

***Bacteroides fragilis* interactions with the collagen type I
component of the host extracellular matrix**

Bruna Patrícia Gomes de Vasconcelos Galvão

Thesis Presented for the Degree of
DOCTOR OF PHILOSOPHY
in the Department of Molecular and Cell Biology
UNIVERSITY OF CAPE TOWN
September 2010

ENDORSEMENT BY THESIS SUPERVISOR

I, Valerie R. Abratt, as supervisor of the candidate B.P.G.V. Galvão, certify that I approve of the submission of this dissertation for examination.

Signed,

Associate Professor V. R. Abratt
University of Cape Town

Date (DD-MM-YYYY)

ACKNOWLEDGEMENTS

I would like to thank my supervisor and mentor Associate Professor Valerie Abratt for her unwavering support throughout this thesis. Thank you for convincing me there was a light at the end of the tunnel and that the journey, however bumpy, was worth taking. You have inspired me to want to be the ambitious and enthusiastic scientist I believe I have become.

Thank you also to my co-supervisor Dr. Suhail Rafudeen for always being there to answer my questions and for all the technical and intellectual input. Your kindness and support will always be remembered.

Many thanks to Dr. Lynthia Paul, Dr. Rob Huddy, Dr. Jason Van Rooyen and all the students past and present of the Anaerobe unit and MCB for their scientific input and endless entertainment.

Special thanks to Di James for giving me a job, letting me play with her dogs, eat her avos and for bringing me into her home. You are a very kind and special person.

To my family and specially my parents for always believing I could do whatever I set my mind to. I hope I have made you proud. Obrigada mummy. Sem ti nada disto era possivel. Amo-te muito.

Thanks to all my mates for all their support and for keeping me sane over the years. Especially my MCB peeps Grantikins, Nano, Pillay, Eluse, Jeff, Laurita, Shammers, Dr. Patel etc., etc.

Finally, to my darling Nick for being there through the good and bad. For putting up with me and for always making me laugh. You are simply the best!

I also acknowledge the financial assistance from the University of Cape Town's International Scholarship. The National research foundation for the MSc Africa Scholarship and the PhD Grant Holders linked Scholarship.

CONTENTS

ABSTRACT	1
ABREVIATIONS	3
CHAPTER 1 General introduction.....	5
CHAPTER 2 <i>Bacteroides fragilis</i> 638R: overview of protease activity and discovery of collagen degrading proteases.....	35
CHAPTER 3 Collagen adhesion and degradation in <i>B.fragilis</i> clinical isolates.....	56
CHAPTER 4 Bioinformatic analysis of the LC MS/MS results of the putative collagen Type I adhesins.....	79
CHAPTER 5 Molecular characterisation of the putative collagen adhesin gene <i>cbp</i> of <i>B. fragilis</i> GSH18.....	93
CHAPTER 6 General conclusions and future research.....	111
REFERENCES	116

ABSTRACT

Bacteroides fragilis is an opportunistic pathogen which can cause life threatening infections in humans and animals. The increasing incidence of antibiotic resistance by the bacterium has prompted the search for novel therapeutic targets. The idea of antivirulence over antimicrobial drug development is becoming increasingly popular and invites detailed research into the virulence factors of target pathogens. The ability to adhere to, and degrade components of the extracellular matrix, such as collagen, is critical in bacterial colonisation and host cell invasion. The aim of this dissertation was to specifically identify the collagen type I adhesins and proteases in the common research strain, *B. fragilis* 638R, as well as in 23 clinical isolates from Groote Schuur Hospital (GSH), Cape Town.

Initial zymography studies in *B. fragilis* 638R, using co-polymerised collagen SDS-PAGE, showed that the bacterium produced two collagenases (~45 and ~37 kDa). A candidate gene for the predominant ~45 kDa protease was identified in the published *B. fragilis* 638R genome annotation using bioinformatic analysis, and a gene specific mutation created in order to confirm function. The 45 kDa collagen protease was, however, still present in the mutant indicating that the candidate gene did not encode the observed enzyme. Particle agglutination assays (PAA) were used to compare the collagen binding ability of *B. fragilis* 638R and 23 clinical isolates. *B. fragilis* 638R. Five of the clinical strains were found to bind collagen more strongly than *B. fragilis* 638R. These were also analysed by zymography and collagen Far Western (FW) detection and found to have identical collagen adhesin and protein profiles. One of these 5 good collagen adherent clinical isolates, GSH18, was therefore chosen for further analysis. The two collagenases (~37 and 45 kDa) were found to be present in both the outer-membrane (OMP) and cell free extracts (CFE) fractions. A collagen adhesin was detected in the CFE at ~45 kDa, and an adhesin doublet of ~34 and ~33 kDa was observed in the OMP fraction.

Two-dimensional SDS-PAGE was employed to identify all collagen adhesins present in the OMP fraction, and collagen affinity chromatography was used to purify putative collagen adhesins in both the OMP and CFE fractions. Mass spectroscopy was then used for

identification of the proteins of interest. The 34 kDa adhesin was found to be identical to the hypothetical protein encoded by the ORF BF0586 of *B. fragilis* 9343, named here *cbp* (collagen binding protein). Functionality of the protein was confirmed in *B. fragilis* GSH18 by targeted insertional mutagenesis of the *cbp* gene and demonstrating the loss of the 34 kDa adhesin protein band. Heterologous expression of the Cbp protein in *Escherichia coli* followed by collagen Far Western analysis, confirmed that Cpb was able to adhere to collagen. This study is the first to report the presence of collagen type I proteases and adhesins in *B. fragilis* and to assign function to a specific gene encoding a collagen binding protein.

ABBREVIATIONS

°C	degree(s) Celsius
aa	amino acid(s)
A	adenosine
Amp	ampicillin
bp	base pair(s)
BSA	bovine serum albumin
C	cytosine
C-	carboxy-(terminal)
CFE	cell free extract
d-	distilled
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DTT	dithiothreitol
EDTA	ethylenediaminetetra acetic acid
<i>et al.</i>	<i>et alia</i>
<i>g</i>	acceleration due to gravity
g	gram(s)
G	guanine
G+C	guanine and cytosine
h	hour(s)
IPTG	isopropyl β -D-1-thiogalactopyranoside
Kan	kanamycin
kb	kilobase pairs
kDa	kilodalton(s)
kV	kilovolt(s)
LB	Luria-Bertani medium
LC MS/MS	liquid chromatography tandem mass spectrometry
M	molar
MCS	multiple cloning site
mg	milligram
MIC	minimum inhibitory concentration
min	minute(s)
ml	millilitre(s)
mM	millimolar
MS/MS	tandem mass spectrometry

MTZ	metronidazole
MW	molecular weight
N-	amino-(terminal)
n	number of samples
NCBI	National Centre for Biotechnology Information
nm	nanometre(s)
nM	nanomolar
OD _x	optical density at <i>x</i> nanometres
ORF	open reading frame
ori	origin of replication
<i>P</i>	p-value; indicating probability
p	plasmid
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
RNA	ribonucleic acid
rRNA	ribosomal RNA
rpm	revolutions per minute
RT	room temperature
s	second(s)
SDS	sodium dodecyl sulphate
spp.	species
T	thymine
TAE	Tris-acetate-EDTA electrophoresis buffer
TCA	trichloroacetic acid
TMB	3,3',5,5'-tetramethylbenzidine
U	units
UV	ultraviolet
w/v	weight per volume
WT	wild-type
<i>x</i> ^R	resistance to <i>x</i>
<i>x</i> ^S	sensitive to <i>x</i>
α	alpha
λ	lambda
μg	microgram(s)
μl	microlitre(s)
μM	micromolar

CHAPTER 1

General introduction

1.1 INTRODUCTION TO <i>BACTEROIDES FRAGILIS</i>	6
1.1.1 PHYLOGENY OF <i>BACTEROIDES</i>	6
1.1.2 METABOLIC PROPERTIES OF <i>BACTEROIDES FRAGILIS</i>	7
1.1.2.1 CARBOHYDRATE METABOLISM	7
1.1.2.2 NITROGEN METABOLISM	7
1.1.3 ANTIBIOTIC RESISTANCE.....	9
1.1.4 PATHOGENICITY AND VIRULENCE FACTORS.....	10
1.1.4.1 CAPSULAR POLYSACCHARIDES	11
1.1.4.2 LIPOPOLYSACCHARIDES (LPS).....	12
1.1.4.3 NEURAMINIDASES	13
1.1.4.4 HEMOLYSINS.....	13
1.1.4.5 <i>B. FRAGILIS</i> TOXIN (BFT).....	14
1.2 BACTERIAL INTERACTIONS WITH THE HOST EXTRACELLULAR MATRIX (ECM)	17
1.2.1 THE HOST ECM	17
1.2.2 ADHESION AND PROTEOLYSIS.....	18
1.2.2.1 ADHESINS.....	18
1.2.2.2 PROTEASES.....	19
1.2.2.3 INHIBITION OF ADHESION AND PROTEOLYSIS AS A POTENTIAL THERAPEUTIC METHOD.....	22
1.2.3 ADHESION TO, AND PROTEOLYSIS OF, SPECIFIC COMPONENTS OF THE HUMAN ECM	23
1.2.3.1 GLYCOPROTEINS	23
1.2.3.2 PROTEOGLYCANS	24
1.2.3.3 ELASTIN	26
1.3 COLLAGENS	26
1.3.1 COLLAGEN STRUCTURE AND TISSUE TROPISM.....	27
1.3.2 COLLAGEN ADHESINS IN BACTERIA	29
1.3.3 BACTERIAL COLLAGENASES	29
1.3.4 <i>B. FRAGILIS</i> AND COLLAGEN	31
1.4 AIMS AND OBJECTIVES	33

1.1 INTRODUCTION TO *BACTEROIDES FRAGILIS*

Bacteroides fragilis is an anaerobic, Gram negative, rod-shaped bacterium. It is a normal inhabitant of the human colon where it has a commensal relationship with its host. In this capacity it is involved in assisting in digestion and in the maturation and development of the host immune system (Mazmanian *et al.*, 2005; Mazmanian and Kasper, 2006). However, under certain circumstances, some strains can become opportunistic pathogens. These are known to be able to cause abscess formation, soft-tissue infections, and bacteraemias (Pumbwe *et al.*, 2006). *B. fragilis* is an organism of great interest due to its involvement in these life-threatening infections, and particularly due to the increased emergence of resistance to several of the antibiotics used in treatment in the clinical setting (Snydman *et al.*, 2010).

1.1.1 PHYLOGENY OF *BACTEROIDES*

Members of the genus *Bacteroides* were originally defined obligately anaerobic gram-negative, nonmotile, non-sporeforming, pleomorphic rods with the guanine-cytosine (G+C) content of the group was ranging widely between 39 and 49% (Shah *et al.*, 2009). The original characterisation of this group, based predominantly on morphological characteristics, grouped together organisms that varied substantially based on their physiological and biochemical characteristics. Clarification of the phylogeny of the genus has come a long way since the advent of technologies such as DNA-DNA hybridization and 16S rRNA gene sequence comparison. The outcome of these analyses resulted in the *Bacteroides* genus being deemed equivalent to the *Bacteroides* cluster of the *Bacteroides* sub-group in what is now known as the “*Bacteroidetes*” phylum, previously the cytophaga-flavobacter-bacteroides phylum (Shah *et al.*, 2009). The species that conform to this *Bacteroides* genus include the ten members of the informal taxonomic group, “*B. fragilis* group” (Shah *et al.*, 1989), including *B. fragilis*, as well as a number of other species.

1.1.2 METABOLIC PROPERTIES OF *BACTEROIDES FRAGILIS*

1.1.2.1 CARBOHYDRATE METABOLISM

B. fragilis is able to degrade both simple and complex oligosaccharides. In the human intestine, it uses undigested dietary sugars as well as host glycoproteins, such as mucin, as carbon sources. Host cell surface glycoproteins and glycolipids contain mannose, galactose and *N*-acetyl-D-glucosamine (NAG) and other groups which can serve as carbon sources at the sites of infection (Baughn and Malamy, 2002; Brigham and Malamy, 2005).

B. fragilis is not known to produce heme and so it was originally believed that an exogenous source of heme was an important growth requirement since heme is an essential component of the cytochrome and is thus necessary for the function of the cytochrome *b*-dependent electron transport system which is responsible for the generation of ATP from glucose. It has, however, since been discovered that *B. fragilis* has both a heme-dependent and heme-independent pathway for α -ketoglutarate biosynthesis, which utilises enzymes of the reductive and oxidative branch of the Krebs cycle, respectively (Baughn and Malamy, 2002).

There are other instances where carbon source as well environmental conditions, including oxidative stress, have been found to influence the carbon metabolism in *B. fragilis*. For example, Spence *et al.* (2006) described a starch utilisation operon, consisting of four genes, which was induced by starch, maltose and maltooligosaccharides, as well as by the presence of oxygen. It was, therefore, named the *osu* (oxygen-induced starch utilisation operon) (Spence *et al.*, 2006).

1.1.2.2 NITROGEN METABOLISM

Nitrogen assimilation in bacteria is done mainly by the incorporation of ammonia into glutamate and glutamine, which are the principal nitrogen donors of most cells. There are two basic pathways for this incorporation: the glutamine synthetase/glutamate synthase (GS/GOGAT)

pathway and the glutamate dehydrogenase (GDH) pathway. These have been extensively studied in *Escherichia coli* (Merrick and Edwards, 1995). The route of choice for the assimilation of nitrogen is dependent on the nitrogen source available. Under conditions of high ammonia (NH₃) concentration, in *E. coli*, GDH is more active. During nitrogen limitation (usually under low ammonia conditions) the GS/GOGAT pathway is the more active one. In this case two sequential reactions take place. The first is the conversion of glutamate and ammonia into glutamine by glutamine synthetase type I (GSI) and the second is the transfer of the amide group from the glutamine to α -ketoglutarate by glutamate synthase (GOGAT) (Merrick and Edwards, 1995).

Nitrogen metabolism in *B. fragilis* has certain novel features compared to *E. coli*. In contrast to *E. coli*, the glutamate dehydrogenase of *B. fragilis* serves as the main enzyme for the incorporation of ammonia under high and low ammonia conditions (Abrahams and Abratt, 1998). Moreover, this bacterium has not one but two distinct GDH enzymes for this purpose. These are GdhA (*gdhA*), an NAD(P)H-dependent GDH, and GdhB (*gdhB*), which is NADH-dependent (Abrahams and Abratt 1998, Abrahams 2002). These are regulated in different ways according to the nitrogen source. Another difference between the two bacteria is that, unlike the *E. coli* which has a type I glutamine synthetase, *B. fragilis* has a type III glutamate synthetase (Southern *et al.*, 1986; van Rooyen *et al.*, 2006). This type of GS was first discovered in this bacterium and is the only GS described for *B. fragilis*.

In addition to ammonia, the human large intestine contains a large amount of proteinaceous substances found as a mixture of endogenous and dietary proteins and peptides (Macfarlane *et al.*, 1988). These are metabolized by the intestinal microbiota via hydrolysis by proteases and peptidases. The resulting products are smaller peptides, amino acids, gases, and toxic substances like ammonia, phenols and indoles. *B. fragilis* has been shown to produce a number of proteases. In one study three proteases, termed P₁, P₂ and P₃, were found to be cell surface associated (Gibson and Macfarlane, 1988a). P₂ is a 52 kDa metalloprotease and P₃ is a 34 kDa cysteine protease. Both proteins appear to be endopeptidases associated with the cell surface whose function may be the initiation of proteolysis at this location. P₁ is likely to be a 73 kDa serine exopeptidase protease with a strong ability to hydrolyse dipeptide substrates. P₁ is located in the

periplasm and cytoplasm, its function possibly being further breakdown of peptides made by outer membrane proteases.

Although the exact role of protease activity in the nutrition of *B. fragilis* is uncertain, it is clear that it is a significant one. This is seen, for example, by the fact that proteolytic activity can support growth on casein as the sole nitrogen source (Gibson and Macfarlane, 1988b). Protease activity in *B. fragilis* had been suggested to be a possible factor in the pathogenicity of the bacterium. This involvement of proteases in virulence is reviewed in greater detail later in this Chapter.

