budd *et al.*, 2004; Robert-Genthon *et al.*, 2013). The gene encoding the α2m identified in this study is directly adjacent to an OM protein and, moreover, the α2m is predicted to be exported to the OM with its active domains surface exposed rather than facing the PP. The presence of this protease inhibitor on the surface of OMV, and the ability of OMV to diffuse in the extracellular medium, suggests that proteases which may harm *B. fragilis* cells could be trapped before they reach their targets. This may provide protection from host or bacterial proteases in the gastro-intestinal tract or aid bacterial survival in the peritoneal cavity, for example, by inhibiting thrombin, the enzyme responsible for fibrin production, thus interfering with abscess formation.

4.3.6 *B. fragilis* fibrinogen-binding protein in the OMV

Fibrinogen is a large (350kDa), mammalian glycoprotein made up of 3 chains (α, β and γ). During abscess formation fibrinogen is cleaved by thrombin to produce fibrin, which then contributes to the fibrous network of the abscess (Henschen *et al.*, 1983). *B. fragilis* cells have been shown to bind the β-chain of fibrinogen via a 54kDa protein, BF1705 (NCTC 9343) (Houston *et al.*, 2010). The binding of fibrinogen by bacteria is considered a virulence attribute as it sequesters the glycoprotein before it can be cleaved by thrombin, therefore reducing the total amount of fibrin available and slowing or preventing the formation of the

abscess. This, in turn, allows the bacteria to disseminate into the bloodstream resulting in bacteraemia. Additionally, binding fibrinogen increases clumping of bacterial cells, thus increasing their resistance to phagocytosis by macrophages. Concentrated supernatants of various *B. fragilis* strains also exhibit fibrinogenolytic activity (Houston *et al.*, 2010). The unrergulated degradation of fibrinogen further reduces the availability of fibrin during abscess formation.

BF1705 was detected in the OMV sample by MS, however it was not considered enriched in the OMV as it had an associated p value of >0.05. This lack of enrichment may be an artefact of OM contamination in the PP sample. Since it is a known OM protein with a potential role in pathogenesis, concentrated supernatant from WT, pGB920 and Δ*1705 B. fragilis* were analysed by Western blotting using anti-1705 antibodies, and BF1705 could be identified in the first two strains but not the latter (Figure 25 C). Whilst the protein could be detected by Western blotting in concentrated supernatants, it should be noted that concentrated supernatant is not necessarily representative of purified OMV. The presence of BF1705 on the surface of the OMV increases the total surface area capable of binding fibrinogen and thus improves the ability of the bacterium to inhibit abscess formation. Fibrinogen-binding by OMV may also contribute to bacterial clumping. Supernatants of *B. fragilis* are also known to have fibrinogenolytic activity, the agent of which is unknown. This study identified several proteases which may be capable of degrading fibrinogen (Table 9).

4.3.7 BfUbb in the OMV

BfUbb is a 76aa (8.6kDa), prokaryotic homologue of ubiquitin (described in more detail in Chapter 3). It contains a periplasmic signal sequence and has previously been detected in concentrated supernatants of *B. fragilis* (Patrick *et al.*, 2011). It was also detected by Western blotting in concentrated supernatant produced in this study (Figure 26 C), however

it was not identified in the OMV by MS. Whilst the concentrated supernatant may not be representative of pure OMV it is still likely that BfUbb is present in the OMV. There is a risk that small proteins such as BfUbb are not detected by this particular approach to MS as they produce far fewer peptides than larger, more abundant proteins. Improved gel-free approaches, e.g. PTAG, may be required to detect such proteins by MS.

4.3.8 Future work and considerations

The list of 88 enriched OMV proteins in this study is most likely an incomplete proteome as it is missing at least 2 proteins which can be detected in the concentrated supernatants of *B. fragilis* NCTC 9343 (BF1705 and BfUbb). The list also contains an unprecedented number of cytoplasmic proteins. It may be that *B. fragilis* produces complex OMV, containing fractions of the IM and cytoplasm. TEM of thin sections of *B. fragilis* culture was used to demonstrate the production of simple OMV (Lutton *et al.*, 1991; Patrick *et al.*, 1996) and could also identify the production of complex OMV from this species. The cytoplasmic proteins may also be indicative of contamination in both the isolation of OMV and extraction of the PP. The OMV isolation method used in this study resulted in a significant loss of protein and may have led to the contamination of the sample by a large number of cytoplasmic proteins. An improved method might use log-phase bacterial culture rather than late log-phase/early stationary-phase bacterial culture to avoid contamination from lysed cells. It is also possible that the high pressures experienced by the OMV during TFF is causing premature lysis and leading to loss of proteins. Removing this step or replacing it with a technique that exerts less pressure on the OMV may improve the yield. An ultracentrifugation step after concentration of the supernatant may remove small protein contaminants and increase the concentration of OMV prior to density centrifugation. It is important that the concentration of the sample be determined at each stage to identify which techniques result in the most

protein loss. Finally, the addition of protease inhibitors to the filtered supernatant would ensure that little-to-no protein is lost due to proteolytic digestion.

Comparing the OMV sample to the PP means that the proteins most enriched in the OMV will be OM proteins, however this may not be representative. A comparison of the OMV with not only the PP but the OM and whole cells of *B. fragilis*, using quantitative MS, would provide a better indication of which proteins are enriched in OMV. In this study label-free quantitation was performed using spectral counting, however this is not considered to be the most sensitive measure of protein quantitation and so the spectral counts were not incorporated into the analysis. Alternative tagging of multiple protein samples, as with iTRAQ and TMT, allows proper quantitation using the ratios of the tags for each protein (Kreimer *et al.*, 2015) and would be more appropriate . Additionally, there appears to be significant contamination of the PP sample, which will have affected which proteins are considered enriched in the OMV. A different method of periplasmic extraction, e.g. by freeze-thaw, may be required to generate a purer sample.

The presence of proteins of interest identified in the OMV must be confirmed using biochemical and/or genetic techniques. The generation of deletion mutants for genes of interest allows comparison with the WT strain, e.g. by SDS-PAGE and Western blotting with antibodies raised against the chosen protein. The activity of proteases can be confirmed without generating deletion mutants. OMV proteins could be separated by SDS-PAGE, followed by excision of bands which represent proteins the same size as suspected proteases. These bands could then be tested for various proteolytic activity. Alternatively, zymography can be used to determine which bands exhibit proteolytic activity. This has already been done for *B. fragilis* OMV (Elhenawy *et al.*, 2014), however it does not indicate the exact nature of the individual proteases and therefore further analysis is needed. Armstrong and Quigley (1999) describe a series of tests which can be used to detect α2m activity in a sample. These tests are based on the principle that a protease trapped by an

α2m would not be able to cleave large peptide substrates but would be able to cleave smaller substrates and would also be immune to other protease inhibitors. This approach could be used to confirm the presence of the α2m on the surface of the OMV, on the assumption that an α2m in the lumen of the OMV would not confer protease inhibitor activity. However, the known proteolytic activity of the OMV may interfere with the assay. Alternatively, the band representing the suspected α2m could be excised from a gel, providing a less complex sample for analysis.

The analysis of the proteome can be taken beyond simply producing a list of proteins. A systems biology approach to investigate the protein-protein interactions within the OMV and between OMV proteins and other bacterial and host proteins would provide a better insight into the function and mechanism of biogenesis of OMV (Choi *et al.*, 2015). An alternative "top-down" approach to proteomics would allow characterisation of post-translational modifications, which can significantly impact protein function and localisation. For example, many of the known *B. fragilis* glycoproteins are OM proteins such as OmpA and TonB (Fletcher *et al.*, 2011), several of which were identified in this study (Appendix VII), and BfUbb contains potential target sequences for glycosylation (Patrick *et al*., 2011). Since more than half of the extra-cytoplasmic proteins of *B. fragilis* are candidates for O-glycosylation (Fletcher *et al.*, 2011), it is likely that the OMV contain a large number of glycoproteins. Finally, the lipid profile of the OMV and OM of *B. fragilis* can be determined by lipidomics, possibly providing further information about the origin and function of OMV (Kreimer *et al.*, 2015).

5. Anti-bacterial potential of *B. fragilis* OMV and BfUbb

5.1 Introduction

5.1.1 Antibiotic potential of OMV

One of the many possible functions of OMV is the killing or inhibition of growth of competing bacteria (Kulp & Kuehn, 2010). The Gram-negative soil bacterium *Myxococcus xanthus* is known to selectively "prey" on other bacterial species, specifically by causing lysis upon cell-cell contact. *M. xanthus* expresses a large number of hydrolytic enzymes and secondary metabolites with antibiotic activity, which were found to be enriched in the OMV of this species, along with the potentially lytic protease MepA. It may be that OMV of *M. xanthus* cause lysis of target bacteria upon contact (Berleman *et al.*, 2014). *P. aeruginosa* expresses a murein hydrolase, an autolysin that is associated with the peptidoglycan layer of the periplasm and is a normal part of cell wall turnover (Li *et al.*, 1996). Expression of this enzyme is closely associated with the presence of the B band species of LPS, the only species found in naturally formed *P. aeruginosa* OMV (Kadurugamuwa & Beveridge, 1995). These OMV are able to fuse with the surface of other bacteria, probably through electrostatic interactions between the LPS of the OMV and the surface polysaccharides of the target cell (Kadurugamuwa & Beveridge, 1996; Kadurugamuwa & Beveridge, 1999). This means the OMV could potentially deliver the autolysin, resulting in peptidoglycan degradation and cell lysis (Li *et al.*, 1996). OMV production in *P. aeruginosa* is greater in the presence of sub-lethal concentrations of aminoglycosides, such as gentamicin. Gentamicin is highly electropositive, and it has been hypothesised that it can disrupt the salt bridges between electronegative LPS molecules on the cell surface. This would cause destabilisation of the LPS, allowing the membrane to bleb and artificial OMV to form. Where this occurs, lesions in the peptidoglycan layer may lead to deregulation of autolysins and further destabilisation of the membrane, resulting in increased OMV production (Kadurugamuwa & Beveridge, 1997).

The OMV formed during this process contain gentamicin (Kadurugamuwa & Beveridge, 1995) and are more effective than naturally formed OMV at inhibiting the growth of both Gram-positive and Gram-negative bacteria, presumably due to the synergistic effects of the hydrolases and the antibiotic (Kadurugamuwa & Beveridge, 1996). The OMV of a psychrotrophic strains of *Pseudomonas syringae* (Lz4W) can also carry antibiotics. These OMV are able to protect cells from the action of colistin and melilitin (but not streptomycin), specifically by absorbing the antibiotics into their lumen. Mass spectrometry analysis showed that the antibiotics are not degraded, just sequestered (Kulkarni *et al.*, 2014), suggesting that they could potentially be delivered to other bacteria. OMV are of biotechnological interest as they have the potential to deliver a varied cocktail of antibiotic proteins and compounds, which could prevent the development of single-gene resistance in pathogenic target bacteria (Berleman *et al.*, 2014).

5.1.2 Antibiotic potential of ubiquitin

Amongst its many roles, ubiquitylation is involved in the cellular responses to infection, for example through signalling in the inflammatory response pathway (Corn & Vucic, 2014) and lysosomal-based microbe killing (Alonso *et al.*, 2007). Interestingly, ubiquitin and ubiquitin-like proteins have also been shown to display potent antimicrobial activity *in vitro*. A ubiquitin protein isolated from a Pacific oyster is expressed as a fusion with the ribosomal subunit S27, and cleaved post-translation. This cleaved product was able to inhibit the growth of both Gram-positive and Gram-negative bacterial pathogens via a non-lytic mechanism (Seo *et al.*, 2013). Pikt, a ubiquitin-like protein of the yeast *Pichia anomala*, is a killer toxin which is able to inhibit microbial growth without disrupting membrane integrity (De Ingeniis *et al.*, 2009). The best described ubiquitin-like antimicrobial is ubiquicidin, first identified as a post-translational cleavage product of the protein Fau in the cytoplasm of mouse macrophages stimulated with interferon-γ. Ubiquicidin is 74aa long and shares 38% identity with ubiquitin,

and was able to inhibit the growth of intracellular pathogens *Listeria monocytogenes* and *Salmonella enterica* Typhimurium, as well as other Gram-positive and Gram-negative pathogens. This activity was found to be as potent as many mammalian defensins (Hiemstra *et al.*, 1999; Howell *et al.*, 2003).

Some studies have not been able to replicate the inhibitory effects of whole ubiquitin (Svensson *et al.*, 2005), however others indicate that it is peptide fragments of ubiquitin that are responsible for antimicrobial activity. The C-terminal peptide, Ub_{65-76}, is a potent antifungal which is able to form holes in the cell wall of fungi and accumulate in the cytoplasm where it inhibits growth. The N-terminal peptide, Ub_{1-34}, cannot cross the cell wall but can work synergistically with Ub_{65-76} to inhibit growth, presumably entering the cell via the holes made by the C-terminal peptide (Kieffer *et al.*, 2003). Whilst whole ubiquitin was not able to inhibit growth of *Mycobacterium tuberculosis*, samples of ubiquitin pre-incubated with lysosomal proteinases could, suggesting that the digested products were antibacterial (Alonso *et al.*, 2007). Furthermore, ubiquitin has been identified in biological fractions, extracted from insect cells (Svensson *et al.*, 2005) and colonic mucus (Antoni *et al.*, 2013), that demonstrate antibacterial activity. Ubiquicidin could also be identified in fractions of non-inflammatory colonic mucosa with antimicrobial activity (Howell *et al.*, 2003). Ubiquitin and ubiquicidin, as extra- or intracellular antimicrobial proteins, are therefore considered part of the innate immune system in animals (Hiemstra *et al.*, 1999; Howell *et al.*, 2003), however the exact mechanism of the antimicrobial activity of ubiquitin and its peptide fragments is unknown (Alonso *et al.*, 2007; De Ingeniis *et al.*, 2009; Kieffer *et al.*, 2003; Seo *et al.*, 2013)

5.1.3 Aims

B. fragilis OMV are believed to carry a homologue of eukaryotic ubiquitin (Figure 26 C) (Patrick *et al.*, 2011). Since both eukaryotic ubiquitin and the OMV of various bacteria have been shown to exhibit antibiotic activity, it may be that the OMV of *B. fragilis* are able to kill or inhibit growth of other bacteria in a BfUbb-dependent manner. Such an action would benefit the bacterium in multiple ways: there would be fewer bacterial species competing for available nutrients, and killing or inhibiting pathogens would prevent inflammation of the gut or death of the host, thus preserving the environment. The following experiments were designed to determine whether *B. fragilis* OMV exhibit antibacterial activity and, if so, whether BfUbb contributes to this activity.

5.2 Results

5.2.1 Wild-type *B. fragilis* concentrated supernatant restricts growth of *S. enterica* Typhimurium

Concentrated supernatants of WT and Δ*ubb* strains of *B. fragilis* were tested for their ability to inhibit the growth of the Gram-negative gut pathogen *S. enterica* serovar Typhimurium SL3144. 5ml aliquots of LB were inoculated with 20µl overnight culture of *S. enterica* Typhimurium and 500µl 1mg/ml concentrated supernatant (suspended in PBS) of the WT and Δ*ubb B. fragilis* strains. As a control, 5ml aliquots of LB were inoculated with 20µl overnight culture of *S. enterica* Typhimurium and 500µl PBS. All 3 culture conditions were prepared in triplicate. Cultures were then incubated at 37ºC with shaking for a total of 8h. At time points T0, T2, T4, T6 and T8, 100µl of each culture were used for serial dilutions (10^{-1}-10^{-3}) in PBS. 100µl of an appropriate dilution was plated onto LB agar and grown aerobically for 48h at 37ºC, after which colonies were counted and colony-forming units (cfu)/ml calculated (Appendix VIII) and the averages plotted on a growth curve (Figure 27).

In cultures treated with WT concentrated supernatant, *S. enterica* Typhimurium grew slower than in those treated with PBS or Δ*ubb* concentrated supernatant. One-tailed, paired sample T-tests (assuming unequal variances) were used to compare the cfu/ml of the experimemtal cultures against the control cultures at each time point, with a pre-decided critical p value of 0.05. This analysis indicated that between T2 and T8 the cfu/ml in the "WT OMV" treated cultures was ~2-4 -fold lower than that of the PBS treated cultures (p values of <0.028) (Appendix VIII). The cfu/ml in the "Δ*ubb* OMV" treated cultures, however, was never significantly lower than that in the control, the greatest fold difference being 1.24 at time point T4 (p value of 0.209) (Appendix VIII).

The implication of these results is that the WT concentrated supernatant has the ability to inhibit the growth of other Gram-negative bacteria, and that this ability may be associated with the presence of BfUbb in the OMV. However, carriage of the *ubb* gene is not the only difference between WT and Δ*ubb* strains of *B. fragilis*. Firstly, the Δ*ubb* mutation was generated using the plasmid pGB920 (Patrick *et al.*, 2011), and since the strain demonstrates resistance to 5µg/ml tetracycline the plasmid is probably still present. Secondly, in this study, the WT and Δ*ubb* strains were grown in different antibiotics: WT in 20µg/ml gentamicin, and Δ*ubb* in 10µg/ml erythromycin. During preparation of concentrated supernatants, 400ml of culture supernatant was concentrated using a centrifugal concentrator with a 175kDa cut-off filter to a dead-stop volume of ~1.5ml. The concentrate was then resuspended in 30ml PBS whilst still on the filter, and centrifuged back to a volume of ~1.5ml. This step was repeated and the final ~1.5ml samples retrieved, both at a concentration of ~2mg/ml. These were then diluted to 1mg/ml for this study. Some antibiotic will be retained in the concentrated supernatants of WT and Δ*ubb B. fragilis*, at a final concentration of ~0.05µg/ml gentamicin and ~0.025µg/ml erythromycin, respectively, in the undiluted samples. This would result in ~0.025µg/ml gentamicin and ~0.0125µg/ml erythromycin in the WT and Δ*ubb B. fragilis* concentrated supernatants used in this study. The inhibition of *S. enterica* Typhimurium growth may be related to the presence of gentamicin rather than the action of BfUbb. Additionally, in this experiment the concentrated supernatants were added to stationary/lag phase bacterial cells, i.e. cells that are not actively growing, and the observed effects may not apply to growing cells.

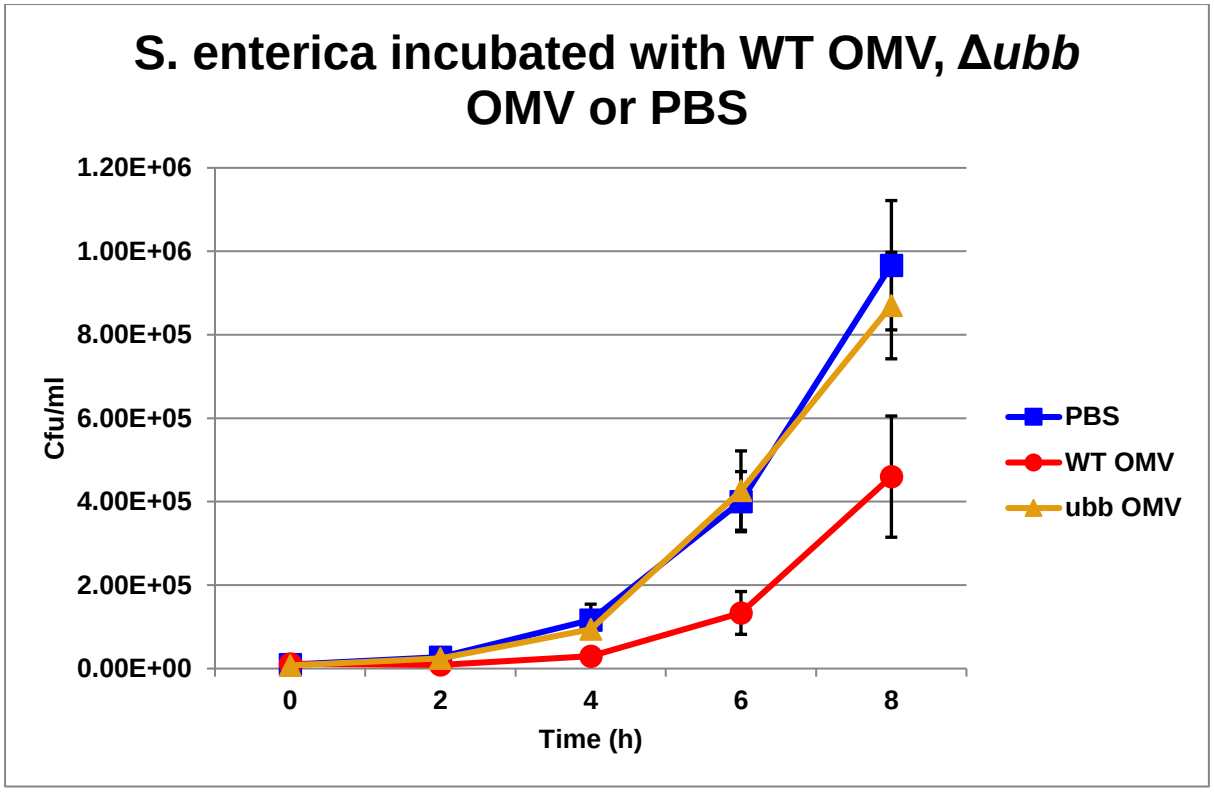


Figure 27 Graph showing the log_{10} mean cfu/ml against time (with standard deviations) in cultures of *S. enterica* Typhimurium incubated with either PBS, WT concentrated supernatant (WT OMV) or Δ*ubb* concentrated supernatant (ubb OMV). At T0, 20µl overnight culture was added to 5ml LB along with 500µl of either PBS, 1mg/ml WT concentrated supernatant or 1mg/ml Δ*ubb* concentrated supernatant, all in triplicate. Cultures were incubated at 37ºC with shaking for 8h. At time points T0, T2, T4, T6 and T8, 100µl of each culture was serially diluted (10^{-1}-10^{-3}) in PBS and 100µl of the appropriate dilution was spread onto LB plates. After 48h colonies were counted and cfu/ml calculated (no. of colonies x 10 x dilution) and the mean and standard deviations plotted. 500µl of each test sample was also added to 5ml LB without *S. enterica* Typhimurium and grown for 8h before plating. No growth was detected (data not shown).

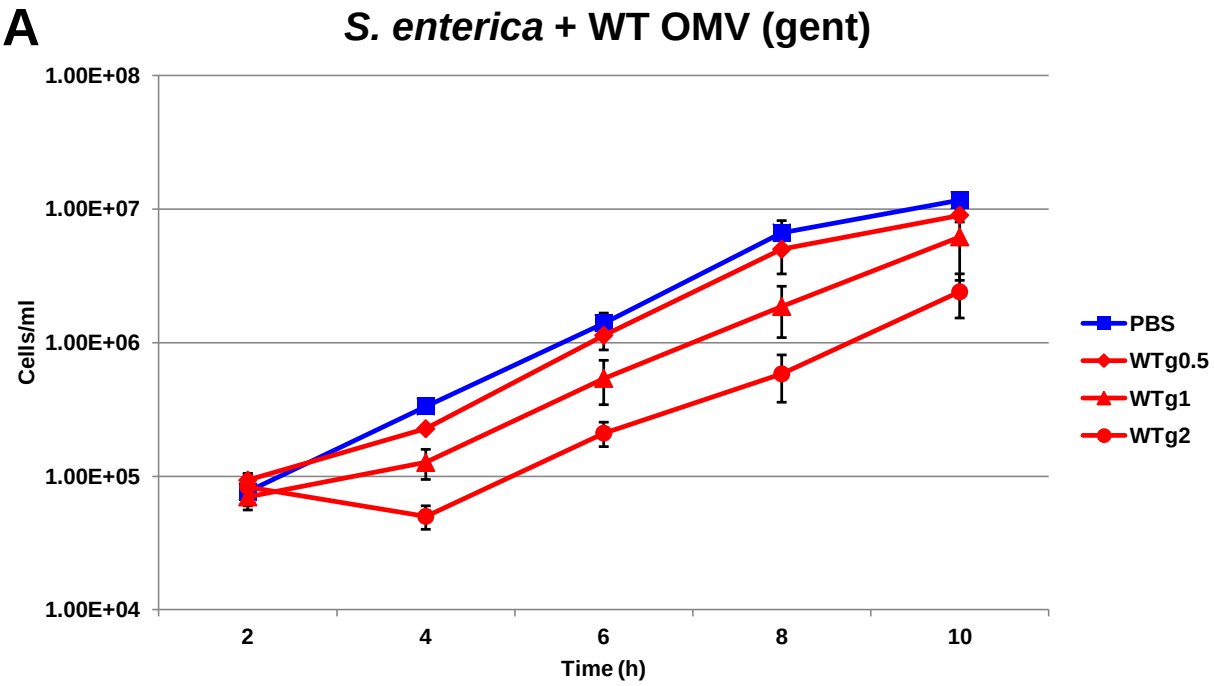
5.2.2 Antibiotic activity of *B. fragilis* OMV is partially dependent on BfUbb

It was necessary to determine whether the ability of *B. fragilis* OMV to inhibit growth of *S. enterica* Typhimurium was related to the presence of the *ubb* gene, the pGB920 plasmid, or gentamicin. Concentrated supernatants were prepared of WT *B. fragilis* grown in the presence (WTg) or absence (WT) of 20µg/ml gentamicin, and of *B. fragilis* pGB920 (grown in 5µg/ml tetracycline, to maintain the plasmid) and *B. fragilis* Δ*ubb* (grown in 10µg/ml erythromycin, to prevent contamination with the WT strain). These were then diluted in PBS to 3 different concentrations: 0.5, 1 and 2mg/ml. 5ml aliquots of LB were inoculated with 5µl overnight culture (T0) and incubated at 37ºC with shaking for 2h (to allow bacteria to leave lag phase), after which (T2) 500µl of each test sample were added to cells in triplicate. At time points T2, T4, T6, T8 and T10, 100µl of each culture were used for serial dilutions (10^{-1}-10^{-6}) and 10µl of each dilution was spotted onto an LB agar plate. Plates were incubated at 37ºC for 48h, then colonies were counted and cfu/ml calculated (Appendix VIII) and the averages plotted on a growth curve.

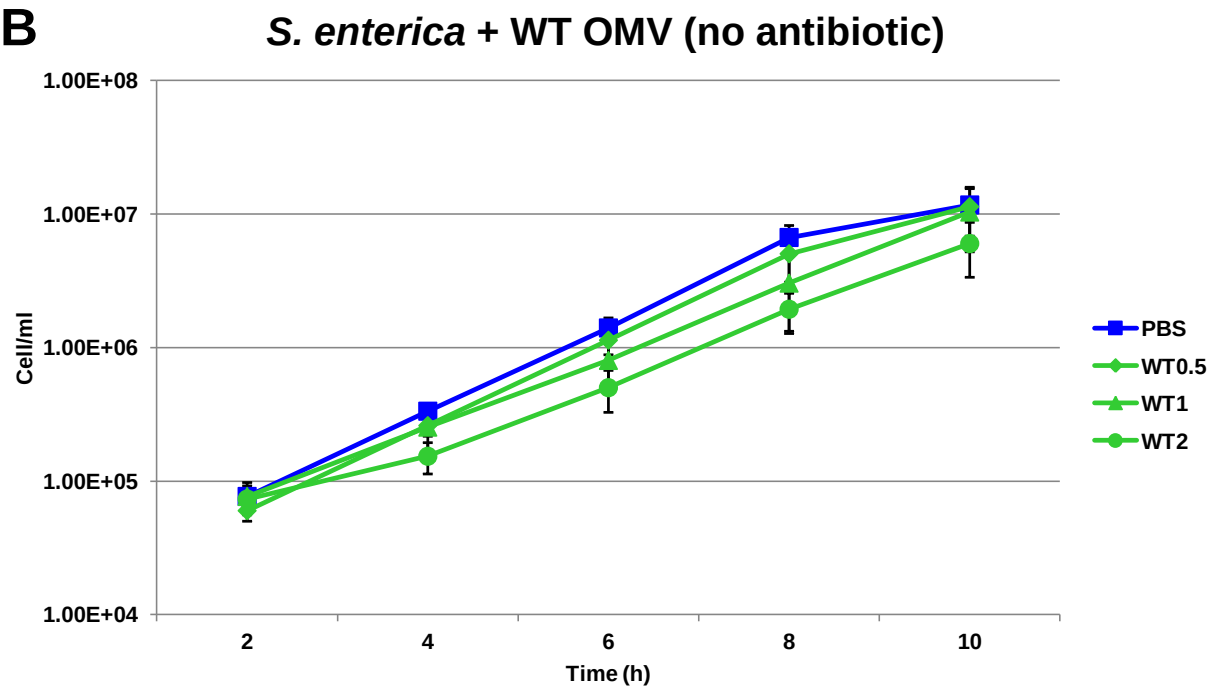
One-tailed, paired sample T-tests (assuming unequal variances) were used to compare the cfu/ml of each test culture with that of the PBS control at every time point (T2-T8), with a pre-decided critical p value of 0.05 (Appendix VIII). Consistent with previous results (Figure 27), concentrated supernatant of WT *B. fragilis* grown in gentamicin delays growth of *S. enterica* Typhimurium (Figure 28 A). T-test analysis indicated that between T4 and T10 the cfu/ml of the WTg_2 treated cultures was 4.9-11.4 -fold lower than that of the PBS treated cultures (p values of <0.009). The cfu/ml of the WTg_1 treated culture was also significantly lower than that of the control: 1.9-3.6 -fold lower between T4 and T10 (p values of <0.044). The cfu/ml of the $WTg_{0.5}$ treated cultures was 1.3-fold lower than that of the control by T10 (p value of 0.02). There were 1.6-fold fewer cfu/ml in the WTg_2 treated cultures at T4 than at T2 (p value of 0.046) suggesting that cells are being killed rather than just inhibited. The concentrated supernatant of WT *B. fragilis* grown without antibiotic also inhibits bacterial growth, but to a

lesser extent than those grown in gentamicin (Figure 28 B), suggesting that the antibiotic is responsible for some antibacterial activity. The T-test analysis showed that the cfu/ml of the WT_2 treated cultures was 1.9-3.4 -fold lower than the control between T4 and T10 (p values of <0.024). The cfu/ml of the WT_1 treated cultures was 1.8-fold lower and 2.2-fold lower than that of the control at T6 and T8, respectively (p values of <0.031), however there was no significant difference at T10. The cfu/ml of the $WT_{0.5}$ treated cultures was not significantly lower than that of the control at any time point other than T4 (1.28-fold lower, p value of 0.040). Growth was also inhibited by concentrated supernatant of *B. fragilis* pGB920 (Figure 28 C). The cfu/ml of the pGB_2 treated cultures was 1.7-2.3 -fold lower than that of the control between T4 and T10 (p values of <0.037), whilst the pGB_1 treated samples had 1.3-fold, 1.5-fold and 1.4-fold lower cfu/ml than that of the control at T4, T6 and T10, respectively (p values of <0.036). The cfu/ml of the $pGB_{0.5}$ treated cultures was not significantly lower than that of the control at any time point other than T4 (1.2-fold difference, p value of 0.025). There was no significant inhibition of growth in cultures treated with concentrated supernatant from *B. fragilis* Δ*ubb* at any time point (Figure 28 D). The greatest difference between the cfu/ml of Δubb_2 treated cultures and that of the control was 1.1-fold at time point T2 (p value of 0.346)

A


S. enterica + WT OMV (gent)
1.00E+08
1.00E+07
1.00E+06
1.00E+05
1.00E+04
Cells/ml
2
4
6
8
10
Time (h)
PBS
WTg0.5
WTg1
WTg2

B


S. enterica + WT OMV (no antibiotic)
1.00E+08
1.00E+07
1.00E+06
1.00E+05
1.00E+04
Cell/ml
2
4
6
8
10
Time (h)
PBS
WT0.5
WT1
WT2

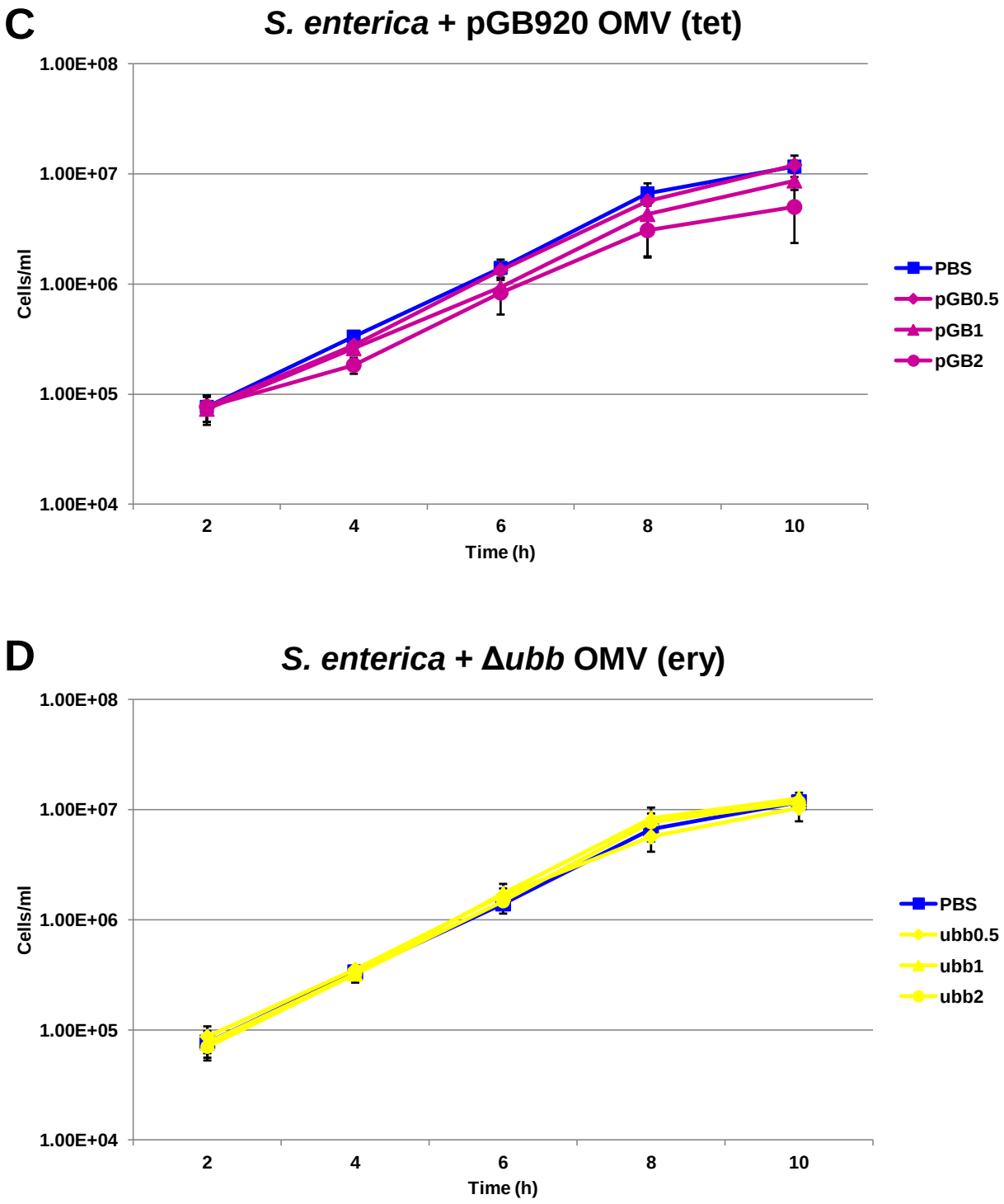


Figure 28 Graphs showing the $\log_{10}$ mean cfu/ml against time (with standard deviations) in cultures of *S. enterica* Typhimurium incubated with either PBS or the concentrated supernatants of WT *B. fragilis* (WT), WT *B. fragilis* grown in gentamicin (WTg), *B. fragilis* pGB920 grown in tetracycline (pGB), or *B. fragilis* Δ*ubb* grown in erythromycin (ubb), at various concentrations. 5µl overnight culture were added to 5ml LB and cultures were grown at 37ºC with shaking for 2h. At this point (T2), 500µl of a test

sample (PBS, or 0.5, 1 or 2ml/ml of WT, WTg, pGB or *ubb* concentrated supernatant) were added and the cultures grown for another 8h. All cultures were prepared in triplicate. At T2, and every 2h until T10, 100µl of each culture were taken and serially diluted (10^{-1}-10^{-6}) in PBS and 10µl of each dilution were spotted onto an LB plate. After 48h colonies were counted and cfu/ml calculated (no. of colonies x 100 x dilution) and the mean and standard deviations plotted. 500µl of 0.5mg/ml of each test sample were added to 5ml LB without *S. enterica* Typhimurium and grown for 10h before spotting. No growth was detected (data not shown).