1.1.3 ANTIBIOTIC RESISTANCE

The antibiotics used for the treatment of *B. fragilis* infections include carbapenems (e.g. imipenem), piperacillin or piperacillin/tazobactam, moxifloxacin, cefoxitin, clindamycin (CLIN), chloramphenicol, metronidazole (MTZ) and recently tigecycline (TIG) (Hawser *et al.*, 2010; Snyderman *et al.*, 2010). Worldwide surveys show the development of resistance to many of these antimicrobial agents. Susceptibility to carbapenems, piperacillin/tazobactam, moxifloxacin, CLIN, MTZ and TIG was tested on strains from various European countries, collected during the period of 2007 to 2008 (Hawser *et al.*, 2010). The agents were active against more than 87 % of isolates except for CLIN which showed up to 22.6 % resistance. MTZ and TIG were the only drugs to which more than 90 % susceptibility was observed. Similarly, an American survey showed an increase in resistance to clindamycin of about 6 % in the 1980's to over 35 % in the current decade (Snyderman *et al.*, 2010). Most other agents remained highly effective with no resistance being observed for chloramphenicol, and only 2 isolates of *B. fragilis* displaying resistance to MTZ.

Metronidazole remains the best antibiotic for the treatment of *B. fragilis* infections in the Hong Kong and Korea with no incidence of resistance reported between 1997 and 2007 (Chan *et al.*, 2010; Roh *et al.*, 2009). In Korea the resistance to CLIN was shown to be extremely high (51 %), while sensitivity to chloramphenicol remained 100 % (Roh *et al.*, 2009).

Very few antibiotic susceptibility surveys have been done of *Bacteroides* spp. in South Africa. The most recent study on antibiotic susceptibility of anaerobic isolates, showed that about 39 % of *Bacteroides* spp. tested were resistant to CLIN which is in line with the other studies discussed above (Meyer *et al.*, 2006). The resistance to MTZ on the other hand was staggering. Of the invasive strains of *Bacteroides* tested, 22 % showed MTZ^R. This is a great contrast from studies done in Cape Town (Koch *et al.*, 1998) and Johannesburg (Van de Westhuyzen and Chalkey, 1992) in the 1990s, which found no resistance to MTZ.

This combined information shows that the threat of the antibiotic resistance is real and wide spread, and highlights the need for a search for alternative treatment solutions. Virulence factors are thought to have good potential as targets for novel vaccines and antimicrobial agents (Barczak and Hung, 2009; Stubben *et al.*, 2009). The theory of antivirulence, instead of antimicrobial compound development is thought to have certain advantages (Barczak and Hung, 2009). It is hypothesised that, because the target is not essential for survival in the host, there should be less, or no, selection for mutations that lead to resistance. Furthermore, since virulence factors are often very specific to an organism, and their targeting drugs are not likely to be lethal to other bacteria, there should be minimal impact on the rest of the host microbiota, therefore preventing secondary infection by drug-resistant bacteria.

It is, therefore, of interest to review what is known about *B. fragilis* pathogenesis and virulence factors with a view to identifying potential therapeutic targets for treating infections.

1.1.4 PATHOGENICITY AND VIRULENCE FACTORS

Although *B. fragilis* plays a significant role as a commensal, it has the ability to become a significant opportunistic pathogen. It makes up only about 0.5% of the total intestinal microflora, yet it is the predominant isolate from clinical infections (Kalka-Moll *et al.*, 2001). *B. fragilis* has been isolated from a number of anatomical sites in the human body including the abdominal cavity, liver, brain, pelvis, lungs and the bloodstream (Pumbwe *et al.*, 2006). Infection takes

place by the entry of the gut flora into damaged areas in the intestinal wall. Damage usually occurs due to gastrointestinal surgery, perforated appendicitis or ulcer, or gut trauma (Brook, 2002). Colonisation of the previously sterile site is done by different bacteria, with aerobic bacteria being most active during the early stages of infection. The aerobes damage tissue and also reduce the oxidation-reduction potential of the colonised site allowing for the subsequent proliferation of the anaerobes (Pumbwe *et al.*, 2006). Certain strains of *B. fragilis* have also been shown to cause gut disease, in the form of diarrhoea, in farm animals and humans (Moncrief *et al.*, 1998).

A number of virulence factors thought to be responsible for these infections have been identified. These include capsular polysaccharides, lipopolysaccharides (LPS), neuraminidases (sialidases), haemolysins and enterotoxins (*B. fragilis* toxin; BFT). These are reviewed below.

1.1.4.1 CAPSULAR POLYSACCHARIDES

B. fragilis forms a distinctive polysaccharide (PS) capsule which has been extensively studied (Patrick, 2002; Pumbwe *et al.*, 2006). It has been linked to abscess formation and has also been shown to confer resistance to phagocytosis and killing (Coyne *et al.*, 2000; Reid and Patrick, 1984; Tzianabos *et al.*, 1993).

Three types of capsule have been observed, namely the micro-capsule (MC), the small capsule (SC) and the large capsule (LC). Each has different virulence characteristics and is believed to be produced by the expression of different PS loci (Patrick *et al.*, 2009). Variable expression of polysaccharides forming the MC has been implicated as a virulence determinant. Ten annotated regions of the genome sequence of *B. fragilis* NCTC9343 have been implicated in the production of extracellular polysaccharides (PS A – J) (Patrick *et al.*, 2009). Of these eight are related to variable MC expression (PS A – H). Recently these loci have also been identified in the *B. fragilis* strains YCH46 and 638R (Patrick *et al.*, 2010). A very interesting mechanism of site-specific inversion of promoter sequences occurs, causing seven of these loci to be switched “ON” or “OFF”, resulting in within-strain phase and antigenic variation (Coyne *et al.*, 2003;

Patrick *et al.*, 2003). Similarly, formation of the LC involves the variable expression of three genes, a putative initiating glycosyltransferase WbaP homologue, a putative Wza polysaccharide export/assembly protein homologue and a putative polysaccharide co-polymerase Wzs homologue (Patrick *et al.*, 2010). Further levels of regulation of the expression of PS synthetic genes have been attributed to the presence of two genes, *upxY* and *upxZ*, found at the beginning of each PS locus, where x will be one of a-h depending on the locus they precede (Chatzidaki-Livanis *et al.*, 2010). The UpxY products are NusG^{SP} transcriptional antiterminator factors and the UpxZ products are unique proteins without any known orthologues. The UpxZ proteins were shown to interact with UpxY proteins to prevent anti-termination therefore indirectly inhibiting the transcription of the PS loci (Chatzidaki-Livanis *et al.*, 2010).

Injection of capsule alone, or even purified PS A and B, into animal models caused the formation of abscesses, confirming its involvement in this process (Coyne *et al.*, 2000). The large capsule conferred resistance to phagocytosis and killing whereas populations displaying predominantly the microcapsule were eliminated (Reid and Patrick, 1984). Despite these differences, the predominant feature of the polysaccharides remains the presence of a zwitterionic motif, shown to be necessary for their activities (Tzianabos *et al.*, 1993). Interestingly, the zwitterionic polysaccharide is also important in the ability of *B. fragilis* to function as a commensal, as it has been shown to direct the maturation of the developing host immune system (Mazmanian *et al.*, 2005). PS A has been shown to modulate the expression of anti-inflammatory cytokines IL-10 and TGF- β 2 by host Foxp3⁺ regulatory Tcells (Tregs), resulting in decreased inflammatory response in a mice models of intestinal inflammatory colitis (Round and Mazmanian, 2010). These conflicting consequences of the effects of the capsular polysaccharide of *B. fragilis* in the host illustrate this bacterium's dual role as a gut commensal and opportunistic pathogen.

1.1.4.2 LIPOPOLYSACCHARIDES (LPS)

These molecules are part of the cell wall of gram negative bacteria. When LPS comes into contact with host cells, it induces a strong immune response which can be toxic to the host cell.

Toll-like receptors (TLRs) recognise LPS and initiate the response of the innate immune system by stimulating signal transduction pathways via adaptor molecules (Mancuso *et al.*, 2005).

The exact nature *B. fragilis* LPS molecule is still controversial, as it does not conform to the classic enteric LPS paradigm (Patrick, 2002). Various biochemical differences have been noted, with the major issue revolving around the presence or absence of an enteric O-antigen-like polysaccharide (Patrick, 2002). Initially the *B. fragilis* LPS was found to use TLR4 to induce cell activation, in the presence of MD2 protein and CD14, which meant that *B. fragilis* LPS used the same receptors and co-receptors as other enterobacterial LPSs (Mancuso *et al.*, 2005). However, it has recently been reported that the LPS, whether purified or in crude extract or killed whole cells, interacts with TLR2 and not TLR4 (Alhawi *et al.*, 2009).

1.1.4.3 NEURAMINIDASES

Neuraminidases, otherwise known as sialidases, are enzymes that cleave the *N*-acetyl neuraminic acid (NANA [sialic acid]) residues of complex sugars and sugar containing molecules. It is believed that bacterial neuraminidases may be virulence factors in some pathogenic bacteria (Godoy *et al.*, 1993). Neuraminidases have been found in many *B. fragilis* strains, where a *nanH* gene encodes a neuraminidase with two possible functions for these for these enzymes have been suggested (Godoy *et al.*, 1993; Pumbwe *et al.*, 2006). The first is that neuraminidase activity may be important for cell growth and multiplication, and the second is that it may be involved in adhesion to host cells. The removal of sialic acid by neuraminidases is thought to expose adhesion sites to which lectin-like adhesins can bind (Domingues *et al.*, 1992). Further information on the adhesin properties of *B. fragilis* is discussed later.

1.1.4.4 HEMOLYSINS

Most *B. fragilis* clinical isolates have been classified as non-hemolytic. However, it has been shown that this bacterium has 10 putative hemolysin genes (Robertson *et al.*, 2006). Hemolysins

are cytotoxic enzymes that damage cell membranes leading to cell lysis. The mechanisms by which hemolysins act include surfactant, pore formation and enzymatic action. Hemolysins HlyA, HlyB and HlyIII of *B. fragilis* have been shown to confer hemolytic activity to a non-hemolytic *E. coli* strain and, HlyA and HlyB were shown to work together to enhance each others activity against erythrocytes (Robertson *et al.*, 2006).

1.1.4.5 *B. FRAGILIS* TOXIN (BFT)

Toxigenic strains of *B. fragilis*, termed enterotoxigenic *B. fragilis* (ETBF), produce a 20 kDa extracellular secretory toxin called *B. fragilis* toxin (BFT) or fragilysin (Sears, 2001). This protein is not found in nontoxigenic *B. fragilis* strains (NTBF). ETBF strains have been found in association with diarrhea in children particularly under the age of 10 (Caceres *et al.*, 2000). The first study in adults showed ETBF in approximately 27% of diarrhea patients and in 12% of healthy adult controls (Zhang *et al.*, 1999).

The BFT protein is produced as a holotoxin with a calculated molecular weight of about 44.5 kDa. The predicted structure of this holotoxin consists of a signal peptide, important for targeting the holotoxin to the bacterial membrane, and a proprotein toxin domain (Sears, 2009). This second domain is thought to be important for correct protein folding and the export of the mature, active toxin of about 20 kDa. The mature toxin has been shown to be a zinc metalloprotease (Moncrief *et al.*, 1995).

Three BFT encoding genes, termed *bft-1*, *bft-2* and Korea-*bft* or *bft-3*, have been discovered (Sears, 2009). Although all these genes are unique, they show great nucleotide homology. Each *bft* gene is found within as pathogenicity island (Buckwold *et al.*, 2007). A pathogenicity island being defined as a region of DNA that has a G+C content that is different to that of the rest of the organism's chromosome and that contains more than one putative virulence gene. The *B. fragilis* pathogenicity island (BfPAI) has been found to be integrated within a putative conjugative transposon (CTn) termed CTn86 (discovered in the ETBF strain 86-5443-2-2) (Buckwold *et al.*, 2007). This genetic element has been shown to be able to transfer between *B. fragilis* strains, i.e.

ETBF to NTBF, but so far no evidence of interorganism transfer has been seen (Sears *et al.*, 2009). It has, however, been hypothesized the BfPAI may have been acquired by ETBF strains from other organisms by the by means of conjugative transposition (Franco *et al.*, 2004).

Based on the information available, a model of the mechanism of action of BFT and how ETBF strains cause intestinal secretion was postulated (Figure. 1.1) (Sears 2009). The model describes by Sears (2009) proposes that the first step in the process involves the ETBF strain attachment to the apical membrane of intestinal epithelial cells followed by the secretion of BFT. The toxin then binds to the colonic epithelium cell (CEC) receptor therefore initiation a complex cascade of signal transduction events leading to the the activation of various inflammatory response molecules, including IL-8 and TGF- β , and to the rapid cleavage of tight junction protein, E-cadherin. This cleavage is a two-step process by which the BFT releases the ectodomain of the E-cadherin and then the intracellular domain is processed by the host cell presenilin-1/ γ -secretase. Cleavage of the intracellular domain, resulting in changes in cell morphology and membrane function and correlates with the onset of colonic inflammation in an in vivo murine model (Rhee *et al.*, 2009).

An initial report by Toprak *et al.* (2006) showed an association between the colonization by ETBF strains and colorectal carcinoma. The putative carcinogenic potential of BFT has been supported by a number of findings. Figure 1.1 shows how the cleavage of E-cadherin induces the nuclear localization of β -catenin causing the upregulation in the transcription and translation of a β -catenin-regulated oncogene (*c-myc*) (Sears, 2009). BFT has also been shown to induce cellular proliferation in HT29/C1cells and to make these highly mobile, a phenomenon consistent with observations that a decrease in E-cadherin in tumour cells increases their metastatic potential (Sears, 2009). Finally, BFT also induces p38 mitogen-activated protein kinase (p38 MAPK) and cyclooxygenase-2 (COX-2) signaling which in turn has been shown to induce the production of the antiapoptotic protein cIAP2 (Sears, 2009). The possibility of the certain *B. fragilis* strains having oncogenic potential further demonstrates the importance of this organism as a health threat.

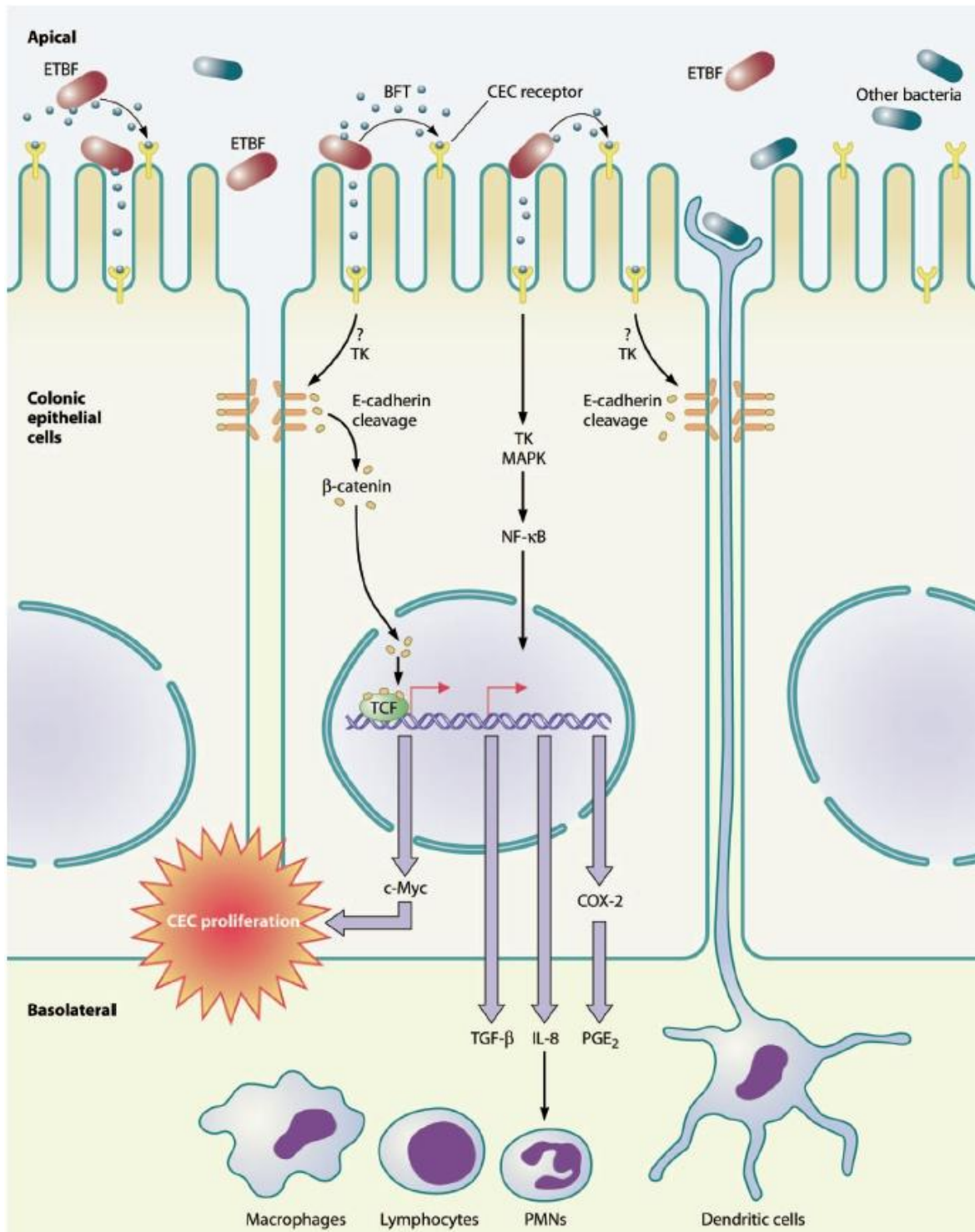


Figure 1.1. Model for the pathogenesis of ETBF (Sears, 2001). See text for details.