These results confirm that concentrated supernatant of WT *B. fragilis* is able to inhibit the growth of *S. enterica* Typhimurium. T-test analysis indicated that between T4 and T10 the cfu/ml of the WT_2 treated cultures was 1.9-4 -fold lower than that of the Δ*ubb*$_2$ treated cultures (p values of <0.022). This suggests that some inhibitory activity can be attributed to the presence BfUbb. In this study, supernatants were isolated from 800ml *B. fragilis* culture, concentrated to 1.5ml by centrifugal concentrator, then resuspended in 30ml PBS and concentrated back down to 1.5ml twice. Approximately 4mg total protein was retrieved in each sample and diluted to 2mg/ml for this experiment. The final concentrations of antibiotics in WTg_2, pGB_2 and Δ*ubb*$_2$ samples was therefore expected to be ~0.025µg/ml gentamicin, ~0.006µg/ml tetracycline and ~0.0125µg/ml erythromycin, respectively. The cfu/ml of the WTg_2 treated cultures was 2.4-3.3 -fold lower than that of the WT_2 treated cultures between T4 and T8 (p values of <0.047), indicating that additional inhibitory activity was associated with the presence of gentamicin at 0.025µg/ml. The greatest difference in cfu/ml in pGB_2 treated cultures and that of the WT_2 treated cultures is 1.2-fold at time point T8 (p value of 0.333). This suggests that there is no significant inhibitory activity associated either with the presence tetracycline at 0.006µg/ml or with the presence of the pGB920 plasmid in *B. fragilis* cells during OMV production. There was no significant inhibition of growth in the Δ*ubb*$_2$ treated cultures, suggesting that no inhibitory activity can be associated with either erythromycin at 0.0125µg/ml or with *B. fragilis* Δ*ubb* OMV at 2mg/ml.

5.3 Discussion

5.3.1 OMV-associated BfUbb demonstrates antibacterial activity

BfUbb is believed to be packaged into OMV of *B. fragilis* (Patrick *et al*., 2011), via which it could be delivered to other bacterial cells (Kadurugamuwa & Beveridge, 1996). Given the potential of both OMV (Li *et al.*, 1996) and ubiquitin homologues (Kieffer *et al.*, 2003) to demonstrate antibacterial activity, concentrated supernatants of WT *B. fragilis* and Δ*ubb B. fragilis* were tested for their ability to inhibit growth of *S. enterica* Typhimurium compared to a PBS control. The initial results indicated that not only does WT concentrated supernatant significantly inhibit bacterial growth, but it does so in a BfUbb-dependent manner (Figure 27). A second experiment was performed, this time including concentrated supernatant from WT *B. fragilis* grown without antibiotics, to determine whether inhibition of growth is due to association with gentamicin, and *B. fragilis* pGB920, to confirm that loss of antibacterial activity is associated with the Δ*ubb* mutation rather than presence of the plasmid. These results showed that concentrated supernatant of WT *B. fragilis* grown without antibiotics was also able to significantly inhibit growth of *S. enterica* Typhimurium (Figure 28 A), as was *B. fragilis* pGB920 concentrated supernatant (Figure 28 C), whilst *B. fragilis* Δ*ubb* concentrated supernatant still showed no inhibitory effects (Figure 28 D). The level of antibacterial activity displayed by the concentrated supernatants of WT *B. fragilis* and *B. fragilis* pGB920 was concentration-dependent, suggesting a cause and effect relationship between the presence of *B. fragilis*-derived extracellular material containing BfUbb and the inhibition of growth in *S. enterica* Typhimurium. It should be noted, however, that although the total protein content of each sample is the same, the size, number and nature of the OMV produced by each strain cannot be assessed. Differences in antibacterial activity could therefore be attributed to differences in OMV production rather than content.

The antibacterial effect of the concentrated supernatants of WT *B. fragilis* grown without gentamicin (and *B. fragilis* pGB920) at the highest concentration (500µl of 2mg/ml total protein) was only inhibitory, rather than lethal (Figure 28 A and C). This is consistent with the known antibacterial effects of ubiquitin and ubiquicidin, which are bacteriostatic rather than bacteriocidal (De Ingeniis *et al.*, 2009; Hiemstra *et al.*, 1999; Seo *et al.*, 2013). Although it cannot be assumed that the concentrated supernatants used in these experiments contain only pure OMV, the samples were concentrated and washed using a centrifugal filter that should have prevented the accumulation of any proteins or molecules (<175kDa) that were not closely associated with OMV. It is therefore likely that BfUbb is OMV-associated and that the OMV may responsible for delivery of BfUbb to the bacterial cells. Purified ubiquitin, or at least ubiquitin fragments, are capable of antimicrobial activity (Kieffer *et al.*, 2003; Seo *et al.*, 2013) and free BfUbb may be capable of the same effects, however the OMV provide a protective environment for antimicrobial proteins in the extracellular environment and allow localised delivery of a high concentration of protein. As with eukaryotic ubiquitin, the putative antibacterial mechanism of action of BfUbb against *S. enterica* Typhimurium is unknown, however the C-terminal fragment of ubiquitin is capable of inhibiting bacterial growth on its own (Kieffer *et al.*, 2003). A small band, believed to be the product of proteolytic cleavage, can be detected by anti-BfUbb antibodies in whole cell extracts of *B. fragilis* (Figure 8 B) (Patrick *et al.*, 2011), which suggests that fragments of BfUbb may also be exported in OMV, thus be able to act on other bacterial cells. It is also unknown whether BfUbb-containing OMV would have the same effect on Gram-positive bacteria.

The ability to control populations of competing bacteria is an important aspect of many commensal species, for example through the secretion of bacteriocins (Avelar *et al.*, 1999). *S. enterica* Typhimurium has the potential not only to compete with *B. fragilis* for resources, but to disrupt its environment by causing diarrhoea in the host (Poppe *et al.*, 1998), thus flushing out resident gut bacteria. It would therefore be beneficial to *B. fragilis* to constitutively secrete OMV containing a bacteriostatic protein capable of preventing colonisation of the gut by pathogenic bacteria which may cause inflammation and diarrhoea.

5.3.2 Additional antibacterial activity of *B. fragilis* concentrated supernatants is due to sample-associated gentamicin

Whilst BfUbb seems responsible for a certain amount of antibacterial activity in *B. fragilis* OMV, the concentrated supernatants of WT *B. fragilis* grown in gentamicin demonstrated significantly greater inhibitory activity than that of WT *B. fragilis* grown without antibiotics (Figure 28 A and B) (Appendix VIII). The minimum inhibitory concentration (MIC, the lowest concentration required to inhibit overnight growth) for gentamicin against various *S. enterica* strains is between 0.13 and 1µg/ml, depending on the strain (Stock & Wiedemann, 2000). The final expected concentration of gentamicin in the concentrated supernatants is ~0.025µg/ml, 10-100 -fold lower than the MIC, and would still be capable of delaying growth of *S. enterica* Typhimurium, but not inhibiting it entirely. This phenomenon is not observed with concentrated supernatants containing erythromycin and tetracycline (Figure 28 C and D). The MIC for erythromycin against *S. enterica* Typhimurium is 16-128µg/ml, whilst the expected concentration of erythromycin in the concentrated supernatants of *B. fragilis* Δ*ubb* is 0.0125µg/ml. This is a 1,000-10,000 -fold difference and erythromycin is therefore unlikely to affect growth at this concentration. Accordingly, there is no significant difference between the cfu/ml of Δ*ubb*$_2$ treated cultures and the PBS treated cultures. The concentration of tetracycline in the concentrated supernatants of *B. fragilis* pGB920 (0.006µg/ml) is 100-

1,000 -fold lower than the MIC of tetracycline against *S. enterica* Typhimurium (0.5-4µg/ml). Again, this does not appear to be a high enough concentration to delay growth, since the cfu/ml of pGB_2 treated cultures was never significantly lower than the WT_2 treated cultures.

5.3.3 Future work and considerations

These results indicate a role for OMV-associated BfUbb in the control of bacterial populations in the gastro-intestinal tract, however the exact nature of this role is unclear. A repeat of the above experiments, testing the antibiotic potential of *B. fragilis* OMV against other bacterial species, both Gram-positive and Gram-negative, and both pathogenic and commensal, would indicate the spectrum of antimicrobial activity of this protein. Additionally, each species should be incubated with purified BfUbb to determine whether the protein is able to inhibit growth as effectively as seen here when not associated with the OMV. BfUbb may also have anti-fungal capability, and could be screened against fungal pathogens both in purified and OMV-associated form. As with eukaryotic ubiquitin, the proteolytic products of BfUbb may be just as effective, if not more effective, at inhibiting microbial growth than the whole protein. Genetic constructs allowing expression and purification of N- and C-terminal peptides of BfUbb could be used to test the antimicrobial activity of these products. Electron micrographs of the bacterial cells of the 3 strains used in this and future studies could determine whether there are any differences in the size, number or nature of the OMV formed by each of these strains. The level of BfUbb expression and packaging in all strains and growth conditions should also be assessed, e.g. by Western blotting or reverse transcription qPCR.

Here, antibiotics were used in the cultures of *B. fragilis* from which concentrated supernatants were prepared: gentamicin to prevent the contamination of the WT culture with facultative anaerobes, tetracycline to maintain the presence of plasmid in the *B. fragilis*

pGB920 culture, and erythromycin to prevent contamination of the *B. fragilis* Δ*ubb* culture with the WT strain. To compare the antibacterial activity of just the OMV from WT, pGB920, and Δ*ubb* strains, concentrated supernatants should be prepared from cultures grown without antibiotics. The use of *B. fragilis* defined medium and careful culturing techniques should prevent contamination of the WT, a mini-prep can be used to confirm the presence of the pGB920 plasmid and PCR can be used to confirm the presence or absence of the *ubb* gene.

6. General Discussion

6.1 Biochemical characterisation of ubiquitin and BfUbb

Some of the biochemical properties of ubiquitin and its homologues have made these proteins difficult to characterise. As a β-grasp fold protein it is thermodynamically stable with a T_M >86ºC (Piana *et al.*, 2013), and at temperatures above this, the thermally denatured ubiquitin monomers form aggregates (Herberhold & Winter, 2002). Melting curves are therefore difficult to generate for ubiquitin-like proteins, meaning the strength of ubiquitin/BfUbb-E2 enzyme bonds could not be analysed by thermal denaturation fluorescence. Ubiquitin is also problematic when used as a ligand in SPR, as it is difficult to conjugate to a CM-5 chip (Hartmann-Petersen & Gordon, 2005). The SPR data generated during investigation of BfUbb performed in parallel to this study was inconclusive due to the low concentration of E2 analyte (Appendix I), however there was also difficulty in generating enough resonance units on the CM-5 chip to elicit a response. The problems with using such methods are exacerbated by the fact that most UBDs have low to moderate affinity for ubiquitin (K_d = 2-500μM) (Grabbe & Dikic, 2009), making these interactions difficult to detect. Isothermal titration calorimetry (ITC) is more accurate than SPR, and has been used successfully in ubiquitin-binding experiments (Roth *et al.*, 2007), however it requires much greater quantities of protein (Wear & Walkinshaw, 2006), making it unsuitable for this study, in which a limited amount of each E2 enzyme was available. Since the two techniques complement each other, SPR and ITC are best used in parallel to characterise specific protein-protein or protein-ligand interactions (Jecklin *et al.*, 2009), and could therefore be applied once a specific E2 enzyme, E3 ligase or target protein has been identified.

There exists a wide variety of ubiquitin transfer assays, used to determine ubiquitylation via a particular enzyme cascade (i.e. a specific combination of E1, E2 and E3 enzymes). The most common are those which use band shift (on SDS-PAGE) as an indication of

ubiquitylation (Berndsen & Wolberger, 2011), like the type used here (Figure 13 and Figure 14). However, these methods are laborious and cannot quantify how much ubiquitin is bound to a protein. Other assays use techniques such as Förster resonance energy transfer (FRET) to detect interactions between ubiquitin and substrate. This requires fluorescent labelling of both proteins, which can interfere with the interactions (Berndsen & Wolberger, 2011), especially with proteins as small as ubiquitin and BfUbb. This problem may also be encountered when using the YFP-tagged ubiquitin created in this study, for example in ligand-binding and competition assays. Some ubiquitin transfer assays monitor enzymatic activity, such as the release of AMP during ubiquitin activation (Berndsen & Wolberger, 2011), however these methods cannot be applied in the study of BfUbb since the lack of the terminal glycine (Patrick *et al.*, 2011) means that it cannot be activated or transferred as ubiquitin is (Wickliffe *et al.*, 2011). Taken together, these factors of BfUbb, i.e. its small size, thermodynamic stability, the lack of a terminal glycine and the impracticality of high through-put ligand-binding assays, mean that biochemical characterisation of this protein may prove time consuming, expensive and difficult.

6.2 Ubiquitin homologues lacking a C-terminal double glycine motif

Previously, BfUbb was shown to form covalent bonds with E1 activating enzyme (Patrick *et al.*, 2011) (Figure 9 B) and in this study BfUbb was shown to form covalent bonds with various E2 enzymes (Figure 13 and Figure 14). Many of these bonds were DTT sensitive, indicating that they were cysteine disulphide bridges, however some were resistant to DTT, suggesting there may be a different interaction. BfUbb lacks the C-terminal double glycine (GG) motif (Patrick *et al.*, 2011), which is considered to be essential for ubiquitin activation and conjugation of ubiquitin to proteins (Grabbe & Dikic, 2009; Wilkinson & Audhya, 1981). Ubiquitin homologues which lack this motif, and which are consequently considered incapable of covalent conjugation via ubiquitylation, are referred to as ubiquitin domain

proteins (UDPs) (Luders *et al.*, 2003). Most bacterial homologues of ubiquitin lack the GG motif and could be considered UDPs. A BLASTP search using human ubiquitin against the bacterial taxid database identifies ~100 proteins with significant homology. Some of these are ULDs within larger proteins, whilst at least one, a *Burkholderia gladioli* protein, is an almost identical copy of ubiquitin (96% identity across 100% of query, e value of 10^{-44}), and includes the GG motif. The rest are largely UDPs, usually trunctuated at the C-terminus and with N-terminal extensions. Several of these can be found in bacterial species known to associate closely with eukaryotes, including the resident rumen commensal *Ruminococcus albus* (88% identity across 98% of query, e value of 10^{-38}) and the plant pathogen *Pectobacterium carotovorum* (94% identity across 94% of the query, e value of 10^{-43}). It is possible that these UDPs are involved in bacterial-host interactions, however none have yet been assayed for their ability to interact with the ubiquitin cascade or conjugate to other proteins. Similarly, small proteins which are cross-reactive with antibodies against eukaryotic ubiquitin have been identified in Archaea such as *Natronococcus occultus* (Nercessian *et al.*, 2002), but have not been characterised. A UBL identified in *Anabaena variabilis* was found to cross-react with ubiquitin antibody and was able to conjugate to various proteins when incubated with *A. variabilis* cell extract, suggesting conjugative machinery exists within this species. However, only the N-terminal sequence of this protein was confirmed and conjugation could not be replicated (Durner & Boger, 1995). Furthermore, no proteins with significant homology to ubiquitin can be identified in *Anabaena* sp. by BLASTP search, so it is unclear whether *A. variabilis* expresses a true conjugative bacterial UBL. A bacterial UDP, YukD, found in *Bacillus subtilis* has a strikingly similar 3D structure to eukaryotic ubiquitin (root mean square deviation of 2.2Å) but has a 6aa C-terminal trunctuation, meaning it lacks the GG motif. YukD could not conjugate to any *B. subtilis* proteins when incubated with *B. subtilis* cell extract (van den Ent & Lowe, 2005), however it has never been incubated with eukaryotic cell extract and may be able to interfere with the ubiquitin cascade as BfUbb does (Patrick *et al.*, 2011).

There is only one well documented case of a UDP acting as a protein modifier. Hub1 (UBL5), a eukaryotic UDP which is highly conserved from yeast to mammals, shares 22% identity with ubiquitin (at the lower limit of accepted homology) and has a different pattern of conserved lysines, however it forms the same β-grasp fold. It is 4aa shorter than ubiquitin at the C-terminus and ends in a double tyrosine (YY) motif followed by a non-conserved amino acid. Multiple studies have shown that Hub1 is able to bind proteins in an ATP-independent and SDS-resistant manner (Dittmar *et al.*, 2002; Luders *et al.*, 2003). It was originally thought that the YY motif was required for covalent conjugation, however this was shown not to be the case. The bonds were also not disulphide bridges formed between cysteine residues as they were resistant to boiling in the presence of DTT, and it is now thought that Hub1 binds proteins, such as splicesosome component Snu66, in a non-covalent but SDS-resistant manner (Luders *et al.*, 2003; van der Veen & Ploegh, 2012; Yashiroda & Tanaka, 2004). The ability of Hub1 to form attachments non-covalently presents a case for UDPs to bypass ubiquitin cascade-like machinery in order to affect protein function. This, rather than covalent binding via its C-terminal cysteine, may be how BfUbb interacts with host cell proteins.

6.3 Potential roles of BfUbb in the host eukaryotic cell

Although the exact mechanism of interaction between BfUbb and the E1 and E2 enzymes remains unknown, BfUbb may have a role in disrupting the ubiquitin cascade in the mammalian epithelial cells. Since BfUbb is known to inhibit ubiquitylation of lysozyme (Patrick *et al.*, 2011) (Figure 9 B) its most likely role is as a suicide substrate, binding one or more of the enzymes of the ubiquitin cascade to prevent ubiquitin binding and transfer. Ubiquitin vinyl methylester (UbVME), a suicide substrate of deubiquitinating enzymes (DUBs), mimics ubiquitin but forms a thioether linkage rather than a thioester linkage with the active site, leaving the DUB enzymatically inert (Misaghi *et al.*, 2005). The results of this study suggest the formation of disulphide bridges between BfUbb and the E1 or E2 enzymes

(Figure 14), possibly between the C-terminal cysteine of BfUbb and the active site cysteine of the enzyme. This would, similar to UbVME, prevent further transfer of BfUbb to the E2 and any eventual substrate, since this transfer requires the presence of thioester bonds (Dye & Schulman, 2007).

In the ubiquitin transfer experiments performed in this study, BfUbb seemed to demonstrate a preference for Ube2M, Ube2R1 and Ube2R2 (Figure 14), all of which are involved in cell cycle progression (Ceccarelli *et al.*, 2011; Cukras *et al.*, 2014; Semplici *et al.*, 2002). Exactly why BfUbb shows a preference for these enzymes is unknown. It may be that the DUBs of these three proteins bind more tightly to BfUbb than the DUBs of other E2 enzymes. Non-covalent interactions via the hydrophobic patch surrounding isoleucine 44 on the surface of ubiquitin play an important role in the stabilisation of ubiquitin within the E2 enzyme active site (Wickliffe *et al.*, 2011) and the non-covalent interactions of UBLs can be strong enough to resist separation by SDS-PAGE (Luders *et al.*, 2003). Perhaps, then, the strength of the non-covalent interactions between BfUbb and these E2 enzymes can explain the resistance of these bonds to reducing environments (Figure 14).

E2 enzymes involved in cell cycle progression are targets for inhibition in the development of anti-cancer drugs. Unregulated mitosis is a common feature in tumorigenesis, and often cells become cancerous as a result of uncontrolled degradation of one or more checkpoint proteins. The inhibition of the E2 enzymes responsible for the degradation of these checkpoint proteins should prevent mitotic progression and protect against tumorigenesis (Ceccarelli *et al.*, 2011; Harper & King, 2011). The presence in the epithelial cell of a ubiquitin-related suicide substrate which preferentially binds such E2 enzymes could therefore protect the host from colonic cancer. Three samples containing *B. fragilis* from a single patient with colonic cancer were screened for the presence of the *ubb* gene by qPCR. The gene could not be detected in any of these samples (Table 5), however the results of this study are inconclusive due to the small sample size, and a larger scale screening could

determine whether there is a negative correlation between presence of BfUbb-expressing *B. fragilis* in the GI tract and the development of colonic cancer. The prevention of chronic disease in the host is believed to be beneficial to commensal and mutualistic GI bacteria (Huang *et al.*, 2011b; Mazmanian *et al.*, 2008; Wexler, 2007) and the composition of the gut microbiome is believed to affect the rate and severity of tumorigenesis in mice with colon cancer (Zackular *et al.*, 2013). It is therefore not unlikely that a common mutualistic bacterium of the GI tract could express a protein that protects against the development of cancer in epithelial cells.

One of the key differences between ubiquitin and BfUbb is that BfUbb lacks K11 (Patrick *et al.*, 2011), one of the highly conserved lysines involved in chain formation. K11 linkages are important for inflammatory pathways (Corn & Vucic, 2014) and essential for mitotic progression (Jin *et al.*, 2008). Furthermore, pre-incubation of BfUbb with E1 activating enzyme seemed to increase the number of bonds formed between BfUbb and most E2 enzymes (Figure 14). It is therefore tempting to suggest a model in which BfUbb is passed through the ubiquitin cascade and is bound to substrates to prevent the extension of K11-specific chains, thus inhibiting inflammation and/or cell cycle progression. However, the only mechanism by which BfUbb could be transferred from one enzyme to another is thiol-disulphide exchange. This would require the deprotonation of the thiol group of the active site cysteine in the E2 enzyme so that it could displace the existing disulphide bond between BfUbb and the E1 active site cysteine (Sevier & Kaiser, 2002). This could only occur in an oxidising environment, which the cytoplasm of the eukaryotic cell is not, and this is therefore unlikely to be occurring *in vivo*.

The theory that BfUbb acts as a suicide substrate by binding the active site cysteine of an E1 or E2 enzyme via a disulphide bridge is problematic considering that the cytoplasm is a reducing environment. In order for these covalent bonds to occur, BfUbb and the target enzyme or substrate would need to coexist in a non-reducing compartment of the cell.

Additionally, in order to outcompete ubiquitin, BfUbb would need to exist at a relatively high local concentration. Given that the OMV are the mechanism by which BfUbb may be delivered to the host cell, the endosome is a likely compartment in which BfUbb could function. Multiple studies have shown that OMV can be internalised by mammalian cells via endocytosis (Bomberger *et al.*, 2009; Chatterjee & Chaudhuri, 2011; Vidakovics *et al.*, 2010), suggesting that their contents will be contained within an endosome and possibly directed to the lysosome (Tanno & Komada, 2013). To escape degradation in the lysosome, bacterial proteins must lyse the endosome or be redirected to another part of the cell (Bomberger *et al.*, 2009). Ubiquitin and UBDs play an important role in the determining the fate of endocytosed proteins, both during normal protein turnover and during infection (Fujita *et al.*, 2013; Tanno & Komada, 2013). The endosome would provide a non-reducing compartment (Austin *et al.*, 2005) in which BfUbb could exist at a high enough concentration to interfere with the ubiquitin system. BfUbb may therefore bind one or more UBDs to prevent trafficking and destruction of other OMV proteins. Alternatively, it may conjugate to either host or bacterial endocytosed proteins in the the endosome and target these proteins for destruction or recycling back to the plasma membrane. For direct binding to UBDs the E1, E2 and E3 enzymes may not be required, and indeed it is worth noting that the binding of BfUbb to E1 and E2 enzymes observed in this study may be coincidental to its actual function. Much more work needs to be done to determine if and how BfUbb interacts with the ubiquitin cascade, or any other ubiquitin-binding proteins, *in vivo*.

There appears to be an alternative role for BfUbb outside of the host epithelial cell. OMV carrying BfUbb were found to delay the growth of *S. enterica* Typhmurium, significantly more so than OMV from a *B. fragilis* Δ*ubb* mutant or a PBS control (Figure 28). This suggests that BfUbb has antibacterial activity, which is consistent with the observed bacteriostatic activity of eukaryotic ubiquitin and other UBLs, the exact mechanism of which is unknown (Alonso *et al.*, 2007; De Ingeniis *et al.*, 2009; Howell *et al.*, 2003; Kieffer *et al.*, 2003; Seo *et al.*, 2013). The production of antimicrobials is an almost universal trait within bacteria, and is an aspect

of "bacterial interference", in which one organism interferes with the biological functioning of another, usually to the benefit of the former (Ji *et al.*, 1997). An antibiotic-producing bacterial species which has already colonised an environment can prevent invasion by a new, sensitive species. Conversely, an antibiotic-producing bacterial species which reaches a new environment can reduce the growth of existing, sensitive species, thus allowing colonisation. If BfUbb is as broad-spectrum an inhibitor as other UBLs (Howell *et al.*, 2003; Seo *et al.*, 2013), BfUbb-producing *B. fragilis* could prevent colonisation of the mucosa by bacterial species which would not only compete with *B. fragilis* for space and nutrients, but which may cause serious disease in the host (Berg, 1996). Additionally, the antimicrobial activity of multiple eukaryotic UBLs and BfUbb suggests that inhibition of bacterial growth may be one of the roles, if not the primary role, of ubiquitin homologues found in other bacteria.

6.4 Characterisation of outer membrane vesicles

The past few years have seen a significant increase in the number of studies which aim to characterise OMV of various Gram-negative bacteria using proteomics (Choi *et al.*, 2015). This requires both the efficient isolation of OMV and an appropriate method for proteomic analysis. Although multiple authors agree that density centrifugation is an essential step in OMV isolation (Bauman & Kuehn, 2006; Kalra *et al.*, 2013; Kreimer *et al.*, 2015; Lee *et al.*, 2008), a wide range of methods are used by different groups (Figure 20) and it is rare for two studies to use the same combination of techniques. This is problematic since there is evidence that the method of isolation used can significantly affect the proteomic content of the OMV (Choi *et al.*, 2015). Most studies favour the extraction of naturally formed OMV, rather than chemically-induced OMV (Figure 20), since the latter are not considered representative of the former (Kulp & Kuehn, 2010). However, the specific isolation techniques used may also affect the content of the vesicles (Tauro *et al.*, 2012). It is

therefore difficult to say which method will isolate OMV that are most representative of those produced by the cell during undisturbed growth. Whichever combination of techniques is eventually used; however, it should be a standard for all proteomic studies of OMV, otherwise the different proteomes cannot be reliably compared. The best method of proteomic analysis of OMV is also still being debated (Figure 21). When taking into consideration the benefits and drawbacks of the approaches discussed in section 4.1.2, the best method would arguably be gel-free (Lee *et al.*, 2008; van de Waterbeemd *et al.*, 2013), using a tagging method for protein quantification (Cooper *et al.*, 2010; Kreimer *et al.*, 2015; Li *et al.*, 2012), and an ultra-low flow fractionation method prior to high-sensitivity MS (Kreimer *et al.*, 2015). Of course, time, cost and availability of resources will be limiting factors for most studies, and proteomics approaches in OMV will most likely continue to vary.

Characterisation of OMV often includes determination of OMV size, either by TEM or DLS. In this study a Zetasizer APS (Malvern Instruments, UK) was used determine the average size of OMV in the concentrated supernatant samples (Figure 22). This technique suggested an average diameter of ~250nm, whereas TEM had previously indicated that OMV of *B. fragilis* were ~50-100nm (Patrick *et al.*, 1996). When the two methods have been used in parallel in other studies they have been in agreement (Choi *et al.*, 2011; Jang *et al.*, 2014). In this study, aspiration through a needle during sample collection by the Zetasizer APS may have perturbed the capsular polysaccharide of the OMV. Whole cells of mutants lacking CPS aggregate in solution (Patrick *et al.*, 2009) and aggregation of OMV would explain the larger average diameter and the inconsistency between readings. Certain other types of DLS apparatus analyse samples in cuvettes rather than aspirating the samples into a chamber, and this is probably more appropriate for analysis of OMV. Whilst DLS is a cheap, fast alternative to TEM in determining the size of the OMV it cannot provide information on the nature of the OMV, e.g. whether they contain one or two lipid bilayers (Perez-Cruz *et al.*, 2013) and, in the case of *B. fragilis* whether they are derived from LC-, SC- or MC-

expressing cells (Patrick *et al.*, 1996). It is therefore likely that DLS will continued to be used in parallel with, rather than instead of, TEM.

Despite the continued characterisation of OMV from various species, much remains unknown about their formation and function. Proteomics studies, including this one, and mutational analysis have not yet identified a common protein or set of proteins involved in biogenesis (Kulkarni & Jagannadham, 2014). The lack, thus far, of a clear active mechanism for OMV production could be considered as support for the passive biogenesis theory, in which OMV are produced as a side effect of OM turnover and other regular cell processes (Schwechheimer *et al.*, 2013). However, multiple studies, including this one (Table 8, Figure 24), have reported the enrichment of proteins in OMV samples compared to OM, PP or whole cell samples (Elhenawy *et al.*, 2014; Haurat *et al.*, 2011; Kato *et al.*, 2002; van de Waterbeemd *et al.*, 2013), which supports the theory of an active sorting mechanism. Further studies aimed at characterisation of the OMV of Gram-negative bacteria are required to identify a mechanism for sorting and biogenesis (Choi *et al.*, 2015; Lee *et al.*, 2008). The presence of cytoplasmic proteins in OMV is also a contested issue, with some studies suggesting that efficiently isolated vesicles should not contain cytoplasmic proteins (Kulp & Kuehn, 2010; Scorza *et al.*, 2008), whilst others provide precedence for their inclusion (Pancholi & Chhatwal, 2003; Perez-Cruz *et al.*, 2013). As discussed in section 4.3.3, this study identified a large number of cytoplasmic proteins, suggesting a certain amount of contamination in both the OMV and PP samples. However, it is impossible to determine from individual proteomics studies which, if any, of these proteins occur naturally in *B. fragilis* OMV and which are contaminants. This problem could also potentially be resolved by the development of a standard, high purity procedure for isolation of OMV.

Proteomics has proved to be a powerful tool in biological research, allowing the rapid, large scale identification and putative characterisation of proteins from any cellular or subcellular fraction of any species. There are limitations, however, with possibly the greatest being that a large proportion of proteins in the available databases are uncharacterised (hypothetical).

More than 40% of proteins enriched in *B. fragilis* OMV showed homology only to hypothetical proteins from other bacterial species (Figure 24 B) and previous studies also found that a similar proportion of the OMV proteins of other species were poorly characterised (Altindis *et al.*, 2014; Bauman & Kuehn, 2006; Berlanda Scorza *et al.*, 2012; Elhenawy *et al.*, 2014; Lee *et al.*, 2007; Mullaney *et al.*, 2009; Nally *et al.*, 2005). One or more of these uncharacterised proteins may be conserved between OMV from different species and may even play a role in OMV biogenesis or function, however genetic and biochemical analysis, and the construction of an OMV proteome database, would be required to determine this. Furthermore, even when function can be strongly inferred from sequence identity to known proteins, classical techniques are often required to confirm this function. This study, and others like it, demonstrate that until the function of the majority of "hypothetical" proteins has been determined, the power of proteomics as a tool for the characterisation of OMV, and other biological samples, will be limited.

6.5 Potential roles of *B. fragilis* OMV

It is likely that the OMV of any given species have multiple roles which support the survival and propagation of that species in its environment (Kulkarni & Jagannadham, 2014). *B. fragilis* OMV are already considered to have a role in suppression of the inflammatory response via the delivery of PSA, a function which benefits both bacterium and host by maintaining a stable environment in the GI tract (Shen *et al.*, 2012). They also bind fibrinogen (Houston *et al.*, 2010) and are known to promote haemagglutination of erythrocytes (Patrick *et al.*, 1996), and may therefore play a role in opportunistic pathogenesis. Since the natural reservoir of *B. fragilis* is the GI tract the primary role(s) of *B. fragilis* OMV is most likely to promote survival in this environment rather than to contribute to extra-abdominal infection. This study suggests several additional putative roles of *B. fragilis* OMV.

Some of the most abundant and most enriched proteins in *B. fragilis* OMV in this study are involved in the binding and degradation of common nutrients in GI tract bacteria (Figure 24*,* Table 7, Table 8, Appendix VII). For example, several SusC and SusD proteins, which are sufficient for starch binding in *B. thetaiotaomicron* (Shipman *et al.*, 2000) were identified, as well as a potential SusF, the exact function of which is unknown. SusG, which is responsible for starch degradation (Shipman *et al.*, 2000), was not identified, however. Elhenawy *et al* (2014) found that several hydrolytic enzymes were enriched in the OMV of *B. fragilis*, including xylanase, which plays a role in degradation of hemicellulose (Juturu & Wu, 2012), α-L-fucosidase, which is involved in the processing of complex sugars (Wright *et al.*, 2013), a zinc protease and a dipeptidase. All of these enzymes were identified in this study, as well as several other extra-cytoplasmic proteases (Table 9). OMV capable of binding and degrading various nutrients present in the GI tract could allow *B. fragilis* to remotely process these nutrients, increasing the extracellular concentration of smaller molecules that can be utilised by either the host or the bacterium.

Amongst the most abundant of the 783 proteins identified in *B. fragilis* OMV were an unexpectedly high number of cytoplasmic, metabolic "housekeeping" enzymes, including GAPDH, alcohol dehydrogenase and succinyl coenzyme A synthetase (Table 7, Appendix VII). Despite lacking a signal sequence, these enzymes have been found on the surface of several bacterial pathogens, where they appear to act primarily as adhesins (Pancholi & Chhatwal, 2003). The presence of multiple adhesive proteins on the surface of both whole cells and OMV may serve to promote and maintain colonisation of the gut mucosa by *B. fragilis*, as well as potentially improving binding of OMV to epithelial cells.

The results of this study suggest that the OMV of *B. fragilis* carry a surface-exposed α2m, a macromolecular protease inhibitor found in both eukaryotes and prokaryotes (Kantyka *et al.*, 2010). Depending on its protease specificity, which is determined by the bait region

(Sottrupjensen *et al.*, 1989), this α2m has the potential to trap and inhibit a number of proteases. Periplasmic bacterial α2ms are thought to protect the cell against invading proteases during a breach of the OM (Kantyka *et al.*, 2010; Robert-Genthon *et al.*, 2013). This may also be the primary function of *B. fragilis* α2m, however if it is present on the surface of OMV this particular α2m could also inhibit one of the many proteases in the extracellular medium of the GI tract. For example, it may prevent activation of antimicrobial peptides by host proteases (Antalis *et al.*, 2007), thus protecting itself from the host innate immune system. Alternatively, it may inhibit one of the many host and bacterial proteases involved in promoting the inflammatory response (Antalis *et al.*, 2007), thus protecting the host from inflammation and maintaining its own environment.

B. fragilis OMV are believed to carry BfUbb (Patrick *et al.*, 2011) (Figure 26), which has the potential to interfere with the eukaryotic ubiquitylation pathway (section 3.3.3) and appears to demonstrate antibacterial activity against *S. enterica* Typhimurium (section 5.3.1). *B. fragilis* OMV can be engulfed by DCs (Shen *et al.*, 2012), whilst OMV of other species can be engulfed by epithelial cells (Chatterjee & Chaudhuri, 2011; Kesty & Kuehn, 2004) and are able to fuse with other bacteria (Kadurugamuwa & Beveridge, 1996), thus the OMV of *B. fragilis* provide a mechanism by which BfUbb can be delivered to both eukaryotic and bacterial cells. Despite carrying a lytic murein transglycosylase (Appendix VII), an autolysin which could cause bacteriolysis when deregulated (Li *et al.*, 1996), *B. fragilis* OMV do not appear to demonstrate bacteriolytic activity against *S. enterica* Typhimurium (Figure 28 D). They may have a lytic effect on Gram-positive or other Gram-negative bacteria, however. The combined results of this study suggest that the OMV of *B. fragilis* have the potential to contribute to bacterial survival and host health in multiple ways.