It has already been shown that the ability of bacteria to adhere to the host is considered important for virulence. Proteolytic activity can also have a severe impact on the host, as described in this last example. This thesis focuses on the ability of *B. fragilis* to adhere to and degrade components of the host extracellular matrix and how this may relate to virulence. This is discussed in detail in the following sections.

1.2 BACTERIAL INTERACTIONS WITH THE HOST EXTRACELLULAR MATRIX (ECM)

In addition to the *B. fragilis* virulence factors that have been described in the previous section, the interaction of the bacterium with the host extracellular matrix (ECM) is also thought to play a major role in tissue invasion and pathogenesis. These interactions can involve both adhesion to components of the ECM as well as their degradation as part of a process of colonization and tissue invasion. This aspect is the major thrust of the work described in this thesis and will be reviewed below.

1.2.1 THE HOST ECM

The host ECM is an intricate network of macromolecules that underlies the epithelial and endothelial cells and surrounds connective tissue cells (Alexander *et al.*, 1991). Based on its composition, this biologically active tissue not only serves a structural function, but is also involved in processes like cellular adhesion, migration, proliferation and differentiation (Alexander *et al.*, 1991). The macromolecules that make up the ECM are divided into four main classes, namely glycoproteins, proteoglycans, elastin and collagens (Figure.1.2).

The ECM allows for the adhesion of the host's cells in physiological conditions, but it also serves as a substrate for the attachment of microorganisms (Patti *et al.*, 1994). Attachment is considered a crucial first step in the infective process. The following sections of this review will deal with the molecular mechanisms involved in adhesion and degradation of some of these ECM components. The importance of these interactions and their relevance to bacterial and *B. fragilis* pathogenesis will be highlighted.

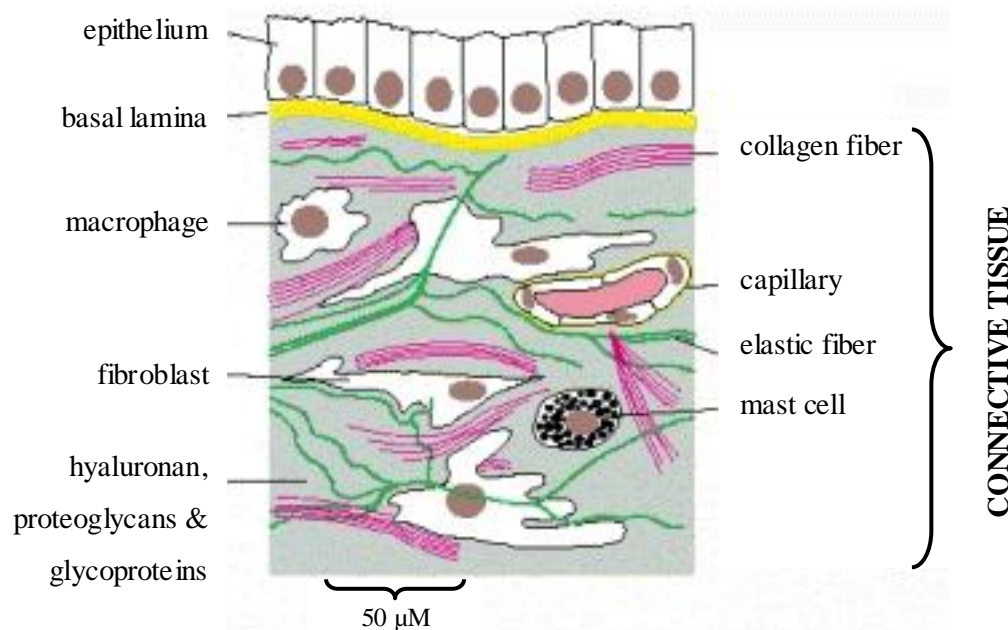


Figure 1.2. The basic structure of the extracellular matrix. (Alberts *et al.*, 2008a).

1.2.2 ADHESION AND PROTEOLYSIS

1.2.2.1 ADHESINS

The attachment of bacteria to the ECM is mediated by what are known as microbial surface components recognising adhesive matrix molecules (MSCRAMMs) (Patti *et al.*, 1994). As the name implies, these adhesins are expressed on the surface of bacteria. In gram negative bacteria, like *B. fragilis*, these are often part of surface structures such as fimbriae and pili, although major surface proteins, like YadA of certain *Yersinia* species, can also mediate adhesion (Patti *et al.*, 1994). Certainly in *B. fragilis*, surface appendages have been shown to mediate adhesion to human cells lines (Pumbwe *et al.*, 2006). MSCRAMMs are able to bind components of the ECM that are either full time ECM residents, like collagen and laminin, or part-time molecules like fibrinogen and fibronectin (Patti *et al.*, 1994). The adhesion must be of high affinity and often a high degree of specificity is found, although adhesins with the ability to bind multiple ECM components have been reported.

Although schematic models of MSCRAMMs have been developed, based on the characteristics of *Staphylococcus aureus* adhesins, there is not a great deal of information available about the general structural features of these adhesins (Patti *et al.*, 1994). Some conserved domains have been identified, and they include a signal sequence at the amino-terminal end of the protein, as well as a cell wall spanning domain, a hydrophobic membrane spanning domain and a positively charged carboxyterminus (Figure. 1.3). A unique or variable region is found in between the signal sequence and the rest of the conserved domains. This is thought to be specific for each adhesin.

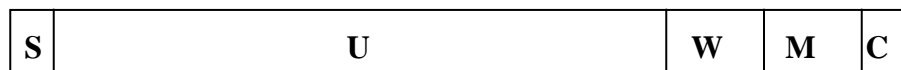


Figure 1.3. Model of MSCRAMM domains. (**S**) signal sequence; (**U**) unique sequence; (**W**) cell wall spanning domain; (**M**) hydrophobic membrane spanning domain; (**C**) and a positively charged carboxyterminus. (Based on models described by Patti *et al.*, 1994).

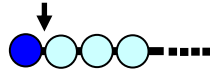
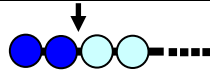
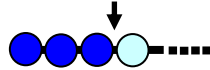
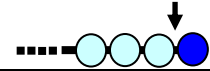
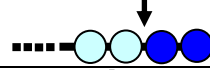
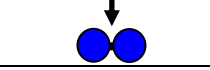

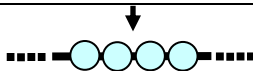
1.2.2.2 PROTEASES

Proteases are enzymes that hydrolyse proteins into their amino acid components. Those enzymes produced by bacteria can be highly specific to a particular substrate or have the ability to degrade a plethora of proteins. Bacteria can use proteases for nutrient generation by degrading the host tissue, like the ECM, but this destruction also allows for the bacteria to invade and colonise the tissue (Supuran *et al.*, 2001; Travis and Potempa, 2000). Proteases are important virulence factors not only because of their involvement in tissue destruction, but also because they help the microorganism to evade the host defences and deregulate important host cell processes, therefore making them critical for the survival and proliferation of the pathogen (Lebrun *et al.*, 2009).

Extensive research has been done on proteases and their classification is well described. Three levels of protease classification have been established by the International Union of Biochemistry and Molecular Biology (IUBMB). The first is based on the type of reaction that is catalysed, the second on the sequence specificity and the third on the chemical nature of the catalytic site (Barrett, 2001; Rao *et al.*, 1998).

The type of reaction catalysed relates to the position of bond hydrolysed, namely whether it is near one of the ends of the peptide chain or internal to the peptide. These are called exo- and endopeptidases respectively, peptidase being an alternative name for protease (Table 1). The complete digestion of proteins into their amino acid components requires the action of both endo- and exopeptidases (generally in that order). This is due to the fact that protein molecules, which are usually hundreds of amino acids long, have only two ends on which exopeptidases can act. Therefore, the initial protein degradation is done by endopeptidases, with exopeptidases completing the breakdown into amino acids.

Table 1. Types of Protease reactions

Protease	Action
Exopeptidases	
Aminopeptidase	
Dipeptidyl-peptidase	
Tripeptidyl-peptidase	
Carboxypeptidase	
Peptidyl-peptidase	
Dipeptidase	
Omega peptidase	
Endopeptidases	
Serine, Cysteine, Aspartic, Threonine & Metalloproteases	

Arrows show the cleavage sites. Dark blue circles are terminal amino acids. Light blue circles are internal amino acids. Dark blue triangles denote a variable number of terminal amino acids.

The sequence specificity of proteases relates to the preference of different enzymes for specific amino acid residues surrounding the target peptide bond.

In addition, five catalytic types of proteases are known, namely serine, cysteine, aspartic, threonine and metalloproteases (Barrett, 2001). Serine proteases have a catalytic site in which the side-chain hydroxyl group of a serine residue forms a transient acyl enzyme intermediate during catalysis. The 'catalytic triad' of these enzymes is usually composed of a histidine, which acts together with the serine, and an aspartate residue. They do not normally require cofactors and have optimal activities at neutral or slightly alkaline pH values. Threonine proteases have a threonine residue that is involved in forming the transient acyl intermediate. Similarly to the two previously mentioned catalytic types, cysteine or thiol proteases have a cysteine residue in which the thiol group is the nucleophilic group. Aspartic proteases have two aspartic residues, one in each of the two lobes of the enzyme which activate a water molecule which is directly involved in the hydrolysis of the peptide bond. Metalloprotease activity is dependent on the presence of a divalent metal ion bound at the active site with the general motif of His-Glu-Xaa-Xaa-His (HEXXH).

Unfortunately the IUBMB protease classification only allows for a small number of groups under which proteases can be classified. This results in enzymes that have significant differences being grouped together. The classification is also not useful in predicting the properties of proteases. The MEROPS system (Rawlings *et al.*, 2010) overcomes this by grouping proteases into families based on statistically significant similarities between amino acids believed to be important for catalysis. Families that appear to have evolved from a single peptidase, as determined by three-dimensional structure comparisons or amino acid sequence motifs, are then clustered into clans (Barrett, 2001).

1.2.2.3 INHIBITION OF ADHESION AND PROTEOLYSIS AS A POTENTIAL THERAPEUTIC METHOD

The relevance of adhesins and proteases as virulence factors has been discussed. Inhibition of both the adhesive and proteolytic properties of organisms has been examined as a potential therapeutic method, and a few strategies for preventing adhesion have been reviewed by Barczak and Hung (2009). In gram positive bacteria, inhibition of sortases is considered a promising strategy as these enzymes are responsible for the localization and anchoring of proteins, including adhesins, to the cell wall. However, no *in vivo* studies have been performed. In Gram negative bacteria, where adhesion is primarily attributed to the pili and fimbriae, inhibition of the formation of these structures and/or competition for binding sites are the main strategies proposed. The effect of pilicides, which prevent pili formation, has been assessed *in vitro* and shown to prevent hemagglutination of *E. coli* and a number of synthetic and natural ligands of fimbriae have been shown to block adhesion (Dobrindt and Hacker, 2008).

The inhibition of proteolytic virulence targets is a much more developed field. The importance of proteases to bacteria as virulence factors, as well as in cell regulatory mechanisms and for nutrition, makes them attractive drug targets. It also appears that the structure of various bacterial proteases, where they are known, show significant differences from those of the human host (Travis and Potempa, 2000). This is of extreme importance in drug design making proteases attractive targets. Although inhibitors of viral proteases are in use, like for example the HIV-1 protease inhibitor, there is still very little progress in the design of bacterial protease inhibitors for use as antibiotics (Supuran *et al.*, 2001; Travis and Potempa, 2000). At least one bacterial collagenase inhibitor study has been conducted in which potent *Clostridium histolyticum* collagenase inhibitors were analysed *in vitro* (Scozzafava *et al.*, 2000).

1.2.3 ADHESION TO, AND PROTEOLYSIS OF, SPECIFIC COMPONENTS OF THE HUMAN ECM

1.2.3.1 GLYCOPROTEINS

Glycoproteins are proteins covalently bound to oligosaccharide chains (Alberts *et al.*, 2008b). Examples of these include laminin and fibronectin. Laminin occurs in over 15 isoforms and together with collagen type IV (see later sections) form networks that are the central part of the basal lamina, or basement membrane (Sasaki *et al.*, 2004). The basal lamina is a specific ECM that underlies epithelia and endothelia. Fibronectin has two forms, a soluble one found in plasma (pFn) and an insoluble form (Alexander *et al.*, 1991). It is essential for the adhesion of eukaryotic cells and is involved in wound healing (Alexander *et al.*, 1991)

B. fragilis has been shown to adhere to both laminin and fibronectin (Ferreira *et al.*, 2006; Pauer *et al.*, 2009; Szoke *et al.*, 1997). Laminin binding was attributed to adhesin(s) found in the outer-membrane fraction which were hypothesised to be glycoproteins due to the loss of adhesion observed after treatment with meta-sodium periodate (Ferreira *et al.*, 2006). An adhesin of approximately 102 kDa, with the ability to bind pFn was purified and protein sequencing revealed it to be a putative TonB-dependent outer membrane protein (Pauer *et al.*, 2009).

To date, no proteases with the ability to degrade either of these ECM glycoproteins have been discovered for *B. fragilis*. However, various other pathogens do degrade both laminin and fibronectin. Some examples include the anaerobic bacteria *Clostridium difficile* and *Clostridium perfringens* which produce a cysteine and a metalloprotease, respectively, which degrade both substrates (Janoir *et al.*, 2007; Jin *et al.*, 1996).

The gram-negative oral pathogen, *Porphyromonas gingivalis*, like various *Clostridium* species, is also a notorious protease producer which causes extensive tissue damage. The main virulence factors of this bacterium are known as gingipains (O'Brien-Simpson *et al.*, 2003). These are arginine and lysine specific proteases encoded by the genes *rgpA*, *rgpB* and *kgp*. They have been

reported to degrade laminin, fibronectin, fibrinogen and collagens type I, III, IV and V. Furthermore, the RgpA and Kgp proteases have C-terminal domains that are responsible for adhesion to laminin and fibronectin as well as fibrinogen and collagen type V (see below). This extraordinary organism is a good model for many of the ECM adhesion and degradation properties to be described in this thesis.

1.2.3.2 PROTEOGLYCANS

These molecules are highly glycosylated proteins composed of a core protein covalently attached to glycosaminoglycans (GAGs) (Alexander *et al.*, 1991). These include molecules like heparin and fibrinogen, both of which are essential components of blood clotting mechanism and wound healing. Fibrinogen is the soluble precursor to insoluble fibrous protein fibrin which binds platelets to form a clot which temporarily plugs a damaged area of a blood vessel (Gannong, 2001). The clotting and fibrinolytic system is kept in balance by cascades of activators and inhibitors, many of which are proteases zymogens (Figure. 1.4). Bacterial proteases have been shown to activate many of these during pathogenesis, leading to deregulation of the system (Bergmann and Hammerschmidt, 2007; Maeda, 1996).

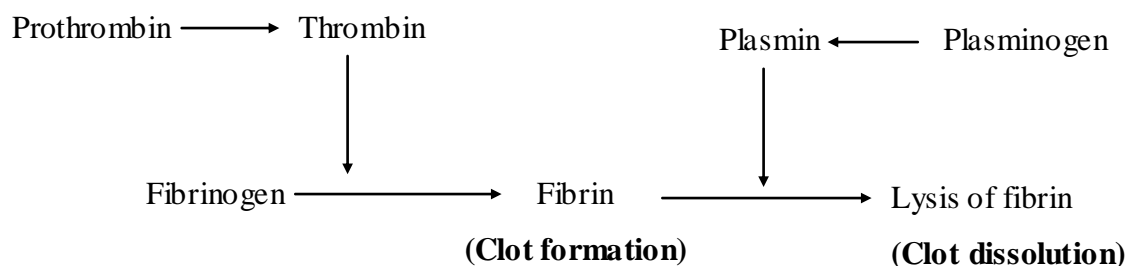


Figure 1.4. Some steps in the formation and dissolution of fibrin clots. (See text for details).