6.6 Biotechnological potential of BfUbb and the OMV of *B. fragilis*

The function or functions of BfUbb, as indicated by the results of this study, could be adapted for treatment or prevention of disease in the GI tract. E2 conjugating enzymes involved in the ubiquitin-dependent control of mitotic progression are established targets for anti-cancer therapies (Harper & King, 2011). If it can be determined that BfUbb is able to inhibit these specific E2 enzymes, it could be developed as a drug for the prevention or treatment of tumorigenesis. Additionally, depending on the spectrum and effectivity of BfUbb as an antibacterial protein, it could be considered a novel antibiotic in the treatment of gastro-intestinal bacterial infections. The use of OMV as vaccines is of ongoing interest in biotechnology as they represent a non-pathogenic alternative to using whole cells for delivery of antigens (Altindis *et al.*, 2014; Kolodziejek *et al.*, 2013; van de Waterbeemd *et al.*, 2013). Using *B. fragilis* OMV as a vector for antigen delivery would probably promote an immune reaction to *B. fragilis* itself, which would be detrimental to the health of the vaccinated individual colonised by the bacterium. These OMV may therefore be put to best use as a delivery mechanism for antibiotics and drugs targeted to the GI tract, such as BfUbb itself. In fact, if BfUbb acts via one or more of the mechanisms proposed in this study, deliberate colonisation of the GI tract with BfUbb-expressing *B. fragilis* may prove effective in preventing the development of colon cancer or infection by certain GI pathogens.

7. Conclusions

This study aimed to characterise both the bacterial ubiquitin homologue, BfUbb, and the outer membrane vesicles of *B. fragilis*. BfUbb was found to multimerise spontaneously *in vitro*, most likely via the formation of disulphide bridges between cysteine residues in the C-terminal tail. It was also found to form covalent bonds with multiple E2 conjugating enzymes of the eukaryotic ubiquitylation pathway. These bonds were DTT-sensitive, and therefore also likely to be disulphide bridges, however pre-incubation of BfUbb with E1 activating enzyme promoted the formation of DTT-resistant bonds between BfUbb and E2 enzymes Ube2M, Ube2R1 and Ube2R2 via an unknown mechanism. These three E2 enzymes are involved in ubiquitin-dependent progression of cell cycle checkpoints (Ceccarelli *et al.*, 2011; Cukras *et al.*, 2014; Semplici *et al.*, 2002) and as such are targets for anti-cancer therapies (Harper & King, 2011). Covalent binding of BfUbb to the active site of these enzymes *in vivo* could inhibit cell cycle progression, provided BfUbb were able to accumulate and function within the host cells. There is no evidence for this thus far, however. Nickel affinity pull-down of his-tagged BfUbb incubated with epithelial cell extract did not identify any binding partners. So far the only proteins known to form covalent bonds with BfUbb are enzymes of the ubiquitylation cascade.

Two plasmids were constructed which allowed the expression and purification of a native form of BfUbb, for use in X-ray crystallography, and a YFP-tagged form of BfUbb, for use in *in vitro* assays. Expression from both plasmids was demonstrated in *E. coli*, however the proteins have not yet been purified and utilised. Three additional plasmids were constructed which allow for the expression of his-tagged BfUbb, YFP-tagged BfUbb and a YFP-control in mammalian cells. This would allow observation of the effects of the presence of BfUbb in epithelial cells, as well as the subcellular localisation, if any, of the *B. fragilis* protein. These plasmids have not yet been successfully established in Caco-2 cells, however.

qPCR was used to screen for the presence of *B. fragilis* and the *ubb* gene in clinical samples from patients with abscesses or gastro-intestinal diseases. Due to the small scale of the study no correlation could be established between the presence or absence of BfUbb and the development of any particular disease. The exact role, if any, of BfUbb in the host epithelial cells, and its phenotypic effect on the host, has yet to be determined.

The OMV of *B. fragilis* were isolated from culture supernatant using a combination of techniques that should have ensured high sample purity. Samples of OMV and periplasmic extract were analysed via LTQ-Orbitrap mass spectrometry and a proteome of the OMV was generated. Within this proteome were proteins that suggested a role for the OMV in nutrient acquisition, proteolysis, protease inhibition and adhesion to epithelial cells. A previously characterised fibrinogen-binding protein was identified but with low reliability, and BfUbb was not identified at all. Both proteins could be detected by Western blotting in concentrated supernatants of WT *B. fragilis*, however, suggesting that the proteomic techniques used in this study may have a limited capacity for accurate protein identification. Moreover, the purity of the sample, and consequently the efficiency of the OMV isolation procedure, was called into question due to the large number of cytoplasmic proteins identified, many of which are likely to be contaminants. There is a clear need for a standardised, efficient isolation procedure in the proteomic analysis of OMV from various bacterial species (Choi *et al.*, 2015). The mechanism of OMV biogenesis in *B. fragilis*, and indeed in all Gram-negative bacteria, has yet to be elucidated.

The antibiotic potential of *B. fragilis* OMV and BfUbb was tested by incubating inoculates of *S. enterica* Typhimurium with concentrated supernatants of WT and Δ*ubb B. fragilis*. BfUbb-containing OMV were found to inhibit growth of *S. enterica* Typhimurium, significantly more so than both Δ*ubb* OMV and a PBS control, suggesting that BfUbb has some antibacterial activity. The secretion of antimicrobials is a feature of most characterised bacterial species (Ji *et al.*, 1997) and the antibiotic activity of ubiquitin and ubiquitin-like proteins has been previously established (Alonso *et al.*, 2007; De Ingeniis *et al.*, 2009; Howell *et al.*, 2003; Kieffer *et al.*, 2003; Seo *et al.*, 2013). BfUbb may provide *B. fragilis* with the ability to inhibit colonisation of invading bacteria in the gastro-intestinal mucosa.

The results of this study suggest a variety of roles for both BfUbb and the OMV of *B. fragilis*. Interestingly, many of these putative functions would serve to benefit the host as well as the bacterium. *B. fragilis* is usually described as a commensal of the GI tract, i.e. an organism that benefits from its host without significantly effecting it (Wexler, 2007). However, given the accumulation of evidence suggesting that *B. fragilis* provides significant health benefits for its host, it seems prudent that this host-bacterial relationship should instead be described as mutualism (Backhed *et al.*, 2005).

References

Aguilera, L., Toloza, L., Gimenez, R., Odena, A., Oliveira, E., Aguilar, J., Badia, J. & Baldoma, L. (2014). Proteomic analysis of outer membrane vesicles from the probiotic strain Escherichia coli Nissle 1917. *Proteomics* **14**, 222-229.

Alonso, S., Pethe, K., Russell, D. G. & Purdy, G. E. (2007). Lysosomal killing of Mycobacterium mediated by ubiquitin-derived peptides is enhanced by autophagy. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 6031-6036.

Altindis, E., Fu, Y. & Mekalanos, J. J. (2014). Proteomic analysis of Vibrio cholerae outer membrane vesicles. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E1548-E1556.

Antalis, T. M., Shea-Donohue, T., Vogel, S. N., Sears, C. & Fasano, A. (2007). Mechanisms of disease: protease functions in intestinal mucosal pathobiology. *Nature Clinical Practice Gastroenterology & Hepatology* **4**, 393-402.

Antoni, L., Nuding, S., Weller, D., Gersemann, M., Ott, G., Wehkamp, J. & Stange, E. F. (2013). Human colonic mucus is a reservoir for antimicrobial peptides. *Journal of Crohn's & colitis* **7**, e652-664.

Armstrong, P. B. & Quigley, J. P. (1999). alpha(2)-macroglobulin: an evolutionarily conserved arm of the innate immune system. *Developmental and Comparative Immunology* **23**, 375-390.

Ashida, H., Nakano, H. & Sasakawa, C. (2013). Shigella IpaH0722 E3 Ubiquitin Ligase Effector Targets TRAF2 to Inhibit PKC-NF-kappa B Activity in Invaded Epithelial Cells. *Plos Pathogens* **9**.

Ashida, H., Kim, M. & Sasakawa, C. (2014). Exploitation of the host ubiquitin system by human bacterial pathogens. *Nature Reviews Microbiology* **12**, 399-413.

Austin, C. D., Wen, X. H., Gazzard, L., Nelson, C., Scheller, R. H. & Scales, S. J. (2005). Oxidizing potential of endosomes and lysosomes limits intracellular cleavage of disulfide-based antibody-drug conjugates. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 17987-17992.

Austinprather, S. L. & Booth, S. J. (1984). EVIDENCE FOR A MEMBRANE-BOUND FORM OF A BACTERIOCIN OF BACTEROIDES-UNIFORMIS TL-1. *Canadian Journal of Microbiology* **30**, 268-272.

Avelar, K. E. S., Pinto, L. J. F., Antunes, L. C. M., Lobo, L. A., Bastos, M. C. F., Domingues, R. & Ferreira, M. C. D. (1999). Production of bacteriocin by Bacteroides fragilis and partial characterization. *Letters in Applied Microbiology* **29**, 264-268.

Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. (2005). Host-bacterial mutualism in the human intestine. *Science* **307**, 1915-1920.

Bade, V. N., Nickels, J., Keusekotten, K. & Praefcke, G. J. K. (2012). Covalent Protein Modification with ISG15 via a Conserved Cysteine in the Hinge Region. *Plos One* **7**.

Bauman, S. J. & Kuehn, M. J. (2006). Purification of outer membrane vesicles from Pseudomonas aeruginosa and their activation of an IL-8 response. *Microbes and Infection* **8**, 2400-2408.

Berg, R. D. (1996). The indigenous gastrointestinal microflora. *Trends in Microbiology* **4**, 430-435.

Berlanda Scorza, F., Colucci, A. M., Maggiore, L. & other authors (2012). High yield production process for Shigella outer membrane particles. *PloS one* **7**, e35616-e35616.

Berleman, J. E., Allen, S., Danielewicz, M. A. & other authors (2014). The lethal cargo of Myxococcus xanthus outer membrane vesicles. *Frontiers in Microbiology* **5**.

Berndsen, C. E. & Wolberger, C. (2011). A spectrophotometric assay for conjugation of ubiquitin and ubiquitin-like proteins. *Analytical Biochemistry* **418**, 102-110.

Bishop, D. G. & Work, E. (1965). AN EXTRACELLULAR GLYCOLIPID PRODUCED BY ESCHERICHIA COLI GROWN UNDER LYSINE-LIMITING CONDITIONS. *Biochemical Journal* **96**, 567-&.

Bomberger, J. M., MacEachran, D. P., Coutermarsh, B. A., Ye, S. Y., O'Toole, G. A. & Stanton, B. A. (2009). Long-Distance Delivery of Bacterial Virulence Factors by Pseudomonas aeruginosa Outer Membrane Vesicles. *Plos Pathogens* **5**.

Budd, A., Blandin, S., Levashina, E. A. & Gibson, T. J. (2004). Bacterial alpha(2)-macroglobulins: colonization factors acquired by horizontal gene transfer from the metazoan genome? *Genome Biology* **5**.

Ceccarelli, D. F., Tang, X., Pelletier, B. & other authors (2011). An Allosteric Inhibitor of the Human Cdc34 Ubiquitin-Conjugating Enzyme. *Cell* **145**, 1075-1087.

Cerdeno-Tarraga, A. M., Patrick, S., Crossman, L. C. & other authors (2005). Extensive DNA inversions in the B. fragilis genome control variable gene expression. *Science* **307**, 1463-1465.

Chatterj.Sn & Das, J. (1967). ELECTRON MICROSCOPIC OBSERVATIONS ON EXCRETION OF CELL-WALL MATERIALS BY VIBRIO CHOLERAE. *Journal of General Microbiology* **49**, 1-&.

Chatterjee, D. & Chaudhuri, K. (2011). Association of cholera toxin with Vibrio cholerae outer membrane vesicles which are internalized by human intestinal epithelial cells. *Febs Letters* **585**, 1357-1362.

Chen, Y. Z., McPhie, D. L., Hirschberg, J. & Neve, R. L. (2000). The amyloid precursor protein-binding protein APP-BP1 drives the cell cycle through the S-M checkpoint and causes apoptosis in neurons. *Journal of Biological Chemistry* **275**, 8929-8935.

Choi, D.-S., Kim, D.-K., Kim, Y.-K. & Gho, Y. S. (2015). Proteomics of extracellular vesicles: Exosomes and ectosomes. *Mass Spectrometry Reviews* **34**, 474-490.

Choi, D. S., Kim, D. K., Choi, S. J. & other authors (2011). Proteomic analysis of outer membrane vesicles derived from Pseudomonas aeruginosa. *Proteomics* **11**, 3424-3429.

Cooper, B., Feng, J. A. & Garrett, W. M. (2010). Relative, Label-free Protein Quantitation: Spectral Counting Error Statistics from Nine Replicate MudPIT Samples. *Journal of the American Society for Mass Spectrometry* **21**, 1534-1546.

Corn, J. E. & Vucic, D. (2014). Ubiquitin in inflammation: the right linkage makes all the difference. *Nature Structural & Molecular Biology* **21**, 297-300.

Costa, T. R. D., Felisberto-Rodrigues, C., Meir, A., Prevost, M. S., Redzej, A., Trokter, M. & Waksman, G. (2015). Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nature Reviews Microbiology* **13**, 343-359.

Cukras, S., Morffy, N., Ohn, T. & Kee, Y. (2014). Inactivating UBE2M Impacts the DNA Damage Response and Genome Integrity Involving Multiple Cullin Ligases. *Plos One* **9**.

De Ingeniis, J., Raffaelli, N., Ciani, M. & Mannazzu, I. (2009). Pichia anomala DBVPG 3003 Secretes a Ubiquitin-Like Protein That Has Antimicrobial Activity. *Applied and Environmental Microbiology* **75**, 1129-1134.

De, S. N. (1959). ENTEROTOXICITY OF BACTERIA-FREE CULTURE-FILTRATE OF VIBRIO-CHOLERAE. *Nature* **183**, 1533-1534.

Dittmar, G. A. G., Wilkinson, C. R. M., Jedrzejewski, P. T. & Finley, D. (2002). Role of a ubiquitin-like modification in polarized morphogenesis. *Science* **295**, 2442-2446.

Domingues, R., Souza, W., Moraes, S. R., Avelar, K. E. S., Hirata, R., Fonseca, M. E. F. & Ferreira, M. C. D. (1997). Surface vesicles: A possible function in commensal relations of Bacteroides fragilis. *Zentralblatt Fur Bakteriologie-International Journal of Medical Microbiology Virology Parasitology and Infectious Diseases* **285**, 509-517.

Durner, J. & Boger, P. (1995). UBIQUITIN IN THE PROKARYOTE ANABAENA-VARIABILIS. *Journal of Biological Chemistry* **270**, 3720-3725.

Dye, B. T. & Schulman, B. A. (2007). Structural mechanisms underlying posttranslational modification by ubiquitin-like proteins. *Annual Review of Biophysics and Biomolecular Structure* **36**, 131-150.

Eitel, Z., Soki, J., Urban, E., Nagy, E. & Anaerobic, E. S. G. (2013). The prevalence of antibiotic resistance genes in Bacteroides fragilis group strains isolated in different European countries. *Anaerobe* **21**, 43-49.

Elhenawy, W., Debelyy, M. O. & Feldman, M. F. (2014). Preferential Packing of Acidic Glycosidases and Proteases into Bacteroides Outer Membrane Vesicles. *Mbio* **5**.

Espinosa-Urgel, M. & Kolter, R. (1998). Escherichia coli genes expressed preferentially in an aquatic environment. *Molecular Microbiology* **28**, 325-332.

Finley, D. (2009). Recognition and Processing of Ubiquitin-Protein Conjugates by the Proteasome. *Annual Review of Biochemistry* **78**, 477-513.

Fiocca, R., Necchi, V., Sommi, P., Ricci, V., Telford, J., Cover, T. L. & Solcia, E. (1999). Release of Helicobacter pylori vacuolating cytotoxin by both a specific secretion pathway and budding of outer membrane vesicles. Uptake of released toxin and vesicles by gastric epithelium. *Journal of Pathology* **188**, 220-226.

Fletcher, C. M., Coyne, M. J., Villa, O. F., Chatzidaki-Livanis, M. & Comstock, L. E. (2009). A General O-Glycosylation System Important to the Physiology of a Major Human Intestinal Symbiont. *Cell* **137**, 321-331.

Fletcher, C. M., Coyne, M. J. & Comstock, L. E. (2011). Theoretical and Experimental Characterization of the Scope of Protein O-Glycosylation in Bacteroides fragilis. *Journal of Biological Chemistry* **286**, 3219-3226.

Fujimura, K. E., Slusher, N. A., Cabana, M. D. & Lynch, S. V. (2010). Role of the gut microbiota in defining human health. *Expert Review of Anti-Infective Therapy* **8**, 435-454.

Fujita, N., Morita, E., Itoh, T. & other authors (2013). Recruitment of the autophagic machinery to endosomes during infection is mediated by ubiquitin. *Journal of Cell Biology* **203**, 115-128.

Gibson, F. C., Onderdonk, A. B., Kasper, D. L. & Tzianabos, A. O. (1998). Cellular mechanism of intraabdominal abscess formation by Bacteroides fragilis. *Journal of Immunology* **160**, 5000-5006.

Grabbe, C. & Dikic, I. (2009). Functional Roles of Ubiquitin-Like Domain (ULD) and Ubiquitin-Binding Domain (UBD) Containing Proteins. *Chemical Reviews* **109**, 1481-1494.

Grenier, D. & Mayrand, D. (1987). FUNCTIONAL-CHARACTERIZATION OF EXTRACELLULAR VESICLES PRODUCED BY BACTEROIDES-GINGIVALIS. *Infection and Immunity* **55**, 111-117.

Grifantini, R., Bartolini, E., Muzzi, A., Draghi, M., Frigimelica, E., Berger, J., Randazzo, F. & Grandi, G. (2002). Gene expression profile in Neisseria meningitidis and Neisseria lactamica upon host-cell contact - From basic research to vaccine development. *Microarrays, Immune Responses and Vaccines* **975**, 202-216.

Haase, R., Richter, K., Pfaffinger, G., Courtois, G. & Ruckdeschel, K. (2005). Yersinia outer protein P suppresses TGF-beta-activated kinase-1 activity to impair innate immune signaling in Yersinia enterocolitica-infected cells. *Journal of Immunology* **175**, 8209-8217.

Haglund, K. & Dikic, I. (2005). Ubiquitylation and cell signaling. *Embo Journal* **24**, 3353-3359.

Harper, J. W. & King, R. W. (2011). Stuck in the Middle: Drugging the Ubiquitin System at the E2 Step. *Cell* **145**, 1007-1009.

Hartmann-Petersen, R. & Gordon, C. (2005). Quantifying protein-protein interactions in the ubiquitin pathway by surface plasmon resonance. In Ubiquitin and Protein Degradation, Pt B, pp. 164-177. Edited by R. J. Deshaies.

Haurat, M. F., Aduse-Opoku, J., Rangarajan, M., Dorobantu, L., Gray, M. R., Curtis, M. A. & Feldman, M. F. (2011). Selective Sorting of Cargo Proteins into Bacterial Membrane Vesicles. *Journal of Biological Chemistry* **286**, 1269-1276.

Henschen, A., Lottspeich, F., Kehl, M. & Southan, C. (1983). COVALENT STRUCTURE OF FIBRINOGEN. *Annals of the New York Academy of Sciences* **408**, 28-43.

Herberhold, H. & Winter, R. (2002). Temperature- and pressure-induced unfolding and refolding of ubiquitin: A static and kinetic Fourier transform infrared spectroscopy study. *Biochemistry* **41**, 2396-2401.

Herrmann, J., Lerman, L. O. & Lerman, A. (2007). Ubiquitin and ubiquitin-like proteins in protein regulation. *Circulation Research* **100**, 1276-1291.

Hiemstra, P. S., van den Barselaar, M. T., Roest, M., Nibbering, P. H. & van Furth, R. (1999). Ubiquicidin, a novel murine microbicidal protein present in the cytosolic fraction of macrophages. *Journal of Leukocyte Biology* **66**, 423-428.

Horstman, A. L. & Kuehn, M. J. (2000). Enterotoxigenic Escherichia coli secretes active heat-labile enterotoxin via outer membrane vesicles. *Journal of Biological Chemistry* **275**, 12489-12496.

Houston, S., Blakely, G. W., McDowell, A., Martin, L. & Patrick, S. (2010). Binding and degradation of fibrinogen by Bacteroides fragilis and characterization of a 54 kDa fibrinogen-binding protein. *Microbiology-Sgm* **156**, 2516-2526.

Howell, S. J., Wilk, D., Yadav, S. P. & Bevins, C. L. (2003). Antimicrobial polypeptides of the human colonic epithelium. *Peptides* **24**, 1763-1770.

Huang, A. M., Kao, Y.-T., Toh, S. & other authors (2011a). UBE2M-mediated p27(KiP1) degradation in gemcitabine cytotoxicity. *Biochemical Pharmacology* **82**, 35-42.

Huang, J. Y., Lee, S. M. & Mazmanian, S. K. (2011b). The human commensal Bacteroides fragilis binds intestinal mucin. *Anaerobe* **17**, 137-141.

Imkamp, F., Striebel, F., Sutter, M., Oezcelik, D., Zimmermann, N., Sander, P. & Weber-Ban, E. (2010). Dop functions as a depupylase in the prokaryotic ubiquitin-like modification pathway. *Embo Reports* **11**, 791-797.

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

Jang, K. S., Sweredoski, M. J., Graham, R. L. J., Hess, S. & Clemons, W. M. (2014). Comprehensive proteomic profiling of outer membrane vesicles from Campylobacter jejuni. *Journal of Proteomics* **98**, 90-98.

Jecklin, M. C., Schauer, S., Dumelin, C. E. & Zenobi, R. (2009). Label-free determination of protein-ligand binding constants using mass spectrometry and validation using surface plasmon resonance and isothermal titration calorimetry. *Journal of Molecular Recognition* **22**, 319-329.

Ji, G. Y., Beavis, R. & Novick, R. P. (1997). Bacterial interference caused by autoinducing peptide variants. *Science* **276**, 2027-2030.

Jin, L., Williamson, A., Banerjee, S., Philipp, I. & Rape, M. (2008). Mechanism of ubiquitin-chain formation by the human anaphase-promoting complex. *Cell* **133**, 653-665.

Jones, D. T., Taylor, W. R. & Thorton, J. M. (1994). A MODEL RECOGNITION APPROACH TO THE PREDICTION OF ALL-HELICAL MEMBRANE-PROTEIN STRUCTURE AND TOPOLOGY. *Biochemistry* **33**, 3038-3049.

Juturu, V. & Wu, J. C. (2012). Microbial xylanases: Engineering, production and industrial applications. *Biotechnology Advances* **30**, 1219-1227.

Kadurugamuwa, J. L. & Beveridge, T. J. (1995). VIRULENCE FACTORS ARE RELEASED FROM PSEUDOMONAS-AERUGINOSA IN ASSOCIATION WITH MEMBRANE-VESICLES DURING NORMAL GROWTH AND EXPOSURE TO GENTAMICIN - A NOVEL MECHANISM OF ENZYME-SECRETION. *Journal of Bacteriology* **177**, 3998-4008.

Kadurugamuwa, J. L. & Beveridge, T. J. (1996). Bacteriolytic effect of membrane vesicles from Pseudomonas aeruginosa on other bacteria including pathogens: Conceptually new antibiotics. *Journal of Bacteriology* **178**, 2767-2774.

Kadurugamuwa, J. L. & Beveridge, T. J. (1997). Natural release of virulence factors in membrane vesicles by Pseudomonas aeruginosa and the effect of aminoglycoside antibiotics on their release. *Journal of Antimicrobial Chemotherapy* **40**, 615-621.

Kadurugamuwa, J. L. & Beveridge, T. J. (1999). Membrane vesicles derived from Pseudomonas aeruginosa and Shigella flexneri can be integrated into the surfaces of other Gram-negative bacteria. *Microbiology-Uk* **145**, 2051-2060.

Kalra, H., Adda, C. G., Liem, M., Ang, C.-S., Mechler, A., Simpson, R. J., Hulett, M. D. & Mathivanan, S. (2013). Comparative proteomics evaluation of plasma exosome isolation techniques and assessment of the stability of exosomes in normal human blood plasma. *Proteomics* **13**, 3354-3364.

Kantyka, T., Rawlings, N. D. & Potempa, J. (2010). Prokaryote-derived protein inhibitors of peptidases: A sketchy occurrence and mostly unknown function. *Biochimie* **92**, 1644-1656.

Karlsson, F. H., Ussery, D. W., Nielsen, J. & Nookaew, I. (2011). A Closer Look at Bacteroides: Phylogenetic Relationship and Genomic Implications of a Life in the Human Gut. *Microbial Ecology* **61**, 473-485.

Kato, S., Kowashi, Y. & Demuth, D. R. (2002). Outer membrane-like vesicles secreted by Actinobacillus actinomycetemcomitans are enriched in leukotoxin. *Microbial Pathogenesis* **32**, 1-13.

Kelley, L. A. & Sternberg, M. J. E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**, 363-371.

Kesty, N. C. & Kuehn, M. J. (2004). Incorporation of heterologous outer membrane and periplasmic proteins into Escherichia coli outer membrane vesicles. *Journal of Biological Chemistry* **279**, 2069-2076.

Kieffer, A. E., Goumon, Y., Ruh, O., Chasserot-Golaz, S., Nullans, G., Gasnier, C., Aunis, D. & Metz-Boutigue, M. H. (2003). The N- and C-terminal fragments of ubiquitin are important for the antimicrobial activities. *Faseb Journal* **17**, 776-+.

Kolodziejek, A. M., Caplan, A. B., Bohach, G. A., Paszczynski, A. J., Minnich, S. A. & Hovde, C. J. (2013). Physiological Levels of Glucose Induce Membrane Vesicle Secretion and Affect the Lipid and Protein Composition of Yersinia pestis Cell Surfaces. *Applied and Environmental Microbiology* **79**, 4509-4514.

Komander, D. & Rape, M. (2012). The Ubiquitin Code. *Annual Review of Biochemistry, Vol 81* **81**, 203-229.

Kreimer, S., Belov, A. M., Ghiran, I., Murthy, S. K., Frank, D. A. & Ivanov, A. R. (2015). Mass-spectrometry-based molecular characterization of extracellular vesicles: lipidomics and proteomics. *Journal of proteome research* **14**, 2367-2384.

Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology* **305**, 567-580.

Kubori, T., Hyakutake, A. & Nagai, H. (2008). Legionella translocates an E3 ubiquitin ligase that has multiple U-boxes with distinct functions. *Molecular Microbiology* **67**, 1307-1319.

Kuehn, M. J. & Kesty, N. C. (2005). Bacterial outer membrane vesicles and the host-pathogen interaction. *Genes & Development* **19**, 2645-2655.

Kulkarni, H. M. & Jagannadham, M. V. (2014). Biogenesis and multifaceted roles of outer membrane vesicles from Gram-negative bacteria. *Microbiology-Sgm* **160**, 2109-2121.

Kulkarni, H. M., Swamy, C. V. B. & Jagannadham, M. V. (2014). Molecular Characterization and Functional Analysis of Outer Membrane Vesicles from the Antarctic Bacterium Pseudomonas syringae Suggest a Possible Response to Environmental Conditions. *Journal of Proteome Research* **13**, 1345-1358.

Kulp, A. & Kuehn, M. J. (2010). Biological Functions and Biogenesis of Secreted Bacterial Outer Membrane Vesicles. In Annual Review of Microbiology, Vol 64, 2010, pp. 163-184. Edited by S. Gottesman & C. S. Harwood.

Larsen, C. N. & Wang, H. L. (2002). The ubiquitin superfamily: Members, features, and phylogenies. *Journal of Proteome Research* **1**, 411-419.

Lau, A. W., Fukushima, H. & Wei, W. Y. (2012). The Fbw7 and Beta-TRCP E3 ubiquitin ligases and their roles in tumorigenesis. *Frontiers in Bioscience-Landmark* **17**, 2197-2212.

Le Bihan, T., Grima, R., Martin, S., Forster, T. & Le Bihan, Y. (2010). Quantitative analysis of low-abundance peptides in HeLa cell cytoplasm by targeted liquid chromatography/mass spectrometry and stable isotope dilution: emphasising the distinction between peptide detection and peptide identification. *Rapid Communications in Mass Spectrometry* **24**, 1093-1104.

Le Bihan, T., Martin, S. F., Chirnside, E. S., van Ooijen, G., Barrios-Llerena, M. E., O'Neill, J. S., Shliaha, P. V., Kerr, L. E. & Millar, A. J. (2011). Shotgun proteomic analysis of the unicellular alga Ostreococcus tauri. *Journal of Proteomics* **74**, 2060-2070.

Lee, C. S. & Lee, J. (2010). Evaluation of new gyrB-based real-time PCR system for the detection of B. fragilis as an indicator of human-specific fecal contamination. *Journal of Microbiological Methods* **82**, 311-318.

Lee, E.-Y., Bang, J. Y., Park, G. W. & other authors (2007). Global proteomic profiling of native outer membrane vesicles derived from Escherichia coli. *Proteomics* **7**, 3143-3153.

Lee, E. Y., Choi, D. S., Kim, K. P. & Gho, Y. S. (2008). PROTEOMICS IN GRAM-NEGATIVE BACTERIAL OUTER MEMBRANE VESICLES. *Mass Spectrometry Reviews* **27**, 535-555.

Leidecker, O., Matic, I., Mahata, B., Pion, E. & Xirodimas, D. P. (2012). The ubiquitin E1 enzyme Ube1 mediates NEDD8 activation under diverse stress conditions. *Cell Cycle* **11**, 1142-1150.

Li, Z., Adams, R. M., Chourey, K., Hurst, G. B., Hettich, R. L. & Pan, C. L. (2012). Systematic Comparison of Label-Free, Metabolic Labeling, and Isobaric Chemical Labeling for Quantitative Proteomics on LTQ Orbitrap Velos. *Journal of Proteome Research* **11**, 1582-1590.

Li, Z. S., Clarke, A. J. & Beveridge, T. J. (1996). A major autolysin of Pseudomonas aeruginosa: Subcellular distribution, potential role in cell growth and division, and secretion in surface membrane vesicles. *Journal of Bacteriology* **178**, 2479-2488.

Liu, C. Y., Huang, Y. T., Liao, C. H., Yen, L. C., Lin, H. Y. & Hsueh, P. R. (2008). Increasing trends in antimicrobial resistance among clinically important anaerobes and Bacteroides fragilis isolates causing nosocomial infections: Emerging resistance to carbapenems. *Antimicrobial Agents and Chemotherapy* **52**, 3161-3168.

Lottenberg, R., Broder, C. C., Boyle, M. D. P., Kain, S. J., Schroeder, B. L. & Curtiss, R. (1992). CLONING, SEQUENCE-ANALYSIS, AND EXPRESSION IN ESCHERICHIA-COLI OF A STREPTOCOCCAL PLASMIN RECEPTOR. *Journal of Bacteriology* **174**, 5204-5210.

Luders, J., Pyrowolakis, G. & Jentsch, S. (2003). The ubiquitin-like protein HUB1 forms SDS-resistant complexes with cellular proteins in the absence of ATP. *Embo Reports* **4**, 1169-1174.

Lutton, D. A., Patrick, S., Crockard, A. D., Stewart, L. D., Larkin, M. J., Dermott, E. & McNeill, T. A. (1991). FLOW CYTOMETRIC ANALYSIS OF WITHIN-STRAIN VARIATION IN POLYSACCHARIDE EXPRESSION BY BACTEROIDES-FRAGILIS BY USE OF MURINE MONOCLONAL-ANTIBODIES. *Journal of Medical Microbiology* **35**, 229-237.

Manning, A. J. & Kuehn, M. J. (2011). Contribution of bacterial outer membrane vesicles to innate bacterial defense. *Bmc Microbiology* **11**.

Manning, A. J. & Kuehn, M. J. (2013). Functional Advantages Conferred by Extracellular Prokaryotic Membrane Vesicles. *Journal of Molecular Microbiology and Biotechnology* **23**, 131-141.

Martin, H. H., Heilmann, H. D. & Preusser, H. J. (1972). STATE OF RIGID-LAYER IN CELL-WALLS OF SOME GRAM-NEGATIVE BACTERIA. *Archiv Fur Mikrobiologie* **83**, 332-&.

Mashburn, L. M. & Whiteley, M. (2005). Membrane vesicles traffic signals and facilitate group activities in a prokaryote. *Nature* **437**, 422-425.

Mashburn-Warren, L., Howe, J., Brandenburg, K. & Whiteley, M. (2009). Structural Requirements of the Pseudomonas Quinolone Signal for Membrane Vesicle Stimulation. *Journal of Bacteriology* **191**, 3411-3414.

Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. (2005). An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122**, 107-118.

Mazmanian, S. K., Round, J. L. & Kasper, D. L. (2008). A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453**, 620-625.

McBroom, A. J. & Kuehn, M. J. (2007). Release of outer membrane vesicles by gram-negative bacteria is a novel envelope stress response. *Molecular Microbiology* **63**, 545-558.

McCaig, W. D., Koller, A. & Thanassi, D. G. (2013). Production of Outer Membrane Vesicles and Outer Membrane Tubes by Francisella novicida. *Journal of Bacteriology* **195**, 1120-1132.

Meggersee, R. & Abratt, V. (2015). The occurrence of antibiotic resistance genes in drug resistant Bacteroides fragilis isolates from Groote Schuur Hospital, South Africa. *Anaerobe* **32**, 1-6.

Misaghi, S., Galardy, P. J., Meester, W. J. N., Ovaa, H., Ploegh, H. L. & Gaudet, R. (2005). Structure of the ubiquitin hydrolase UCH-L3 complexed with a suicide substrate. *Journal of Biological Chemistry* **280**, 1512-1520.

Modun, B., Morrissey, J. & Williams, P. (2000). The staphylococcal transferrin receptor: a glycolytic enzyme with novel functions. *Trends in Microbiology* **8**, 231-237.

Moncrief, J. S., Obiso, R., Barroso, L. A., Kling, J. J., Wright, R. L., Vantassell, R. L., Lyerly, D. M. & Wilkins, T. D. (1995). THE ENTEROTOXIN OF BACTEROIDES-FRAGILIS IS A METALLOPROTEASE. *Infection and Immunity* **63**, 175-181.

Mossner, E., Boll, M. & Pfleiderer, G. (1980). PURIFICATION OF HUMAN AND BOVINE ALKALINE-PHOSPHATASES BY AFFINITY-CHROMATOGRAPHY. *Hoppe-Seylers Zeitschrift Fur Physiologische Chemie* **361**, 543-549.

Mullaney, E., Brown, P. A., Smith, S. M., Botting, C. H., Yamaoka, Y. Y., Terres, A. M., Kelleher, D. P. & Windle, H. J. (2009). Proteomic and functional characterization of the outer membrane vesicles from the gastric pathogen Helicobacter pylori. *Proteomics Clinical Applications* **3**, 785-796.

Munro, P., Flatau, G. & Lemichez, E. (2007). Bacteria and the ubiquitin pathway. *Current Opinion in Microbiology* **10**, 39-46.

Nally, J. E., Whitelegge, J. P., Aguilera, R., Pereira, M. M., Blanco, D. R. & Lovett, M. A. (2005). Purification and proteomic analysis of outer membrane vesicles from a clinical isolate of Leptospira interrogans serovar Copenhageni. *Proteomics* **5**, 144-152.

Nercessian, D., de Castro, R. E. & Conde, R. D. (2002). Ubiquitin-like proteins in halobacteria. *Journal of Basic Microbiology* **42**, 277-283.

Pancholi, V. & Fischetti, V. A. (1992). A MAJOR SURFACE PROTEIN ON GROUP-A STREPTOCOCCI IS A GLYCERALDEHYDE-3-PHOSPHATE-DEHYDROGENASE WITH MULTIPLE BINDING-ACTIVITY. *Journal of Experimental Medicine* **176**, 415-426.

Pancholi, V. & Chhatwal, G. S. (2003). Housekeeping enzymes as virulence factors for pathogens. *International Journal of Medical Microbiology* **293**, 391-401.

Patrick, S., Reid, J. H. & Coffey, A. (1986). CAPSULATION OF INVITRO AND INVIVO GROWN BACTEROIDES SPECIES. *Journal of General Microbiology* **132**, 1099-1109.

Patrick, S., McKenna, J. P., Ohagan, S. & Dermott, E. (1996). A comparison of the haemagglutinating and enzymic activities of Bacteroides fragilis whole cells and outer membrane vesicles. *Microbial Pathogenesis* **20**, 191-202.

Patrick, S., Parkhill, J., McCoy, L. J., Lennard, N., Larkin, M. J., Collins, M., Sczaniecka, M. & Blakely, G. (2003). Multiple inverted DNA repeats of Bacteroides fragilis that control polysaccharide antigenic variation are similar to the hin region inverted repeats of Salmonella typhimurium. *Microbiology-Sgm* **149**, 915-924.

Patrick, S., Houston, S., Thacker, Z. & Blakely, G. W. (2009). Mutational analysis of genes implicated in LPS and capsular polysaccharide biosynthesis in the opportunistic pathogen Bacteroides fragilis. *Microbiology-Sgm* **155**, 1039-1049.

Patrick, S., Blakely, G. W., Houston, S. & other authors (2010). Twenty-eight divergent polysaccharide loci specifying within- and amongst-strain capsule diversity in three strains of Bacteroides fragilis. *Microbiology-Sgm* **156**, 3255-3269.

Patrick, S., Jobling, K. L., O'Connor, D., Thacker, Z., Dryden, D. T. F. & Blakely, G. W. (2011). A unique homologue of the eukaryotic protein-modifier ubiquitin present in the bacterium Bacteroides fragilis, a predominant resident of the human gastrointestinal tract. *Microbiology-Sgm* **157**, 3071-3078.

Perez-Cruz, C., Carrion, O., Delgado, L., Martinez, G., Lopez-Iglesias, C. & Mercade, E. (2013). New Type of Outer Membrane Vesicle Produced by the Gram-Negative Bacterium Shewanella vesiculosa M7(T): Implications for DNA Content. *Applied and Environmental Microbiology* **79**, 1874-1881.

Piana, S., Lindorff-Larsen, K. & Shaw, D. E. (2013). Atomic-level description of ubiquitin folding. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 5915-5920.

Poppe, C., Smart, N., Khakhria, R., Johnson, W., Spika, J. & Prescott, J. (1998). Salmonella typhimurium DT104: A virulent and drug-resistant pathogen. *Canadian Veterinary Journal-Revue Veterinaire Canadienne* **39**, 559-565.

Post, D. M. B., Zhang, D. S., Eastvold, J. S., Teghanemt, A., Gibson, B. W. & Weiss, J. P. (2005). Biochemical and functional characterization of membrane blebs purified from Neisseria meningitidis serogroup B. *Journal of Biological Chemistry* **280**, 38383-38394.

Pumbwe, L., Skilbeck, C. A., Nakano, V., Avila-Campos, M. J., Piazza, R. M. F. & Wexler, H. M. (2007). Bile salts enhance bacterial co-aggregation, bacterial-intestinal epithelial cell adhesion, biofilm formation and antimicrobial resistance of Bacteroides fragilis. *Microbial Pathogenesis* **43**, 78-87.

Rappsilber, J., Mann, M. & Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols* **2**, 1896-1906.

Redondo, M. C., Arbo, M. D. J., Grindlinger, J. & Snydman, D. R. (1995). ATTRIBUTABLE MORTALITY OF BACTEREMIA ASSOCIATED WITH THE BACTEROIDES-FRAGILIS GROUP. *Clinical Infectious Diseases* **20**, 1492-1496.

Reid, J. H. & Patrick, S. (1984). PHAGOCYTIC AND SERUM KILLING OF CAPSULATE AND NON-CAPSULATE BACTEROIDES-FRAGILIS. *Journal of Medical Microbiology* **17**, 247-257.

Renelli, M., Matias, V., Lo, R. Y. & Beveridge, T. J. (2004). DNA-containing membrane vesicles of Pseudomonas aeruginosa PAO1 and their genetic transformation potential. *Microbiology-Sgm* **150**, 2161-2169.

Ribet, D., Hamon, M., Gouin, E. & other authors (2010). Listeria monocytogenes impairs SUMOylation for efficient infection. *Nature* **464**, 1192-U1111.

Robert-Genthon, M., Casabona, M. G., Neves, D., Coute, Y., Ciceron, F., Elsen, S., Dessen, A. & Attree, I. (2013). Unique Features of a Pseudomonas aeruginosa alpha 2-Macroglobulin Homolog. *Mbio* **4**.

Roth, G., Freund, S., Mohrle, B., Wollner, K., Brunjes, J., Gauglitz, G., Wiesmuller, K. H. & Jung, G. (2007). Ubiquitin binds to a short peptide segment of hydrolase UCH-L3: A study by FCS, RIfS, ITC and NMR. *Chembiochem* **8**, 323-331.

Salyers, A. A., Gupta, A. & Wang, Y. P. (2004). Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends in Microbiology* **12**, 412-416.

Sanchez, C. (2011). Cellular microbiology: bacterial pea shooters. *Nature reviews Microbiology* **9**, 562-562.

Schaar, V., Uddback, I., Nordstrom, T. & Riesbeck, K. (2014). Group A streptococci are protected from amoxicillin-mediated killing by vesicles containing beta-lactamase derived from Haemophilus influenzae. *Journal of Antimicrobial Chemotherapy* **69**, 117-120.

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

Schwechheimer, C., Sullivan, C. J. & Kuehn, M. J. (2013). Envelope Control of Outer Membrane Vesicle Production in Gram-Negative Bacteria. *Biochemistry* **52**, 3031-3040.

Scorza, F. B., Doro, F., Rodriguez-Ortega, M. J. & other authors (2008). Proteomics characterization of outer membrane vesicles from the extraintestinal pathogenic Escherichia coli Delta tolR IHE3034 mutant. *Molecular & Cellular Proteomics* **7**, 473-485.

Sears, C. L., Islam, S., Saha, A. & other authors (2008). Association of enterotoxigenic Bacteroides fragilis infection with inflammatory diarrhea. *Clinical Infectious Diseases* **47**, 797-803.

Sears, C. L. (2009). Enterotoxigenic Bacteroides fragilis: a Rogue among Symbiotes. *Clinical Microbiology Reviews* **22**, 349-+.

Seifert, H., Dalhoff, A. & Grp, P. S. (2010). German multicentre survey of the antibiotic susceptibility of Bacteroides fragilis group and Prevotella species isolated from intra-abdominal infections: results from the PRISMA study. *Journal of Antimicrobial Chemotherapy* **65**, 2405-2410.

Semplici, F., Meggio, F., Pinna, L. A. & Oliviero, S. (2002). CK2-dependent phosphorylation of the E2 ubiquitin conjugating enzyme UBC3B induces its interaction with beta-TrCP and enhances beta-catenin degradation. *Oncogene* **21**, 3978-3987.

Seo, J.-K., Lee, M. J., Go, H.-J., Do Kim, G., Do Jeong, H., Nam, B.-H. & Park, N. G. (2013). Purification and antimicrobial function of ubiquitin isolated from the gill of Pacific oyster, Crassostrea gigas. *Molecular Immunology* **53**, 88-98.

Sevier, C. S. & Kaiser, C. A. (2002). Formation and transfer of disulphide bonds in living cells. *Nature Reviews Molecular Cell Biology* **3**, 836-847.

Shen, Y., Torchia, M. L. G., Lawson, G. W., Karp, C. L., Ashwell, J. D. & Mazmanian, S. K. (2012). Outer Membrane Vesicles of a Human Commensal Mediate Immune Regulation and Disease Protection. *Cell Host & Microbe* **12**, 509-520.

Sherwood, J. E., Fraser, S., Citron, D. M., Wexler, H., Blakely, G., Jobling, K. & Patrick, S. (2011). Multi-drug resistant Bacteroides fragilis recovered from blood and severe leg wounds caused by an improvised explosive device (IED) in Afghanistan. *Anaerobe* **17**, 152-155.

Shetty, A., Chen, S., Tocheva, E. I., Jensen, G. J. & Hickey, W. J. (2011). Nanopods: A New Bacterial Structure and Mechanism for Deployment of Outer Membrane Vesicles. *Plos One* **6**.

Shipman, J. A., Berleman, J. E. & Salyers, A. A. (2000). Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron. *Journal of Bacteriology* **182**, 5365-5372.

Sottrupjensen, L., Sand, O., Kristensen, L. & Fey, G. H. (1989). THE ALPHA-MACROGLOBULIN BAIT REGION - SEQUENCE DIVERSITY AND LOCALIZATION OF CLEAVAGE SITES FOR PROTEINASES IN 5 MAMMALIAN ALPHA-MACROGLOBULINS. *Journal of Biological Chemistry* **264**, 15781-15789.

Stentz, R., Horn, N., Cross, K., Salt, L., Brearley, C., Livermore, D. M. & Carding, S. R. (2015). Cephalosporinases associated with outer membrane vesicles released by Bacteroides spp. protect gut pathogens and commensals against beta-lactam antibiotics. *Journal of Antimicrobial Chemotherapy* **70**, 701-709.

Stock, I. & Wiedemann, B. (2000). Natural antibiotic susceptibility of Salmonella enterica strains. *International Journal of Antimicrobial Agents* **16**, 211-217.

Striebel, F., Imkamp, F., Sutter, M., Steiner, M., Mamedov, A. & Weber-Ban, E. (2009). Bacterial ubiquitin-like modifier Pup is deamidated and conjugated to substrates by distinct but homologous enzymes. *Nature Structural & Molecular Biology* **16**, 647-U689.

Striebel, F., Imkamp, F., Ozcelik, D. & Weber-Ban, E. (2014). Pupylation as a signal for proteasomal degradation in bacteria. *Biochimica Et Biophysica Acta-Molecular Cell Research* **1843**, 103-113.

Svensson, I., Calles, K., Lindskog, E., Henriksson, H., Eriksson, U. & Haggstrom, L. (2005). Antimicrobial activity of conditioned medium fractions from Spodoptera frugiperda Sf9 and Trichoplusia ni Hi5 insect cells. *Applied Microbiology and Biotechnology* **69**, 92-98.

Swidsinski, A., Weber, J., Loening-Baucke, V., Hale, L. P. & Lochs, H. (2005). Spatial organization and composition of the mucosal flora in patients with inflammatory bowel disease. *Journal of Clinical Microbiology* **43**, 3380-3389.

Tan, K. S., Chen, Y., Lim, Y.-C., Tan, G.-Y. G., Liu, Y., Lim, Y.-T., MacAry, P. & Gan, Y.-H. (2010). Suppression of Host Innate Immune Response by Burkholderia pseudomallei through the Virulence Factor TssM. *Journal of Immunology* **184**, 5160-5171.

Tanno, H. & Komada, M. (2013). The ubiquitin code and its decoding machinery in the endocytic pathway. *Journal of Biochemistry* **153**, 497-504.

Tashiro, Y., Uchiyama, H. & Nomura, N. (2012). Multifunctional membrane vesicles in Pseudomonas aeruginosa. *Environmental Microbiology* **14**, 1349-1362.

Tauro, B. J., Greening, D. W., Mathias, R. A., Ji, H., Mathivanan, S., Scott, A. M. & Simpson, R. J. (2012). Comparison of ultracentrifugation, density gradient separation, and immunoaffinity capture methods for isolating human colon cancer cell line LIM1863-derived exosomes. *Methods* **56**, 293-304.

Technologies, L. (2011).T-REx System: User Guide. In A Tetracycline-Regulated Expression System for Mammalian Cells. Carlsbad, CA: Life Technologies.

Thornton, R. F., Kagawa, T. F., O'Toole, P. W. & Cooney, J. C. (2010). The dissemination of C10 cysteine protease genes in Bacteroides fragilis by mobile genetic elements. *Bmc Microbiology* **10**.

Troy, E. B. & Kasper, D. L. (2010). Beneficial effects of Bacteroides fragilis polysaccharides on the immune system. *Frontiers in Bioscience-Landmark* **15**, 25-34.

Tusnady, G. E. & Simon, I. (1998). Principles governing amino acid composition of integral membrane proteins: Application to topology prediction. *Journal of Molecular Biology* **283**, 489-506.

Tzianabos, A. O. & Kasper, D. L. (2002). Role of T cells in abscess formation. *Current Opinion in Microbiology* **5**, 92-96.

Valkevich, E. M., Sanchez, N. A., Ge, Y. & Strieter, E. R. (2014). Middle-Down Mass Spectrometry Enables Characterization of Branched Ubiquitin Chains. *Biochemistry* **53**, 4979-4989.

van de Waterbeemd, B., Mommen, G. P. M., Pennings, J. L. A., Eppink, M. H., Wijffels, R. H., van der Pol, L. A. & de Jong, A. (2013). Quantitative Proteomics Reveals Distinct Differences in the Protein Content of Outer Membrane Vesicle Vaccines. *Journal of Proteome Research* **12**, 1898-1908.

van den Ent, F. & Lowe, J. (2005). Crystal structure of the ubiquitin-like protein YukD from Bacillus subtilis. *Febs Letters* **579**, 3837-3841.

van der Veen, A. G. & Ploegh, H. L. (2012). Ubiquitin-Like Proteins. *Annual Review of Biochemistry, Vol 81* **81**, 323-357.

Van Tassell, R. L. & Wilkins, T. D. (1978). ISOLATION OF AUXOTROPHS OF BACTEROIDES-FRAGILIS. *Canadian Journal of Microbiology* **24**, 1619-1621.

Vidakovics, M., Jendholm, J., Morgelin, M., Mansson, A., Larsson, C., Cardell, L. O. & Riesbeck, K. (2010). B Cell Activation by Outer Membrane Vesicles-A Novel Virulence Mechanism. *Plos Pathogens* **6**.

Vieira, J., Vallim, D. C., Ferreira, E. O., Seabra, S. H., Vommaro, R. C., Avelar, K. E. S., De Souza, W., Ferreira, M. C. S. & Domingues, R. (2005). Bacteroides fragilis interferes with iNOS activity and leads to pore formation in macrophage surface. *Biochemical and Biophysical Research Communications* **326**, 607-613.

Vijaykumar, S., Bugg, C. E. & Cook, W. J. (1987). STRUCTURE OF UBIQUITIN REFINED AT 1.8 A RESOLUTION. *Journal of Molecular Biology* **194**, 531-544.

Wear, M. A. & Walkinshaw, M. D. (2006). Thermodynamics of the cyclophilin-A/cyclosporin-A interaction: A direct comparison of parameters determined by surface plasmon resonance using Biacore T100 and isothermal titration calorimetry. *Analytical Biochemistry* **359**, 285-287.

Wexler, H. M. (2007). Bacteroides: the good, the bad, and the nitty-gritty. *Clinical Microbiology Reviews* **20**, 593-+.

Wick, E. C. & Sears, C. L. (2010). Bacteroides spp. and diarrhea. *Current Opinion in Infectious Diseases* **23**, 470-474.

Wickliffe, K. E., Lorenz, S., Wemmer, D. E., Kuriyan, J. & Rape, M. (2011). The Mechanism of Linkage-Specific Ubiquitin Chain Elongation by a Single-Subunit E2. *Cell* **144**, 769-781.

Wilkinson, K. D. & Audhya, T. K. (1981). STIMULATION OF ATP-DEPENDENT PROTEOLYSIS REQUIRES UBIQUITIN WITH THE COOH-TERMINAL SEQUENCE ARG-GLY-GLY. *Journal of Biological Chemistry* **256**, 9235-9241.

Work, E., Knox, K. W. & Vesk, M. (1966). CHEMISTRY AND ELECTRON MICROSCOPY OF AN EXTRACELLULAR LIPOPOLYSACCHARIDE FROM ESCHERICHIA COLI. *Annals of the New York Academy of Sciences* **133**, 438-&.

Wright, D. W., Moreno-Vargas, A. J., Carmona, A. T., Robina, I. & Davies, G. J. (2013). Three dimensional structure of a bacterial alpha-L-fucosidase with a 5-membered iminocyclitol inhibitor. *Bioorganic & Medicinal Chemistry* **21**, 4751-4754.

Wu, S., Rhee, K.-J., Zhang, M., Franco, A. & Sears, C. L. (2007). Bacteroides fragilis toxin stimulates intestinal epithelial cell shedding and gamma-secretase-dependent E-cadherin cleavage (Vol 120, pg 1944, 2007). *Journal of Cell Science* **120**, 3713-3713.

Xu, G. W., Ali, M., Wood, T. E. & other authors (2010). The ubiquitin-activating enzyme E1 as a therapeutic target for the treatment of leukemia and multiple myeloma. *Blood* **115**, 2251-2259.

Xu, P., Duong, D. M., Seyfried, N. T. & other authors (2009). Quantitative Proteomics Reveals the Function of Unconventional Ubiquitin Chains in Proteasomal Degradation. *Cell* **137**, 133-145.

Yao, F., Svensjo, T., Winkler, T., Lu, M., Eriksson, C. & Eriksson, E. (1998). Tetracycline repressor, tetR, rather than the tetR-mammalian cell transcription factor fusion derivatives, regulates inducible gene expression in mammalian cells. *Human Gene Therapy* **9**, 1939-1950.

Yaron, S., Kolling, G. L., Simon, L. & Matthews, K. R. (2000). Vesicle-mediated transfer of virulence genes from Escherichia coli O157 : H7 to other enteric bacteria. *Applied and Environmental Microbiology* **66**, 4414-4420.

Yashiroda, H. & Tanaka, K. (2004). Hub1 is an essential ubiquitin-like protein without functioning as a typical modifier in fission yeast. *Genes to Cells* **9**, 1189-1197.

Zackular, J. P., Baxter, N. T., Iverson, K. D., Sadler, W. D., Petrosino, J. F., Chen, G. Y. & Schloss, P. D. (2013). The Gut Microbiome Modulates Colon Tumorigenesis. *Mbio* **4**.

Zhang, X. G. & Bremer, H. (1995). CONTROL OF THE ESCHERICHIA-COLI RRNB P1 PROMOTER STRENGTH BY PPGPP. *Journal of Biological Chemistry* **270**, 11181-11189.

Zhang, Y., Higashide, W. M., McCormick, B. A., Chen, J. & Zhou, D. (2006). The inflammation-associated Salmonella SopA is a HECT-like E3 ubiquitin ligase. *Molecular Microbiology* **62**, 786-793.

List of Appendices

Appendix I - Surface plasmon resonance analysis of BfUbb-E2 interactions

Approximately 200 response units of the ligand molecules, i.e. bovine ubiquitin (Sigma) or rBfUbb were affixed to a CM-5 chip (GE Healthcare). The analyte proteins (E1 enzyme and 26 E2 enzymes) were diluted to a concentration of 2µM in buffer (20mM Tris; 150nM NaCl; 2mM β-mercaptoethanol; 1% glycerol). SPR analysis was performed on a Biacore T-100 (GE Healthcare) by Dr Martin Wear (Edinburgh Protein Production Facility and Edinburgh Biochemical Characterisation Facility, University of Edinburgh). Analyte was passed over the chip at a rate of 30µl/s with a contact time of 60s. Response rates (Figure A.I.i) were too low to calculate a reliable dissociation constant for most E2 enzymes, however they do appear to indicate that non-covalent interactions between BfUbb and E1 are greater than those between BfUbb and E2 enzymes. An experimental error resulted in sample 12 being twice as concentrated as the other samples, hence the higher than average response rate in this sample.

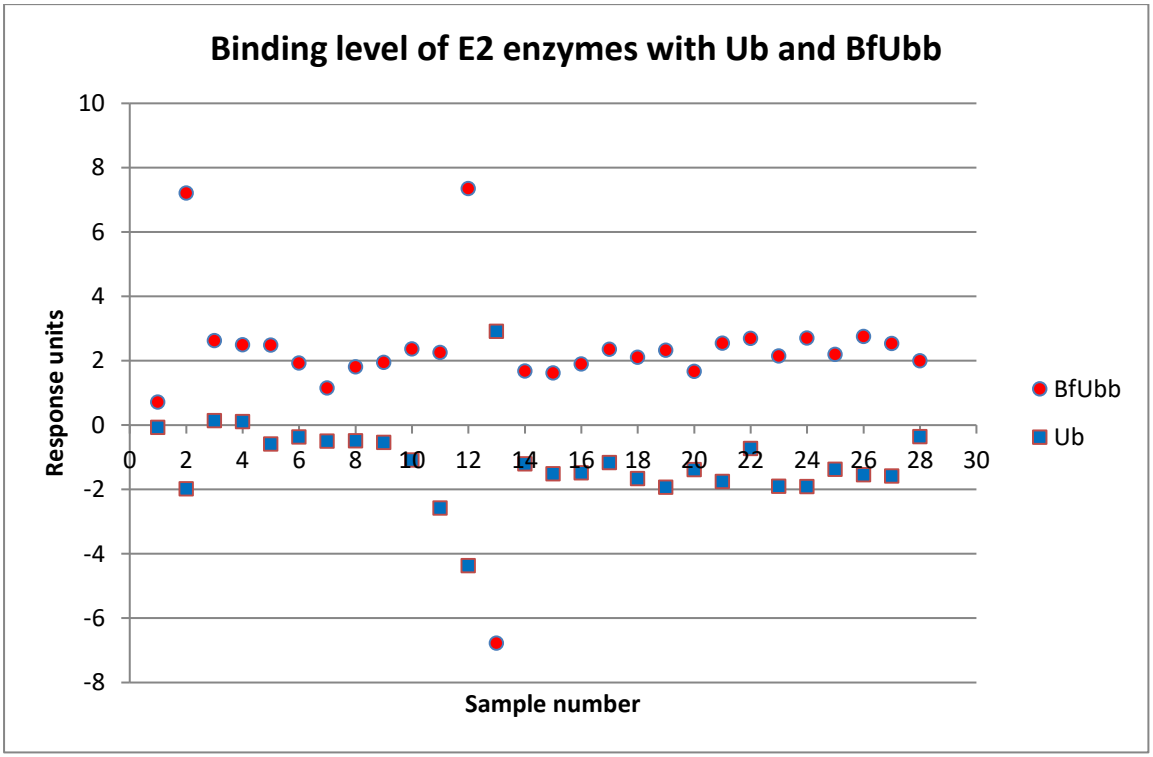


Figure A.I.i Response rates (out of 200) of ubiquitin and rBfUbb interacting with E1 activating enzyme Uba1 and 26 E2 conjugating enzymes.

Appendix II - Sequencing of constructs in pET19b

600ng of pET19b-10xH-BfUbb was mixed with 3.2pmole of either Bfndei or Brbamhi. 600ng of pET19b-10xH-YFP-BfUbb was mixed with 3.2pmole of either Yfndei or Brbamhi. Constructs were quantified and sequenced by the Sequencing and Services Department (Medical Research Council, Dundee, Scotland). The returned sequences were compared with the expected sequence of the construct using Clustal Omega (McWilliam *et al*., 2013) (Figure A.II.i and Figure A.II.ii). Where there are

A

```
pET19b::10xH-BfUbb      CACCTATGCCAAGTTTTTATAAAAAACAGATATGGCTGGACCATAACATTAGAGGTATCA
Result                  ---CATATGCAAGTTTTTATAAAAAACAGATATGGCTGGACCATAACATTAGAGGTATCA
                           *:::  **************************************************

pET19b::10xH-BfUbb      CCTACTGATACTGTAGAAAACGTAAAACAAAAAATTCAGGATAAAGAAGGTTTTCCACCT
Result                  CCTACTGATACTGTAGAAAACGTAAAACAAAAAATTCAGGATAAAGAAGGTTTTCCACCT
                        ************************************************************

pET19b::10xH-BfUbb      GATAAAATAAGGCTTATATATGGAGGAAAACAAATGGAAGATGGACGAACTTTAGCAGAT
Result                  GATAAAATAAGGCTTATATATGGAGGAAAACAAATGGAAGATGGACGAACTTTAGCAGAT
                        ************************************************************

pET19b::10xH-BfUbb      TATAATGTTCAAAAAGACTCAACTATATTAATTTGCATAAGAGATGTCGACTGTTAA
Result                  TATAATGTTCAAAAAGACTCAACTATATTAATTTGCATAAGAGATGTCGACTGTTAA
                        ********************************************************
```

B

```
pET19b::10xH-BfUbb      ---------------TTAACAGTCGACATCTCTTATGCAAATTAATATAGTTGAGTCTTT
Result                  ACGGGACGTTAATGGTTAACAGTCGACATCTCTTATGCAAATTAATATAGTTGAGTCTTT
                                       *******************************************

pET19b::10xH-BfUbb      TTGAACATTATAATCTGCTAAAGTTCGTCCATCTTCCATTTGTTTTCCTCCATATATAAG
Result                  TTGAACATTATAATCTGCTAAAGTTCGTCCATCTTCCATTTGTTTTCCTCCATATATAAG
                        ************************************************************

pET19b::10xH-BfUbb      CCTTATTTTATCAGGTGGAAAACCTTCTTTATCCTGAATTTTTTGTTTTACGTTTTCTAC
Result                  CCTTATTTTATCAGGTGGAAAACCTTCTTTATCCTGAATTTTTTGTTTTACGTTTTCTAC
                        ************************************************************

pET19b::10xH-BfUbb      AGTATCAGTAGGTGATACCTCTAATGTTATGGTCCAGCCATATCTGTTTTTTTATAAAAAC
Result                  AGTATCAGTAGGTGATACCTCTAATGTTATGGTCCAGCCATATCTGTTTTTTTATAAAAAC
                        ************************************************************

pET19b::10xH-BfUbb      TTGCAT-------------------------------------------------------
Result                  TTGCATATGCTTGTCGTCGTCGTCGATATGGCCGCTGCTGTGATGATGATGATGATGATG
                        ******
```

C

```
pET19b::10xH-YFP-BfUbb      CATATGGTGAGCAAGGGCGAGCAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTG
Result                      -TGGGTTGAAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTG
                             : .    .************ ****************************************

pET19b::10xH-YFP-BfUbb      GACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACC
Result                      GACGGCGACGTAAACGGCCACTAGTTCAGCTTGTCCGGGAAGGGCGAGGGCGATGCCACC
                            ********************:******** ******* .********************

pET19b::10xH-YFP-BfUbb      TACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCC
Result                      TACGGCAAGCTGACCCTGAATTTCTTCTGCTCCACCGGCAAGCTGCCCGTGCCTTGGCCC
                            ******************** ***:*****:********************** ******

pET19b::10xH-YFP-BfUbb      ACCCTCGTGACCACCTTCGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATG
Result                      ACCCTCTTGACCACCCTGACCTGGGGCGTGCACTGCTTCGCCCGCTATTCCGACCACTTG
                            ****** ******** * . **. ******** **************  ********:**

pET19b::10xH-YFP-BfUbb      CGCCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATC
Result                      AAGCAGCACGACTTCTTCTAGTCCGCCATGACCGATGGCTACGTCCAGGAGCGCACCATC
                            .. ****************:************.****:************************

pET19b::10xH-YFP-BfUbb      TTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACC
Result                      TTCTTCAAGGACCACCCCAACTACAATACCCACGCCGAAGAGAACTTCGTGGGCGATGTC
                            ************ **  ********* ****.******.*:*** ****:****** . *

pET19b::10xH-YFP-BfUbb      CTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGG
Result                      CTGGTTAACCGAATCAAGCTGAAAGGCAACGACTTCAAGGGGGACGGTAACATACTGATT
                            ***** *****.***.*******.****:***********.****** *****.***.

pET19b::10xH-YFP-BfUbb      CACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAG
Result                      TACTTGCTGGAGTACCACCACATCTTCGACAACGTCTATATCATGGTGGACAA-TTGAAG
                             **::**********.** ***:*: * ******************  *****  :****

pET19b::10xH-YFP-BfUbb      AACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTC
Result                      AACGTCATCTATGTGAACTTCAAGATCAGCAACAATATCGTCGACGGCTTCTTGCA-CTC
                            **** ****:* ***************.**.**** ****: ******: * **** ***

pET19b::10xH-YFP-BfUbb      GCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAA-
Result                      TCCCACCTCTAC-AACTCAATACCCCCTTCGATTACGGCCCCTTGTGA-TGACCGAAACT
                             ** ***:**** *.*: ** ******:***.  ******** **  . **.****.*.

pET19b::10xH-YFP-BfUbb      CCACTACCTGAGCTACCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACAT
Result                      CAACT------GACACCTCTCCTCCTT-GTCAACGACACCAATA-AATGGTCGTCCCTGT
                            *.***      *. ***: *** ** * . ***.***.**** . .*:*  **: *. .*

pET19b::10xH-YFP-BfUbb      GGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAA
Result                      GGTCCTGT-------------------------------------------------------
                            *******

pET19b::10xH-YFP-BfUbb      GATGCAAGTTTTTTATAAAAAACAGATATGGCTGGACCATAACATTAGAGGTATCACCTAC
Result                      ------------------------------------------------------------

pET19b::10xH-YFP-BfUbb      TGATACTGTAGAAAACGTAAAACAAAAAATTCAGGATAAAGAAGGTTTTCCACCTGATAA
Result                      ------------------------------------------------------------

pET19b::10xH-YFP-BfUbb      AATAAGGCTTATATATGGAGGAAAACAAATGGAAGATGGACGAACTTTAGCAGATTATAA
Result                      ------------------------------------------------------------

pET19b::10xH-YFP-BfUbb      TGTTCAAAAAGACTCAACTATATTAATTTGCATAAGAGATGTCGACTGTTAACACTAAGG
Result                      ------------------------------------------------------------

pET19b::10xH-YFP-BfUbb      GATCC
Result                      -----
```

D

```
pET19b::10xH-YFP-BfUbb      ---------------TTAACAGTCGACATCTCTTATGCAAATTAATATAGTTGAGTCTTT
Result                      ACCGGCCCTTTAATGGTTACAGTCGACATCTCTTATGCAAATTAATATAGTTGAGTCTTT
                                          *:******************************************

pET19b::10xH-YFP-BfUbb      TTGAACATTATAATCTGCTAAAGTTCGTCCATCTTCCATTTGTTTTCCTCCATATATAAG
Result                      TTGAACATTATAATCTGCTAAAGTTCGTCCATCTTCCATTTGTTTTCCTCCATATATAAG
                            ************************************************************

pET19b::10xH-YFP-BfUbb      CCTTATTTTATCAGGTGGAAAACCTTCTTTATCCTGAATTTTTTGTTTTACGTTTTCTAC
Result                      CCTTATTTTATCAGGTGGAAAACCTTCTTTATCCTGAATTTTTTGTTTTACGTTTTCTAC
                            ************************************************************

pET19b::10xH-YFP-BfUbb      AGTATCAGTAGGTGATACCTCTAATGTTATGGTCCAGCCATATCTGTTTTTTTATAAAAAC
Result                      AGTATCAGTAGGTGATACCTCTAATGTTATGGTCCAGCCATATCTGTTTTTTTATAAAAAC
                            ************************************************************

pET19b::10xH-YFP-BfUbb      TTGCATCTTGTACAGCTCGTCCATGCCGAGAGTGATCCCGGCGGCGGTCACGAACTCCAG
Result                      TTGCATCTTGTACAGCTCGTCCATGCCGAGAGTGATCCCGGCGGCGGTCACGAACTCCAG
                            ************************************************************

pET19b::10xH-YFP-BfUbb      CAGGACCATGTGATCGCGCTTCTCGTTGGGGTCTTTGCTCAGGGCGGACTGGTAGCTCAG
Result                      CAGGACCATGTGATCGCGCTTCTCGTTGGGGTCTTTGCTCAGGGCGGACTGGGTGCTCAG
                            ************************************************** :******

pET19b::10xH-YFP-BfUbb      GTAGTGGTTGTCGGGCAGCAGCACGGGGCCGTCGCCGATGGGGGTGTTCTGCTGGTAGTG
Result                      GTAGTGGTTGTCGGGCAGCAGCACGGGGCCGTCGCCGATGGGGGTGTTCTGCTGGTAGTG
                            ************************************************************

pET19b::10xH-YFP-BfUbb      GTCGGCGAGCTGCACGCTGCCGTCCTCGATGTTGTGGCGGATCTTGAAGTTCACCTTGAT
Result                      GTCGGCGAGCTGCACGCAGCCGTCCTCGATGTTGTGGCGGATCTTGAAGTTCACCTTGAT
                            *****************:******************************************

pET19b::10xH-YFP-BfUbb      GCCGTTCTTCTGCTTGTCGGCCATGATATAGACGTTGTGGCTGTTGTAGTTGTACTCCAG
Result                      GCCGTTCTTCTGCTTGTCGGCCATGATATAGACGTTGTCGCTGATGGCGTTGTACTCCAG
                            ************************************ ****:** .*************

pET19b::10xH-YFP-BfUbb      CTTGTGCCCCAGGATGTTGCCGTCCTCCTTGAAGTCGATGCCCTTCAGCTCGATGCGGTT
Result                      CTTGTGCCCCAGGATGTTGCCGTCCTCCTTGAAGTCGATGCCCTTCAGCTCGATGCGGTT
                            ************************************************************

pET19b::10xH-YFP-BfUbb      CACCAGGGTGTCGCCCTCGAACTTCACCTCGGCGCGGGTCTTGTAGTTGCCGTCGTCCTT
Result                      CACCAGGGTGTCGCCCTCGAACTTCACCTCGGCGCGGGTCTTGTAGTTGCCGTCGTCCTT
                            ************************************************************

pET19b::10xH-YFP-BfUbb      GAAGAAGATGGTGCGCTCCTGGACGTAGCCTTCGGGCATGGCGGACTTGAAGAAGTCGTG
Result                      GAAGAAGATGGTGCGCTCCTGGACGTAGCCTTCGGGCATGGCGGACTTGAAGAAGTCGTG
                            ************************************************************

pET19b::10xH-YFP-BfUbb      CTGGCGCATGTGGTCGGGGTAGCGGGCGAAGCACTGCACGCCGTAGCCGAAGGTGGTCAC
Result                      CTGCTTCATGTGGTCGGGGTAGCGGGCGAAGCACTGCACGCCCCAGGTCAGGGTGGTCAC
                            ***   ************************************  **   *.*********

pET19b::10xH-YFP-BfUbb      GAGGGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAGCTT
Result                      GAGGGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAGCTT
                            ************************************************************

pET19b::10xH-YFP-BfUbb      GCCGTAGGTGGCATCGCCCTCGCCCTCGCCGGACACGCTGAACTTGTGGCCGTTTACGTC
Result                      GCCGTAGGTGGCATCGCCCTCGCCCTCGCCGGACACGCTGAACTTGTGGCCGTTTACGTC
                            ************************************************************

pET19b::10xH-YFP-BfUbb      GCCGTCCAGCTCGACCAGGATGGGCACCACCCCGGTGAACAGCTCCTCGCCCTTGCTCAC
Result                      GCCGTCCAGCTCGACCAGGATGGGCACCACCCCGGTGAACAGCTCCTCGCCCTTGCTCAC
                            ************************************************************

pET19b::10xH-YFP-BfUbb      CAT-------------------------------------------------------
Result                      CATATGCTTGTCGTCGTCGTCGATATGGCCGCTGCTGTGATGATGATGATGATGATGATG
                            ***
```

Figure A.IV.i Expected sequences of pET19b constructs compared with sequencing results. pET19b::10xH-BfUbb forward (**A**) and reverse (**B**) sequences. pET19b::10xH-YFP-BfUbb forward (**C**) and reverse (**D**) sequences.

Appendix III - Expression of 10xH-YFP-BfUbb in *E. coli* DH5α

Expression of 10xH-YFP-BfUbb was induced in *E. coli* BL21 DE3 pET19b-10xH-YFP-BfUbb with 200µM IPTG for 3h. 10µl each of induced and uninduced cell culture was then mixed with 5µl loading buffer and 1µl DTT. The mixture was boiled at 90ºC for 10 minutes and analysed by SDS-PAGE and Coomassie staining (Figure A.III.i). A band the expected size of 10xH-YFP-BfUbb (36.5kDa) can be seen in both the uninduced and induced culture lanes.