Many bacteria can disrupt either the conversion of fibrinogen to fibrin, or the breakdown of the fibrin clot. Both of these can lead to impaired coagulation and even bleeding, which can allow the bacteria to spread into tissue and into the blood stream causing tissue disruption and

septicaemia (Maeda, 1996). Many bacteria have been shown to bind and/or degrade fibrinogen. In both cases this protein is no longer available for the formation of fibrin. As mentioned in the previous section, gingipains of *P. gingivalis* have been shown to have just such an effect, which leads to binding and invasion of tissue as well as vascular disruption, or bleeding (O'Brien-Simpson *et al.*, 2003).

B. fragilis has also been shown to adhere and degrade fibrinogen. A serine-thiol-like protease of approximately 100 kDa was shown to degrade fibrinogen, casein, azocasein, gelatine and azocoll but had no activity against proteins like fibrin, fibronectin and collagens type I, II and IV (Chen *et al.*, 1995). In an initial study, the *B. fragilis* 20 kDa enterotoxin (BFT) also demonstrated the ability to degrade fibrinogen (Moncrief *et al.*, 1995). However, a subsequent study (Wu *et al.*, 2002) revealed that that BFT protein preparation had in fact not been pure and, although this was never re-tested, the activity described against the protein substrates may not be attributable to BFT. Recently, Houston *et al.* (2010), have reported the discovery of extracellular fibrinogenolytic activity at about 45 and 50 kDa using zymography. These were determined to have metalloprotease activity in *B. fragilis* strains NCTC9343 and 638R, and mild cysteine protease activity in YCH46, demonstrating the variations observed between different strains of the same species. Neither the 100 kDa or the 45 and 50 kDa protease activities have as yet been linked to specific genes. However, a 54 kDa fibrinogen binding protein (BF-FB) has been identified and characterised (Houston *et al.*, 2010).

It is worth mentioning here, in the context clotting and fibrinolysis, that a plasminogen binding protein (Bfp60) has been described in *B. fragilis* (Sijbrandi *et al.*, 2008). Interestingly, this adhesin was also found to bind laminin (Ferreira *et al.*, 2009). Furthermore, the recombinant Bfp60 was able to activate plasminogen to plasmin to levels comparable to the known plasminogen activator, streptokinase. The ability to degrade fibrin and/or prevent its formation, by fibrinogen elimination, can have huge consequences for the pathogenicity of *B. fragilis*. Abscesses, which are characteristic of this bacterium's infection, are mainly composed of a thick fibrin wall, and preventing its formation or disrupting it can allow the bacterium to disseminate in the host (Houston *et al.*, 2010).

1.2.3.3 ELASTIN

Elastin is a fibrous polymeric protein which offers tissue flexibility (Alexander *et al.*, 1991). It is a polymer of cross-linked tropoelastin precursor molecules and is the main component of elastic fibres. The adhesion to and degradation of this substrate by bacteria has been reported in association with various pathologies like cystic fibrosis associated pulmonary infections (Spencer *et al.*, 2010). *Pseudomonas aeruginosa*, infections are the leading cause of death in cystic fibrosis patients. This bacterium produces an endopeptidase, LasA, which is responsible for the degradation of elastin and heparan sulphate proteoglycan syndecan-1 and is considered an important virulence factor (Spencer *et al.*, 2010). LasA has similarities to the *S. aureus* lysostaphin protease which is also known to adhere to elastin (Park *et al.*, 1995). Fibrinogen binding protein A (FnBPA) is another *S. aureus* adhesin that can also bind elastin and is an example of this bacterium's multi-functional MSCRAMMs (Keane *et al.*, 2007).

Little is known of the relationship of *B. fragilis* to elastin. Gibson *et al.* (1988b) reported the strains assayed to be unable to use this protein as a substrate, whereas a more recent study looking at elastase activity of anaerobes from a samples of preterm premature ruptured membranes, found that 15 of the 28 *B. fragilis* strains could degrade elastin (Mikamo *et al.*, 1999).

Collagen, like elastin, is a fibrous protein component of the ECM. In fact, it is the predominant component of the ECM and, as the primary focus of this study, is discussed in detail in the following section.

1.3 COLLAGENS

Collagen is the major component of the ECM and is in fact the most abundant protein in mammals, accounting for 25 to 33% of all proteins (Harrington, 1996; Watanabe, 2004). It is the main component of skin, tendons and cartilage, as well as being the organic component of bones, teeth and cornea. Collagen gives tissues their strength and structure and is found in the

connective tissues of nearly all organs. The ability of bacteria to interact with it is, therefore, important for commensals and especially for pathogens.

1.3.1 COLLAGEN STRUCTURE AND TISSUE TROPISM

Collagens are long (approximately 300 nm), rod-like molecules made up of three parallel polypeptide chains, each of which has the left-handed helix (Cunningham and O'Connor, 1997). This type of helix is seen when proline residues occur sequentially. Proline is an amino acid which is unique amongst the other nineteen amino acids in that it has a pyrrolidine ring structure as its side chain (Figure. 1.5). This cyclic structure imposes constraints on the rotation of the N-C $^{\alpha}$ peptide bond, making this peptide bond one of the most difficult to hydrolyse. Another consequence of this cyclic nature is that there are no functional groups from which H-bonding may occur. Proline is thus the only amino acid that cannot participate in the formation of α -helices or β -sheets.

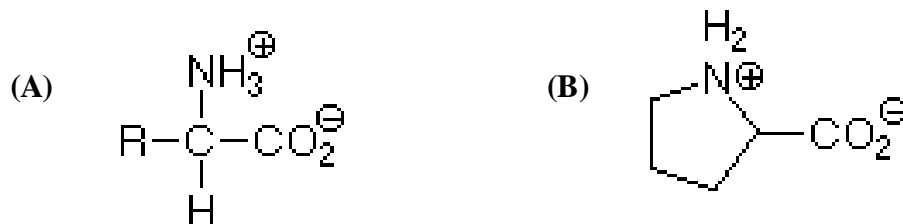


Figure 1.5. Structural characteristics of (A) general amino acids and (B) proline. (Cunningham and O'Connor, 1997)

The three left-handed helices twist around a common axis making a major helix that is slightly right handed. A common amino acid sequence, Gly-X-Y, is found in the triple helix, where the X and Y are often proline and hydroxy-proline (Alexander *et al.*, 1991). The triple helix is stabilized by the presence of hydrogen bonds between the chains as well as by inter- and intramolecular cross-links. Figure 1.6 below shows a schematic view of these levels of organization.

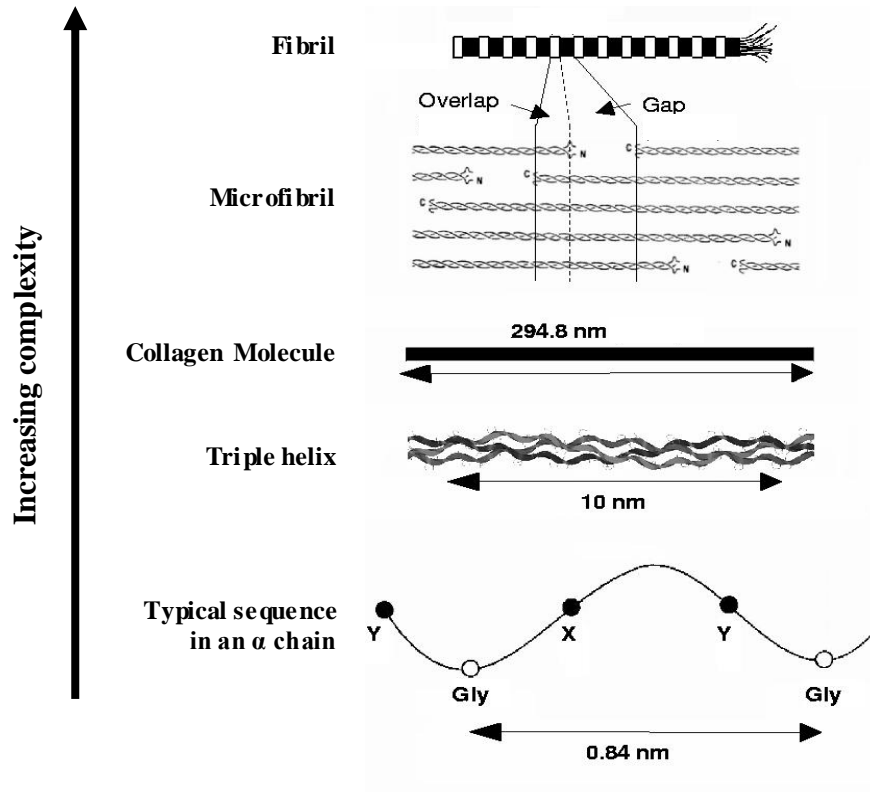


Figure 1.6. Schematic diagram of the levels of organization of collagen structure from the alpha chain to the collagen fibril. Based on previously described data (Orgel *et al.*, 2000)

At least 25 distinct collagen encoding genes and 19 collagen molecules have been discovered in humans (Watanabe, 2004). These differ in terms of their length, degree of hydroxylysine glycosylation and ratios of hydroxylated and non-hydroxylated residues (Harrington, 1996). The classical fibrillar collagens, made of a continuous triple helix with non-helical ends, are types I, II, III, V and IX. Type IV is composed of many more non-helical regions and it assembles to form tetramers, which can in turn interact to form collagen sheets via globular domains (Harrington, 1996). The different types of collagen are distributed in particular areas of the body. The most abundant collagen is type I which has been isolated from various connective tissues from cornea and skin to tendons and bone (Alexander *et al.*, 1991). Other fibrillar types are also found in a variety of tissues. The non-fibrillar, or network-forming, type IV collagen is found in the basal lamina. This variety of structure and tissue tropism means that an equally diverse set of bacterial adhesins and proteases can occur.

1.3.2 COLLAGEN ADHESINS IN BACTERIA

Adhesins with the ability to bind multiple forms of collagen have been described, as well as those specific to a single form (Patti *et al.*, 1994). The MSCRAMMs of *S. aureus* were some of the first to be studied. The Cna, collagen binding protein mediates the adhesion of *S. aureus* to collagen (Foster and Hook, 1998). The “unique” domain (see section 1.2.1) of this 55 kDa adhesin is a 19 kDa domain which has been shown to bind to different collagen types. Structural studies have shown that it binds to multiple binding sites of a collagen triple helix. Interestingly, not all strains of *S. aureus* express Cna.

The *yadA* gene, encoding the YadA protein, has been found on a virulence plasmid of some pathogenic *Yersinia* spp (El and Skurnik, 2001). YadA binds not only to various collagen types (I, III, IV and V), but to other ECM proteins as well, such as laminin and fibronectin. In *Y. enterocolitica*, the YadA mediated collagen adhesion has been found to be necessary in order for virulence to occur (El and Skurnik, 2001).

1.3.3 BACTERIAL COLLAGENASES

The ability of bacteria to adhere to collagen has already been shown to be important for virulence. Collagenolytic enzymes produced by some bacteria have similarly been identified as crucial virulence factors (Watanabe, 2004).

The study of collagenases is relevant to human health due to the large range of diseases associated with collagenolytic bacteria (Harrington, 1996). The focus of this discussion will be on diseases caused by anaerobic bacteria, which often occur in tissue due to poor blood supply to the area. Other predisposing factors to these infections include wound formation after trauma or surgery, malignancy, tissue necrosis and severe aerobic microbial infection. The last factor results in a decrease in the redox potential of the site allowing for the proliferation of anaerobic bacteria. The production of degradative enzymes, like collagenase, also allows bacteria to penetrate further into the tissue.

Collagenolytic bacteria cause diseases in various anatomical sites. Soft tissue is of particular interest to us due to the nature of *B. fragilis* infection sites. Clostridial collagenases were the first to be described and are used as a comparison for any new collagenases. *C. perfringens* causes necrotising diseases like clostridial myonecrosis, otherwise known as gas gangrene. *C. histolyticum* can also cause myonecrosis and is known to produce more than five electrophoretically different collagenases. Several oral pathogens have also been shown to be collagenolytic. *P. gingivalis* which has been discussed throughout this chapter is a good example of a highly collagenolytic pathogen.

When referring to enzymes that cleave collagen, also known as collagenolytic proteases, two basic types are generally recognised. True collagenases digest collagen macromolecules in their fibrillar form, and non-true collagenases recognize and cleave collagen specific sequences which can be found, for example, in gelatin and the synthetic collagenase substrate Pz (4-phenylazobenzoyloxycarbonyl)-Pro-Leu-Gly-Pro-D-Arg (Pz-PLGPR) (Harrington, 1996). The two types are relatively hard to differentiate. In general, it is likely that most collagenases in the human body can degrade both collagen and gelatin. This is due to the fact that, at body temperature, the thermodynamically preferred conformation of collagen is not the helical form but rather a random coil (Watanabe, 2004).

Mammalian matrix metalloproteinases (MMP's) have been extensively studied and a large amount of structural detail is known (Watanabe, 2004). By comparison, although collagenolytic activity has been described for numerous bacteria, detailed molecular information is limited. Collagenases are mostly extracellular or cell surface associated enzymes of various catalytic types. Most described collagenases are either metallo- or serine proteases, although cystein or thiol proteases have also been observed.

Collagen-binding domains have been identified in mammalian matrix metalloproteinases (MMP's) and in some bacterial collagenases, although there are no similarities between the protein sequences of the domains from the two sources (Watanabe, 2004). The collagen binding proteases of *P. gingivalis* have already been mentioned. However, gingipains are not specific to collagen. Examples of specific enzymes include ColH and ColG of *C. histolyticum* (Matsushita

et al., 1999; Yoshihara *et al.*, 1994), ColA of *Clostridium perfringens* (Matsushita *et al.*, 1994) as well as a *Geobacillus collagenovorans* collagen binding collagenase of approximately 105 kDa (Itoi *et al.*, 2006). These double faceted enzymes could be very significant virulence factors, but several other bacterial collagenases lacking binding domains have also been described (Watanabe, 2004).

1.3.4 B. FRAGILIS AND COLLAGEN

Surprisingly little research has been done on the interactions of *B. fragilis* with collagen. The ability of this bacterium to bind to collagen has been tested against abundant fibrillar collagen type I, and a preparation of type I mixed with type III (Szoke *et al.*, 1996; Szoke *et al.*, 1997). The first study showed that strains of *B. fragilis* had the ability to bind collagen type I and that this was associated with proteinaceous receptors on the cell surface of the organisms (Szoke *et al.*, 1996). In the second study, 15 of the 24 *B. fragilis* isolates from infected sites and 9 out of the 13 of strains isolated from faecal material were found to be able to adhere to collagen type I (Szoke *et al.*, 1997). This means that adhesion is likely to be strain specific.

Early studies on the ability of *B. fragilis* to degrade various protein substrates, showed that it was unable to degrade gelatin, collagen type V or azocoll (diazotized collagen) under the conditions tested (Gibson and Macfarlane, 1988b). Later studies revealed the presence of proteases that could degrade gelatin but were refractory to collagens type I, II, III and VI (Patrick, 2002). As mentioned previously a 100 kDa fibrinolytic, serine-thiol-like protease, of *B. fragilis* did show activity against gelatine and azocoll but none against other collagens (Chen *et al.*, 1995). Similarly, the *B. fragilis* enterotoxin (BFT) also showed some activity against azocoll (Moncrief *et al.*, 1998).

Table 1.2: Summary of interaction of *B. fragilis* with components of the ECM.

ECM component	Adhesion	Reference	Degradation	Reference
Fibronectin	(+) Bfp60 (60 kDa)	Szoke <i>et al.</i> , 1997 Ferreira <i>et al.</i> , 2009	(-)	Patrick, 2002; Chen <i>et al.</i> , 1995
Laminin	(+) Putative TonB-dependent outer membrane protein (102 kDa)	Szoke <i>et al.</i> , 1997 Ferreira <i>et al.</i> , 2006; Pauer <i>et al.</i> , 2009	(-)	Patrick, 2002; Chen <i>et al.</i> , 1995
Fibrinogen	(+) BF-FB (54 kDa)	Houston <i>et al.</i> , 2010	(+) (100kDa) BFT (20 kDa)? 45 kDa 50 kDa	Chen <i>et al.</i> , 1995 Moncrief <i>et al.</i> , 1995 Houston <i>et al.</i> , 2010
Elastin	NT		(+)	Gibson <i>et al.</i> , 1988b
Collagens				
Type I	(+)	Szoke <i>et al.</i> , 1997	(-)	
Type II	NT		(-)	
Type III	NT		(-)	Gibson and MacFarlane, 1988b
Type V	NT		(-)	
Type VI	NT		(-)	

(+): Able to adhere/degrade ECM component; (-): Unable to adhere/degrade ECM component; NT: not tested. Where an interaction has been attributed to at least one protein this has been stated.