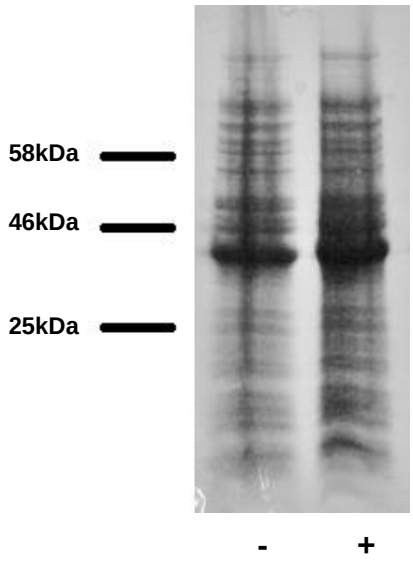


Figure A.III.i Expression of 10xH-YFP-BfUbb in uninduced (-) and induced (+) cultures of *E. coli* DH5α pET19b::10xH-YFP-BfUbb

Appendix IV - Sequencing of constructs in pcDNA4/TO

600ng of pcDNA4/TO-6xH-BfUbb was mixed with 3.2pmole of either Bfhindiii or Brbamhi, 600ng of pcDNA4/TO-YFP-BfUbb was mixed with 3.2pmole of either Yfhindiii or Brbamhi., and 600ng of pcDNA4/TO-YFP was mixed with 3.2pmole of either Yfhindiii or Yrbamhi. The mixtures were sent to the Sequencing and Services Department (Medical Research Council, Dundee, Scotland) for sequencing. The returned sequences were compared to the expected sequence of the construct using Clustal Omega (McWilliam *et al*., 2013) (Figure A.IV.i and Figure A.IV.ii). The 6xH-BfUbb construct was sequenced twice, once including and once excluding the 6xH sequence. The combination of both sequences confirms the full correct sequence of the construct. Where there are mismatches on the forward sequence of the YFP-BfUbb and the YFP constructs, the sequence can be correctly identified on the reverse sequence.

A

```
pcDNA4/TO::6xH-BfUbb     -------------------------AAGCTTCACCACCATCACCACCATATCGGCATGCA
Result                   CCCCCCCCGGAATCTAGCGTTTAACTTAGCTTCACCACCATCACCACCATATCGGCATGCA
                                                 :**********************************

pcDNA4/TO::6xH-BfUbb     AGTTTTTTATAAAAAACAGATATGGCTGGACCATAACATTAGAGGTATCACCTACTGATAC
Result                   AGTTTTTTATAAAAAACAGATATGGCTGGACCATAACATTACAGGGATCAGCTACTGATAC
                         **************************************** *** **** **********

pcDNA4/TO::6xH-BfUbb     TGTAGAAAACGTAAAACAAAAAATTCAGGATAAAGAAGGTTTTCCACCTGATAAAATAAG
Result                   TGTAGAATGCGGATGG-GGAAAATTCAGGATAAAGAAGGTTTTCCCCTGGATAAAATAAG
                         *******:.** *:.. ..**************************.*  ***********

pcDNA4/TO::6xH-BfUbb     GCTTATATATGGAGGAAAACAAATGGAAGATGGACGAACTTTAGCAGATTATAATGTTCA
Result                   GCTGATAAATGCATTAACCTAAATGGAACAGGGATTATCTTTGTGTGTTACACGTGCTGA
                         *** ***:*** *  **.. ******** * ***  *:****.  :*:*:.:..** * *

pcDNA4/TO::6xH-BfUbb     AAAAGACTCAACTATATTAATTTGCAT-----AAGAGATGTCGACTGTTAAGGGATCC--
Result                   GGAGGACTCATGTATATTCACTTTCGTGTTAGATGTGTAGTCCTCTGCCTGTGGTGCCCA
                         ..*.******: ******.* ** *.*     *:*:*::*** :***  :. **: **
```

B

```
pcDNA4/TO::6xHBfUbbREV    ----------ATGCAAGTTTTTATAAAAAACAGATATGGCTGGACCATAACATTAGAGGT 50
Result                   CCATATCGGCATGCAAGTTTTTATAAAAAACGGATATGGCTGGACCATAACATTAGAGGT 120
                                 ********************* ****************************

pcDNA4/TO::6xHBfUbbREV    ATCACCTACTGATACTGTAGAAAACGTAAAACAAAAAATTCAGGATAAAGAAGGTTTTCC 110
Result                   ATCACCTACTGATACTGTAGAAAACGTAAAACAAAAAATTCAGGATAAAGAAGGTTTTCC 180
                         ************************************************************

pcDNA4/TO::6xHBfUbbREV    ACCTGATAAAATAAGGCTTATATATGGAGGAAAACAAATGGAAGATGGACGAACTTTAGC 170
Result                   ACCTGATAAAATAAGGCTTATATATGGAGGAAAACAAATGGAAGATGGACGAACTTTAGC 240
                         ************************************************************

pcDNA4/TO::6xHBfUbbREV    AGATTATAATGTTCAAAAAGACTCAACTATATTAATTTGCATAAGAGATGTC-------- 222
Result                   AGATTATAATGTTCAAAAAGACTCAACTATATTAATTTGCATAAGAGATGTCGACTGTTA 300
                         ************************************************
```

C

```
pcDNA4/TO::YFP-BfUbb      ---------------------------AAGCTTATGGTGAGCAAGGGCGAGCAGCTGTTC
Result                    CCCCCTCCGGGATCTAGCGTTTAACTTAAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTC
                                                     ********************** ********

pcDNA4/TO::YFP-BfUbb      ACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGC
Result                    ACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGC
                          ************************************************************

pcDNA4/TO::YFP-BfUbb      GTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGC
Result                    GTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGC
                          ************************************************************

pcDNA4/TO::YFP-BfUbb      ACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCTTCGGCTACGGCGTG
Result                    ACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTGGGGCGTG
                          ********************************************* * . **. ******

pcDNA4/TO::YFP-BfUbb      CAGTGCTTCGCCCGCTACCCCGACCACATGCGCCAGCACGACTTCTTCAAGTCCGCCATG
Result                    CAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATG
                          ******************************.. ***************************

pcDNA4/TO::YFP-BfUbb      CCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACC
Result                    CCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACC
                          ************************************************************

pcDNA4/TO::YFP-BfUbb      CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATC
Result                    CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATC
                          ************************************************************

pcDNA4/TO::YFP-BfUbb      GACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCAC
Result                    GACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACGCCATCAGCGAC
                          ************************************************ .**:**** **

pcDNA4/TO::YFP-BfUbb      AACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGC
Result                    AACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGC
                          *************  ****************************  ***************

pcDNA4/TO::YFP-BfUbb      CACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATC
Result                    CACAACATCGAGGACGGCTGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATC
                          ******************:*****************************************

pcDNA4/TO::YFP-BfUbb      GGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCGCCCTGAGC
Result                    GGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGC
                          *********************************************:.*****************

pcDNA4/TO::YFP-BfUbb      AAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGG
Result                    AAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGG
                          ************************************************************

pcDNA4/TO::YFP-BfUbb      ATCACTCTCGGCATGGACGAGCTGTACAAGATGCAAGTTTTTATAAAAAACAGATATGGC
Result                    ATCACTCTCGGCATGGACGAGCTGTACAAGATGCAAGTTTTTATAAAAAACAGATATGGC
                          ************************************************************

pcDNA4/TO::YFP-BfUbb      TGGACCATAACATTAGAGGTATCACCTACTGATACTGTAGAAAACGTAAAACAAAAAATT
Result                    TGGACCATAACATTAGAGGTATCACCTACTGATACTGTAGAAAACGTAAAACAAAAAATT
                          ************************************************************

pcDNA4/TO::YFP-BfUbb      CAGGATAAAGAAGGTTTTCCACCTGATAAAATAAGGCTTATATATGGAGGAAAACAAATG
Result                    CAGGATAAAGAAGGTTTTCCACCTGATAAAATAAGGCTTATATATGGAGGAAAACAAATG
                          ************************************************************

pcDNA4/TO::YFP-BfUbb      GAAGATGGACGAACTTTAGCAGATTATAATGTTCAAAAAGACTCAACTATATTAATTTGC
Result                    GAAGATGGACGAACTTTAGCAGATTATAATGTTCAAAAAGACTCAACTATATTAATTTGC
                          ************************************************************

pcDNA4/TO::YFP-BfUbb      ATAAGAGATGTCGACTGTTAACACTAAGGGATCC--------------------------
Result                    ATAAGAGATGTCGACTGTTAACACTAAGGGATCCACTAGTCCAGTGTGGTGGAATTCTGC
                          ***********************************
```

D

```
pcDNA4/TO::YFP-BfUbbREV      ----------------TTAACAGTCGACATCTCTTATGCAAATTAATATAGTTGAGTCT
Result                       CCGCCAGTGCACCTTAGTGTACAGTCGACATCTCTTATGCAAATTAATATAGTTGAGTCT
                                          * :****************************************

pcDNA4/TO::YFP-BfUbbREV      TTTTGAACATTATAATCTGCTAAAGTTCGTCCATCTTCCATTTGTTTTCCTCCATATATA
Result                       TTTTGAACATTATAATCTGCTAAAGTTCGTCCATCTTCCATTTGTTTTCCTCCATATATA
                             ************************************************************

pcDNA4/TO::YFP-BfUbbREV      AGCCTTATTTTATCAGGTGGAAAACCTTCTTTATCCTGAATTTTTTGTTTTACGTTTTCT
Result                       AGCCTTATTTTATCAGGTGGAAAACCTTCTTTATCCTGAATTTTTTGTTTTACGTTTTCT
                             ************************************************************

pcDNA4/TO::YFP-BfUbbREV      ACAGTATCAGTAGGTGATACCTCTAATGTTATGGTCCAGCCATATCTGTTTTTTATAAAA
Result                       ACAGTATCAGTAGGTGATACCTCTAATGTTATGGTCCAGCCATATCTGTTTTTTATAAAA
                             ************************************************************

pcDNA4/TO::YFP-BfUbbREV      ACTTGCATCTTGTACAGCTCGTCCATGCCGAGAGTGATCCCGGCGGCGGTCACGAACTCC
Result                       ACTTGCATCTTGTACAGCTCGTCCATGCCGAGAGTGATCCCGGCGGCGGTCACGAACTCC
                             ************************************************************

pcDNA4/TO::YFP-BfUbbREV      AGCAGGACCATGTGATCGCGCTTCTCGTTGGGGTCTTTGCTCAGGGCGGACTGGTAGCTC
Result                       AGCAGGACCATGTGATCGCGCTTCTCGTTGGGGTCTTTGCTCAGGGCGGACTGGGTGCTC
                             ********************************************************* :****

pcDNA4/TO::YFP-BfUbbREV      AGGTAGTGGTTGTCGGGCAGCAGCACGGGGCCGTCGCCGATGGGGGTGTTCTGCTGGTAG
Result                       AGGTAGTGGTTGTCGGGCAGCAGCACGGGGCCGTCGCCGATGGGGGTGTTCTGCTGGTAG
                             ************************************************************

pcDNA4/TO::YFP-BfUbbREV      TGGTCGGCGAGCTGCACGCTGCCGTCCTCGATGTTGTGGCGGATCTTGAAGTTCACCTTG
Result                       TGGTCGGCGAGCTGCACGCAGCCGTCCTCGATGTTGTGGCGGATCTTGAAGTTGGCCTTG
                             ******************:********************************* .*****

pcDNA4/TO::YFP-BfUbbREV      ATGCCGTTCTTCTGCTTGTCGGCCATGATATAGACGTTGTGGCTGTTGTAGTTGTACTCC
Result                       ATGCCGTTCTTCTGCTTGTCGGCGGTGATATAGACGTTGTCGCTGATGGCGTTGTACTCC
                             ********************** .*************** ****:** .**********

pcDNA4/TO::YFP-BfUbbREV      AGCTTGTGCCCCAGGATGTTGCCGTCCTCCTTGAAGTCGATGCCCTTCAGCTCGATGCGG
Result                       AGCTTGTGCCCCAGGATGTTGCCGTCCTCCTTGAAGTCGATGCCCTTCAGCTCGATGCGG
                             ************************************************************

pcDNA4/TO::YFP-BfUbbREV      TTCACCAGGGTGTCGCCCTCGAACTTCACCTCGGCGCGGGTCTTGTAGTTGCCGTCGTCC
Result                       TTCACCAGGGTGTCGCCCTCGAACTTCACCTCGGCGCGGGTCTTGTAGTTGCCGTCGTCC
                             ************************************************************

pcDNA4/TO::YFP-BfUbbREV      TTGAAGAAGATGGTGCGCTCCTGGACGTAGCCTTCGGGCATGGCGGACTTGAAGAAGTCG
Result                       TTGAAGAAGATGGTGCGCTCCTGGACGTAGCCTTCGGGCATGGCGGACTTGAAGAAGTCG
                             ************************************************************

pcDNA4/TO::YFP-BfUbbREV      TGCTGGCGCATGTGGTCGGGGTAGCGGGCGAAGCACTGCACGCCGTAGCCGAAGGTGGTC
Result                       TGCTGCTTCATGTGGTCGGGGTAGCGGGCGAAGCACTGCACGCCCCAGGTCAGGGTGGTC
                             *****   ************************************  **   *.*******

pcDNA4/TO::YFP-BfUbbREV      ACGAGGGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAGC
Result                       ACGAGGGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAGC
                             ************************************************************

pcDNA4/TO::YFP-BfUbbREV      TTGCCGTAGGTGGCATCGCCCTCGCCCTCGCCGGACACGCTGAACTTGTGGCCGTTTACG
Result                       TTGCCGTAGGTGGCATCGCCCTCGCCCTCGCCGGACACGCTGAACTTGTGGCCGTTTACG
                             ************************************************************

pcDNA4/TO::YFP-BfUbbREV      TCGCCGTCCAGCTCGACCAGGATGGGCACCACCCCGGTGAACAGCTCCTCGCCCTTGCTC
Result                       TCGCCGTCCAGCTCGACCAGGATGGGCACCACCCCGGTGAACAGCTCCTCGCCCTTGCTC
                             ************************************************************

pcDNA4/TO::YFP-BfUbbREV      ACCAT-------------------------------------------------------
Result                       ACCATAAGCTTAAGTTTAAACGCTAGAGTCCGGAGGCTGGATCGGTCCCGGTGTCTTCTA
                             +++++
```

E

```
pcDNA4/TO::YFP      ---------------------------AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTC
Result              CCACCCTCCGGATCTAGCGTTTAACTTAAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTC
                                               *********************************

pcDNA4/TO::YFP      ACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGC
Result              ACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGC
                    ************************************************************

pcDNA4/TO::YFP      GTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGC
Result              GTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGC
                    ************************************************************

pcDNA4/TO::YFP      ACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCTTCGGCTACGGCGTG
Result              ACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTGGGGCGTG
                    ******************************************** * . **. ******

pcDNA4/TO::YFP      CAGTGCTTCGCCCGCTACCCCGACCACATGCGCCAGCACGACTTCTTCAAGTCCGCCATG
Result              CAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATG
                    ******************************.. ***************************

pcDNA4/TO::YFP      CCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACC
Result              CCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACC
                    ************************************************************

pcDNA4/TO::YFP      CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATC
Result              CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATC
                    ************************************************************

pcDNA4/TO::YFP      GACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCAC
Result              GACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACGCCATCAGCGAC
                    ********************************************** .**:**** **

pcDNA4/TO::YFP      AACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGC
Result              AACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGC
                    *************  ****************************  ***************

pcDNA4/TO::YFP      CACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATC
Result              CACAACATCGAGGACGGCTGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATC
                    ******************:*****************************************

pcDNA4/TO::YFP      GGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCGCCCTGAGC
Result              GGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGC
                    ******************************************:.****************

pcDNA4/TO::YFP      AAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGG
Result              AAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGG
                    ************************************************************

pcDNA4/TO::YFP      ATCACTCTCGGCATGGACGAGCTGTACAAGTAAGGATTC----------------------
Result              ATCACTCTCGGCATGGACGAGCTGTACAAGTAAGGATCCACTAGTCCAGTGTGGTGGAAT
                    ************************************ *
```

F

```
pcDNA4/TO::YFPREV      ----------------TTACTTGTACAGCTCGTCCATGCCGAGAGTGATCCCGGCGGCG
Result                 GGGGAACAAGGGGGACCTTACTTGTACAGCTCGTCCATGCCGAGAGTGATCCCGGCGGCG
                                       ****************************************

pcDNA4/TO::YFPREV      GTCACGAACTCCAGCAGGACCATGTGATCGCGCTTCTCGTTGGGGTCTTTGCTCAGGGCG
Result                 GTCACGAACTCCAGCAGGACCATGTGATCGCGCTTCTCGTTGGGGTCTTTGCTCAGGGCG
                       ************************************************************

pcDNA4/TO::YFPREV      GACTGGTAGCTCAGGTAGTGGTTGTCGGGCAGCAGCACGGGGCCGTCGCCGATGGGGGTG
Result                 GACTGGGTGCTCAGGTAGTGGTTGTCGGGCAGCAGCACGGGGCCGTCGCCGATGGGGGTG
                       ****** :***********************************************************

pcDNA4/TO::YFPREV      TTCTGCTGGTAGTGGTCGGCGAGCTGCACGCTGCCGTCCTCGATGTTGTGGCGGATCTTG
Result                 TTCTGCTGGTAGTGGTCGGCGAGCTGCACGCAGCCGTCCTCGATGTTGTGGCGGATCTTG
                       ********************************:*****************************

pcDNA4/TO::YFPREV      AAGTTCACCTTGATGCCGTTCTTCTGCTTGTCGGCCATGATATAGACGTTGTGGCTGTTG
Result                 AAGTTGGCCTTGATGCCGTTCTTCTGCTTGTCGGCGGTGATATAGACGTTGTCGCTGATG
                       ***** .***************************** .*************** ****:**

pcDNA4/TO::YFPREV      TAGTTGTACTCCAGCTTGTGCCCCAGGATGTTGCCGTCCTCCTTGAAGTCGATGCCCTTC
Result                 GCGTTGTACTCCAGCTTGTGCCCCAGGATGTTGCCGTCCTCCTTGAAGTCGATGCCCTTC
                        .************************************************************

pcDNA4/TO::YFPREV      AGCTCGATGCGGTTCACCAGGGTGTCGCCCTCGAACTTCACCTCGGCGCGGGTCTTGTAG
Result                 AGCTCGATGCGGTTCACCAGGGTGTCGCCCTCGAACTTCACCTCGGCGCGGGTCTTGTAG
                       ************************************************************

pcDNA4/TO::YFPREV      TTGCCGTCGTCCTTGAAGAAGATGGTGCGCTCCTGGACGTAGCCTTCGGGCATGGCGGAC
Result                 TTGCCGTCGTCCTTGAAGAAGATGGTGCGCTCCTGGACGTAGCCTTCGGGCATGGCGGAC
                       ************************************************************

pcDNA4/TO::YFPREV      TTGAAGAAGTCGTGCTGGCGCATGTGGTCGGGGTAGCGGGCGAAGCACTGCACGCCGTAG
Result                 TTGAAGAAGTCGTGCTGCTTCATGTGGTCGGGGTAGCGGGCGAAGCACTGCACGCCCCAG
                       *****************   ************************************  **

pcDNA4/TO::YFPREV      CCGAAGGTGGTCACGAGGGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGCAGATGAAC
Result                 GTCAGGGTGGTCACGAGGGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGCAGATGAAC
                          *.*********************************************************

pcDNA4/TO::YFPREV      TTCAGGGTCAGCTTGCCGTAGGTGGCATCGCCCTCGCCCTCGCCGGACACGCTGAACTTG
Result                 TTCAGGGTCAGCTTGCCGTAGGTGGCATCGCCCTCGCCCTCGCCGGACACGCTGAACTTG
                       ************************************************************

pcDNA4/TO::YFPREV      TGGCCGTTTACGTCGCCGTCCAGCTCGACCAGGATGGGCACCACCCCGGTGAACAGCTCC
Result                 TGGCCGTTTACGTCGCCGTCCAGCTCGACCAGGATGGGCACCACCCCGGTGAACAGCTCC
                       ************************************************************

pcDNA4/TO::YFPREV      TCGCCCTTGCTCACCAT----------------------------------------
Result                 TCGCCCTTGCTCACCATAAGCTTAAGTTTAAACGCTAGAGTCCGGAGGCTGGATCGGTCC
                       *****************
```

Figure A.IV.ii Expected sequences of pcDNA4/TO constructs compared with sequencing results. pcDNA4/TO::6xH-BfUbb with (**A**) and without (**B**) his-tag sequence. pcDNA4/TO::YFP-BfUbb forward (**C**) and reverse (**D**) sequences. pcDNA4/TO::YFP forward (**E**) and reverse (**F**) sequences.

Appendix V - qPCR samples and results

Clinical samples from 37 patients with abscesses or GI diseases were screened for the presence of genes *gyrB*, *psa*, *bft* and *ubb* (Table A.V.i).

Patient	Identity	Infection	Disease Details	Sample type	*gyrB*	*psa*	*bft*	*ubb*
1	LS11	*Bacteroides fragilis*	Abdominal abcess	Pus/Blood	485446	0	0	0
2	LS14	*Bacteroides fragilis*	Perianal abcess	Pus/Blood	12936708	0	0	0
3	LS16	*Bacteroides fragilis*	Ischiorecal abcess	Culture	14278	0	0	163921
4	LS18	*Bacteroides fragilis*	Bartholins abcess	Culture	2068	0	937862	0
5	LS27	*Bacteroides fragilis*	Neoplasm abcess	Pus/Blood	276373	0	0	0
6	LS52	*Bacteroides fragilis*	Perianal abcess	Culture	16652970	0	0	0
7	LS54	*Bacteroides fragilis*	Perianal abcess	Pus/Blood	1257490	0	0	0
8	LS66	*Bacteroides fragilis*	Abdominal abcess	Pus/Blood	11365376	0	0	0
9	LS67	*Bacteroides fragilis*	Perianal abcess	Culture	44228	0	0	33452
10	LS75	*Bacteroides fragilis*	Perianal abcess	Culture	6442	0	11955137	0
11	LS80	*Bacteroides fragilis*	Perianal abcess	Culture	4531898	1976672	696	0
12	LS84	*Bacteroides fragilis*	Diverticular abcess	Pus/Blood	16.1	0	0	0
13	LS98	*Bacteroides fragilis*	Blood culture	Culture	6903352	0	2823	0
14	3496 ii	*Bacteroides fragilis*	Diverticulitis: Biopsy ascending colon	Culture	8063952	0	122	1558257
14	3496 vii	*Bacteroides fragilis*		Culture	81816	0	0	201778
14	3497 iii	*Bacteroides fragilis*	Diverticulitis: Biopsy rectum	Culture	2264	0	0	1011
14	3497 vi	*Bacteroides fragilis*		Culture	4835040	0	8	428242
15	3498 ii	*Bacteroides fragilis*	Irritable Bowel Syndrome: Biopsy ascending colon	Culture	976869	0	0	0
15	3498 vi	*Bacteroides fragilis*		Culture	56.15	0	0	0
15	3498 ix	*Bacteroides fragilis*		Culture	1257490	0	2473146	0
15	3498 x	*Bacteroides fragilis*		Culture	858216	0	4139661	53.6
15	3499 vi	*Bacteroides fragilis*	Irritable Bowel Syndrome: Biopsy rectum	Culture	898008	0	2808798	0
15	3499 x	*Bacteroides fragilis*		Culture	56201	0	355654	0
16	3528 vii	*Bacteroides fragilis*	Quiescent Ulcerative Colitis (pancolitis): Biopsy ascending colon	Culture	8222122	0	6443124	0

16	3529 iii	*Bacteroides fragilis*	Quiescent Ulcerative Colitis (pancolitis): Biopsy rectum	Culture	3683801	0	0	0
16	3529 iv	*Bacteroides fragilis*		Culture	1171042	0	9920461	0
17	3532 x	*Bacteroides fragilis*	Cancer of ascending colon: Biopsy ascending colon	Culture	1009013	0	0	0
17	3533 v	*Bacteroides fragilis*	Cancer of ascending colon: Biopsy rectum	Culture	1772313	0	48	0
17	3533 viii	*Bacteroides fragilis*		Culture	455009	0	23	0
18	3534 ii	*Bacteroides fragilis*	Quiescent Ulcerative Colitis (extensive): Biopsy ascending colon	Culture	10	0	23	0
18	3534 vi	*Bacteroides fragilis*		Culture	0	0	0	0
18	3535 v	*Bacteroides fragilis*	Quiescent Ulcerative Colitis (extensive): Biopsy rectum	Culture	4893	0	180	0
19	3536 i	*Bacteroides fragilis*	Active Ulcerative Colitis (pancolitis): Biopsy ascending colon	Culture	3050	0	11	0
19	3536 ii	*Bacteroides fragilis*		Culture	203	0	0	0
19	3536 vii	*Bacteroides fragilis*		Culture	3254	0	54	0
19	3536 ix	*Bacteroides fragilis*		Culture	628	0	0	0
19	3536 x	*Bacteroides fragilis*		Culture	897	0	9	0
19	3537 ii	*Bacteroides fragilis*	Active Ulcerative Colitis (pancolitis): Biopsy rectum	Culture	4055	0	7	0
20	LS15	*Bacteroides fragilis*	Colostomy patient	Culture	2124596	0	0	15
21	EOC	Unknown	Unknown	Culture	223203	5841857	0	0
22	VH	Unknown	Unknown	Culture	369859	4596319	0	0
23	RS	Unknown	Unknown	Culture	13.34	0	0	0
24	AI	Unknown	Unknown	Culture	13020743	0	0	0
25	CN	Unknown	Unknown	Culture	7.89	0	0	0
26	BE1	Unknown	Unknown	Culture	12125616	0	0	0
27	BE3	Unknown	Unknown	Culture	15710272	0	0	0
28	GNAB85	Unknown	Unknown	Culture	3707731	0	0	0
29	GNAB92	Unknown	Unknown	Culture	1459421	0	0	0
30	JC6	Unknown	Unknown	Culture	7223438	0	0	0
31	JC17	Unknown	Unknown	Culture	2679.8	0	0	0
32	JC19	Unknown	Unknown	Culture	176.6	0	0	0
33	NCTC 9344	Unknown	Unknown	Culture	915622	0	0	0
34	NCTC 10583	Unknown	Unknown	Culture	21.4	0	0	0
35	NCTC 10584	Unknown	Unknown	Culture	11892353	0	0	0

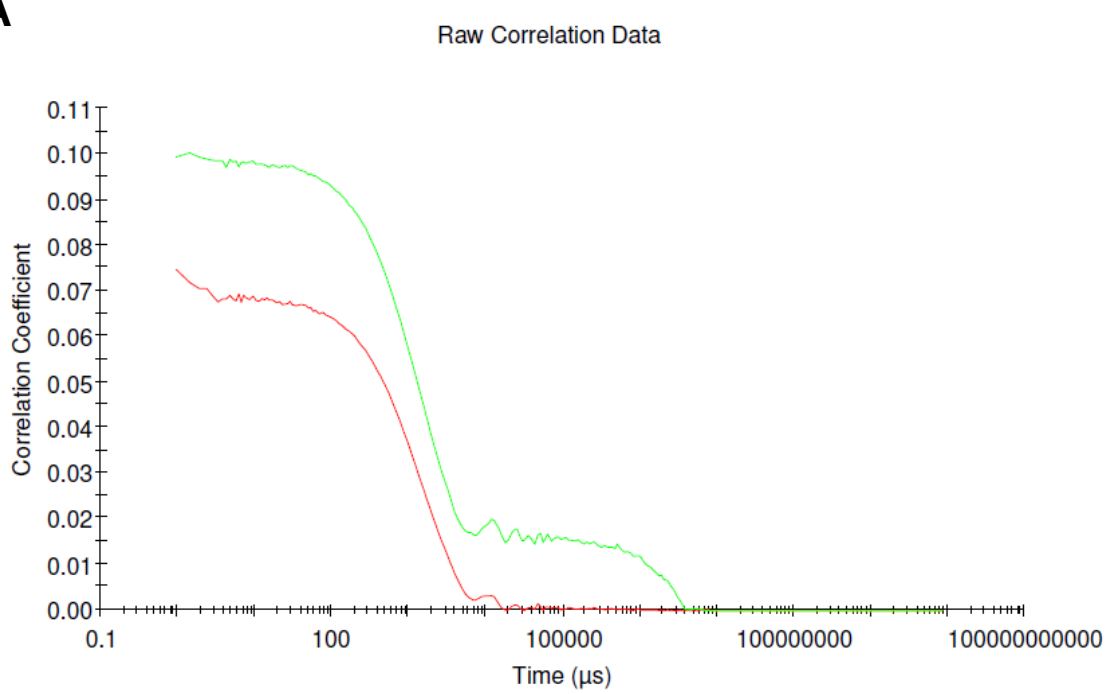
36	ATCC 23745	Unknown	Unknown	Culture	4502649	0	9	0
37	ATCC 29765	Unknown	Unknown	Culture	17998473	0	5	0

Table A.V.i Details of clinical samples used in the qPCR study, including arbitrarily assigned patient number, sample ID, bacterial species, associated disease, sample type, and the number of copies of *gyrB*, *psa*, *bft* and *ubb* in each sample.

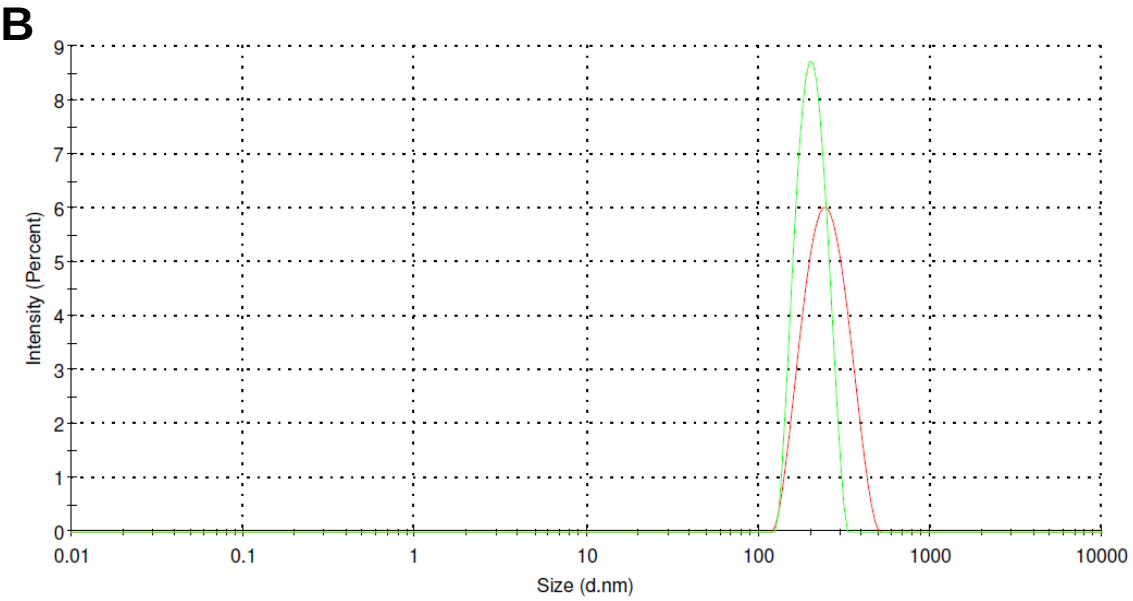
Appendix VI - DLS measurements of WT *B. fragilis* concentrated supernatant

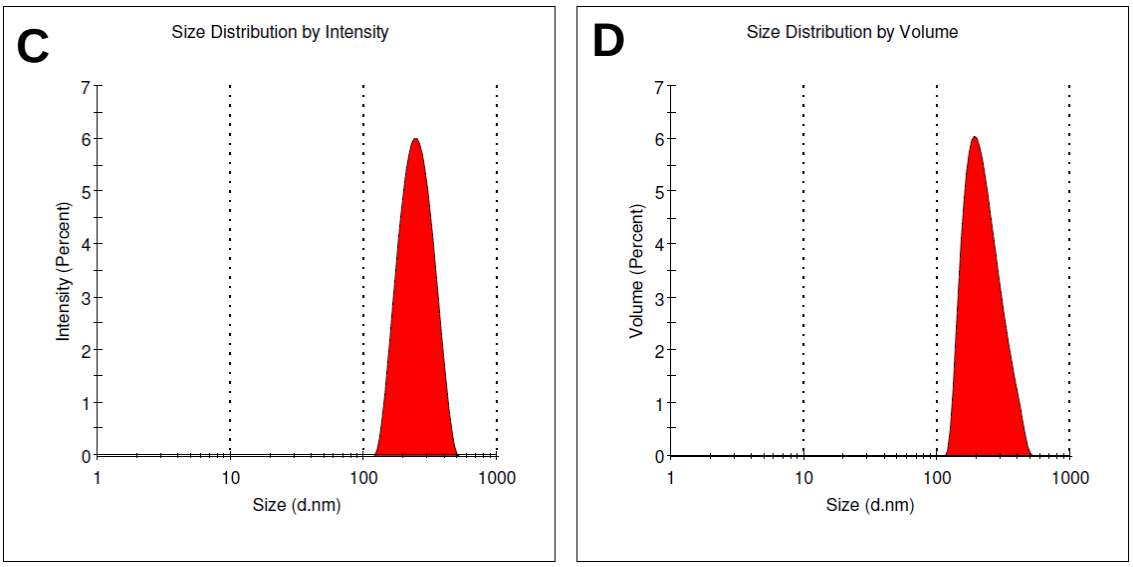
DLS was performed as described in section 2.7.9. There was little consistency between readings 1 (green) and 2 (red), with an apparent increase in average diameter between readings (Figure A.VI.i A and B). The third reading was excluded due to low quality (not shown). The intensity and mass distributions both indicate an average diameter of ~250nm (Figure A.VI.i C and D).

A


Raw Correlation Data
Correlation Coefficient
0.11
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0.1
100
100000
100000000
100000000000
Time (µs)

B


Intensity (Percent)
9
8
7
6
5
4
3
2
1
0
0.01
0.1
1
10
100
1000
10000
Size (d.nm)



E

Intensity Distribution Results

	%Int	Mean (d.n...	Mode (d.n...	StD (d.nm...	%Pd	MW (kDa)*
Peak 1:	100.0	252.9	236.9	69.79	27.6	2.79e5
Peak 2:	0.0	0.000	0.000	0.000	0	0.00
Peak 3:	0.0	0.000	0.000	0.000	0	0.00

Mass Distribution Results

	%Mass	Mean (d.n...	Mode (d.n...	StD (d.nm...	%Pd	MW (kDa)*
Peak 1:	100.0	225.9	188.0	69.94	31.0	2.14e5
Peak 2:	0.0	0.000	0.000	0.000	0	0.00
Peak 3:	0.0	0.000	0.000	0.000	0	0.00

Figure A.VII.i DLS analysis of concentrated supernatant of *B. fragilis* OMV.

Appendix VII - Proteins identified in the OMV of *B. fragilis*

188 proteins were identified in the OMV of *B. fragilis* (Figure A.VII.i). 88 of these proteins were enriched in the OMV, 84 were enriched in the PP and 16 were not significantly enriched in either sample.