In this chapter a review of the importance of *B. fragilis* as commensal and an opportunistic pathogen has been presented. As a commensal it has been shown to aid in digestion and to stimulate the host immune system (Patrick, 2002; Round and Mazmanian, 2010). Our interest, however, is focused on the threat that the bacterium poses to human health and on the importance of discovering novel therapeutic strategies, such as anti-virulence drugs, to combat the disease caused by the microorganism. The ability of bacteria to bind and degrade components of the host ECM is considered an important virulence trait (Pauer *et al.*, 2009). Several of these interactions have been studied in *B. fragilis* and are summarized in Table 1.2. *B. fragilis* has been shown to bind most of the ECM components, and to degrade some these. It, therefore, seems unlikely that it should have such a limited relationship with the predominant ECM component, collagen. We

hypothesise that the interactions with collagen may be collagen type and bacterial strain specific. Only about 60 to 70 % of the strains tested in a previous study were found to adhere to collagen type I (Szoke *et al.*, 1996; Szoke *et al.*, 1997). A similar situation may occur for the production of collagen proteases and it is, therefore, relevant to further explore the ability of the bacterium to interact with this substrate.

1.4 AIMS AND OBJECTIVES

In this dissertation the general aim was to investigate further the relationship of *B. fragilis* with the collagen type that it has been shown to adhere to, namely type I, with the specific objectives being the following:

Objective 1: to determine if *B. fragilis* produces any proteases with the ability to digest collagen type I.

Objective 2: to assess the relative collagen affinity of several South African clinical isolates of *B. fragilis* in order to find the best strain for further investigation.

Objective 3: to investigate the presence of specific collagen type I proteases and adhesins using molecular techniques.

Objective 4: to identify the various collagenases and collagen type I adhesins using bioinformatic and proteomic techniques, taking advantage of the availability of the full genome sequences of three different *B. fragilis* strains to specifically identify genes encoding these proteins.

Objective 5: to genetically and functionally characterise a gene encoding a putative collagen adhesin.

CHAPTER 2

***Bacteroides fragilis* 638R: Overview of protease activity and discovery of collagen degrading proteases**

2.1 INTRODUCTION.....	36
2.2 MATERIALS AND METHODS	38
2.2.1 BACTERIAL STRAINS AND GROWTH CONDITIONS	38
2.2.2 PROTEOLYTIC ACTIVITY <i>B. FRAGILIS</i>	39
2.2.3 STATISTICAL ANALYSIS.....	40
2.2.4 PROTEIN SAMPLE PREPARATIONS.....	40
2.2.5 PROTEIN QUANTITATION AND SEPARATION.....	39
2.2.6 COLLAGEN ZYMOGRAPHY.....	42
2.2.7 BIOINFORMATIC ANALYSIS.....	42
2.2.8 DNA TECHNIQUES.....	43
2.2.9 GENERAL POLYMERASE CHAIN REACTION (PCR).....	43
2.2.10 INSERTIONAL MUTAGENESIS.....	43
2.2.11 CONFIRMATION OF INSERTION.....	44
2.3 RESULTS AND DISCUSSION	45
2.3.1 <i>B. FRAGILIS</i> GENERAL PROTEOLYTIC ACTIVITY.....	45
2.3.2 COLLAGEN PROTEASES.....	46
2.3.3 BIOINFORMATIC SEARCH FOR COLLAGENASE CANDIDATES.....	48
2.3.4 MUTAGENESIS OF CLG CANDIDATE.....	50
2.3.5 COLLAGEN PROTEASE PROFILE OF MUTANT AND WILD-TYPE.....	52
2.4 CONCLUSION.....	53

2.1 INTRODUCTION

B. fragilis has previously been demonstrated to adhere to collagen type I (Szoke *et al.*, 1997). The first aim of this study, therefore, was to test whether *B. fragilis* also had the ability to degrade collagen type I and to identify the proteases responsible for this activity. Various methods exist for the identification of bacterial protease genes. Choosing an appropriate strategy depends on the type of protease and the substrate of interest, and proteases can range from highly substrate-specific enzymes to those with the ability to hydrolyze numerous substrates. Identification of the more promiscuous types can be done by screening for activity using a general protein substrate such as casein or its dye conjugated version, azocasein. Generally, there are two types of collagenases: those known as true collagenases, which only degrade fibrillar collagen, and non-true collagenases that recognize and cleave collagen specific sequences which can be found in relatively inexpensive substrates such as gelatin (Harrington, 1996).

Traditionally, the simplest way of identifying genes encoding functional proteases is by creating a gene bank of the subject bacterium in *E. coli* and screening clones for their ability to degrade a general protease substrate, such as casein, on an agar plate assay (Takahashi *et al.*, 1991). Gelatin plates could also be used as a more specific substrate for screening for collagenase genes. This method can, however, be somewhat limiting as it can only be used to detect non-true collagenases. Furthermore, expression of *B. fragilis* genes in *E. coli* is often weak or altogether unsuccessful. This is thought to be due to differences in the promoter structures between the two bacteria (Bayley *et al.*, 2000). Various *B. fragilis* promoters have been shown to have a -7 region with a consensus sequence of TAnnTTTG, and upstream, at about -33, a TTTG sequence is also found. This differs from the typical -10 and -35 promoter consensus sequences of *E. coli*. For these reasons, in this study, the production of a gene bank and expression of possible recombinant proteases was not thought to be the most appropriate approach to the identification of collagenase genes.

Another commonly employed strategy is the creation of mutants of the bacterium of interest, followed by screening for the loss of activity (Bonifait *et al.*, 2010). Again, there are limitations associated with the screening method. Not only does one require a relatively inexpensive and simple assay to screen large numbers of mutants, but if there are multiple proteases with similar substrate specificities being produced by the bacterium, these may mask the absence of a mutated one.

An alternative approach is to combine functional analysis of the protease activity with bioinformatic prediction of the genes involved. In using this strategy, the first step in identifying the collagenase genes of *B. fragilis* required the assessment of the bacterium's ability to degrade collagen. Commonly used collagenous substrates include various types of fibrillar collagens, as well as gelatin and synthetic peptide analogs of collagen like phenylazobenzoyloxycarbonyl-L-leucylglycyl-L-prolyl-D-arginine (PZ-PLGPA) and 2-furanacryloyl-L-leucylglycyl-L-prolyl-alanine (FALGPA). Zymograms, or activity gels, that have been either co-polymerised, overlaid or soaked with the substrate of choice are a particularly useful technique to use to show the ability of a bacterium to degrade the substrate, and for identifying the number and relative protein sizes of the proteases present. Collagen has been used for this type of assay and this is an excellent approach for identifying proteases that can degrade the fibrillar form of this substrate (Phillips *et al.*, 2000).

Once the sizes of the proteins of interest have been determined, identification of the putative genes responsible for their production can be done by two methods. One is the isolation and purification of the proteins of interest, followed by the determination of the protein sequence by methods like N-terminal amino acid sequencing or MALDI-TOF/TOF MS (MS/MS) analysis. Alternatively, bioinformatic gene mining for putative protease genes of appropriate sizes to those showing activity can be done. Both of these approaches were considered viable options for this study due to the availability of the complete genome sequence and annotation of various *B. fragilis* strains, including *B. fragilis* 638R which is used in this study.

The aims of the work reported in this chapter were, therefore, to use protein gel zymography to determine whether *B. fragilis* expressed proteases with the ability to degrade collagen Type I and to determine their protein sizes. This information was then used to examine the published *B. fragilis* genome sequence to identify potential candidate genes for further study.

2.2 MATERIALS AND METHODS

2.2.1 BACTERIAL STRAINS AND GROWTH CONDITONS

All *B. fragilis* strains described in Table 2.1 were routinely cultured in Difco brain heart infusion, supplemented with hemin (5 µg/ml), menadione (0.5 µg/ml) and cysteine (0.5 µg/ml) (BHIS) agar and broth (Holdeman and Moore, 1972). *B. fragilis* was grown under anaerobic conditions in an anaerobic chamber (Model 1024, Forma Scientific Inc., Marietta, Ohio) with atmospheric conditions of oxygen-free N₂, CO₂ and H₂ (85:10:5 by vol). For the selection of transconjugants following mutagenesis (see Section 2.2.11 below), *B. fragilis* 638R mutants were cultured on BHIS agar including gentamycin (200 µg/ml), rifampicin (20 µg/ml) and erythromycin (10 µg/ml). *E. coli* strains were grown in Luria-Bertani (LB) broth or agar, aerobically, at 37°C (Maniatis *et al.*, 1982). Strains harbouring plasmids were supplemented with ampicillin (100 µg/µl).

Table 2.1. Description of bacterial strains and plasmids

Strain/Plasmid	Genotype/phenotype	Source/reference
Plasmids		
pGERM	pUC19-based suicide vector	(Shoemaker <i>et al.</i> , 2000)
pGERMclg-int	pGERM containing the BF638R internal fragment	This study
Strains		
<i>Bacteroides fragilis</i>		
638R	Clinical strain, Rif ^R Gent ^R	(Privitera <i>et al.</i> , 1979)
638R clg⁻	638R derivative, clg ⁻ , Gent ^R Erm ^R	This study
<i>Escherichia coli</i>		
S17-1	RP4-2-Tc::Mu aph::Tn7recA ⁻ , Strep ^R	(Simon <i>et al.</i> , 1983)
S17-1 pGERMclg-int	RP4-2-Tc::Mu aph::Tn7recA ⁻ , Strep ^R Amp ^R	This study
S17-1 pGERMclos-int	RP4-2-Tc::Mu aph::Tn7recA ⁻ , Strep ^R Amp ^R	This study

Rif^R, Gent^R, Erm^R, Strep^R and Amp^R denote resistance to rifampin, gentamycin, erythromycin, streptomycin and ampicillin respectively.

2.2.2 PROTEOLYTIC ACTIVITY OF *B. FRAGILIS*

B. fragilis 638R was grown in BHIS broth for a period of 40 h. Samples were collected at 16, 24 and 40 h, and bacterial cells were harvested in pre-weighed eppendorf tubes by centrifugation at 14 000 rpm (16 000 g) in an Eppendorf 5415C bench top centrifuge (15 min). The cell pellets were then dried for 16 h at 65°C and the dry cell mass weighed. At these same time points, 2 ml samples were collected and stored at -20°C until the 40 h growth period was complete. Each sample was then thawed and cells were collected by centrifugation, washed in 0.1 M sodium phosphate buffer pH7.4 and resuspended in 2 ml of the same buffer.

The cell bound proteolytic activity of *B. fragilis* was analyzed against azocasein using a modified method based on Gibson and Macfarlane (1988). Each sample collected was divided into four 0.5 ml aliquots (three technical replicates and one blank). To three of these tubes 0.1 ml 0.1 M sodium phosphate buffer pH7.4 and 0.15 ml azocasein solution (10 mg/ml azocasein made up in 0.1M sodium phosphate buffer pH 7.4) was added and

incubated for two hours at 37°C. The reaction was then stopped by addition of 0.5 ml of 10% Trichloroacetic acid (TCA). After 30 min at room temperature (RT) the tubes were centrifuged and 0.5 ml of the supernatant was added to 0.5ml of 1M NaOH and the OD_{450nm} was measured. The fourth tube contained the sodium phosphate buffer pH7.4 and azocasein solution added after the 2 hour incubation period, just before the addition of TCA and was used as a blank. One unit of protease activity per mg of dry cell (U/mg) was defined as an increase in absorbance of 0.1 OD units at 450 nm after 2 h at 37 °C.

2.2.3 STATISTICAL ANALYSIS

The growth and azocasein experiments were performed on three biological replicates. The data for these are represented as the mean of the three replicates and error bars of the standard error are displayed. To analyse the experimental variance an unpaired, two-tailed Student's t-test was used. The means were only considered significantly different when a *P* value of 0.05 or less was obtained by the test.

2.2.4 PROTEIN SAMPLE PREPARATIONS

B. fragilis 638R was grown anaerobically for 16 h to late log/early stationary phase in BHIS broth. The cells were harvested by centrifugation using a Beckman JA-14 rotor (10 000 rpm or 9600 g, 4°C, 15 min), and washed twice in 1 x phosphate buffered saline (PBS), pH7.4, by centrifugation using a Beckman JA-14 rotor (10 000 rpm or 9600 g, 4°C, 5 min each). The pellets were then resuspended to a tenth of the original culture volume with PBS and the cells were disrupted by sonication using a Misonic sonicator 3000 at a power output of 3 W for 5 rounds of 30 s. Cell debris was removed by centrifugation in an Eppendorf 5415C bench top centrifuge (14 000 rpm or 16 000 g, 4°C, 10 min) and the supernatant cell free extract (CFE) was collected.

After the cells were harvested for the preparation of CFEs, the cell culture supernatants were filtered using a 0.2µm cellulose acetate syringe filter to remove any residual bacterial cells. The filtrate was then concentrated approximately 20 fold by pressure

filtration through a 10 kDa cut-off membrane in an Amicon stirred ultracentrifugation cell (Millipore). This preparation was designated the extracellular fraction (ExC). The protein samples were at all times kept on ice or at 4°C to avoid loss of enzymatic activity or auto-proteolytic events.

2.2.5 PROTEIN QUANTITATION AND SEPARATION

All protein samples used at defined concentrations were quantitated using the Bradford method (Bradford, 1976) with the Bio-Rad Protein Assay Dye Reagent Concentrate according to the manufacturer's instructions (Bio-Rad Laboratories, cat # 1500- 0006) using bovine serum albumin (BSA) (Fermentas) as a standard.

Protein samples were mixed 1:1 with 2 x SDS gel loading buffer without β -mercaptoethanol (0.125M Tris-HCl [pH 6.8], 4% SDS, 20% glycerol, 0.002% bromophenol blue) and incubated at room temperature for 10 min prior to application to the acylamide gels. Electrophoresis was done in SDS 10% (w/v) polyacrylamide gels (SDS-PAGE) without β -mercaptoethanol (Laemmli, 1970), using a Mini-PROTEAN Electrophoresis System (Bio-rad) at a fixed current of 15 mA per gel, and at approximately 100 V.

After electrophoresis the separated proteins were visualised by staining with Coomassie Brilliant Blue R-250. The molecular weights of the separated proteins was estimated by comparison to the prestained PageRuler™ molecular weight marker (Fermentas; #SM0671).

2.2.6 COLLAGEN ZYMOGRAPHY

Collagen zymograms were performed using collagen type I from calf skin (Elastin Products Company Inc.) co-polymerised to SDS polyacrylamide gels (SDS-PAGE) based

on previously described methods (Phillips *et al.*, 2000). Briefly, collagen type I was co-polymerised to a concentration of 150 µg/ml in an SDS-PAGE without β-mercaptoethanol. After electrophoresis, the SDS was removed from the gel by washing the gel twice with 2.5% (w/v) Triton-X100 in distilled water, twice with 2.5% (w/v) Triton-X100 in 50 mM Tris-HCl, pH 7.5 and twice in 50 mM Tris-HCl, pH 7.5. The gel was then incubated at 37°C for 48 h in Activation buffer (0.1 M Tris-HCl, pH 7.6; 0.4 M NaCl; 1.97 mM CaCl₂). The SDS removal and activation procedure described is based on a method by Houston *et al.* (personal communication). Bands of activity were viewed as areas of clearing after the gel was stained with Coomassie brilliant blue (Laemmli, 1970).

2.2.7 BIOINFORMATIC ANALYSIS

The genome annotations of *B. fragilis* 638R (Wellcome Trust Sanger Institute – www.sanger.ac.uk) and *B. fragilis* NCTC 9343 (National Centre for Biotechnology Information – www.ncbi.nih.gov) were searched for the presence of putative protease encoding genes. The predicted protein sequence of these were then submitted for BLAST analysis on the MEROPS peptidase database (Rawlings *et al.*, 2010). The protein sequences for *B. fragilis* 638R and *Porphyromonas gingivalis* W83 were obtained from the Wellcome Trust Sanger Institute website and National Centre for Biotechnology Information respectively. Protein sequence alignments were performed using DNAMAN version 4.13 (Lynnon BioSoft).

2.2.8 DNA TECHNIQUES

Genomic DNA (gDNA) of *B. fragilis* 638R was extracted by the high-salt buffer total DNA extraction method (Steffens *et al.*, 2010). Plasmids were extracted from *E. coli* strains using the BioSpin Plasmid Extraction Kit (BioFlux). DNA obtained was quantitated using a Nanodrop spectrophotometer (NanoDrop® ND-1000). Analysis of gDNA, plasmid or polymerase chain reaction (PCR) DNA products was routinely done

by electrophoresis on 0.8% or 1% (w/v) agarose gels in Tris-acetate buffer (TAE) with ethidium bromide and visualized on a GelDoc-XR UV System (Bio-Rad) (Sambrook *et al.*, 2001). Plasmids were transformed into *E. coli* S17-1 CaCl₂ competent cells (Dagert and Ehrlich, 1979).

2.2.9 GENERAL POLYMERASE CHAIN REACTION (PCR) TECHNIQUES

All PCR analyses were performed using either Supertherm TAQ polymerase (Southern Cross Biotechnology) or Kapa ReadyMix (Inqaba Biotech) as per manufacturers' instructions, and the thermal cycling was performed using a GeneAmp® PCR system 9700 (Applied Biosystems). The standard PCR parameters used consisted of an initial 5 min denaturation step at 95°C, followed by 25 cycles of denaturation for 30 sec at 95°C, annealing for 30 sec at primer pair specific temperatures, and elongation at 72°C for variable time periods. The thermal cycling was then concluded with a 5 min elongation step at 72°C.

2.2.10 INSERTIONAL MUTAGENESIS

An internal fragment of the *clg* (BF638R4511) gene, of approximately 0.58 kb, was amplified by PCR with primers ColI-F & ColI-R (Table 2.2) at an annealing temperature of 51.1°C, for an elongation period of 30 sec. The resulting PCR products were cloned into the *Sma*I site of the suicide vector pGERM (Shoemaker *et al.*, 2000) by blunt cloning using standard protocols (Maniatis *et al.*, 1982). Confirmation of insert ligations (to form the plasmids pGERM_{clg-int}) was done by sequencing (section 2.2.13) with primer M13F (Table 2.2). The plasmids were transformed into *E. coli* S17 CaCl₂ competent cells and then transferred from these transformants into *B. fragilis* by conjugation (Shoemaker *et al.*, 2000).

Table 2.2 Description of PCR primers

Name	Primer	*Hyb T°	Fragment size (kb)	Description
ColI-F	5'-GCTATCCA GGCA GGA GC-3'	51.1°C	0.58 kb	Primer pair amplifies an internal fragment of BF638R4511 of <i>B. fragilis</i> 638R; <i>clg</i> .
ColI-R	5'-CGGACGGTGTAA GCA CG-3'	51.1°C		
ClgF	5'-GAAA GCCATATTGATGCGGGCTACC-3'	57.2°C	2.1 kb	Primer pair amplifies the full length BF638R4511 gene (<i>clg</i>) of <i>B. fragilis</i> 638R
ClgR	5'-GAAAAGCGGGGTCA GTGACATCC-3'	57.0°C		
M13F	5'-CGCCA GGGTTTTCCCA GTCA CGA C-3'	60.5°C	variable	Universal primers (Yanisch-Perron <i>et al.</i> , 1985)
M13R	5'-GAGCGGATAA CAATTTCA CACA GG-3'	53.7°C		

*Hyb T°. Hybridization temperature

2.2.11 CONFIRMATION OF INSERTION

The pGERM plasmid carries the erythromycin resistance (Erm^R) gene which is used as a selective marker for transconjugants. *B. fragilis* colonies displaying the Erm^R phenotype were analysed for the presence of the inserted plasmid by PCR. This was performed using the M13F and R primers, which bind to the pGERM vector, together with primers for the full length target gene, ClgF and ClgR, at an annealing temperature of 55°C, and for an elongation period of 30 sec. Sequencing of these PCR product with the M13 primers was done by Macrogen Inc., Seoul, Korea.

2.3 RESULTS AND DISCUSSION

2.3.1 *B. FRAGILIS* GENERAL PROTEOLYTIC ACTIVITY

In order to determine at which growth stage to assay for collagenase activity a measure of collagenase activity over time would have to be determined. Unfortunately our attempts

at using the traditional quantitative assays for collagenolytic activity, Azocoll and Pz-peptide assays, were not successful due to various technical difficulties, including the inability to dissolve the insoluble Azocoll which lead to non-reproducible activity readings (data not shown). Instead a general protease assay using azocasein was performed to establish a general protease activity trend. In a previous study of proteolytic activity in *B. fragilis* (Gibson and Macfarlane, 1988b) it was shown that this bacterium displayed the greatest cell bound proteolytic activity between late-log/early stationary phase (about 16h growth), and late stationary phase (about 32 h of growth). To establish whether there was any significant difference in the general proteolytic activity over a similar period of growth, assays of the cell bound azocasein activity of *B. fragilis* grown to 16, 24 and 40 hours were performed (Figure. 2.1 B). It appeared that the highest proteolytic activity occurred at 16 h, but statistical analysis showed that there was no significant difference ($P > 0.05$) between the time points analyzed (16 h to 24 h, $P = 0.25$; 16 h to 40 h, $P = 0.13$; 24 h to 40 h, $P = 0.30$). *B. fragilis* 16 h cultures were, therefore, used for further analysis.

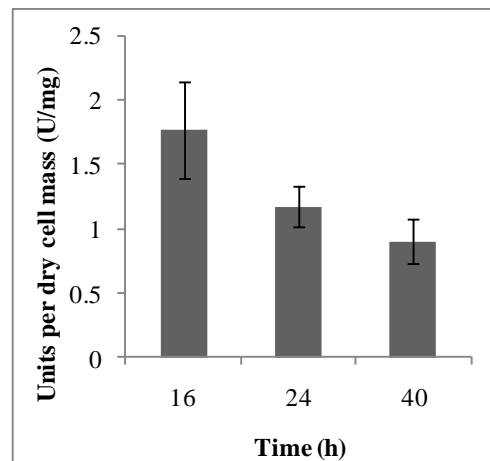


Figure 2.1 General proteolytic activity of *B. fragilis* 638R in BHIS broth. Mean values \pm SD ($n=3$).

2.3.2 COLLAGEN PROTEASES

B. fragilis had been reported to be unable to degrade collagen, as tested against azocoll, the azodye conjugated collagen substrate, and collagen type V (Gibson and Macfarlane, 1988b). However, the proteolytic toxin (BFT) produced by some *B. fragilis* enterotoxigenic strains (ETBFs), did have the ability to degrade azocoll (Moncrief *et al.*, 1995). Furthermore, *B. fragilis* is known to be able to bind to collagen Type I (Szoke *et al.*, 1997) which means that some interaction between the bacterium and this ECM component does occur. Our attempts to use azocoll as a substrate for the reproducible quantitative analysis of the ability of *B. fragilis* to degrade collagen were not successful due to technical difficulties encountered in standardizing the amount of the insoluble azocoll substrate in the reaction mixture (data not shown). A qualitative approach for identifying any collagen degrading proteases was, therefore, used. This involved collagen Type I zymography, whereby two cellular fractions, the CFE and ExC fractions, were electrophoresed on an SDS-PAGE gel co-polymerised with the substrate (Figure.2.2 A, B).

In order to preserve the activity of any putative proteases present in the protein samples several precautions were undertaken. Firstly the proteins samples were maintained at 4°C or on ice throughout the sample preparation process. Secondly, no β -mercaptoethanol was used in order to prevent the reduction of disulphide bond which could result in the loss of formation of enzymatically important protein multimers. The samples were also not boiled prior to electrophoresis, a common practice before sample application to SDS-PAGE gels. Further more all gels were electrophoresed at 4°C to avoid any heat denaturation of the proteins. Denaturation of the sample proteins by the SDS detergent occurs during electrophoresis. However, it has been shown that renaturation upon removal of the SDS is possible by washing the gels with Triton X-100. This was performed on the zymograms, therefore allowing for the proteolytic activity to be detected. This renaturation procedure has been shown to be effective in several instances, including in a recent publication by Houston *et al.* (2010) where detection of fibrinogen degrading enzymes of *B. fragilis* was demonstrated. In order to be able to match the size

of the electrophoresed proteins in the all gels were run without the use of β -mercaptoethanol.

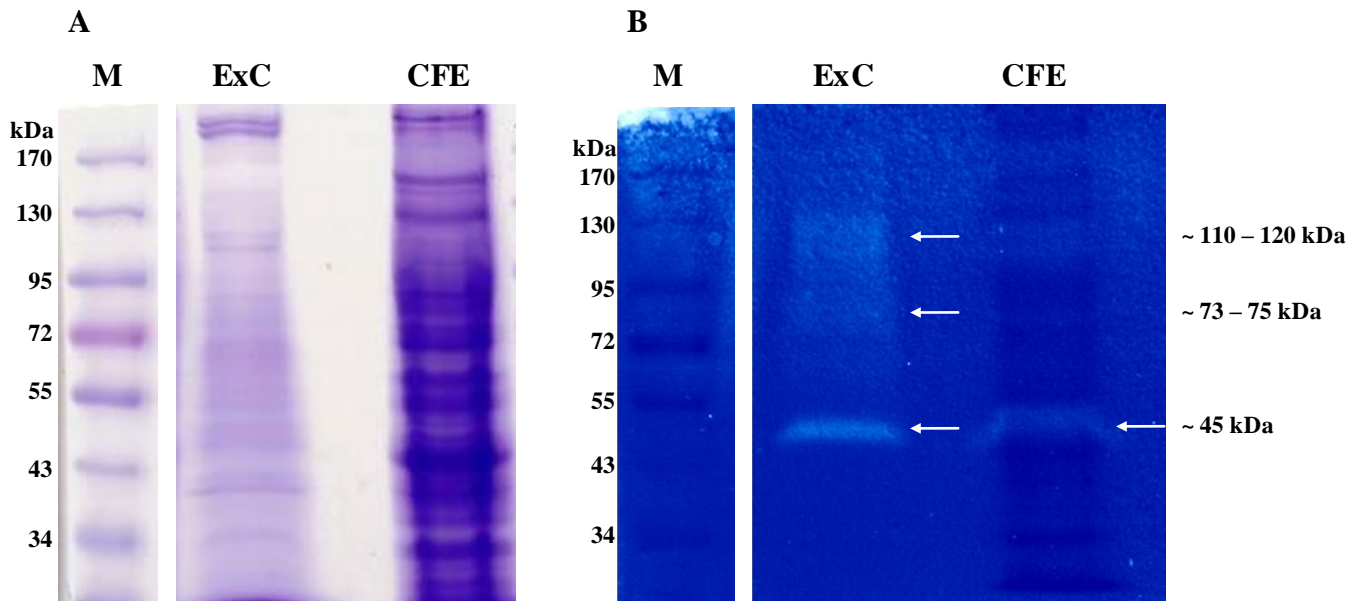


Figure 2.2 Collagen Zymography *B. fragilis* 638R proteins. (A) Coomassie stained SDS – PAGE. (B) Collagen Zymogram. Lanes: (M) PageRuler™ #SM0671; (ExC) 6.5 μ g of extracellular fraction; (CFE) 75 μ g cell free extract.

Collagen is a large molecule and cannot be internalized, meaning that any collagenase enzymes would have to be either membrane bound or secreted. For this reason the concentrated cell supernatant (ExC fraction) and the CFE, containing both the cytoplasmic and membrane fractions, were used initially in the zymographic analysis. In both these fractions (Figure 2.2. B), a distinct protease activity band was seen at approximately 45 kDa. Another two areas of activity were observed in the ExC fraction at approximately 73 – 75 kDa and 110 – 120 kDa but these could not be observed as discrete bands like the ~45 kDa one. It is important to note that although only 6.5 μ g of the ExC fraction was loaded on the zymogram, a far more distinct ~45 kDa protease band was observed in this fraction, as compared to in the CFE fraction which had 75 μ g of protein. This indicates that the collagenases make up a larger proportion of the ExC fraction.

2.3.3 BIOINFORMATIC SEARCH FOR COLLAGENASE CANDIDATES

The genomes of various *B. fragilis* strains have been sequenced and annotated, including that of the *B. fragilis* 638R strain. Although there were at least three areas of collagen degradation observed in the zymograms, the ~45 kDa collagenase band was the only discrete band which allowed for relatively accurate estimation of the protein size. For this reason, the genome annotation was searched in order to find possible gene candidates for the ~45 kDa collagenase. Putative protein sequences corresponding to genes annotated as putative protease encoding genes were compared to the predicted protein sequences of other known proteases in the NCBI database. These were narrowed down to those annotated as putative endopeptidases with predicted protein sizes between 43 and 55 kDa (Table 2.3).

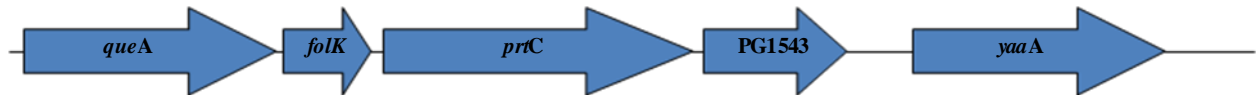
Table 2.3 Endopeptidases of *B. fragilis* 638R with a MW range between 43 and 55 kDa

PROTEIN NAME	LOCUS TAG BF 638R	MW (kDa)	MEROPS DESIGNATION	% IDENTITY (ID) TO FAMILY/ SUBFAMILY TYPE PROTEASE
putative exported peptidase	BF638R0332	49.5	Metallo-peptidase subfamily M23B	8.49% ID to lysostaphin of <i>Staphylococcus simulans</i> (Thumm and Gotz, 1997)
putative peptidase	BF638R2765	42.5	Cysteine peptidase family C11	12.64% ID to Clostripain of <i>Clostridium histolyticum</i> (Dargatz et al., 1993)
putative protease	BF638R3514	51.1	Metallo-peptidase subfamily M50B	4.82% ID to Stage IV sporulation protein FB of <i>Bacillus subtilis</i> (Cutting et al., 1991)
putative protease	BF638R3965	47.9	Metallo-peptidase subfamily M16B	15.43% ID to Mitochondrial-processing peptidase subunit beta of <i>Saccharomyces cerevisiae</i> (Witte et al., 1988)
putative peptidase	BF638R4511	47.4	family of peptidases of unknown catalytic mechanism U32	49.41% ID to PrtC of <i>Porphyromonas gingivalis</i> W83 (Kato et al., 1992)

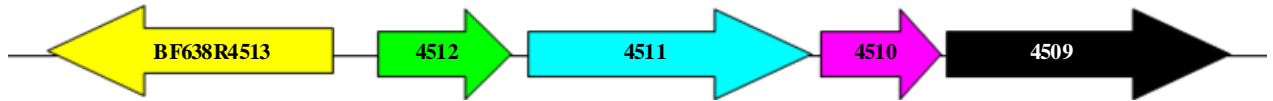
Of the five putative protease gene candidates, BF638R4511, had the highest predicted amino acid identity was to the type protease, PrtC, of *P. gingivalis*. This protease is a collagenase of *P. gingivalis*, known to degrade collagen Type I (Kato et al., 1992; Takahashi et al., 1991). A comparison of the genomic context of the *prtC* gene from *P. gingivalis* and the BF638R4511 gene from *B. fragilis* 638R revealed that both are found in gene clusters, with some of the genes sharing similar putative functions (Figure. 2.3).

A

P. gingivalis W83



B. fragilis 638R



B

<i>P. gingivalis</i> W83 ORF	ANNOTATION	<i>B. fragilis</i> 638R ORF	ANNOTATION	% IDENTITY
PG1540 (<i>queA</i>)	Putative S-adenosylmethionine: tRNA ribosyl transferase isomerase (40 kDa)	BF638R4513	Putative TIM-barrel protein, dihydrouridine synthase (36 kDa)	9.18 %
PG1541 (<i>folk</i>)	Putative 2-amino-4-hydroxy-6-hydroxymethyl-dihydropteridine pyrophosphokinase (16.8 kDa)	BF638R4512	Hypothetical protein of unknown function (17 kDa)	12.74 %
PG1542 (<i>prtC</i>)	Collagenase (46.5 kDa)	BF638R4511	Putative peptidase (47.3 kDa)	67.38 %
PG1543	Putative thioesterase family protein (16 kDa)	BF638R4510	Putative thioesterase protein (15 kDa)	45.00 %
PG1544 (<i>yaaA</i>)	Hypothetical protein of unknown function (29 kDa)	BF638R4509	Putative DNA processing Smf-like protein (40.8 kDa)	8.56 %

Figure 2.3 Comparison of the genomic context of the putative *B. fragilis* 638R collagenase gene, *clg*, to the *P. gingivalis* W83, *prtC* collagenase gene. (A) Arrangement of genes in *P. gingivalis* W83 and *B. fragilis* 638R; (B) Comparison of gene annotation and identity. Predicted protein size given in brackets.

The first four genes in the *P. gingivalis* cluster encode very similar predicted protein sizes to the first four genes in the cluster of *B. fragilis* 638R. The *queA* gene of *P. gingivalis* and BF638R4513 of *B. fragilis*, encode proteins with similarities to enzymes involved in the processing of tRNA, although they share only about 9 % amino acid identity, as determined by the multiple sequence alignment function of DNAMAN. The next gene pair shares about 13 % protein sequence identity but BF638R4512 has no similarity to any protein of known function.