NCTC 9343 Accession No.	Confidence score	Fold Change OMV/PP	Description	BLASTP Notes	Functional group	Subcellular Origin
YP_210059.1	110.27	301.968	TonB-dependent receptor [Bacteroides fragilis YCH46]		Signal transduction	Outer membrane
YP_210687.1	70.55	87.627	hypothetical protein BF1082 [Bacteroides fragilis YCH46]	Hypothetical protein	Unknown	Unknown location
YP_210300.1	2864.27	87.186	surface membrane protein, partial [Bacteroides fragilis NCTC 9343]	Probable SusC/RagA	Nutrient	Outer membrane
YP_210172.1	304.74	75.377	hypothetical protein BF0506 [Bacteroides fragilis YCH46]	Hypothetical protein	Unknown	Extracellular
YP_210578.1	93.54	55.293	hypothetical protein BF0965 [Bacteroides fragilis YCH46]	Hypoyhetical protein	Unknown	Unknown location
YP_211023.1	372.56	54.761	hypothetical protein BF1363 [Bacteroides fragilis NCTC 9343]	Hypothetical protein	Unknown	Unknown location
YP_210845.1	191.74	51.398	hypothetical protein BF1209 [Bacteroides fragilis YCH46]	Hypothetical protein	Unknown	Unknown location
YP_212392.1	320.23	45.978	hypothetical protein BF2776 [Bacteroides fragilis NCTC 9343]	Hypothetical	Unknown	Unknown location
YP_213166.1	3369.03	45.911	hypothetical protein BF3567 [Bacteroides fragilis NCTC 9343]	Possible xylanase	Nutrient	Extracellular
YP_213613.1	244.2	37.548	hypothetical protein BF4211 [Bacteroides fragilis YCH46]	Unknown OMP	Unknown	Outer membrane
YP_213037.1	229.26	35.095	hypothetical protein BF3629 [Bacteroides fragilis YCH46]	Hypothetical protein	Unknown	Unknown location
YP_211269.1	76.27	31.486	hypothetical protein BF1616 [Bacteroides fragilis YCH46]	Possible ATP/GTP-binding subunit	Unknown	Unknown location
YP_212656.1	899.09	29.364	hypothetical protein BF3042 [Bacteroides fragilis NCTC 9343]	Probable lipoprotein, LigD superfamily	DNA/RNA	Unknown location
YP_211022.1	227.07	27.822	hypothetical protein BF1423 [Bacteroides fragilis YCH46]	Hypothetical protein	Unknown	Unknown location
YP_212584.1	121.26	24.199	hypothetical protein BF3135 [Bacteroides fragilis YCH46]	Hypothetical OM protein	Unknown	Outer membrane
YP_211011.1	145.99	23.952	hypothetical protein BF1415 [Bacteroides fragilis YCH46]	Possible alpha-2-Macroglobulin	Protease inhibitor	Outer membrane
YP_213040.1	587.05	22.292	hypothetical protein BF3632 [Bacteroides fragilis YCH46]	Hypothetical protein	Unknown	Unknown location
YP_211278.1	144.57	22.076	hypothetical protein BF1628 [Bacteroides fragilis YCH46]	Possible peptidase C10, streptopain	Protease/ Peptidase	Unknown location
YP_213780.1	326.59	21.475	TonB-dependent receptor [Bacteroides fragilis YCH46]		Signal transduction	Outer membrane
YP_210306.1	1405.18	21.364	hypothetical protein BF0670 [Bacteroides fragilis YCH46]	Probable SusC/RagA	Nutrient	Outer membrane
YP_209713.1	400.79	21.302	hypothetical protein pBF9343.20c [Bacteroides fragilis NCTC 9343]	Hypothetical protein	Unknown	Unknown location

YP_210305.1	2159.73	21.117	hypothetical protein BF0669 [Bacteroides fragilis YCH46]	Unknown outer membrane protein	Unknown	Outer membrane
YP_210567.1	122.84	19.148	hypothetical protein BF0953 [Bacteroides fragilis YCH46]	Possible starch binding protein	Nutrient	Inner membrane, periplasmic side
YP_211693.1	81.51	19.147	Na(+)-translocating NADH-quinone reductase subunit A [Bacteroides fragilis NCTC 9343]		Transport	Cytoplasm
YP_210010.1	137.79	18.023	hypothetical protein BF0273 [Bacteroides fragilis NCTC 9343]	Hypothetical protein	Unknown	Cytoplasm
YP_210206.1	104	17.275	conserved hypothetical protein [Bacteroides fragilis 3_1_12]	Hypothetical protein	Unknown	Unknown location
YP_213035.1	752.42	16.548	hypothetical protein BF3429 [Bacteroides fragilis NCTC 9343]	Hypothetical protein	Unknown	Outer membrane
YP_210964.1	111.62	15.674	hypothetical protein BF1302 [Bacteroides fragilis NCTC 9343]	Hypothetical protein	Unknown	Cytoplasm
YP_211819.1	296.29	15.190	hypothetical protein BF2139 [Bacteroides fragilis YCH46]	Probable SusC/RagA, OM	Nutrient	Outer membrane
YP_211224.1	629.55	15.125	hypothetical protein BF1567 [Bacteroides fragilis YCH46]	Probable secretion protein/cell surface adhesion	Transport	Outer membrane
YP_209856.1	73.68	14.849	pyrogenic exotoxin B [Bacteroides fragilis YCH46]	Cysteine protease	Protease/ Peptidase	Unknown location
YP_213132.1	261.11	14.384	biopolymer transport protein [Bacteroides fragilis YCH46]		Transport	Inner membrane
YP_213262.1	186.13	13.948	hypothetical protein BF3899 [Bacteroides fragilis YCH46]	Hypothetical protein	Unknown	Unknown location
YP_211519.1	250.1	13.733	hypothetical protein BF1820 [Bacteroides fragilis YCH46]	Hypothetical protein	Unknown	Unknown location
YP_210615.1	803.5	13.218	hypothetical protein BF0922 [Bacteroides fragilis NCTC 9343]	Hypothetical protein	Unknown	Unknown location
YP_210304.1	2390.77	12.771	hypothetical protein BF0594 [Bacteroides fragilis NCTC 9343]	Probable SusC/RagA	Nutrient	Outer membrane
YP_210307.1	1027.72	12.414	hypothetical protein BF0597 [Bacteroides fragilis NCTC 9343]	Probable SusD	Nutrient	Outer membrane
YP_213235.1	195.31	12.321	hypothetical protein BF3871 [Bacteroides fragilis YCH46]	Hypothetical protein	Unknown	Unknown location
YP_212758.1	682.82	11.901	hypothetical protein BF3306 [Bacteroides fragilis YCH46]	Probable SusD	Nutrient	Outer membrane
YP_213039.1	476.61	11.901	hypothetical protein BF3631 [Bacteroides fragilis YCH46]	Hypothetical protein	Unknown	Unknown location
YP_211637.1	496.34	11.013	TonB-dependent outer membrane protein [Bacteroides fragilis YCH46]	Receptor	Signal transduction	Outer membrane
YP_210296.1	917.36	10.996	hypothetical protein BF0662 [Bacteroides fragilis YCH46]	Possible lipoprotein	Unknown	Inner membrane
YP_213818.1	901.33	10.795	hypothetical protein BF4460 [Bacteroides fragilis YCH46]	Probable SusC/RagA	Nutrient	Outer membrane
YP_211590.1	1638.1	10.790	hypothetical protein BF1957 [Bacteroides fragilis NCTC 9343]	Possible starch binding protein	Nutrient	Outer membrane
YP_211589.1	2819.92	10.494	hypothetical protein BF1894 [Bacteroides fragilis YCH46]	Probable SusC/RagA	Nutrient	Outer membrane
YP_213038.1	1713.53	10.237	hypothetical protein BF3432 [Bacteroides fragilis NCTC 9343]	Possible xylanase, possible lipoprotein	Nutrient	Outer membrane
YP_210444.1	163.88	10.166	UDP-GlcNAc 2-epimerase [Bacteroides fragilis NCTC 9343]		Biosynthesis	Cytoplasm
YP_213878.1	117.35	9.692	outer membrane protein [Bacteroides fragilis NCTC 9343]	Probable SusD/RagB	Nutrient	Unknown location
YP_213581.1	537.86	9.390	50S ribosomal protein L4 [Bacteroides fragilis YCH46]		DNA/RNA	Cytoplasm

YP_210073.1	102.78	9.343	hypothetical protein BF0393 [Bacteroides fragilis YCH46]	Unknown membrane protein	Unknown	Unknown location
YP_213149.1	328.99	9.083	hypothetical protein BF3759 [Bacteroides fragilis YCH46]	Probable OmpB	Unknown	Outer membrane
YP_211674.1	203.73	8.448	hypothetical protein BF2045 [Bacteroides fragilis NCTC 9343]	Hypothetical protein	Unknown	Outer membrane
YP_211583.1	598.25	8.441	conserved hypothetical protein [Bacteroides fragilis 3_1_12]	Hypothetical protein	Unknown	Unknown location
YP_210677.1	539.6	8.348	hypothetical protein BF1073 [Bacteroides fragilis YCH46]	Unknown outer membrane protein	Unknown	Outer membrane
YP_213779.1	586.47	8.292	hypothetical protein BF4417 [Bacteroides fragilis YCH46]	Possible Tol-pal protein	Transport	Outer membrane
YP_213865.1	362.03	7.655	peptidylprolyl isomerase [Bacteroides fragilis YCH46]		Protein folding	Inner membrane, periplasmic side
YP_213819.1	787.72	7.603	hypothetical protein BF4461 [Bacteroides fragilis YCH46]	SusD-like	Nutrient	Outer membrane
YP_210170.1	257.78	7.177	outer membrane protein [Bacteroides fragilis NCTC 9343]		Unknown	Outer membrane
YP_212741.1	592.26	7.095	propionyl-CoA carboxylase subunit beta [Bacteroides fragilis YCH46]		Metabolic	Cytoplasm
YP_212757.1	529.07	6.881	hypothetical protein BF3305 [Bacteroides fragilis YCH46]	Probable SusF	Nutrient	Extracellular
YP_212330.	145.36	6.345	ribonucleotide-diphosphate reductase subunit alpha [Bacteroides fragilis YCH46]	SirA	DNA/RNA	Cytoplasm
YP_212258.1	427.51	5.867	hypothetical protein BF2635 [Bacteroides fragilis NCTC 9343]	Possible peptidase S41	Protease/ Peptidase	Unknown location
YP_213107.1	62.68	5.854	glucose-1-phosphate thymidylyltransferase [Bacteroides fragilis YCH46]		Metabolic	Cytoplasm
YP_211325.1	1629.32	5.849	hypothetical protein BF1680 [Bacteroides fragilis YCH46]	Hypothetical	Unknown	Unknown location
YP_211557.1	286.06	5.705	hypothetical protein BF1924 [Bacteroides fragilis NCTC 9343]	Hypothetical protein	Unknown	Inner membrane, periplasmic side
YP_210087.1	178.49	5.690	FKBP-type peptidylprolyl isomerase [Bacteroides fragilis YCH46]		Protein folding	Inner membrane, periplasmic side
YP_211041.1	597.21	5.457	hypothetical protein BF1383 [Bacteroides fragilis NCTC 9343]	Hypothetical protein	Unknown	Unknown location
YP_213397.1	1313.04	5.015	tryptophanyl-tRNA synthetase [Bacteroides fragilis YCH46]		DNA/RNA	Cytoplasm
YP_213328.1	85.57	4.818	conserved hypothetical protein [Bacteroides fragilis 3_1_12]	Possible biopolymer transport/protein channel protein	Transport	Inner membrane
YP_209830.1	359.41	4.595	conserved hypothetical protein [Bacteroides fragilis 3_1_12]	Hypothetical protein	Unknown	Unknown location
YP_213263.1	128.6	4.558	exopolyphosphatase [Bacteroides fragilis YCH46]		Metabolic	Cytoplasm
YP_212625.1	648.93	4.538	hypothetical protein BF3011 [Bacteroides fragilis NCTC 9343]	Probable peptidase S46	Protease/ Peptidase	Unknown location, not cytoplasmic
YP_211146.1	427.35	4.275	hypothetical protein BF1500 [Bacteroides fragilis NCTC 9343]	Hypothetical protein	Unknown	Unknown location
YP_211327.1	357.49	4.253	major outer membrane protein OmpA [Bacteroides fragilis YCH46]	OMP	Transport	Outer membrane
YP_211064.1	642.26	4.152	peptidase [Bacteroides fragilis YCH46]	Related to peptidase C69	Protease/ Peptidase	Unknown location, not cytoplasmic
YP_213143.1	210.31	4.080	citrate synthase [Bacteroides fragilis NCTC 9343]		Biosynthesis	Cytoplasm
YP_210506.1	645.92	4.052	glycosylhydrolase [Bacteroides fragilis YCH46]	Probable alpha-L-fucosidase	Metabolic	Unknown location

YP_212164.1	97.92	3.932	cobalt chelatase [Bacteroides fragilis NCTC 9343]		Biosynthesis	Cytoplasm
YP_213364.1	224.81	3.345	subtilisin-like serine protease [Bacteroides fragilis YCH46]		Protease/ Peptidase	Unknown location, not cytoplasmic
YP_210980.1	135.16	3.102	redox-sensitive transcriptional activator [Bacteroides fragilis 3_1_12]		Stress response	Cytoplasm
YP_210999.1	79.38	2.977	acetyl esterase [Bacteroides fragilis YCH46]	XynC	Metabolic	Inner membrane, periplasmic side
YP_212189.1	287.15	2.914	hypothetical protein BF2536 [Bacteroides fragilis YCH46]	Hypothetical protein	Unknown	Unknown location
YP_210396.1	195.61	2.881	lytic murein transglycosylase [Bacteroides fragilis NCTC 9343]		PG cleavage	Unknown location
YP_210628.1	2408.44	2.634	zinc protease [Bacteroides fragilis NCTC 9343]	Related to peptidase M16	Protease/ Peptidase	Periplasm
YP_210985.1	2411.22	2.315	formate acetyltransferase [Bacteroides fragilis YCH46]		Metabolic	Cytoplasm
YP_212118.1	2989.98	2.237	hypothetical protein BF2494 [Bacteroides fragilis NCTC 9343]	Hypothetical protein	Unknown	Periplasm
YP_210678.1	2677.4	2.182	hypothetical protein BF1074 [Bacteroides fragilis YCH46]	Hypothetical protein	Unknown	Unknown location
YP_212494.1	63.9	2.127	hypothetical protein BF2880 [Bacteroides fragilis NCTC 9343]	Possible ATPase AAA	ATPase	Cytoplasm

Table A.VII.i The 88 proteins significantly enriched in the OMV of *B. fragilis*. Shown is the accession number in NCTC 9343, the Mascot confidence score, the fold increase in OMV over PP, the identity of the protein as determined by Mascot, the identity of the protein as determined by BLAST, the predicted functional group of the protein and the predicted subcellular origin of the protein. Proteases are highlighted in yellow.

Accession No.	Confidence Score	Mascot Predicted Identity	Average Abundance
YP_097602.1	971.09	hypothetical protein BF0319 [Bacteroides fragilis YCH46]	268917993.7
YP_101476.1	3830.51	elongation factor Tu [Bacteroides fragilis YCH46]	192885908.6
YP_099239.1	894.28	hypothetical protein BF1957 [Bacteroides fragilis YCH46]	167747374.9
YP_101460.1	4950.24	elongation factor G [Bacteroides fragilis YCH46]	164439714.3
YP_099848.1	1175.71	phenylalanyl-tRNA synthetase subunit beta [Bacteroides fragilis YCH46]	151078016.7
YP_100577.1	2484.25	fructose-bisphosphate aldolase [Bacteroides fragilis YCH46]	109236286.4
YP_100610.1	6822.2	pyruvate-flavodoxin oxidoreductase [Bacteroides fragilis YCH46]	91176650.31
YP_098251.1	2993.35	glyceraldehyde 3-phosphate dehydrogenase [Bacteroides fragilis YCH46]	72189970.59
YP_098327.1	3259.65	30S ribosomal protein S1 [Bacteroides fragilis YCH46]	61955437.23
YP_101034.1	2672.24	ketol-acid reductoisomerase [Bacteroides fragilis YCH46]	52041183.3
YP_101717.1	4358.11	phosphoenolpyruvate carboxykinase [Bacteroides fragilis YCH46]	40692256.84
ZP_07811280.1	2006.14	two-component system response regulator [Bacteroides fragilis 3_1_12]	37039824.25
YP_100507.1	910.2	lipoprotein [Bacteroides fragilis YCH46]	33887483.49
YP_099179.1	1256.72	OmpA family outer membrane protein [Bacteroides fragilis YCH46]	32911486.23
YP_101286.1	2795.66	elongation factor Ts [Bacteroides fragilis YCH46]	31652651.88

YP_097970.1	782.16	ribosome recycling factor [Bacteroides fragilis YCH46]	29806455.33
YP_101472.1	771	50S ribosomal protein L1 [Bacteroides fragilis YCH46]	29563233.59
YP_212870.1	956.07	hypothetical protein BF3256 [Bacteroides fragilis NCTC 9343]	27740151.49
YP_213015.1	2824.92	glutamate dehydrogenase [Bacteroides fragilis NCTC 9343]	27020265.58
YP_097891.1	3591.9	bifunctional aspartokinase I/homoserine dehydrogenase I [Bacteroides fragilis YCH46]	25855278.22
YP_100485.1	1723.63	hypothetical protein BF3206 [Bacteroides fragilis YCH46]	25474155.36
YP_098509.1	2996.26	molecular chaperone DnaK [Bacteroides fragilis YCH46]	25191319.63
YP_100486.1	2810.87	hypothetical protein BF3207 [Bacteroides fragilis YCH46]	24710100.22
YP_097785.1	1513.18	outer membrane protein OmpH [Bacteroides fragilis YCH46]	24224412.94
YP_099822.1	175.04	RNA methyltransferase [Bacteroides fragilis YCH46]	23755364.08
YP_098935.1	1918.9	galactokinase [Bacteroides fragilis YCH46]	23547600.66
YP_212794.1	960.38	hypothetical protein BF3181 [Bacteroides fragilis NCTC 9343]	23451753.13
YP_099876.1	1840.64	UDP-glucose 4-epimerase [Bacteroides fragilis YCH46]	23414909.25
YP_212720.1	3492.98	phosphoglycerate kinase [Bacteroides fragilis NCTC 9343]	23367210.12
YP_210264.1	2551.72	TonB dependent receptor [Bacteroides fragilis NCTC 9343]	22716217.33
YP_100395.1	2498.67	glucosamine-6-phosphate deaminase-like protein [Bacteroides fragilis YCH46]	22475397.77
YP_100698.1	1708.04	adenylosuccinate synthetase [Bacteroides fragilis YCH46]	21794176.59
YP_100882.1	2047.16	malic enzyme [Bacteroides fragilis YCH46]	19951901.99
YP_098481.1	774.13	50S ribosomal protein L25 [Bacteroides fragilis YCH46]	18238734.51
YP_100330.1	1783.92	aspartokinase [Bacteroides fragilis YCH46]	17845209.3
YP_098944.1	829.9	hypothetical protein BF1662 [Bacteroides fragilis YCH46]	17437909.67
YP_100703.1	762.01	AraC family transcriptional regulator [Bacteroides fragilis YCH46]	16549885.51
YP_097878.1	2352.91	aspartate aminotransferase [Bacteroides fragilis YCH46]	15829428.36
YP_098472.1	1681.01	phosphopyruvate hydratase [Bacteroides fragilis YCH46]	14029105.33
YP_100457.1	594.57	rubrerythrin [Bacteroides fragilis YCH46]	13729475.15
YP_213596.1	897.71	50S ribosomal protein L10 [Bacteroides fragilis NCTC 9343]	13481312.66
YP_212348.1	588.59	30S ribosomal protein S16 [Bacteroides fragilis NCTC 9343]	13392593.14
YP_101232.1	1773.07	triosephosphate isomerase [Bacteroides fragilis YCH46]	13134421.53
YP_097958.1	2514.44	anaerobic fumarate hydratase class I [Bacteroides fragilis YCH46]	12787165.12
YP_210853.1	1104.32	hypothetical protein BF1184 [Bacteroides fragilis NCTC 9343]	12295134.73
YP_100642.1	1194.66	aminopeptidase [Bacteroides fragilis YCH46]	12126267.38
YP_100673.1	2772.58	molecular chaperone GroEL [Bacteroides fragilis YCH46]	12110789.09
YP_101254.1	1914.97	malate dehydrogenase [Bacteroides fragilis YCH46]	11997683.14
YP_101439.1	712.79	50S ribosomal protein L15 [Bacteroides fragilis YCH46]	11558038.21
YP_101445.1	221.39	30S ribosomal protein S14 [Bacteroides fragilis YCH46]	11177089.19
YP_100379.1	2694.75	diphosphate--fructose-6-phosphate 1-phosphotransferase [Bacteroides fragilis YCH46]	11123356.69
YP_212085.1	1567.87	8-amino-7-oxononanoate synthase [Bacteroides fragilis NCTC 9343]	10876505.12
YP_099970.1	815.34	hypothetical protein BF2686 [Bacteroides fragilis YCH46]	10822073.35
YP_101462.1	546.24	30S ribosomal protein S12 [Bacteroides fragilis YCH46]	10434522.39
YP_101461.1	1029.25	30S ribosomal protein S7 [Bacteroides fragilis YCH46]	10400165.01
YP_101355.1	1836.66	FKBP-type peptidylprolyl isomerase [Bacteroides fragilis YCH46]	10397056
YP_210947.1	1132.23	outer membrane protein [Bacteroides fragilis NCTC 9343]	10101132.98
YP_101470.1	677.61	50S ribosomal protein L7/L12 [Bacteroides fragilis YCH46]	9955730.357
YP_101287.1	2364.85	30S ribosomal protein S2 [Bacteroides fragilis YCH46]	9512015.696

YP_210272.1	1138.47	threonine synthase [Bacteroides fragilis NCTC 9343]	9282428.354
YP_101734.1	1319.27	GTP-binding elongation factor family protein TypA/BipA [Bacteroides fragilis YCH46]	9277554.348
YP_213153.1	1826.71	dihydroxy-acid dehydratase [Bacteroides fragilis NCTC 9343]	9273729.78
YP_101598.1	1932.66	polynucleotide phosphorylase [Bacteroides fragilis YCH46]	9262494.898
YP_100988.1	964.61	dTDP-glucose 4,6-dehydratase [Bacteroides fragilis YCH46]	9033149.954
YP_211784.1	899.61	hypothetical protein BF2161 [Bacteroides fragilis NCTC 9343]	9013351.555
YP_212191.1	1796.9	pyruvate phosphate dikinase [Bacteroides fragilis NCTC 9343]	8930334.66
YP_210113.1	1376.46	aspartate-semialdehyde dehydrogenase [Bacteroides fragilis NCTC 9343]	8858569.431
YP_101454.1	257.25	30S ribosomal protein S19 [Bacteroides fragilis YCH46]	8848063.575
YP_099847.1	703.97	hypothetical protein BF2564 [Bacteroides fragilis YCH46]	8811287.55
YP_097854.1	264.9	hypothetical protein BF0571 [Bacteroides fragilis YCH46]	8280344.706
YP_097855.1	361.51	hypothetical protein BF0572 [Bacteroides fragilis YCH46]	8156735.491
YP_101311.1	2187.4	glutamine-dependent carbamyl phosphate synthetase [Bacteroides fragilis YCH46]	8092319.026
YP_101443.1	717.13	50S ribosomal protein L6 [Bacteroides fragilis YCH46]	8064922.827
YP_101220.1	1302.15	phospho-2-dehydro-3-deoxyheptonate aldolase [Bacteroides fragilis YCH46]	7987226.472
YP_101088.1	2178.27	glucose-6-phosphate isomerase [Bacteroides fragilis YCH46]	7926799.48
YP_099673.1	1122.54	hypothetical protein BF2390 [Bacteroides fragilis YCH46]	7881976.875
YP_101288.1	535.76	30S ribosomal protein S9 [Bacteroides fragilis YCH46]	7727314.601
YP_101449.1	805.75	30S ribosomal protein S17 [Bacteroides fragilis YCH46]	7466340.966
YP_210299.1	1033.94	hypothetical protein BF0589 [Bacteroides fragilis NCTC 9343]	7438333.178
YP_101473.1	619.7	50S ribosomal protein L11 [Bacteroides fragilis YCH46]	7288114.482
YP_101434.1	613.65	30S ribosomal protein S13 [Bacteroides fragilis YCH46]	7216386.074
YP_210271.1	1028.31	cofactor-independent phosphoglycerate mutase [Bacteroides fragilis NCTC 9343]	7166523.793
YP_101756.1	3070.14	hypothetical protein BF4485 [Bacteroides fragilis YCH46]	7066161.122
YP_099485.1	1197.64	serine hydroxymethyltransferase [Bacteroides fragilis YCH46]	7007150.494
YP_213164.1	678.39	hypothetical protein BF3562 [Bacteroides fragilis NCTC 9343]	7006715.464
YP_101458.1	1219.8	50S ribosomal protein L3 [Bacteroides fragilis YCH46]	6996162.402
YP_100028.1	2250.17	serine protease [Bacteroides fragilis YCH46]	6939960.761
YP_209901.1	1099.03	cysteinyl-tRNA synthetase [Bacteroides fragilis NCTC 9343]	6781033.907
YP_213783.1	1163.78	arginyl-tRNA synthetase [Bacteroides fragilis NCTC 9343]	6518690.219
YP_099705.1	1526.35	magnesium chelatase subunit I [Bacteroides fragilis YCH46]	6509437.24
YP_101432.1	939.34	30S ribosomal protein S4 [Bacteroides fragilis YCH46]	6447415.812
YP_101038.1	922.77	acetolactate synthase large subunit [Bacteroides fragilis YCH46]	6307463.994
YP_213085.1	1291.34	oxidoreductase [Bacteroides fragilis NCTC 9343]	6180609.854
YP_098934.1	1971.72	transketolase [Bacteroides fragilis YCH46]	5834782.091
YP_097845.1	1057.29	transcriptional repressor [Bacteroides fragilis YCH46]	5786197.386
YP_212624.1	1112.99	hypothetical protein BF3010 [Bacteroides fragilis NCTC 9343]	5772189.063
YP_210841.1	1978.32	heat shock ClpB protein [Bacteroides fragilis NCTC 9343]	5675915.003
YP_213382.1	1065.41	beta-galactosidase [Bacteroides fragilis NCTC 9343]	5597199.198
YP_209837.1	366.78	dipeptidyl peptidase IV [Bacteroides fragilis NCTC 9343]	5595436.493
YP_211656.1	1490.59	hypothetical protein BF2024 [Bacteroides fragilis NCTC 9343]	5489876.415
YP_210392.1	1792.8	alanyl-tRNA synthetase [Bacteroides fragilis NCTC 9343]	5482603.342
YP_100269.1	996.62	NAD-dependent nucleotide-diphosphate-sugar epimerase [Bacteroides fragilis YCH46]	5444832.242

YP_210849.2	992.68	recombinase A [Bacteroides fragilis NCTC 9343]	5396318.54
YP_211991.1	1810.67	isoleucyl-tRNA synthetase [Bacteroides fragilis NCTC 9343]	5359915.042
YP_101444.1	716.11	30S ribosomal protein S8 [Bacteroides fragilis YCH46]	5343247.123
YP_097545.1	1286.19	transcription elongation factor NusA [Bacteroides fragilis YCH46]	5329099.065
YP_101430.1	1083.37	50S ribosomal protein L17 [Bacteroides fragilis YCH46]	5307324.252
YP_099909.1	655.42	small heat shock protein [Bacteroides fragilis YCH46]	5282399.637
YP_100585.1	1048.45	hypothetical protein BF3307 [Bacteroides fragilis YCH46]	5133896.284
YP_101365.1	604.98	3-oxoacyl-ACP synthase [Bacteroides fragilis YCH46]	5073373.238
YP_101468.1	3206.98	DNA-directed RNA polymerase subunit beta' [Bacteroides fragilis YCH46]	5056283.147
YP_211701.1	1357.36	D-3-phosphoglycerate dehydrogenase [Bacteroides fragilis NCTC 9343]	5046665.14
YP_101446.1	951.86	50S ribosomal protein L5 [Bacteroides fragilis YCH46]	5030961.295
YP_098938.1	1088.66	mannose-6-phosphate isomerase [Bacteroides fragilis YCH46]	4943229.676
YP_100593.1	1302.39	phosphoribosylformylglycinamidine synthase [Bacteroides fragilis YCH46]	4906728.757
YP_213165.1	383.28	hypothetical protein BF3563 [Bacteroides fragilis NCTC 9343]	4882388.143
YP_101292.1	1035.6	asparaginyl-tRNA synthetase [Bacteroides fragilis YCH46]	4874768.656
YP_097309.1	1366.78	coenzyme A transferase [Bacteroides fragilis YCH46]	4858928.202
YP_212278.1	1262.39	carbamoyl-phosphate synthase large subunit [Bacteroides fragilis NCTC 9343]	4844743.859
YP_098586.1	833.99	peptidase T [Bacteroides fragilis YCH46]	4768870.292
YP_097502.1	1397.42	3-oxoacyl-ACP synthase [Bacteroides fragilis YCH46]	4679959.367
YP_101394.1	706.78	3-oxoacyl-ACP reductase [Bacteroides fragilis YCH46]	4677177.891
YP_099299.1	1647.62	phosphoserine aminotransferase [Bacteroides fragilis YCH46]	4674802.849
YP_101363.1	977	GTP-binding protein EngA [Bacteroides fragilis YCH46]	4649634.107
YP_213196.1	1080.56	lysyl-tRNA synthetase [Bacteroides fragilis NCTC 9343]	4544546.325
YP_101138.1	536.89	30S ribosomal protein S6 [Bacteroides fragilis YCH46]	4534846.394
YP_101431.1	932.54	DNA-directed RNA polymerase subunit alpha [Bacteroides fragilis YCH46]	4457712.757
ZP_07811680.1	438.17	50S ribosomal protein L22 [Bacteroides fragilis 3_1_12]	4403507.307
YP_098745.1	734.31	hypothetical protein BF1460 [Bacteroides fragilis YCH46]	4338971.307
YP_098400.1	300.86	hypothetical protein BF1116 [Bacteroides fragilis YCH46]	4273517.385
YP_213523.1	1074.67	hypothetical protein BF3942 [Bacteroides fragilis NCTC 9343]	4129056.842
YP_211657.1	696.53	hypothetical protein BF2025 [Bacteroides fragilis NCTC 9343]	4127073.362
YP_098968.1	1047.14	threonyl-tRNA synthetase [Bacteroides fragilis YCH46]	4068020.477
YP_210692.1	1091.34	hypothetical protein BF1004 [Bacteroides fragilis NCTC 9343]	4043147.962
YP_098483.1	246.83	nitrogen utilization substance protein [Bacteroides fragilis YCH46]	4007139.336
YP_212087.1	1220.25	aspartyl-tRNA synthetase [Bacteroides fragilis NCTC 9343]	3979232.904
YP_100721.1	1047.24	3-isopropylmalate dehydrogenase [Bacteroides fragilis YCH46]	3978215.349
YP_101091.1	683.78	asparagine synthetase AsnA [Bacteroides fragilis YCH46]	3958570.115
YP_098969.1	379.65	translation initiation factor IF-3 [Bacteroides fragilis YCH46]	3933149.458
YP_101469.1	2170.11	DNA-directed RNA polymerase subunit beta [Bacteroides fragilis YCH46]	3882545.214
YP_099717.1	1189.04	prolyl-tRNA synthetase [Bacteroides fragilis YCH46]	3822171.828
YP_213385.1	217.89	hyaluronidase [Bacteroides fragilis NCTC 9343]	3739191.3
YP_100329.1	522.89	diaminopimelate decarboxylase [Bacteroides fragilis YCH46]	3669274.559
YP_213043.1	582.68	peptidase [Bacteroides fragilis NCTC 9343]	3657512.295
ZP_07811678.1	716.11	50S ribosomal protein L2 [Bacteroides fragilis 3_1_12]	3624734.275

YP_101823.1	1813.08	succinate dehydrogenase flavoprotein subunit [Bacteroides fragilis YCH46]	3590469.217
YP_098253.1	1079.51	GMP synthase [Bacteroides fragilis YCH46]	3464300.792
YP_098319.1	780.51	thioredoxin reductase [Bacteroides fragilis YCH46]	3464116.026
YP_101137.1	387.93	30S ribosomal protein S18 [Bacteroides fragilis YCH46]	3418332.244
YP_213724.1	592.56	aminopeptidase [Bacteroides fragilis NCTC 9343]	3385796.651
YP_213794.1	477.64	hypothetical protein BF4230 [Bacteroides fragilis NCTC 9343]	3355611.947
YP_209819.1	1288.17	tyrosyl-tRNA synthetase 1 [Bacteroides fragilis NCTC 9343]	3308116.909
YP_101289.1	350.03	50S ribosomal protein L13 [Bacteroides fragilis YCH46]	3296918.822
YP_097586.1	619.6	cell division protein FtsA [Bacteroides fragilis YCH46]	3245833.147
YP_097695.1	928.7	glycyl-tRNA synthetase [Bacteroides fragilis YCH46]	3172367.22
YP_098403.1	178.29	hypoxanthine guanine phosphoribosyltransferase [Bacteroides fragilis YCH46]	3164025.76
YP_101352.1	1250.84	ATP-dependent DNA helicase RecQ [Bacteroides fragilis YCH46]	3111100.927
YP_211065.1	2330.97	phosphoglucomutase [Bacteroides fragilis NCTC 9343]	3110032.607
YP_212163.1	736.94	iron transport-like protein [Bacteroides fragilis NCTC 9343]	3051399.596
YP_097846.1	281.02	50S ribosomal protein L19 [Bacteroides fragilis YCH46]	3024933.582
YP_101442.1	479.22	50S ribosomal protein L18 [Bacteroides fragilis YCH46]	3004903.796
YP_098294.1	356.65	50S ribosomal protein L21 [Bacteroides fragilis YCH46]	2975557.529
YP_097304.1	518.54	rubrerythrin [Bacteroides fragilis YCH46]	2962376.963
YP_101850.1	774.96	GTP-dependent nucleic acid-binding protein EngD [Bacteroides fragilis YCH46]	2903112.351
YP_098499.1	1123.1	hypothetical protein BF1215 [Bacteroides fragilis YCH46]	2878457.914
YP_101477.1	475.7	sigma-54 modulation protein [Bacteroides fragilis YCH46]	2877103.75
YP_213927.1	1111.27	leucyl-tRNA synthetase [Bacteroides fragilis NCTC 9343]	2840426.977
YP_100822.1	827.01	O-acetylhomoserine (thiol)-lyase [Bacteroides fragilis YCH46]	2837914.702
YP_097344.1	635.17	S-adenosylmethionine synthetase [Bacteroides fragilis YCH46]	2810792.375
YP_098930.1	767.85	2-ketoisovalerate ferredoxin reductase [Bacteroides fragilis YCH46]	2802155.714
YP_101448.1	241.84	50S ribosomal protein L14 [Bacteroides fragilis YCH46]	2792249.134
YP_099924.1	654.17	glycerophosphodiester phosphodiesterase [Bacteroides fragilis YCH46]	2779053.429
YP_213503.1	1232.36	hypothetical protein BF3922 [Bacteroides fragilis NCTC 9343]	2771648.311
YP_097762.1	1199.73	acetate kinase [Bacteroides fragilis YCH46]	2722763.773
YP_100905.1	535.31	hypothetical protein BF3628 [Bacteroides fragilis YCH46]	2721558.888
YP_097341.1	532.92	S-adenosylmethionine:tRNA ribosyltransferase-isomerase [Bacteroides fragilis YCH46]	2715462.912
YP_100566.1	832.82	lactoylglutathione lyase and related protein [Bacteroides fragilis YCH46]	2709259.675
YP_100047.1	843.51	thiol peroxidase [Bacteroides fragilis YCH46]	2669972.119
YP_100851.1	349.87	hypothetical protein BF3574 [Bacteroides fragilis YCH46]	2668121.939
YP_098334.1	899.29	valyl-tRNA synthetase [Bacteroides fragilis YCH46]	2656886.809
YP_101459.1	331.16	30S ribosomal protein S10 [Bacteroides fragilis YCH46]	2617417.188
YP_210673.1	2085.08	preprotein translocase subunit SecA [Bacteroides fragilis NCTC 9343]	2610628.55
YP_213825.1	753.43	S-adenosyl-L-homocysteine hydrolase [Bacteroides fragilis NCTC 9343]	2592580.993
YP_098065.1	1287.26	aminoacyl-histidine dipeptidase [Bacteroides fragilis YCH46]	2563341.165
YP_100725.1	1174.12	2-isopropylmalate synthase [Bacteroides fragilis YCH46]	2542971.844
YP_210004.1	549.78	hypothetical protein BF0267 [Bacteroides fragilis NCTC 9343]	2516741.495
YP_101452.1	658.42	30S ribosomal protein S3 [Bacteroides fragilis YCH46]	2475662.435
YP_213182.1	683.16	peptidase [Bacteroides fragilis NCTC 9343]	2470119.717