The genes of interest, *prtC* and BF638R4511, have about 67 % protein sequence identity based on the predicted protein sequence from the fully sequenced and annotated genome of *P. gingivalis* W83 available on NCBI. It should be noted that this was a greater percentage identity than that seen when BF638R4511 was compared to the protein sequence lodged on MEROPS. This is because the MEROPS database reflects the protein sequence predicted from the studies done by Kato *et al* in 1992, before the genome sequence was available. In that study, the *PrtC* was found to be produced from a cloned gene isolated from a genomic library of *P. gingivalis*. The predicted sequence from this clone gave a protein of 330 amino acids (AA), as opposed to the 414 AA size predicted in the genome sequence. In that study (Kato *et al.*, 1992), the protein observed to have collagenase activity was approximately 35 kDa in size, but no protein sequencing was ever performed, so the actual size of the gene encoding this enzyme remains unclear. In the study reported in this thesis, the high sequence similarity between the two gene clusters, including a 45 % protein identity between the genes down stream of *prtC* and BF638R4511 (both putative thioesterases), prompted the continuation of the studies into this gene as a putative collagenase candidate in *B. fragilis*.

2.3.4 MUTAGENESIS OF CLG CANDIDATE

Insertional mutagenesis of the putative collagenase gene, BF638R4511 (referred to hereafter as *clg*) was the first step in establishing whether it was responsible for the production of the ~45 kDa protease activity band seen on the collagen zymograms. This previously described method of targeted gene disruption (Shoemaker *et al.*, 2000) involves a homologous recombination event between the chromosomal copy of the target gene and an internal fragment of it, carried on a suicide vector (pGERM). This recombination event results in the integration of the pGERM vector into the gene, causing its disruption (Figure. 2.4 A). Since the pGERM based plasmid (carrying the *ermG* gene) cannot replicate in *B. fragilis*, only strains where it has integrated into the genome will display the Erm^{R} phenotype. The resulting *B. fragilis* 638R transconjugants are, therefore, selected for by growth on media containing erythromycin.

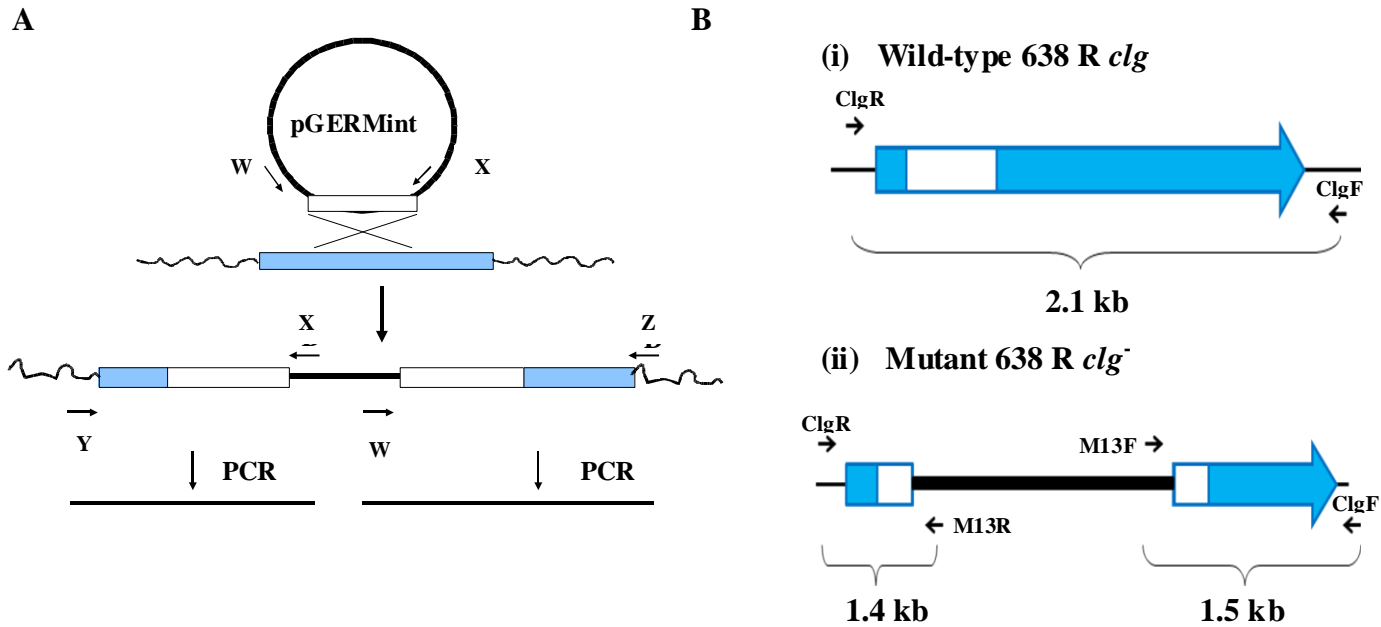


Figure 2.4 Schematic diagram of the targeted gene disruption. (A) Illustration of the principle of gene disruption by homologous recombination. The recombinant event occurs between the cloned, internal fragment of the target gene (white box on pGERM with insert, pGERMint) and the chromosomal copy of the gene (blue). The internal fragment is therefore duplicated on each side of the inserted pGERM. The black area represents the pGERM plasmid DNA and the primer binding sites are shown. (B) Expected PCR products of (i) Wild-type *B. fragilis* 638R *clg* and (ii) *B. fragilis* 638R *clg*⁻ mutant.

Confirmation of insertion into the gene of interest is done by PCR using M13 primers, which bind to the pGERM plasmid sequence only, and primers for the full length gene, in this case ClgF and ClgR. The wild-type strain should only produce a product with primers ClgF and ClgR and the mutant strain only with the appropriate combination of full length and M13 primers (Figure. 2.4 B).

In this study, the predicted 2.1 kb PCR product was produced with the wild-type *B. fragilis* 638R DNA (Figure. 2.5, lane 3) and not with the *clg*⁻ mutant DNA (lane 2). Similarly PCR using primers M13R with ClgR, and M13F with ClgF, only gave products with the mutant (lanes 4 and 6) and not the wild-type DNA (lanes 5 and 7). Sequencing

of the mutant PCR products (Figure. 2.5., lanes 4 and 6) with primers M13R and M13F confirmed the junction between the pGERM inserted vector and the interrupted *clg* gene.

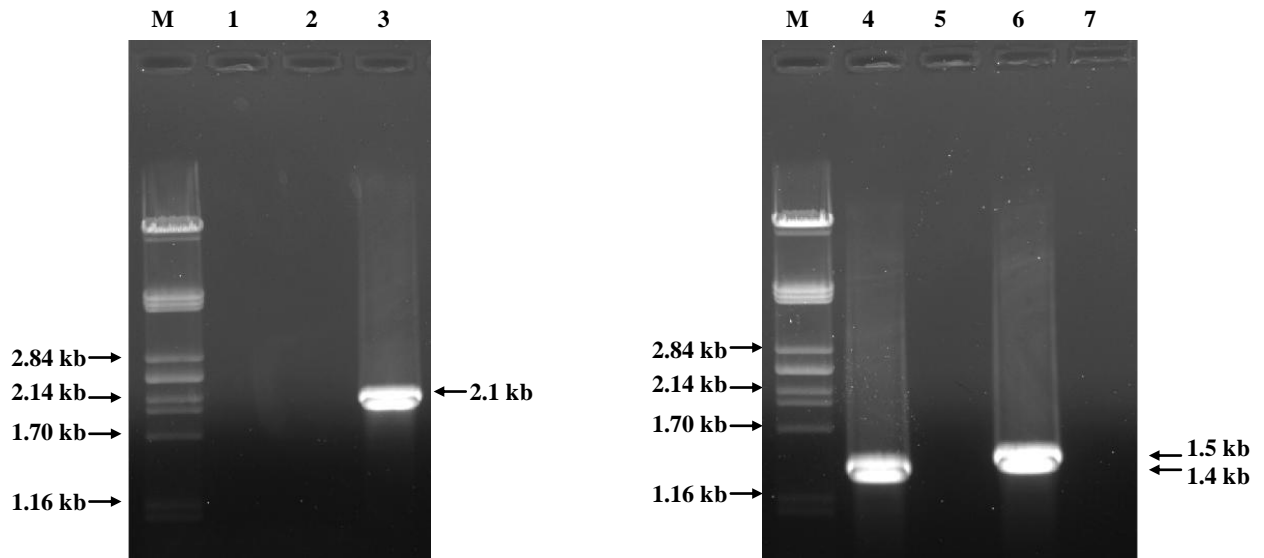


Figure 2.5 PCR of DNA from wild-type *B. fragilis* 638R and *B. fragilis* 638R *clg*⁻ mutant for confirmation of insertion. **Lanes:** (M) molecular weight marker of λ -phage DNA digested with *Pst*I; Primers ClgR and ClgF with (1) no DNA template, (2) mutant DNA, (3) wild-type DNA; Primers ClgR and M13R with (4) mutant DNA, (5) wild-type DNA; Primers ClgF and M13F with (6) mutant DNA, (7) wild-type DNA.

2.3.5 COLLAGEN PROTEASE PROFILE OF MUTANT AND WILD-TYPE

In order to detect any changes in the collagen degradation profile of *B. fragilis* 638R caused by the interruption of the *clg* gene, collagen zymography was performed. The CFE and ExC fractions of the *B. fragilis* 638R *clg*⁻ mutant were compared to those of the wild-type strain (Figure. 2.6).

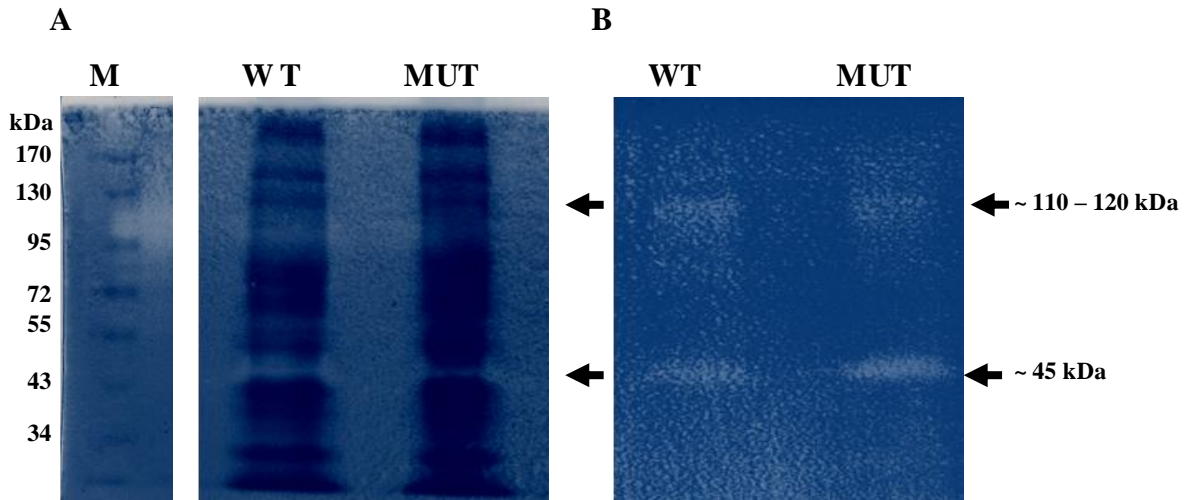


Figure 2.6 Collagen zymogram comparison of *B. fragilis* 638R wild-type (WT) and *B. fragilis* 638R *clg*⁻ mutant (MUT). (A) Cell free extracts (75 µg protein), (B) Extracellular fractions (6.25 µg protein). M: PageRuler™ #SM0671

No changes in the collagenolytic profiles of *B. fragilis* 638R were observed when the putative collagenase gene, *clg*, was interrupted. Both the wild-type (WT) and *clg*⁻ mutant (MUT) produced zones of collagen hydrolysis at about 112 – 120 kDa and 45 kDa. This result indicates that the BF638R4511 ORF does not produce the collagenase of ~45 kDa described in this study, under the growth and assay conditions used.

2.4 CONCLUSION

As a starting point to this thesis we looked to test whether *B. fragilis* could after all degrade collagen. The approach taken was one not previously used in testing for the ability of *B. fragilis* to degrade this substrate. By the use of zymographic techniques, it was possible to demonstrate for the first time that *B. fragilis* has the ability to produce a specific group of collagen degrading enzymes. At least two reproducible zones of collagen degradation were observed using collagen Type I zymography. These were of about 110 to 120 kDa, and of 45 kDa. An area of degradation at about 73 to 75 kDa was also seen but its presence was not easily reproducible. Although there have been reports

of proteases of 100 kDa (Chen *et al.*, 1995), 73 kDa (Gibson and Macfarlane, 1988a), and 45 kDa (Houston *et al.*, 2010) in *B. fragilis*, none of the genes encoding for these enzymes have been found. In future studies, each of these collagenases should be investigated in detail, but in this dissertation only the distinct ~45 kDa collagenase band, found both extracellularly and in the cell free extracts, was analysed further.

Bioinformatic analysis of the annotated genome of *B. fragilis* appeared to be the simplest approach to find the putative protease gene candidates for the collagenase of ~45 kDa but it did not prove fruitful here. The putative *B. fragilis* 638R gene homologue, of the *prtC* collagenase gene of *P. gingivalis*, which appeared to be a strong candidate, was shown not to be responsible for the production of the 45 kDa collagenase of interest. This genomic approach may have been too narrow and overly dependent on the detection of conserved protease domains and homology with other published protease enzymes. Numerous genes are annotated as hypothetical proteins, representing a reservoir of putative proteins whose function was not easily assigned during the annotation process.

With this in mind it became clear that the next step in trying to identify the genes encoding the proteases observed in this study was by the isolation and purification of the proteases followed by protein sequencing (mass spectrometry). The one dimensional separation of the protein fractions in this chapter showed that a large number of proteins were present, making it impossible to assign the protease activity to a particular protein band. An alternative approach would be to use active cellular fractions with fewer proteins and to separate these by 2-dimensional gel electrophoresis (2D SDS-PAGE). The ExC fraction would have been a good option for this purpose but it required the concentration of unmanageable volumes of extracellular supernatant. As an alternative, further fractionation of the CFE, which also contained the predominant ~45 kDa protease, was to be attempted. This fraction represents the whole cell lysate, including the cytoplasmic, inner and outer membranes.

As previously discussed, both proteases and molecules that facilitate bacterial adhesion to host ECM components have been described as putative virulence factors. MSCRAMMs

or cell surface adhesins are usually isolated from the outer membrane protein fraction (OMP). The next chapter describes the broadening of the study to include the identification of collagen type I adhesins. The use of the OMP fraction, therefore, would be ideal for a comparison of the collagen binding and degrading proteins present in *B. fragilis* with a view to identifying any collagenases with collagen binding domains.

CHAPTER 3

Collagen Adhesion and Degradation in *B. fragilis* clinical isolates

3.1 INTRODUCTION.....	57
3.2 MATERIALS AND METHODS	59
3.2.1 BACTERIAL STRAINS AND GROWTH CONDITIONS	59
3.2.2 16S rRNA AND <i>RECA</i> GENE SEQUENCING FOR IDENTIFICATION OF GROOTE SCHUUR CLINICAL ISOLATES.....	60
3.2.3 PARTICLE AGGLUTINATION ASSAYS (PAA).....	60
3.2.3.1 LATEX BEAD PREPARATION.....	61
3.2.3.2 BACTERIAL PREPARATION AND PAA.....	61
3.2.4 PROTEIN SAMPLE PREPARATION	62
3.2.4.1 CELL FREE EXTRACTS (CFES).....	62
3.2.4.2 OUTER MEMBRANE PROTEIN FRACTION (OMP)	62
3.2.5 COLLAGEN FAR WESTERN ANALYSIS.....	63
3.2.6 COLLAGEN ZYMOGRAPHY.....	63
3.2.7 COLLAGEN AFFINITY CHROMATOGRAPHY.....	64
3.2.7.1 COUPLING OF COLLAGEN TO CYANOGEN BROMIDE- ACTIVATED SEPHAROSE.....	64
3.2.7.2 AFFINITY PURIFICATION OF COLLAGEN ADHESINS.....	64
3.2.8 TWO-DIMENSIONAL (2D) GEL ELECTROPHORESIS.....	65
3.3 RESULTS AND DISCUSSION.....	66
3.3.1 IDENTIFICATION OF STRAINS.....	66
3.3.2 PARTICLE AGGLUTINATION ASSAYS.....	67
3.3.3 COLLAGEN ADHESINS AND PROTEASES OF CLINICAL STRAINS.....	68
3.3.4 AFFINITY PURIFICATION OF COLLAGEN ADHESINS.....	70
3.3.5 TWO-DIMENSIONAL SEPARATION OF COLLAGEN ADHESINS.....	73
3.4 CONCLUSION.....	76

3.1 INTRODUCTION

The ability of a *B. fragilis* 638R laboratory strain to degrade an important ECM protein, collagen type I, was investigated as described in Chapter 2. Although quantitative assays for collagenase activity were not successful due to various technical difficulties, two distinct collagenase bands were identified qualitatively using zymography. This showed that *B. fragilis* 638R could, in fact, degrade this form of collagen.