YP_098780.1	548.59	hemagglutinin [Bacteroides fragilis YCH46]	2459745.5
YP_210301.1	737.53	outer membrane receptor protein [Bacteroides fragilis NCTC 9343]	2437683.215
YP_213213.1	969.05	ribonucleoside reductase [Bacteroides fragilis NCTC 9343]	2412895.629
YP_100724.1	984.04	isopropylmalate isomerase large subunit [Bacteroides fragilis YCH46]	2394092.727
YP_212180.1	595.2	superoxide dismutase [Bacteroides fragilis NCTC 9343]	2375693.027
YP_099272.1	775	ABC transporter ATP-binding protein [Bacteroides fragilis YCH46]	2367573.181
YP_098928.1	742.03	ketoisovalerate oxidoreductase subunit VorA [Bacteroides fragilis YCH46]	2327170.925
YP_100674.1	681.18	co-chaperonin GroES [Bacteroides fragilis YCH46]	2302498.072
YP_101313.1	481.79	major outer membrane protein OmpA [Bacteroides fragilis YCH46]	2286289.634
YP_099519.1	580.09	hypothetical protein BF2238 [Bacteroides fragilis YCH46]	2271616.19
YP_101483.1	1091.72	glutamyl-tRNA synthetase [Bacteroides fragilis YCH46]	2264831.634
YP_211154.1	429.22	choloylglycine hydrolase [Bacteroides fragilis NCTC 9343]	2231305.198
YP_213188.1	987.61	methylmalonyl-CoA mutase small subunit [Bacteroides fragilis NCTC 9343]	2229108.589
YP_099104.1	374.32	GDP-mannose 4,6-dehydratase [Bacteroides fragilis YCH46]	2185825.076
YP_101076.1	1207.49	methylmalonyl-CoA mutase [Bacteroides fragilis YCH46]	2148034.738
YP_210619.1	961.4	bifunctional 2-polyprenylphenol hydroxylase/glutamate synthase subunit beta [Bacteroides fragilis NCTC 9343]	2136522.755
YP_099029.1	486.98	hypothetical protein BF1748 [Bacteroides fragilis YCH46]	2066987.384
YP_209978.1	1207.96	phosphoglyceromutase [Bacteroides fragilis NCTC 9343]	2043070.427
YP_212115.1	2152.81	heat shock protein 90 [Bacteroides fragilis NCTC 9343]	2040197.374
YP_097816.1	466.16	argininosuccinate synthase [Bacteroides fragilis YCH46]	2028326.428
YP_211702.1	740.54	hypothetical protein BF2074 [Bacteroides fragilis NCTC 9343]	2009880.084
YP_100411.1	320.4	hypothetical protein BF3132 [Bacteroides fragilis YCH46]	2005509.734
YP_100010.1	1192.92	alpha-glucan phosphorylase [Bacteroides fragilis YCH46]	1974465.309
YP_100498.1	613.12	bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/ 5,10-methylene-tetrahydrofolate cyclohydrolase [Bacteroides fragilis YCH46]	1945657.708
YP_098002.1	658.62	magnesium chelatase subunit I [Bacteroides fragilis YCH46]	1941090.933
YP_098216.1	435.81	oxidoreductase [Bacteroides fragilis YCH46]	1862165.236
YP_098871.1	610.16	aspartate aminotransferase [Bacteroides fragilis YCH46]	1855499.658
YP_098196.1	417.74	phosphoribosylformylglycinamidine cyclo-ligase [Bacteroides fragilis YCH46]	1832128.958
YP_101273.1	649.14	branched-chain amino acid aminotransferase [Bacteroides fragilis YCH46]	1808107.407
YP_098256.1	563.98	enoyl-ACP reductase [Bacteroides fragilis YCH46]	1777541.064
YP_099820.1	678.82	hypothetical protein BF2537 [Bacteroides fragilis YCH46]	1712299.868
YP_099539.1	1016.93	malonyl CoA-ACP transacylase [Bacteroides fragilis YCH46]	1659573.62
ZP_07812042.1	479.73	succinate dehydrogenase/fumarate reductase iron-sulfur subunit [Bacteroides fragilis 3_1_12]	1638879.497
YP_101174.1	595.13	phosphoglucomutase [Bacteroides fragilis YCH46]	1599510.163
YP_097877.1	1098.41	GTP cyclohydrolase II [Bacteroides fragilis YCH46]	1585540.537
YP_097580.1	1343.41	DNA gyrase B subunit [Bacteroides fragilis YCH46]	1564621.336
YP_101753.1	903.94	pyruvate kinase [Bacteroides fragilis YCH46]	1542453.339
YP_099382.1	297.37	hypothetical protein BF2101 [Bacteroides fragilis YCH46]	1525925.235
YP_099451.1	638.82	phosphoribosylglycinamide formyltransferase 2 [Bacteroides fragilis YCH46]	1511344.053
YP_211618.1	403.95	outer membrane protein [Bacteroides fragilis NCTC 9343]	1506943.785
YP_101046.1	494.92	peptidyl-prolyl cis-trans isomerase [Bacteroides fragilis YCH46]	1502360.36
YP_101037.1	270.28	acetohydroxyacid synthase small subunit [Bacteroides fragilis YCH46]	1496468.228

YP_101736.1	377.17	transcriptional regulator [Bacteroides fragilis YCH46]	1491651.454
YP_213729.1	1047.65	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase [Bacteroides fragilis NCTC 9343]	1476632.875
YP_100003.1	471.68	RNA-binding protein [Bacteroides fragilis YCH46]	1476408.783
YP_210617.1	513.64	pantoate--beta-alanine ligase [Bacteroides fragilis NCTC 9343]	1467651.652
YP_101712.1	516.19	2-oxoglutarate ferredoxin oxidoreductase subunit beta [Bacteroides fragilis YCH46]	1458361.714
ZP_07807811.1	723.95	adenylate kinase [Bacteroides fragilis 3_1_12]	1444729.378
YP_101694.1	568.51	DNA-binding protein HU-beta [Bacteroides fragilis YCH46]	1442796.117
YP_211643.1	528.51	hypothetical protein BF2011 [Bacteroides fragilis NCTC 9343]	1416558.172
YP_100496.1	998.48	signal recognition protein [Bacteroides fragilis YCH46]	1408990.629
YP_100328.1	611.64	ferritin [Bacteroides fragilis YCH46]	1404318.369
YP_100469.1	657.78	ATP phosphoribosyltransferase [Bacteroides fragilis YCH46]	1403404.551
YP_099459.1	580.06	ATP synthase F0F1 subunit alpha [Bacteroides fragilis YCH46]	1400913.671
YP_210434.1	660.43	DegT/DnrJ/EryC1/StrS aminotransferase family O-antigen-like protein [Bacteroides fragilis NCTC 9343]	1399071.899
YP_099712.1	208.63	50S ribosomal protein L28 [Bacteroides fragilis YCH46]	1382255.665
YP_212294.1	710.68	anthranilate phosphoribosyltransferase [Bacteroides fragilis NCTC 9343]	1369822.703
YP_213737.1	628.25	carboxy-terminal processing protease [Bacteroides fragilis NCTC 9343]	1363083.833
YP_099419.1	784.98	hypothetical protein BF2138 [Bacteroides fragilis YCH46]	1356061.906
YP_099920.1	196.95	carbamoyl phosphate synthase small subunit [Bacteroides fragilis YCH46]	1346260.078
YP_097786.1	497.2	cationic outer membrane protein [Bacteroides fragilis YCH46]	1335493.713
YP_211077.1	245.62	pyridine nucleotide-disulfide oxidoreductase [Bacteroides fragilis NCTC 9343]	1320555.401
YP_097777.1	302.08	ornithine carbamoyltransferase [Bacteroides fragilis YCH46]	1288469.654
YP_098203.1	868.44	phosphoribosylaminoimidazole-succinocarboxamide synthase [Bacteroides fragilis YCH46]	1261303.396
YP_100258.1	638.62	TPR repeat-containing protein [Bacteroides fragilis YCH46]	1257812.146
YP_098595.1	477.61	peptide chain release factor 3 [Bacteroides fragilis YCH46]	1247899.617
YP_099934.1	265.44	tryptophan synthase subunit beta [Bacteroides fragilis YCH46]	1219983.876
YP_212372.1	295.5	hypothetical protein BF2756 [Bacteroides fragilis NCTC 9343]	1213350.227
YP_100544.1	310.6	phenylalanyl-tRNA synthetase subunit alpha [Bacteroides fragilis YCH46]	1203816.867
YP_097585.1	596.48	cell division protein FtsZ [Bacteroides fragilis YCH46]	1202018.251
YP_098268.1	737.82	glucosamine-6-phosphate deaminase [Bacteroides fragilis YCH46]	1196437.485
YP_210072.1	392.98	beta-glucosidase [Bacteroides fragilis NCTC 9343]	1188745.063
ZP_07812118.1	504.88	phosphoribosylaminoimidazole carboxy formyltransferase [Bacteroides fragilis 3_1_12]	1188647.824
YP_101294.1	622.71	adenylosuccinate lyase [Bacteroides fragilis YCH46]	1187604.976
YP_097552.1	404.68	aminotransferase [Bacteroides fragilis YCH46]	1171514.318
YP_100518.1	900.9	aminoacyl-histidine dipeptidase [Bacteroides fragilis YCH46]	1164580.402
YP_213801.1	560.52	2-oxoglutarate ferredoxin oxidoreductase subunit [Bacteroides fragilis NCTC 9343]	1148950.48
YP_101348.1	516.42	peptidyl-prolyl cis-trans isomerase [Bacteroides fragilis YCH46]	1132722.366
YP_100492.1	350.3	transcription termination factor Rho [Bacteroides fragilis YCH46]	1130150.391
YP_100263.1	409.79	DNA polymerase III subunit beta [Bacteroides fragilis YCH46]	1118692.916
YP_101595.1	538.45	histidine triad (HIT) family protein [Bacteroides fragilis YCH46]	1102453.77
YP_099936.1	290.77	anthranilate synthase component II [Bacteroides fragilis YCH46]	1091388.216
YP_097584.1	604.82	hypothetical protein BF0301 [Bacteroides fragilis YCH46]	1079452.017
YP_211225.1	334.38	hypothetical protein BF1582 [Bacteroides fragilis NCTC 9343]	1077223.145

YP_099236.1	566.49	endopeptidase Clp ATP-binding chain B [Bacteroides fragilis YCH46]	1073408.041
YP_101243.1	482.63	pyridoxine 5'-phosphate synthase [Bacteroides fragilis YCH46]	1062703.55
YP_213287.1	452.19	hypothetical protein BF3697 [Bacteroides fragilis NCTC 9343]	1056980.422
YP_099711.1	201.24	50S ribosomal protein L33 [Bacteroides fragilis YCH46]	1047149.606
YP_099580.1	253.22	ATP-dependent RNA helicase [Bacteroides fragilis YCH46]	1046269.68
YP_099547.1	458.94	thioredoxin peroxidase [Bacteroides fragilis YCH46]	1042609.896
YP_098965.1	236.73	peptide deformylase [Bacteroides fragilis YCH46]	1039960.976
YP_210196.1	263.09	acetylornithine aminotransferase [Bacteroides fragilis NCTC 9343]	1035406.848
YP_213853.1	490.58	methionyl-tRNA synthetase [Bacteroides fragilis NCTC 9343]	1030644.143
YP_211940.1	228.65	hypothetical protein BF2318 [Bacteroides fragilis NCTC 9343]	1029081.699
YP_213285.1	443.38	nucleotidyltransferase [Bacteroides fragilis NCTC 9343]	1025390.445
YP_097769.1	252.17	elongation factor P [Bacteroides fragilis YCH46]	1021969.428
YP_097503.1	148.7	acyl carrier protein [Bacteroides fragilis YCH46]	1006735.361
YP_210302.1	592.3	surface membrane protein, partial [Bacteroides fragilis NCTC 9343]	1000603.005
YP_099089.1	222.88	2',3'-cyclic-nucleotide 2'-phosphodiesterase [Bacteroides fragilis YCH46]	990632.1576
YP_099720.1	494.49	protease [Bacteroides fragilis YCH46]	978863.2397
YP_209815.1	356.28	uroporphyrinogen-III synthase [Bacteroides fragilis NCTC 9343]	977952.4079
YP_097793.1	413.04	orotate phosphoribosyltransferase [Bacteroides fragilis YCH46]	970583.8829
YP_213051.1	323.04	hypothetical protein BF3446 [Bacteroides fragilis NCTC 9343]	969677.7864
YP_098943.1	304.65	short-chain dehydrogenase [Bacteroides fragilis YCH46]	965412.3004
YP_099901.1	296.41	hypothetical protein BF2617 [Bacteroides fragilis YCH46]	962417.0045
YP_099448.1	909.96	hypothetical protein BF2167 [Bacteroides fragilis YCH46]	935356.6685
YP_097994.1	562	thioredoxin [Bacteroides fragilis YCH46]	934476.0514
ZP_07809161.1	759.96	ATP synthase F1 sector subunit beta [Bacteroides fragilis 3_1_12]	925536.8536
YP_101451.1	123.69	50S ribosomal protein L16 [Bacteroides fragilis YCH46]	892679.9243
YP_098291.1	1014.87	seryl-tRNA synthetase [Bacteroides fragilis YCH46]	892352.3366
YP_098927.1	579.33	ketoisovalerate oxidoreductase subunit VorA [Bacteroides fragilis YCH46]	891276.5183
YP_098250.1	487.97	peptidyl-dipeptidase [Bacteroides fragilis YCH46]	887176.7016
YP_098754.1	271.03	nicotinate-nucleotide pyrophosphorylase [Bacteroides fragilis YCH46]	870971.9104
YP_212281.1	1024.08	asparagine synthetase B [Bacteroides fragilis NCTC 9343]	868645.1298
YP_210526.1	577.71	orotidine 5'-phosphate decarboxylase [Bacteroides fragilis NCTC 9343]	866317.1541
ZP_07809209.1	247.46	aspartate carbamoyltransferase subunit catalytic [Bacteroides fragilis 3_1_12]	864312.8164
YP_101474.1	438.35	transcription anti-termination protein [Bacteroides fragilis YCH46]	852836.2009
YP_101735.1	161.55	30S ribosomal protein S15 [Bacteroides fragilis YCH46]	843724.5326
YP_099113.1	614.29	phosphoenolpyruvate phosphomutase [Bacteroides fragilis YCH46]	827485.9464
YP_212110.1	533.51	dihydrodipicolinate synthase [Bacteroides fragilis NCTC 9343]	825431.0447
YP_099025.1	390.03	ABC transporter ATP-binding protein [Bacteroides fragilis YCH46]	816177.2159
YP_213224.1	381.48	50S ribosomal protein L9 [Bacteroides fragilis NCTC 9343]	808133.6336
YP_210827.1	577.21	alpha-glucosidase [Bacteroides fragilis NCTC 9343]	807156.8796
YP_099135.1	430.84	6-phosphogluconate dehydrogenase [Bacteroides fragilis YCH46]	800215.1005
YP_098487.1	522.79	hypothetical protein BF1203 [Bacteroides fragilis YCH46]	795715.4552
YP_101221.1	435.33	prephenate dehydratase [Bacteroides fragilis YCH46]	793648.9995
YP_098482.1	118.33	hypothetical protein BF1198 [Bacteroides fragilis YCH46]	790572.9728

YP_098191.1	627.44	bifunctional UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase/(3R)-hydroxymyristoyl-ACP dehydratase [Bacteroides fragilis YCH46]	789699.8611
YP_101594.1	605.63	mannose-1-phosphate guanylyltransferase [Bacteroides fragilis YCH46]	781345.8086
YP_212478.1	683.83	2-amino-3-ketobutyrate CoA ligase [Bacteroides fragilis NCTC 9343]	778680.2362
YP_098837.1	347.18	DNA-binding protein HU [Bacteroides fragilis YCH46]	778460.8046
YP_100650.1	431.07	electron transfer flavoprotein subunit beta [Bacteroides fragilis YCH46]	773279.0131
YP_211187.1	337.06	hypothetical protein BF1542 [Bacteroides fragilis NCTC 9343]	766820.8998
YP_210815.1	343.19	hypothetical protein BF1145 [Bacteroides fragilis NCTC 9343]	764248.2292
YP_213605.1	687.58	peptidase [Bacteroides fragilis NCTC 9343]	763218.5968
YP_100860.1	211.35	arsenate reductase [Bacteroides fragilis YCH46]	761295.0418
YP_211719.1	218.38	hypothetical protein BF2093 [Bacteroides fragilis NCTC 9343]	755287.312
YP_099682.1	293.06	dihydroorotate dehydrogenase electron transfer subunit [Bacteroides fragilis YCH46]	754051.633
YP_101441.1	164.04	30S ribosomal protein S5 [Bacteroides fragilis YCH46]	751623.1435
YP_098587.1	159.22	glycine cleavage system aminomethyltransferase T [Bacteroides fragilis YCH46]	748388.0096
YP_213898.1	171.72	hypothetical protein BF4343 [Bacteroides fragilis NCTC 9343]	747966.5041
YP_213380.1	579.04	ribosomal large subunit pseudouridine synthase [Bacteroides fragilis NCTC 9343]	747635.7772
YP_101353.1	393.95	ATP-dependent protease ATP-binding subunit ClpX [Bacteroides fragilis YCH46]	732615.9332
YP_097737.1	775.69	glutamate decarboxylase [Bacteroides fragilis YCH46]	728343.7799
YP_097761.1	783.1	phosphate acetyltransferase [Bacteroides fragilis YCH46]	720141.7091
YP_100022.1	163.77	periplasmic protein [Bacteroides fragilis YCH46]	719454.2577
YP_098274.1	232.54	aminotransferase [Bacteroides fragilis YCH46]	716989.1943
YP_101716.1	397.53	uracil phosphoribosyltransferase [Bacteroides fragilis YCH46]	712368.2066
YP_099585.1	547.44	ATP-dependent protease [Bacteroides fragilis YCH46]	708734.1573
YP_098469.1	138.2	hypothetical protein BF1185 [Bacteroides fragilis YCH46]	708701.6565
YP_097978.1	520.63	uridylate kinase [Bacteroides fragilis YCH46]	706989.5035
YP_100529.1	379.35	fructose-bisphosphate aldolase [Bacteroides fragilis YCH46]	702664.5636
YP_098459.1	274.62	purine nucleoside phosphorylase II [Bacteroides fragilis YCH46]	701794.4159
YP_213284.1	237.21	nucleotidyltransferase [Bacteroides fragilis NCTC 9343]	700451.7807
YP_099567.1	359.45	hypothetical protein BF2286 [Bacteroides fragilis YCH46]	692769.7064
YP_100652.1	702.25	acyl-CoA dehydrogenase [Bacteroides fragilis YCH46]	691946.0561
YP_101447.1	529.08	50S ribosomal protein L24 [Bacteroides fragilis YCH46]	689075.6584
YP_098303.1	214.11	2-dehydro-3-deoxyphosphooctonate aldolase [Bacteroides fragilis YCH46]	687865.4747
YP_212851.1	467.59	histidyl-tRNA synthetase [Bacteroides fragilis NCTC 9343]	672518.5865
YP_098547.1	168.05	hypothetical protein BF1263 [Bacteroides fragilis YCH46]	665136.8461
YP_099484.1	210.36	hypothetical protein BF2203 [Bacteroides fragilis YCH46]	659611.9601
YP_099814.1	247.89	thiamine biosynthesis protein ThiC [Bacteroides fragilis YCH46]	659014.8795
YP_212505.1	176.95	imidazole glycerol phosphate synthase subunit HisH [Bacteroides fragilis NCTC 9343]	650015.09
YP_098529.1	654.98	catalase [Bacteroides fragilis YCH46]	647268.785
YP_101358.1	304.58	RNA-binding protein RbpA [Bacteroides fragilis YCH46]	636353.2054
YP_100262.1	481.85	DNA polymerase III subunit epsilon [Bacteroides fragilis YCH46]	636100.6892
ZP_07807375.1	635.58	GTP-binding protein [Bacteroides fragilis 3_1_12]	635596.5683
YP_100651.1	270.68	electron transfer flavoprotein subunit alpha [Bacteroides fragilis YCH46]	631843.8348

YP_211817.1	191.35	hypothetical protein BF2194 [Bacteroides fragilis NCTC 9343]	628662.087
YP_101827.1	607.03	TonB-dependent outer membrane protein [Bacteroides fragilis YCH46]	626973.7977
YP_101025.1	411.65	6-phosphofructokinase [Bacteroides fragilis YCH46]	625020.6861
YP_099557.1	563.97	DnaK suppressor protein [Bacteroides fragilis YCH46]	622835.1906
YP_097546.1	871.55	translation initiation factor IF-2 [Bacteroides fragilis YCH46]	619869.0195
YP_213484.1	208.82	GTP-binding protein Era [Bacteroides fragilis NCTC 9343]	615392.328
YP_099927.1	182.56	L,L-diaminopimelate aminotransferase [Bacteroides fragilis YCH46]	610184.0844
YP_097515.1	471.44	nitroreductase [Bacteroides fragilis YCH46]	608799.513
YP_100024.1	283.65	clostripain-related protein [Bacteroides fragilis YCH46]	605063.9945
YP_211284.1	186.21	hypothetical protein BF1646 [Bacteroides fragilis NCTC 9343]	603098.5652
YP_097758.1	347.5	hypothetical protein BF0475 [Bacteroides fragilis YCH46]	597230.6592
YP_101351.1	229.28	inosine-5'-monophosphate dehydrogenase [Bacteroides fragilis YCH46]	586824.841
YP_211645.1	253.17	hypothetical protein BF2013 [Bacteroides fragilis NCTC 9343]	580195.6993
YP_098478.1	104.84	hypothetical protein BF1194 [Bacteroides fragilis YCH46]	578131.9018
YP_210861.2	222.87	tryptophan synthase subunit beta [Bacteroides fragilis NCTC 9343]	576536.9883
YP_212441.1	427.49	dihydroorotate dehydrogenase 2 [Bacteroides fragilis NCTC 9343]	571869.7639
YP_098172.1	328.18	hypothetical protein BF0887 [Bacteroides fragilis YCH46]	571377.2695
YP_097367.1	897.82	phosphoribosylamine--glycine ligase [Bacteroides fragilis YCH46]	558376.5307
YP_099902.1	258.13	hypothetical protein BF2618 [Bacteroides fragilis YCH46]	558100.7946
YP_101253.1	420.61	Tricorn-like protease [Bacteroides fragilis YCH46]	556388.5529
YP_098863.1	355.42	hypothetical protein BF1579 [Bacteroides fragilis YCH46]	541467.2803
YP_099940.1	297.24	tryptophan synthase subunit alpha [Bacteroides fragilis YCH46]	533262.3508
YP_098405.1	413.47	GTPase ObgE [Bacteroides fragilis YCH46]	531639.1953
YP_101092.1	267.07	uracil-DNA glycosylase [Bacteroides fragilis YCH46]	525833.4216
YP_210123.1	574.9	glutaminase [Bacteroides fragilis NCTC 9343]	525796.121
YP_098897.1	283.86	hypothetical protein BF1615 [Bacteroides fragilis YCH46]	511492.1315
YP_100813.1	335.55	inosine 5-monophosphate dehydrogenase [Bacteroides fragilis YCH46]	511269.9082
YP_100332.1	519.01	bifunctional phosphoribosyl-AMP cyclohydrolase/phosphoribosyl-ATP pyrophosphatase protein [Bacteroides fragilis YCH46]	509219.1338
YP_100466.1	337.9	imidazole glycerol-phosphate dehydratase/histidinol phosphatase [Bacteroides fragilis YCH46]	506457.4955
YP_099291.1	217.58	aminopeptidase [Bacteroides fragilis YCH46]	504455.6695
YP_098189.1	380.02	hypothetical protein BF0904 [Bacteroides fragilis YCH46]	504033.1645
YP_101433.1	240.69	30S ribosomal protein S11 [Bacteroides fragilis YCH46]	502750.825
YP_211735.1	134.65	amino acid transferase [Bacteroides fragilis NCTC 9343]	495636.5386
YP_097605.1	533.05	hemolysin A [Bacteroides fragilis YCH46]	491832.7663
YP_098971.1	237.96	50S ribosomal protein L20 [Bacteroides fragilis YCH46]	491125.6955
YP_100730.1	372.05	purine nucleoside phosphorylase [Bacteroides fragilis YCH46]	490356.8602
YP_098267.1	602.81	flavoprotein [Bacteroides fragilis YCH46]	484733.3611
YP_097848.1	171.91	5'-nucleotidase [Bacteroides fragilis YCH46]	478339.5262
YP_098055.1	205.27	stationary phase survival protein SurE [Bacteroides fragilis YCH46]	476184.8985
YP_212860.1	413.4	peptidase [Bacteroides fragilis NCTC 9343]	474134.8709
YP_211343.1	328.24	surface antigen [Bacteroides fragilis NCTC 9343]	468762.3259
YP_098008.1	178.75	aminopeptidase [Bacteroides fragilis YCH46]	466586.4231
YP_212245.1	206.19	hypothetical protein BF2622 [Bacteroides fragilis NCTC 9343]	463300.6083

YP_101861.1	357.87	quinolinate synthetase [Bacteroides fragilis YCH46]	461809.6253
YP_101270.1	219.74	Mrp/Nbp35 family ATP-binding protein [Bacteroides fragilis YCH46]	459708.4507
YP_097609.1	200.79	hypothetical protein BF0326 [Bacteroides fragilis YCH46]	459561.9328
YP_213145.1	720.5	aconitate hydratase [Bacteroides fragilis NCTC 9343]	455260.2405
YP_213305.1	354.45	aminotransferase [Bacteroides fragilis NCTC 9343]	444893.1457
YP_098840.1	90.86	single-strand binding protein [Bacteroides fragilis YCH46]	441349.6306
YP_101787.1	578.06	transcriptional regulator [Bacteroides fragilis YCH46]	436912.1258
YP_100382.1	167.78	hypothetical protein BF3103 [Bacteroides fragilis YCH46]	434164.7074
YP_211962.1	211.6	oxidoreductase [Bacteroides fragilis NCTC 9343]	431006.466
YP_100408.1	168.76	dihydroorotase [Bacteroides fragilis YCH46]	422426.6787
YP_099921.1	513.77	amidophosphoribosyltransferase [Bacteroides fragilis YCH46]	422193.9846
YP_212331.1	282.59	ribonucleotide-diphosphate reductase subunit beta [Bacteroides fragilis NCTC 9343]	420041.8952
YP_211451.1	295.65	beta-galactosidase [Bacteroides fragilis NCTC 9343]	419465.4729
YP_100467.1	84.42	histidinol-phosphate aminotransferase [Bacteroides fragilis YCH46]	419055.8127
YP_099136.1	216.55	glucose-6-phosphate 1-dehydrogenase [Bacteroides fragilis YCH46]	418077.4059
YP_212885.1	314.99	bacterioferritin-like protein [Bacteroides fragilis NCTC 9343]	416500.356
ZP_07810793.1	186.71	conserved hypothetical protein [Bacteroides fragilis 3_1_12]	411212.0593
YP_210407.1	594.23	hypothetical protein BF0704 [Bacteroides fragilis NCTC 9343]	410732.8736
YP_211902.1	628.07	ribose-phosphate pyrophosphokinase [Bacteroides fragilis NCTC 9343]	408107.2562
YP_211660.1	193.48	aminotransferase [Bacteroides fragilis NCTC 9343]	405109.5419
YP_209759.1	211.44	iron transport receptor protein [Bacteroides fragilis NCTC 9343]	404731.5759
YP_213864.1	371.07	FKBP-type peptidylprolyl isomerase [Bacteroides fragilis NCTC 9343]	399676.9425
YP_101479.1	189.24	30S ribosomal protein S21 [Bacteroides fragilis YCH46]	394796.1851
YP_100723.1	478.07	3-isopropylmalate dehydratase small subunit [Bacteroides fragilis YCH46]	393845.2154
YP_101162.1	203.25	glutathione peroxidase [Bacteroides fragilis YCH46]	391986.8826
YP_100264.1	118.43	hypothetical protein BF2982 [Bacteroides fragilis YCH46]	390111.3973
YP_101129.1	310.37	methylglyoxal synthase [Bacteroides fragilis YCH46]	385270.056
YP_100454.1	316.75	thioredoxin [Bacteroides fragilis YCH46]	382717.0411
YP_100025.1	97.59	hypothetical protein BF2741 [Bacteroides fragilis YCH46]	381659.7322
YP_099935.1	399.81	anthranilate synthase component I [Bacteroides fragilis YCH46]	379449.6725
YP_097550.1	379.77	ABC transporter ATP-binding protein [Bacteroides fragilis YCH46]	375342.5155
YP_210158.1	338.34	hypothetical protein BF0436 [Bacteroides fragilis NCTC 9343]	374522.1622
YP_099444.1	353.68	3-methyl-2-oxobutanoate hydroxymethyltransferase [Bacteroides fragilis YCH46]	368650.9465
YP_101492.1	360.06	L-asparaginase II [Bacteroides fragilis YCH46]	367660.1674
YP_209891.1	297.29	outer membrane protein [Bacteroides fragilis NCTC 9343]	362159.3925
YP_212863.1	180.13	hypothetical protein BF3250 [Bacteroides fragilis NCTC 9343]	358945.1765
YP_100717.1	230.79	UDP-glucuronic acid epimerase [Bacteroides fragilis YCH46]	358106.5469
YP_101179.1	408.89	methionyl-tRNA formyltransferase [Bacteroides fragilis YCH46]	356029.3866
YP_210204.1	391.8	malate dehydrogenase [Bacteroides fragilis NCTC 9343]	355617.0374
YP_211257.1	203.45	adenosylmethionine-8-amino-7-oxononanoate aminotransferase [Bacteroides fragilis NCTC 9343]	355586.6289
YP_101596.1	624.03	transcription elongation factor GreA [Bacteroides fragilis YCH46]	354476.7334
YP_101750.1	133.7	ribosome-binding factor A [Bacteroides fragilis YCH46]	344944.8212
YP_100983.1	177.55	hypothetical protein BF3706 [Bacteroides fragilis YCH46]	343873.6974

YP_100731.1	186.08	thiamine-monophosphate kinase [Bacteroides fragilis YCH46]	341664.8285
YP_099839.1	297.4	GTP-binding protein LepA [Bacteroides fragilis YCH46]	338121.8613
ZP_07809463.1	190.36	conserved hypothetical protein [Bacteroides fragilis 3_1_12]	328781.9452
YP_211654.1	243.19	hypothetical protein BF2022 [Bacteroides fragilis NCTC 9343]	326856.9566
YP_209848.1	521.84	deoxyribose-phosphate aldolase [Bacteroides fragilis NCTC 9343]	312227.3473
YP_101818.1	129.33	carboxy-terminal processing protease [Bacteroides fragilis YCH46]	309927.1642
YP_212850.1	222.43	hypothetical protein BF3237 [Bacteroides fragilis NCTC 9343]	309264.0826
YP_099696.1	278.57	universal stress protein UspA [Bacteroides fragilis YCH46]	306991.7818
YP_100497.1	216.15	thiol:disulfide interchange protein [Bacteroides fragilis YCH46]	304176.7317
YP_099024.1	383.67	heat shock protein GrpE [Bacteroides fragilis YCH46]	304006.0864
YP_101437.1	139.95	methionine aminopeptidase [Bacteroides fragilis YCH46]	303193.4545
ZP_07807614.1	291.89	peptide chain release factor 1 [Bacteroides fragilis 3_1_12]	302340.8378
YP_210609.1	239.64	FNR-like protein [Bacteroides fragilis NCTC 9343]	295920.411
YP_212290.1	539.68	alcohol dehydrogenase [Bacteroides fragilis NCTC 9343]	290527.6629
YP_100530.1	495.86	phosphoglyceromutase [Bacteroides fragilis YCH46]	290236.9168
YP_099545.1	238.04	phosphodiesterase I [Bacteroides fragilis YCH46]	289909.6865
YP_100597.1	205.34	hypothetical protein BF3319 [Bacteroides fragilis YCH46]	287945.4839
YP_101389.1	377.58	cyclophilin type peptidyl-prolyl cis-trans isomerase [Bacteroides fragilis YCH46]	287625.3118
YP_097591.1	195.95	UDP-N-acetylmuramoylalanine--D-glutamate ligase [Bacteroides fragilis YCH46]	283942.8164
YP_101036.1	223.44	acyl-ACP thioesterase [Bacteroides fragilis YCH46]	279123.3613
YP_210148.1	340.52	alkaline phosphatase [Bacteroides fragilis NCTC 9343]	274983.2914
YP_097607.1	111.82	deoxyuridine 5'-triphosphate nucleotidohydrolase [Bacteroides fragilis YCH46]	274772.3528
YP_212295.1	234.82	indole-3-glycerol phosphate synthase [Bacteroides fragilis NCTC 9343]	272129.4107
YP_100976.1	89.91	1-deoxy-D-xylulose 5-phosphate reductoisomerase [Bacteroides fragilis YCH46]	270796.9882
YP_100397.1	195.29	xanthan lyase [Bacteroides fragilis YCH46]	270207.0626
YP_101781.1	170.82	UDP-N-acetyl-D-mannosamine dehydrogenase [Bacteroides fragilis YCH46]	269726.3932
YP_101691.1	356.5	DNA topoisomerase I [Bacteroides fragilis YCH46]	266657.5242
YP_098598.1	84.51	haloacid dehalogenase [Bacteroides fragilis YCH46]	265431.0795
YP_098364.1	177.35	4-hydroxythreonine-4-phosphate dehydrogenase [Bacteroides fragilis YCH46]	264675.7635
YP_100026.1	293.08	RNA polymerase sigma factor RpoD [Bacteroides fragilis YCH46]	262749.5482
ZP_07809230.1	664.26	ribose-phosphate pyrophosphokinase [Bacteroides fragilis 3_1_12]	261188.9022
YP_097656.1	147.46	long-chain-fatty-acid-CoA ligase [Bacteroides fragilis YCH46]	257209.0406
YP_100042.1	470.35	glutaminyl-tRNA synthetase [Bacteroides fragilis YCH46]	254026.2875
YP_098602.1	227	naphthoate synthase [Bacteroides fragilis YCH46]	252769.4027
YP_098270.1	178.05	hypothetical protein BF0986 [Bacteroides fragilis YCH46]	252181.2913
YP_212683.1	128.93	hypothetical protein BF3070 [Bacteroides fragilis NCTC 9343]	251660.508
YP_211281.1	146.33	hypothetical protein BF1643 [Bacteroides fragilis NCTC 9343]	251072.0645
YP_099164.1	278.51	glycerate dehydrogenase [Bacteroides fragilis YCH46]	249910.4603
YP_101012.1	151.17	transcriptional regulator [Bacteroides fragilis YCH46]	248847.8942
YP_100391.1	190.19	hypothetical protein BF3112 [Bacteroides fragilis YCH46]	248398.3007
YP_098870.1	201.17	hypothetical protein BF1586 [Bacteroides fragilis YCH46]	248202.6379
YP_098074.1	213.37	arginine/ornithine transport system ATPase [Bacteroides fragilis YCH46]	245547.0853
YP_213028.1	113.95	DNA-binding protein [Bacteroides fragilis NCTC 9343]	243651.0846

YP_212572.1	133.65	hypothetical protein BF2958 [Bacteroides fragilis NCTC 9343]	240925.715
YP_100806.1	294.2	pyruvate carboxylase subunit A [Bacteroides fragilis YCH46]	232572.2861
YP_097811.1	239.1	acetyl-coenzyme A synthetase [Bacteroides fragilis YCH46]	228045.9193
ZP_07811021.1	134.58	conserved hypothetical protein [Bacteroides fragilis 3_1_12]	228022.9089
YP_099886.1	304.07	CTP synthetase [Bacteroides fragilis YCH46]	227111.5428
YP_098286.1	544.83	glycogen synthase [Bacteroides fragilis YCH46]	223998.3584
YP_099524.1	75.17	1,4-alpha-glucan branching enzyme [Bacteroides fragilis YCH46]	223346.187
YP_211315.1	187.47	sulfate adenylyltransferase [Bacteroides fragilis NCTC 9343]	222567.5881
ZP_07810591.1	84.93	thioredoxin [Bacteroides fragilis 3_1_12]	220629.1625
YP_101234.1	178.01	hypothetical protein BF3958 [Bacteroides fragilis YCH46]	218884.8498
YP_101068.1	130.51	chorismate synthase [Bacteroides fragilis YCH46]	217001.4318
YP_212643.1	382.65	histidinol dehydrogenase [Bacteroides fragilis NCTC 9343]	216960.9834
ZP_07809999.1	147.38	conserved hypothetical protein [Bacteroides fragilis 3_1_12]	214998.4943
YP_101798.1	163.45	hypothetical protein BF4527 [Bacteroides fragilis YCH46]	214573.65
ZP_07810468.1	124.43	imidazole glycerol phosphate synthase subunit hisF [Bacteroides fragilis 3_1_12]	213209.2853
YP_209852.1	207.37	adenine phosphoribosyltransferase [Bacteroides fragilis NCTC 9343]	211529.8044
YP_098747.1	159.76	hypothetical protein BF1462 [Bacteroides fragilis YCH46]	210212.716
YP_100036.1	146.13	riboflavin synthase subunit alpha [Bacteroides fragilis YCH46]	209760.5732
YP_097657.1	257.33	peptide chain release factor 2 [Bacteroides fragilis YCH46]	209240.1142
YP_101239.1	140.16	ThiJ family intracellular protease [Bacteroides fragilis YCH46]	207463.8942
YP_211347.1	117.6	LPS-related regulatory protein [Bacteroides fragilis NCTC 9343]	206414.8975
YP_098028.1	254.31	3-dehydroquinate synthase [Bacteroides fragilis YCH46]	199619.3458
YP_099317.1	84.63	xylanase [Bacteroides fragilis YCH46]	198833.3266
YP_213475.1	226.06	ATP-dependent Clp protease proteolytic subunit [Bacteroides fragilis NCTC 9343]	196140.4637
YP_210443.1	212.68	UDP-ManNAc dehydrogenase [Bacteroides fragilis NCTC 9343]	195091.847
YP_099578.1	326.53	UDP-glucose 6-dehydrogenase [Bacteroides fragilis YCH46]	192803.8737
YP_098192.1	113.34	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase [Bacteroides fragilis YCH46]	192495.1459
YP_098960.1	110.85	hypothetical protein BF1679 [Bacteroides fragilis YCH46]	191708.7142
YP_100711.1	271.55	hypothetical protein BF3434 [Bacteroides fragilis YCH46]	190882.4826
YP_101765.1	159.04	acetyl-CoA synthetase [Bacteroides fragilis YCH46]	189818.49
YP_101619.1	119.09	haloacid dehalogenase [Bacteroides fragilis YCH46]	189099.3613
YP_101282.1	308.56	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [Bacteroides fragilis YCH46]	188598.9336
YP_100471.1	179.81	hypothetical protein BF3192 [Bacteroides fragilis YCH46]	185756.5696
YP_101635.1	131.1	glycine cleavage system protein H [Bacteroides fragilis YCH46]	183677.3365
YP_209844.1	90.41	DNA polymerase I [Bacteroides fragilis NCTC 9343]	183371.0251
YP_101436.1	216.21	translation initiation factor IF-1 [Bacteroides fragilis YCH46]	182245.9787
YP_099170.1	80.65	hypothetical protein BF1888 [Bacteroides fragilis YCH46]	181497.7365
YP_101450.1	219.41	50S ribosomal protein L29 [Bacteroides fragilis YCH46]	181483.3991
YP_211282.1	182.47	hypothetical protein BF1644 [Bacteroides fragilis NCTC 9343]	179076.1672
YP_213246.1	81.41	hypothetical protein BF3653 [Bacteroides fragilis NCTC 9343]	177273.8363
YP_210140.1	105.67	hypothetical protein BF0416 [Bacteroides fragilis NCTC 9343]	176971.0191
YP_099813.1	117.06	thiazole synthase [Bacteroides fragilis YCH46]	175917.5822
YP_098616.1	270.65	non-specific DNA-binding protein Dps [Bacteroides fragilis YCH46]	175065.8859

YP_100573.1	152.25	alpha-amylase [Bacteroides fragilis YCH46]	171623.1094
YP_211101.1	365.91	aspartate aminotransferase [Bacteroides fragilis NCTC 9343]	170324.5486
YP_099538.1	308.92	succinyl-CoA synthetase subunit beta [Bacteroides fragilis YCH46]	168155.8285
YP_209762.1	350.93	L-aspartate oxidase [Bacteroides fragilis NCTC 9343]	167742.3201
YP_098177.1	179.28	bifunctional UDP-N-acetylmuramoyl-tripeptide:D-alanyl-D-alanine ligase/alanine racemase [Bacteroides fragilis YCH46]	167573.0455
YP_210416.1	186.02	hypothetical protein BF0713 [Bacteroides fragilis NCTC 9343]	167512.4643
YP_101543.1	128.72	hypothetical protein BF4268 [Bacteroides fragilis YCH46]	166982.7073
YP_101300.1	98.54	hypothetical protein BF4024 [Bacteroides fragilis YCH46]	165259.3923
YP_210420.1	67.15	hypothetical protein BF0717 [Bacteroides fragilis NCTC 9343]	163543.1963
YP_097847.1	197.84	5'-nucleotidase [Bacteroides fragilis YCH46]	160904.5772
YP_101841.1	361.5	ribonuclease R [Bacteroides fragilis YCH46]	160697.2032
YP_101851.1	240.4	oxidoreductase [Bacteroides fragilis YCH46]	157983.9282
YP_098190.1	219.04	UDP-N-acetylglucosamine acyltransferase [Bacteroides fragilis YCH46]	157758.5244
YP_099102.1	176.35	long-chain-fatty-acid-CoA ligase [Bacteroides fragilis YCH46]	156840.8953
YP_097522.1	155.23	arginine decarboxylase [Bacteroides fragilis YCH46]	155263.7435
YP_213340.1	56.85	hypothetical protein BF3750 [Bacteroides fragilis NCTC 9343]	153576.5623
YP_100979.1	155.94	UDP-N-acetylglucosamine 1-carboxyvinyltransferase [Bacteroides fragilis YCH46]	153038.2637
YP_213288.1	178.49	transcriptional regulator [Bacteroides fragilis NCTC 9343]	152732.0074
YP_101659.1	237.8	hypothetical protein BF4387 [Bacteroides fragilis YCH46]	152299.3983
YP_212755.1	226.37	hypothetical protein BF3142 [Bacteroides fragilis NCTC 9343]	151931.1229
ZP_07807995.1	106.23	peptidase [Bacteroides fragilis 3_1_12]	146689.1382
YP_212952.1	117.69	beta-lactamase [Bacteroides fragilis NCTC 9343]	145821.6316
ZP_07809338.1	81.97	AMP nucleosidase [Bacteroides fragilis 3_1_12]	144870.5651
YP_099683.1	425.49	dihydroorotate dehydrogenase 1B [Bacteroides fragilis YCH46]	144238.2852
YP_100808.1	251.97	propionyl-CoA carboxylase subunit beta [Bacteroides fragilis YCH46]	143939.1259
YP_100018.1	113.17	ATP synthase subunit E [Bacteroides fragilis YCH46]	143723.154
YP_211301.1	331.96	aldose 1-epimerase [Bacteroides fragilis NCTC 9343]	143048.3636
YP_100558.1	70.59	hypothetical protein BF3280 [Bacteroides fragilis YCH46]	142083.7366
YP_209709.1	152.31	putative ParA-related protein [Bacteroides fragilis NCTC 9343]	138595.4684
YP_098976.1	126.76	indolepyruvate oxidoreductase subunit IorA [Bacteroides fragilis YCH46]	134786.2113
YP_097548.1	112.08	cysteine desulfurase [Bacteroides fragilis YCH46]	131284.9143
YP_101133.1	168.27	long-chain-fatty-acid-CoA ligase [Bacteroides fragilis YCH46]	129778.3051
YP_209950.1	280.99	L-fucose isomerase [Bacteroides fragilis NCTC 9343]	129738.1122
YP_212985.1	196.53	bifunctional methionine sulfoxide reductase A/B [Bacteroides fragilis NCTC 9343]	129714.889
YP_211923.1	146.02	hypothetical protein BF2301 [Bacteroides fragilis NCTC 9343]	129162.1783
YP_098332.1	157.39	nucleoside triphosphate pyrophosphohydrolase [Bacteroides fragilis YCH46]	126205.4618
YP_212043.1	203.03	hypothetical protein BF2420 [Bacteroides fragilis NCTC 9343]	125124.7239
YP_099523.1	130.46	hypothetical protein BF2242 [Bacteroides fragilis YCH46]	119811.2226
ZP_07811547.1	139.9	phosphoribosylaminoimidazole carboxylase [Bacteroides fragilis 3_1_12]	117849.6269
YP_212681.1	118.94	hypothetical protein BF3068 [Bacteroides fragilis NCTC 9343]	117155.1426
YP_213738.1	186.6	PfkB family carbohydrate kinase [Bacteroides fragilis NCTC 9343]	116314.9147
YP_099497.1	123.45	hypothetical protein BF2216 [Bacteroides fragilis YCH46]	114020.5909
YP_098317.1	205.34	DNA translocase FtsK [Bacteroides fragilis YCH46]	114011.7379

YP_098321.1	127.78	Crp/Fnr family transcriptional regulator [Bacteroides fragilis YCH46]	113600.376
YP_210157.1	356.54	D-alanyl-alanine synthetase A [Bacteroides fragilis NCTC 9343]	113374.8922
YP_100981.1	157.9	glycoprotease [Bacteroides fragilis YCH46]	112265.7767
YP_210566.1	107.82	hypothetical protein BF0871 [Bacteroides fragilis NCTC 9343]	111242.2713
YP_209838.1	326.05	methylase [Bacteroides fragilis NCTC 9343]	111224.1477
YP_101376.1	138.21	rod shape-determining protein MreB [Bacteroides fragilis YCH46]	108514.4387
YP_098401.1	154.79	TonB-dependent outer membrane receptor protein [Bacteroides fragilis YCH46]	108275.2122
YP_213389.1	309.54	outer membrane protein [Bacteroides fragilis NCTC 9343]	108142.6012
YP_098152.1	65.55	NADH dehydrogenase I chain D [Bacteroides fragilis YCH46]	107032.9538
YP_097593.1	185.47	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase [Bacteroides fragilis YCH46]	105075.0413
YP_098284.1	88.46	alpha-amylase [Bacteroides fragilis YCH46]	104566.3082
YP_097459.1	162.15	hypothetical protein BF0176 [Bacteroides fragilis YCH46]	102566.8001
YP_098318.1	82.29	hypothetical protein BF1034 [Bacteroides fragilis YCH46]	100543.241
YP_097538.1	121.77	L-fuculose-1-phosphate aldolase [Bacteroides fragilis YCH46]	100380.2362
YP_210647.1	338.87	glutamine synthetase [Bacteroides fragilis NCTC 9343]	98857.5689
YP_098361.1	120.21	peptidyl-prolyl cis-trans isomerase [Bacteroides fragilis YCH46]	98400.14247
YP_098853.1	94.85	outer membrane vitamin B12 receptor protein [Bacteroides fragilis YCH46]	98127.49063
YP_210325.1	115.91	ribosome-associated GTPase [Bacteroides fragilis NCTC 9343]	96535.76674
YP_099972.1	140.57	TonB-dependent receptor [Bacteroides fragilis YCH46]	96106.06983
YP_213221.1	158.6	hypothetical protein BF3627 [Bacteroides fragilis NCTC 9343]	93522.93923
YP_212044.1	198.96	hypothetical protein BF2421 [Bacteroides fragilis NCTC 9343]	93094.49102
YP_098316.1	84.87	hypothetical protein BF1032 [Bacteroides fragilis YCH46]	91545.3986
YP_101836.1	91.26	hypothetical protein BF4565 [Bacteroides fragilis YCH46]	90636.30743
YP_099694.1	297.21	DNA gyrase A subunit [Bacteroides fragilis YCH46]	89362.76282
YP_098970.1	86.29	50S ribosomal protein L35 [Bacteroides fragilis YCH46]	88887.18548
YP_100375.1	216.84	hypothetical protein BF3096 [Bacteroides fragilis YCH46]	87337.31822
YP_100456.1	83.64	hypothetical protein BF3177 [Bacteroides fragilis YCH46]	85784.47146
YP_097794.1	72.69	hypothetical protein BF0511 [Bacteroides fragilis YCH46]	83899.78185
YP_097813.1	97.11	pyrroline-5-carboxylate reductase [Bacteroides fragilis YCH46]	83373.86337
YP_100655.1	110.71	molecular chaperone DnaJ [Bacteroides fragilis YCH46]	82283.98841
YP_100564.1	59.58	carbonic anhydrase [Bacteroides fragilis YCH46]	82258.80689
ZP_07809488.1	109.31	polysaccharide biosynthesis protein CapD [Bacteroides fragilis 3_1_12]	81215.78843
YP_099697.1	173.52	hypothetical protein BF2414 [Bacteroides fragilis YCH46]	79442.32574
YP_098949.1	108.44	adenylylsulfate kinase [Bacteroides fragilis YCH46]	76994.65637
YP_211652.1	84.16	hypothetical protein BF2020 [Bacteroides fragilis NCTC 9343]	76665.64662
YP_098525.1	106.85	endonuclease IV [Bacteroides fragilis YCH46]	75679.1021
YP_099504.1	66.03	3-deoxy-manno-octulosonate cytidylyltransferase [Bacteroides fragilis YCH46]	74694.30042
YP_211384.1	206.1	hypothetical protein BF1748 [Bacteroides fragilis NCTC 9343]	74642.21321
YP_101410.1	97.62	MotA/TolQ/ExbB proton channel [Bacteroides fragilis YCH46]	73910.51366
YP_100969.1	96.63	anaerobic ribonucleoside triphosphate reductase [Bacteroides fragilis YCH46]	70875.31446
YP_101218.1	256.03	prephenate dehydratase [Bacteroides fragilis YCH46]	67886.91118
YP_101835.1	171.57	collagenase [Bacteroides fragilis YCH46]	67681.38419
YP_213286.1	234.39	haloacid dehalogenase [Bacteroides fragilis NCTC 9343]	67225.97775

YP_099441.1	150.18	phosphatase [Bacteroides fragilis YCH46]	67039.27703
YP_100993.1	248.33	two-component system response regulator [Bacteroides fragilis YCH46]	66959.86984
YP_100707.1	94.48	cysteine synthase A [Bacteroides fragilis YCH46]	65632.14449
YP_098950.1	78.23	sulfate adenylyltransferase subunit 2 [Bacteroides fragilis YCH46]	64555.60089
ZP_07807931.1	59.42	amidophosphoribosyltransferase [Bacteroides fragilis 3_1_12]	61885.95312
YP_099801.1	90.33	precorrin-3B C17-methyltransferase [Bacteroides fragilis YCH46]	60945.68473
YP_098900.1	218.17	dihydrolipoamide dehydrogenase [Bacteroides fragilis YCH46]	60688.13863
YP_098917.1	155.55	cytidine/deoxycytidylate deaminase [Bacteroides fragilis YCH46]	57054.54588
YP_099818.1	79.74	thiamine phosphate pyrophosphorylase [Bacteroides fragilis YCH46]	56298.97063
ZP_07808207.1	236.4	histidine decarboxylase [Bacteroides fragilis 3_1_12]	56216.90051
ZP_07809203.1	138.47	formate-tetrahydrofolate ligase [Bacteroides fragilis 3_1_12]	55032.19555
YP_098987.1	58.67	N-acetylneuraminate lyase [Bacteroides fragilis YCH46]	54818.15086
YP_100230.1	97.12	hypothetical protein BF2948 [Bacteroides fragilis YCH46]	54309.12663
YP_098052.1	67.84	chromosome partitioning protein ParB [Bacteroides fragilis YCH46]	54075.20782
YP_100037.1	185.64	phosphate uptake transcriptional regulator [Bacteroides fragilis YCH46]	53400.62206
ZP_07809888.1	127.58	argininosuccinate lyase [Bacteroides fragilis 3_1_12]	53110.83634
YP_098947.1	110.5	sulfite synthesis pathway protein CysQ [Bacteroides fragilis YCH46]	51824.42491
YP_097470.1	159.54	hypothetical protein BF0187 [Bacteroides fragilis YCH46]	49837.81672
YP_098744.1	103.07	D-lactate dehydrogenase [Bacteroides fragilis YCH46]	49711.45628
YP_211170.1	180	hypothetical protein BF1525 [Bacteroides fragilis NCTC 9343]	49599.38495
YP_212165.1	86.93	transport-like protein [Bacteroides fragilis NCTC 9343]	48625.96616
YP_100473.1	134.71	hypothetical protein BF3194 [Bacteroides fragilis YCH46]	47594.56169
YP_097781.1	232.98	gamma-glutamyl phosphate reductase [Bacteroides fragilis YCH46]	46571.01868
YP_101210.1	134.54	hypothetical protein BF3934 [Bacteroides fragilis YCH46]	46459.01545
ZP_07811550.1	112.18	RNA polymerase sigma-54 [Bacteroides fragilis 3_1_12]	46265.74436
YP_100331.1	114.28	ATP-binding protein involved in cell division [Bacteroides fragilis YCH46]	45537.5532
YP_211653.1	78.24	hypothetical protein BF2021 [Bacteroides fragilis NCTC 9343]	45387.8524
YP_212176.1	162.79	hexokinase [Bacteroides fragilis NCTC 9343]	45065.77706
YP_100260.1	129.89	DNA repair protein RecN [Bacteroides fragilis YCH46]	43454.40921
YP_210497.1	93.87	phosphoglycolate phosphatase [Bacteroides fragilis NCTC 9343]	43100.52119
ZP_07808299.1	171.12	ABC transporter ATP-binding protein [Bacteroides fragilis 3_1_12]	42803.32724
YP_099137.1	96.38	6-phosphogluconolactonase [Bacteroides fragilis YCH46]	40785.26479
YP_212116.1	132.01	negative regulator of genetic competence [Bacteroides fragilis NCTC 9343]	39752.93177
YP_100870.1	54.91	electron transport protein [Bacteroides fragilis YCH46]	37414.24777
YP_212109.1	174.49	DNA ligase [Bacteroides fragilis NCTC 9343]	37153.25372
YP_098213.1	170.93	hypothetical protein BF0928 [Bacteroides fragilis YCH46]	35876.53392
YP_099447.1	112.8	GTP pyrophosphokinase [Bacteroides fragilis YCH46]	35219.89741
YP_097375.1	206.23	octaprenyl-diphosphate synthase [Bacteroides fragilis YCH46]	34237.38566
YP_099115.1	83.24	2-aminoethylphosphonate pyruvate aminotransferase [Bacteroides fragilis YCH46]	33844.53152
YP_098060.1	98.28	hypothetical protein BF0775 [Bacteroides fragilis YCH46]	32702.39773
YP_098353.1	82.01	alkaline phosphatase [Bacteroides fragilis YCH46]	30878.96035
YP_099383.1	100.99	TonB-dependent outer membrane receptor protein [Bacteroides fragilis YCH46]	29495.94333

ZP_07811952.1	117.17	propionyl-CoA carboxylase alpha polypeptide [Bacteroides fragilis 3_1_12]	28962.06411
YP_097570.1	73.29	alpha-galactosidase [Bacteroides fragilis YCH46]	28930.02806
YP_097812.1	168.73	transcriptional regulator [Bacteroides fragilis YCH46]	28193.11323
YP_097881.1	88.11	homoserine O-succinyltransferase [Bacteroides fragilis YCH46]	27647.00789
YP_097311.1	164.85	(dimethylallyl)adenosine tRNA methylthiotransferase [Bacteroides fragilis YCH46]	27023.36075
YP_098498.1	130.57	bacterioferritin co-migratory protein [Bacteroides fragilis YCH46]	26341.10433
YP_101780.1	172.54	UDP-N-acetylglucosamine 2-epimerase [Bacteroides fragilis YCH46]	25849.52317
YP_212581.1	134.7	5-methyltetrahydrofolate--homocysteine methyltransferase [Bacteroides fragilis NCTC 9343]	23940.97527
ZP_07810286.1	80.35	DKI isomerase 1 [Bacteroides fragilis 3_1_12]	22775.20094
YP_098372.1	48.71	sugar kinase [Bacteroides fragilis YCH46]	22267.16075
YP_100994.1	76.5	ABC transporter ATP-binding protein [Bacteroides fragilis YCH46]	22253.73647
YP_209876.1	103.53	fructokinase [Bacteroides fragilis NCTC 9343]	20499.46861
YP_098455.1	84.36	tRNA modification GTPase TrmE [Bacteroides fragilis YCH46]	20440.81946
YP_099918.1	55.57	thioredoxin [Bacteroides fragilis YCH46]	18985.75532
YP_097596.1	78.2	S-adenosyl-methyltransferase MraW [Bacteroides fragilis YCH46]	17900.73237
ZP_07811175.1	238.7	cytidylate kinase [Bacteroides fragilis 3_1_12]	17357.89166
YP_101613.1	63.08	hypothetical protein BF4341 [Bacteroides fragilis YCH46]	16696.35234
YP_101859.1	64.99	deoxyribonucleoside-triphosphatase [Bacteroides fragilis YCH46]	16648.71792
YP_101834.1	77.74	hypothetical protein BF4563 [Bacteroides fragilis YCH46]	15814.77689
YP_211896.1	68.26	hypothetical protein BF2273 [Bacteroides fragilis NCTC 9343]	14889.14793
YP_100569.1	102.11	biotin carboxyl carrier protein [Bacteroides fragilis YCH46]	13375.1994
YP_098275.1	65.71	O-acetylhomoserine sulfhydrylase [Bacteroides fragilis YCH46]	11203.8136
YP_211924.1	100.36	hypothetical protein BF2302 [Bacteroides fragilis NCTC 9343]	10999.63493
YP_098779.1	167.09	cytidine deaminase [Bacteroides fragilis YCH46]	10280.42216
ZP_07811394.1	69.64	conserved hypothetical protein [Bacteroides fragilis 3_1_12]	10251.62692
YP_099530.1	97.75	glutamine synthetase I [Bacteroides fragilis YCH46]	10201.95059
YP_098022.1	86.86	3-phosphoshikimate 1-carboxyvinyltransferase [Bacteroides fragilis YCH46]	10121.96353
YP_100579.1	105.8	50S ribosomal protein L31 [Bacteroides fragilis YCH46]	9461.820064
YP_098160.1	79.4	hypothetical protein BF0875 [Bacteroides fragilis YCH46]	9235.033125
YP_100878.1	51.92	Cu2+ homeostasis protein CutC [Bacteroides fragilis YCH46]	6480.779946
YP_100943.1	108.17	hypothetical protein BF3666 [Bacteroides fragilis YCH46]	5895.34273
YP_213144.1	133.16	isocitrate dehydrogenase [Bacteroides fragilis NCTC 9343]	5748.306096
YP_099462.1	72.68	thioredoxin [Bacteroides fragilis YCH46]	5614.987901
ZP_07809987.1	67.73	topoisomerase IV subunit A [Bacteroides fragilis 3_1_12]	5524.645358
ZP_07810427.1	141.77	conserved hypothetical protein [Bacteroides fragilis 3_1_12]	4608.395189
YP_211379.1	176.87	hypothetical protein BF1743 [Bacteroides fragilis NCTC 9343]	3941.277243
YP_099537.1	100.27	succinyl-CoA synthetase alpha chain [Bacteroides fragilis YCH46]	3769.051072
ZP_07807476.1	105.4	GTP pyrophosphokinase [Bacteroides fragilis 3_1_12]	1677.205995
YP_211521.1	93.38	GDP-L-fucose synthetase [Bacteroides fragilis NCTC 9343]	337.2817895

Table A.VII.ii The 695 proteins identified using 2 or more peptides but which are not significantly enriched in the OMV, sorted according to abundance. Shown are the accession number, Mascot confidence score, Mascot predicted identity and the average raw abundance (across the 3 OMV samples). All of these proteins were also identified in the PP samples.

Appendix VIII - Growth curves of *S. enterica* Typhimurium incubated with *B. fragilis* OMV

Growth curves were conducted as described in section 2.9.1. Cfu/ml for each sample at each time point, along with mean and standard deviation, are listed in Table A.VIII.i (first experiment, 5.2.1) and Table A.VIII.ii (second experiment 5.2.2). T-tests were used to determine the statistical significance of the differences between the cfu/ml of different samples at each time point. The fold difference and calculated p values for each comparison are listed in Table A.VIII.iii (first experiment) and Table A.VIII.iv (second experiment).

	Cfu/ml in the triplicate samples			**Mean**	**Standard Deviation**
	PBS 1	PBS 2	PBS 3		
T0	6900.00	9600.00	11800.00	9433.33	2454.25
T2	26000.00	24000.00	33000.00	27666.67	4725.82
T4	87000.00	104000.00	159000.00	116666.67	37634.20
T6	320000.00	420000.00	460000.00	400000.00	72111.03
T8	810000.00	970000.00	1120000.00	966666.67	155026.88
	WT OMV 1	WT OMV 2	WT OMV 3		
T0	10100.00	13500.00	9200.00	10933.33	2267.89
T2	8000.00	11000.00	7000.00	8666.67	2081.67
T4	29000.00	34000.00	26000.00	29666.67	4041.45
T6	120000.00	190000.00	90000.00	133333.33	51316.01
T8	450000.00	610000.00	320000.00	460000.00	145258.39
	ubb OMV 1	*ubb* OMV 2	*ubb* OMV 3		
T0	5600.00	9100.00	7700.00	7466.67	1761.628035
T2	19000.00	29000.00	25000.00	24333.33	5033.222957
T4	79000.00	113000.00	90000.00	94000.00	17349.35157
T6	330000.00	520000.00	430000.00	426666.67	95043.84953
T8	760000.00	1010000.00	840000.00	870000.00	127671.4533

Table A.VIII.i The Cfu/ml, including mean and standard deviation, from the experiment described in section 5.2.1.

	Cfu/ml				
	PBS				
	1	2	3	Mean	SD
2	70000	60000	100000	76666.66667	20816.66

4	330000	310000	360000	333333.3333	25166.11
6	1500000	1600000	1100000	1400000	264575.1
8	5000000	7000000	8000000	6666666.667	1527525
10	13000000.00	11000000.00	11000000.00	11666666.67	1154701
	WT0.5				
	1	2	3	Mean	SD
0	60000	50000	70000	60000	10000
2	280000	210000	290000	260000	43588.99
4	1400000	900000	1100000	1133333.333	251661.1
6	7000000	3100000	5000000	5033333.333	1950214
8	16000000	7000000	11000000	11333333.33	4509250
	WT1				
	1	2	3	Mean	SD
0	80000	60000	90000	76666.66667	15275.25
2	240000	170000	350000	253333.3333	90737.72
4	800000	500000	1100000	800000	300000
6	2500000	1600000	5000000	3033333.333	1761628
8	9000000	6000000	16000000	10333333.33	5131601
	WT2				
	1	2	3	Mean	SD
0	70000	70000	80000	73333.33333	5773.503
2	160000	110000	190000	153333.3333	40414.52
4	400000	400000	700000	500000	173205.1
6	1800000	1400000	2600000	1933333.333	611010.1
8	5000000	4000000	9000000	6000000	2645751
	WTg0.5				
	1	2	3	Mean	SD
0	100000.00	100000.00	80000.00	93333.33	11547.01
2	240000.00	220000.00	220000.00	226666.67	11547.01
4	1100000.00	1400000.00	900000.00	1133333.33	251661.1
6	4000000.00	7000000.00	4000000.00	5000000.00	1732051
8	8000000.00	10000000.00	9000000.00	9000000.00	1000000
	WTg1				
	1	2	3	Mean	SD
0	80000.00	70000.00	60000.00	70000.00	10000
2	140000.00	90000.00	150000.00	126666.67	32145.5
4	700000.00	320000.00	600000.00	540000.00	196977.2
6	2100000.00	1000000.00	2500000.00	1866666.67	776745.3
8	7000000.00	2600000.00	9000000.00	6200000.00	3274141

	WTg2				
	1	2	3	Mean	SD
0	60000.00	90000.00	100000.00	83333.33	20816.66
2	40000.00	50000.00	60000.00	50000.00	10000
4	160000.00	230000.00	240000.00	210000.00	43588.99
6	350000.00	600000.00	800000.00	583333.33	225462.5
8	1400000.00	2800000.00	3000000.00	2400000.00	871779.8
	pGB0.5				
	1	2	3	Mean	SD
0	100000.00	60000.00	70000.00	76666.67	20816.66
2	300000.00	250000.00	280000.00	276666.67	25166.11
4	1400000.00	1200000.00	1400000.00	1333333.33	115470.1
6	7000000.00	4000000.00	6000000.00	5666666.67	1527525
8	15000000.00	10000000.00	11000000.00	12000000.00	2645751
	pGB1				
	1	2	3	Mean	SD
0	50000.00	80000.00	90000.00	73333.33	20816.66
2	260000.00	230000.00	290000.00	260000.00	30000
4	900000.00	800000.00	1100000.00	933333.33	152752.5
6	4000000.00	1900000.00	7000000.00	4300000.00	2563201
8	9000000.00	7000000.00	10000000.00	8666666.67	1527525
	pGB2				
	1	2	3	Mean	SD
0	60000.00	100000.00	70000.00	76666.67	20816.66
2	150000.00	210000.00	190000.00	183333.33	30550.5
4	500000.00	1100000.00	900000.00	833333.33	305505
6	1600000.00	4000000.00	3600000.00	3066666.67	1285820
8	3000000.00	8000000.00	4000000.00	5000000.00	2645751
	ubb0.5				
	1	2	3	Mean	SD
0	110000.00	80000.00	70000.00	86666.67	20816.66
2	370000.00	360000.00	330000.00	353333.33	20816.66
4	1900000.00	1700000.00	1400000.00	1666666.67	251661.1
6	7000000.00	6000000.00	4000000.00	5666666.67	1527525
8	13000000.00	10000000.00	8000000.00	10333333.33	2516611
	ubb1				
	1	2	3	Mean	SD
0	90000.00	60000.00	80000.00	76666.67	15275.25

2	380000.00	270000.00	320000.00	323333.33	55075.71
4	2000000.00	1300000.00	1900000.00	1733333.33	378593.9
6	10000000.00	6000000.00	9000000.00	8333333.33	2081666
8	13000000.00	11000000.00	14000000.00	12666666.67	1527525
	ubb2				
	1	2	3	Mean	SD
0	80000.00	50000.00	80000.00	70000.00	17320.51
2	350000.00	280000.00	330000.00	320000.00	36055.51
4	1600000.00	1400000.00	1400000.00	1466666.67	115470.1
6	8000000.00	6000000.00	9000000.00	7666666.67	1527525
8	12000000.00	10000000.00	13000000.00	11666666.67	1527525

Table A.VIII.ii The Cfu/ml , including mean and standard deviation, from the experiment described in section 5.2.2.

Control and Experimental Samples	**Time**	**Mean Con**	**Mean Exp**	**Fold Difference Con/Exp**	**p value**	**Significant?**
PBS and WT	0	9433.33	10933.33	0.86	0.2403	No
	2	27666.67	8666.67	3.19	0.0051	Yes
	4	116666.67	29666.67	3.93	0.0278	Yes
	6	400000.00	133333.33	3.00	0.0042	Yes
	8	966666.67	460000.00	2.10	0.0073	Yes
PBS and Δubb	0	9433.33	7466.67	1.26	0.1643	No
	2	27666.67	24333.33	1.14	0.2251	No
	4	116666.67	94000.00	1.24	0.2088	No
	6	400000.00	426666.67	0.94	0.3599	No
	8	966666.67	870000.00	1.11	0.2265	No

Table A.VIII.iii The T-test comparison of the samples from Table A.VIII.i, showing fold difference between control and experimental samples, calculated p value and whether or not the difference is significant.

Control and Experimental Samples	Time	Mean Con	Mean Exp	Fold Diff (Exp/Con)	p value	Significant?
PBS and WTg0.5	2	76666.67	93333.33	0.82	0.1545	No
	4	333333.33	226666.67	1.47	0.0042	Yes
	6	1400000.00	1133333.33	1.24	0.1374	No
	8	6666666.67	5000000.00	1.33	0.1402	No
	10	11666666.67	9000000.00	1.30	0.0200	Yes
PBS and WTg1	2	76666.67	70000.00	1.10	0.3264	No
	4	333333.33	126666.67	2.63	0.0006	Yes
	6	1400000.00	540000.00	2.59	0.0064	Yes
	8	6666666.67	1866666.67	3.57	0.0086	Yes
	10	11666666.67	6200000.00	1.88	0.0443	Yes
PBS and WTg2	2	76666.67	83333.33	0.92	0.3574	No
	4	333333.33	50000.00	6.67	0.0004	Yes
	6	1400000.00	210000.00	6.67	0.0071	Yes
	8	6666666.67	583333.33	11.43	0.0093	Yes
	10	11666666.67	2400000.00	4.86	0.0003	Yes
PBS and WT0.5	2	76666.67	60000.00	1.28	0.1517	No
	4	333333.33	260000.00	1.28	0.0403	Yes
	6	1400000.00	1133333.33	1.24	0.1374	No
	8	6666666.67	5033333.33	1.32	0.1603	No
	10	11666666.67	11333333.33	1.03	0.4557	No
PBS and WT1	2	76666.67	76666.67	1.00	0.5000	No
	4	333333.33	253333.33	1.32	0.1315	No
	6	1400000.00	800000.00	1.75	0.0306	Yes
	8	6666666.67	3033333.33	2.20	0.0277	Yes
	10	11666666.67	10333333.33	1.13	0.3500	No
PBS and WT2	2	76666.67	73333.33	1.05	0.4057	No
	4	333333.33	153333.33	2.17	0.0026	Yes
	6	1400000.00	500000.00	2.80	0.0057	Yes
	8	6666666.67	1933333.33	3.45	0.0106	Yes
	10	11666666.67	6000000.00	1.94	0.0244	Yes
PBS and pGB0.5	2	76666.67	76666.67	1.00	0.5000	No
	4	333333.33	276666.67	1.20	0.0255	Yes
	6	1400000.00	1333333.33	1.05	0.3592	No
	8	6666666.67	5666666.67	1.18	0.2338	No
	10	11666666.67	12000000.00	0.97	0.4277	No
PBS and pGB1	2	76666.67	73333.33	1.05	0.4270	No
	4	333333.33	260000.00	1.28	0.0165	Yes
	6	1400000.00	933333.33	1.50	0.0361	Yes
	8	6666666.67	4300000.00	1.55	0.1282	No
	10	11666666.67	8666666.67	1.35	0.0288	Yes
PBS and pGB2	2	76666.67	76666.67	1.00	0.5000	No
	4	333333.33	183333.33	1.82	0.0016	Yes

	6	1400000.00	833333.33	1.68	0.0367	Yes
	8	6666666.67	3066666.67	2.17	0.0184	Yes
	10	11666666.67	5000000.00	2.33	0.0166	Yes
PBS and ubb0.5	2	76666.67	86666.67	0.88	0.2940	No
	4	333333.33	353333.33	0.94	0.1753	No
	6	1400000.00	1666666.67	0.84	0.1374	No
	8	6666666.67	5666666.67	1.18	0.2338	No
	10	11666666.67	10333333.33	1.13	0.2346	No
PBS and ubb1	2	76666.67	76666.67	1.00	0.5000	No
	4	333333.33	323333.33	1.03	0.3973	No
	6	1400000.00	1733333.33	0.81	0.1434	No
	8	6666666.67	8333333.33	0.80	0.1657	No
	10	11666666.67	12666666.67	0.92	0.2102	No
PBS and ubb2	2	76666.67	70000.00	1.10	0.3462	No
	4	333333.33	320000.00	1.04	0.3151	No
	6	1400000.00	1466666.67	0.95	0.3592	No
	8	6666666.67	7666666.67	0.87	0.2338	No
	10	11666666.67	11666666.67	1.00	0.5000	No
PBS and WTg2	2 and 4	83333.33	50000.00	1.67	0.0457	Yes
WT2 and WTg2	2	73333.33	83333.33	0.88	0.2485	No
	4	153333.33	50000.00	3.07	0.0203	Yes
	6	500000.00	210000.00	2.38	0.0466	Yes
	8	1933333.33	583333.33	3.31	0.0242	Yes
	10	6000000.00	2400000.00	2.50	0.0659	No
WT2 and ubb2	2	70000.00	73333.33	0.95	0.3884	No
	4	320000.00	153333.33	2.09	0.0031	Yes
	6	1466666.67	500000.00	2.93	0.0011	Yes
	8	7666666.67	1933333.33	3.97	0.0066	Yes
	10	11666666.67	6000000.00	1.94	0.0223	Yes
WT2 and pGB2	2	73333.33	76666.67	0.96	0.4057	No
	4	153333.33	183333.33	0.84	0.1835	No
	6	500000.00	833333.33	0.60	0.0970	No
	8	1933333.33	3066666.67	0.63	0.1329	No
	10	6000000.00	5000000.00	1.20	0.3337	No

Table A.VIII.iii T-test comparisons of samples in Table A.VIII.ii, including fold difference between the samples, calculated p values and whether or not the difference is significant.