As discussed in Chapter 1, the ability of bacteria to bind, as well as to degrade, components of the ECM is believed to be important for virulence, with attachment being considered the first step in colonization by both commensal and pathogenic bacteria (Pumbwe *et al.*, 2007). The ability of various *Bacteroides spp.* to bind collagen type I has been previously examined (Szoke *et al.*, 1997). It was shown that more *B. fragilis* strains bound to collagen as compared to other *Bacteroides spp.* although only about 70 % of the *B. fragilis* isolates tested could bind the substrate. It was, therefore, important to determine how effectively the *B. fragilis* 638R lab strain was able to bind collagen as compared to other recently isolated *B. fragilis* clinical strains.

Assessment of collagen binding properties in bacteria has previously been done using the particle agglutination assay (PAA) which involves mixing bacterial suspensions with collagen coated latex beads (Naidu *et al.*, 1988). Adhesion of the bacteria to collagen causes the beads to agglutinate, with the level of agglutination being proportional to the degree of binding.

The ability of *B. fragilis* to bind collagen could also offer a method of purifying the collagenases, since various collagenases have been shown to have collagen binding domains (Chen and Duncan, 2004; Itoi *et al.*, 2006; Watanabe, 2004). If any of the *B. fragilis* collagenases observed in Chapter 2 have such substrate binding domains, then collagen affinity chromatography could be used to purify these enzymes.

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) is another method that may be used to separate proteins for subsequent mass spectroscopy identification. There have been reports of this technique being used in combination with zymography to identify proteases (Pucci-Minafra *et al.*, 2001). There are, however, limitations to this technique as certain proteases do not remain active, or cannot be reactivated after exposure to the denaturing process involved in 2D-PAGE. Similarly, 2D-PAGE has been used to identify bacterial ECM adhesins in combination with Far Western analysis (Yu *et al.*, 2006). Collagen Far western analysis has not been reported for *B. fragilis* and is potentially a good method of correlating the collagenase bands seen in the collagen zymograms to possible collagen adhesin proteins.

This chapter, therefore, describes the evaluation of clinical isolates of *B. fragilis* for their ability to bind to and degrade collagen, and the selection of the best candidate strain for further studies. The use of various new techniques for the isolation of putative collagen adhesins and proteases is also detailed.

3.2 MATERIALS AND METHODS

3.2.1 BACTERIAL STRAINS AND GROWTH CONDITIONS

Twenty-four clinical isolates, preliminarily identified as *B. fragilis*, were sourced from the Groote Schuur Service Laboratory at the Groote Schuur Hospital in Cape Town, Western Cape, South Africa (Table 3.1). Additional reference strains are shown in Table 3.2.

Table 3.1 Bacterial Strains from Groote Schuur Hospital (GSH)

Strain Name	Clinical Identification Number	Site of Isolation
GSH 2	1675193	Intra abdominal
GSH 3	205615	Septic wound-neuro surgery
GSH 5	525057	Retrocaecal – appendicitis
GSH 6	451653	Pus swab – thigh
GSH 7	73849	Peritoneal fld
GSH 8	521205	Pus swab – intra abdominal
GSH 9	222263	Information not available
GSH 10	379509	Information not available
GSH 12	21152 (24/02/03)	Bile
GSH 13	96360	Pus swab – hip abscess
GSH 14	181360 (5/07/03)	Hip abscess
GSH 15	0031574	Pus swab – urology/plastic surgery
GSH 16	263921	Pus swab – site unknown
GSH 17	64699	Pus swab – site unknown
GSH 18	76217 (11/04/03)	Pus swab – nephrectomy
GSH 19	64455	Blood culture
GSH 20	48078	Pus swab – hip
GSH 21	645186	Hip aspirate
GSH 22	321276	Pus swab – urology/plastic surgery
GSH 23	1098582	Pus swab abdominal
GSH 24	1437797	Brain abscess
GSH 25	3256692	Pus swab endocrine gynaecology
GSH 28	1237733	Peri-anal abscess
GSH 30	1651096	Pus swab empyema (lung)

Table 3.2 Other bacterial strains used

Strain	Genotype/phenotype	Reference
<i>Staphylococcus aureus</i>		University of Cape Town
<i>Micrococcus luteus</i>		University of Cape Town
<i>B. fragilis</i> 638R	Clinical strain, Rif ^R Gent ^R	(Privitera <i>et al.</i> , 1979)
<i>B. fragilis</i> MC2	Known laminin binder	(Ferreira <i>et al.</i> , 2006)
<i>B. fragilis</i> P35	Known collagen binder	(Soki <i>et al.</i> , 1999)

Rif^R Gent^R denotes resistance to rifampicin and gentamycin.

All *Bacteroides* strains were grown on BHIS agar and broth without antibiotic, as described in Chapter 2. *Staphylococcus aureus* and *Micrococcus luteus* were grown on nutrient agar (NA) (Merck) and Luria-Bertani (LB) broth, aerobically, at 37 and 30°C respectively.

3.2.2 16S rRNA AND RECA GENE SEQUENCING FOR IDENTIFICATION OF GROOTE SCHUUR CLINICAL ISOLATES

All PCR reactions were performed as describe in Chapter 2. Isolates GSH 5, 13, 18, and 19 were identified by sequencing of the *recA* gene using primers RIF (5'-CAG GTT CGA TAG CAC TGA ATG C-3') and RRA (5'-GGG CAT GCC TAT CGA GTT GG-3') (Steffens *et al.*, 2010). The remainder of the GSH isolates were identified by sequencing of the 16S rRNA gene which was amplified using universal primers F27 (5'-AGAGTTTGATCITGGCTCAG-3') and R5 (5'-ACGGITACCTTGTTACGACTT-3') (Cheneby *et al.*, 2000).

DNA sequencing was performed by Macrogen Inc., Seoul, Korea. The sequences obtained were analysed using BLAST 2.2.17 (Altschul *et al.*, 1997; Schaffer *et al.*, 2001).

3.2.3 PARTICLE AGGLUTINATION ASSAYS (PAA)

The particle agglutination assays were performed as previously described (Naidu *et al.*, 1988) with amendments as described below.

3.2.3.1 LATEX BEAD PREPARATION

One millilitre of 0.8 μm (diameter) latex beads (Sigma-Aldrich, LB8) was mixed with 3 ml of 0.2 M glycine buffer. The mixture was centrifuged at 3 000 rpm (735 g) in an Eppendorf 5415C bench top centrifuge or 30 min, the supernatant was discarded, and the pellet was washed twice with glycine buffer. Finally, the pellet was resuspended in glycine buffer with collagen Type I from calf skin (elastin products company inc.) to a final substrate concentration of 5 mg/ml. The tube was then incubated at 30 °C, and shaken horizontally for approximately 19 h. After centrifugation at 3 000 rpm (735 g) in an Eppendorf 5415C bench top centrifuge or 30 min, the supernatant was discarded and the pellet washed twice with glycine buffer. Finally the pellet was resuspended in 1.5 ml glycine buffer with 0.01 % (w/v) bovine serum albumin (BSA, Fermentas #B14). The tube was shaken for 18 h at 4 °C until the beads were completely resuspended. A control of non-coated beads, in only glycine buffer with 0.01 % (w/v) BSA (no collagen), was also prepared in this way.

3.2.3.2 BACTERIAL PREPARATION AND PAA

Bacterial colonies were resuspended in phosphate buffered saline (PBS) (pH7.4) and adjusted to 4 McFarland standard (or $\text{OD}_{600\text{nm}}$ at 0.669). On a microscope slide, 10 μl of beads and 10 μl of bacterial preparation were evenly mixed and shaken gently for approximately 3 minutes. The controls were: PBS mixed with coated and non-coated beads, in order to determine the effect of PBS on agglutination, and bacteria mixed with non-coated beads to determine if there was any substrate independent agglutination. *S. aureus* was used as a positive control for the binding of collagen and assigned a score of strong binding (3), whereas *M. luteus* was used as a negative control and assigned a score of no binding (1) based on their ability to cause bead agglutination. All experimental strains were scored based on a comparison with the controls.

3.2.4 PROTEIN SAMPLE PREPARATION

3.2.4.1 CELL FREE EXTRACTS (CFES)

B. fragilis strains were grown anaerobically for 16 h to late log/early stationary phase in BHIS broth. The cells were harvested by centrifugation in a Beckman JA-14 rotor (10 000 rpm or 9600 g, 4°C, 15 min each) and cell pellets washed twice in 1 X PBS pH 7.4 by centrifugation in a Beckman JA-14 rotor (10 000 rpm or 9600 g, 4°C, 5 min each). The pellets were then resuspended in 1 ml PBS and the cells were disrupted by sonication using a Misonic sonicator 3000 at a power output of 3 W for 5 rounds of 30 s. Cell debris was removed by centrifugation in an Eppendorf 5415C bench top centrifuge (14 000 rpm or 16 000 g, 4°C, 10 min). The supernatant was collected (CFE) and protein concentrations were determined using Bio-Rad Protein Assay Dye Reagent Concentrate according to the manufacturer's instructions (Bio-Rad Laboratories, cat # 1500 - 0006).

3.2.4.2 OUTER MEMBRANE PROTEIN FRACTION (OMP)

The outer membrane protein extraction protocol was based on a previously described method (Ferreira *et al.*, 2008) with various modifications described here in detail. For the preparation of the OMP fraction, the *B. fragilis* CFE was ultracentrifuged for 1 h at 39 000 rpm (100 000 g) in a Beckman 50Ti rotor. The resulting crude membrane pellet was resuspended in 200 µl of PBS (pH7.4) and mixed with up to 9.8 ml of 0.3 % (w/v) N-Lauroylsarcosine sodium (Sigma, L5125) solution for 1 h at 4 °C to dissolve the inner membrane. The outer membrane was then collected by ultracentrifugation as described above, and the pellets were washed using PBS (pH 7.4). After a final ultracentrifugation step, the OMP pellet was resuspended by vigorous agitation in 500 µl PBS (pH7.4).

3.2.5 COLLAGEN FAR WESTERN ANALYSIS

The Far Western procedure was adapted from previously described methods (Esmay *et al.*, 2003; Ferreira *et al.*, 2006) and is briefly described here. The protein sample (75 µg) was mixed in a 1:1 ratio with 2X SDS gel loading buffer (without β-mercaptoethanol) (0.125 M Tris-HCl [pH6.8], 4 % SDS, 20 % glycerol, 0.002 % bromophenol blue) and incubated at room temperature for 10 min prior to electrophoresis in a 10 % SDS (w/v) polyacrylamide gel without β-mercaptoethanol (Laemmli, 1970). PageRuler™ Prestained Protein Ladder (Fermentas, #SM0671) was used as the molecular weight standard for all protein gels. Samples were run on two separate duplicate gels. One set of separated protein was stained with Coomassie brilliant blue, and the other was transferred on to a nitrocellulose membrane using the Trans-Blot® SD DNA/RNA Blotting Kit from Bio-Rad as per the manufacturer's instructions.

The nitrocellulose blots were blocked overnight by incubating at room temperature (RT) in blocking buffer (1X PBS buffer, pH7.2 with 1.5 % BSA [Roche, 10 735 086 001] and 5% skim milk). This was followed by a 3 h incubation period at RT in PBS-B-T buffer (1X PBS, pH7.2; 1.5 % BSA; 0.1 % Tween) containing 20 µg/ml of collagen Type I from calf skin (Elastin Products Company Inc.). The membrane was then washed three times with PBS-B-T buffer, and incubated with anti-collagen type I (Rabbit) antibody (Rockland, 600-401-103-0.1) at a 1:500 dilution, for 1 h, followed by another 3 washes. Incubation with peroxidase conjugated anti-rabbit IgG (Goat) antibody (1:5000 using Rockland, 611-1302; or 1:2500 using Gene Tex, GTX77060) was then performed for 1 hour at RT, followed by 3 washes with PBS-T buffer (1 X PBS, pH7.2; 0.1 % Tween) and the membrane was finally developed with TMB Membrane Peroxidase substrate (KPL, 50-77-18). In order to confirm the specificity of the anti-collagen type I (Rabbit) antibody a control blot was exposed to the antibody without the collagen binding step.

3.2.6 COLLAGEN ZYMOGRAPHY

Collagen zymography was performed as described in Chapter 2.

3.2.7 COLLAGEN AFFINITY CHROMATOGRAPHY

3.2.7.1 COUPLING OF COLLAGEN TO CYANOGEN BROMIDE-ACTIVATED SEPHAROSE

Cyanogen bromide (CnBr) activated sepharose matrix (Sigma, C9142) was coupled with collagen Type I by the method described below which was adapted from the manufacturer's instructions.

CnBr-activated resin (1 g) was rehydrated for 30 min in cold 1 mM HCl and washed several times in approximately 10 ml volumes of 1 mM HCl to a total volume of 200 ml. The resin was then washed with 5 to 10 volumes of sterile distilled water (dH₂O) followed by a wash with coupling buffer (0.1 M NaHCO₃, 0.5 M NaCl, pH8.5). This buffer was replaced by coupling buffer containing 5 mg/ml of collagen type I, and the suspension was mixed, using an end-over-end stirrer, at 4 °C for 16 to 19 h for coupling to take place. The unreacted collagen was removed by washing 3 times with coupling buffer. Coupling buffer was then replaced by 0.2 M glycine buffer (pH8) and the resin was incubated in this for 2 h at room temperature (about 18 to 25 °C). The resin was then washed alternately with the alkaline, coupling buffer, and acidic acetate buffer (0.1 M sodium acetate, 0.5 M NaCl, pH 4) for four cycles. The buffer was then replaced with 1M NaCl and the column was packed and stored in 1 M NaCl with 0.1% (w/v) sodium azide at 4°C until use.

3.2.7.2 AFFINITY PURIFICATION OF COLLAGEN ADHESINS

The outer membrane protein (OMP) fraction or cell free extract (CFE) of *B. fragilis* GSH18 strain (3-7 µg in 10 ml PBS) were passed three times through the collagen affinity column which was then washed with 20 ml PBS. The first seven column volumes (2 ml) were collected (W1 to W7). Elution of the bound proteins was done in a step-wise manner with 2 ml of 0.15, 0.3, 0.6, 1, 1.5, and 2 M NaCl solution.

The protein content of 200 µl of the OMP (or CFE) samples, the non-binding fraction (NBF) and wash fractions (W1 to W7), and 2 ml of each elution fraction were precipitated using 100 % (w/v) Trichloroacetic acid (TCA) (Isaacson *et al.*, 2006). Each protein pellet was then resuspended in 20 µl of 1X SDS-PAGE loading dye and examined on SDS – 10 % (w/v) polyacrylamide without β-mercaptoethanol.

3.2.8 TWO-DIMENSIONAL (2D) GEL ELECTROPHORESIS

Outer membrane proteins (250 µg per strip) were precipitated using a trichloroacetic acid (TCA) precipitation protocol based on a previously described method (Isaacson *et al.*, 2006). Briefly, samples were mixed with 100% (w/v) TCA at a volume ratio of 4:1 (sample to TCA), placed on ice for 20 min, and the precipitated proteins collected by centrifugation in an Eppendorf 5415C bench top centrifuge (14 000 rpm or 16 000 g, 4°C, 5 min). The pellets were then washed three times with ice-cold 80% (v/v) acetone and dried using a Savant Speed Vac[®] Plus SC210A (Thermo Quest). Pellets were then solubilised in 100 µl of urea lysis buffer (ULB) (8 M urea, 2% Triton X-100, 2% CHAPS, 2 M thiourea) with 2 µl of 1 M NaOH. Once completely solubilised, 2.5 µl of 50% (w/v) dithiothreitol (DTT) with bromophenol blue, 7.5 µl of 20% ampholytes (pH 3-10) (Bio-Lyte 3-10, Bio-Rad) and 15 µl of ULB were added to the mixture which was then applied to a 7 cm pH 3 – 10 IPG strip (Bio-Rad) for passive rehydration. Iso-electric focusing (IEF) was performed in a Protean IEF Cell (Bio-Rad) as follows: 250V, 15 m; 2500V, 2h30m; 4000V, 20 000Vh. Equilibration of focused strips before second dimension separation was done sequentially in equilibration buffer I (6 M urea, 30% (v/v) glycerol, 2% (w/v) SDS, 1% (w/v) DTT, bromophenol blue) and equilibration buffer II (6 M urea, 30% (v/v) glycerol, 2 % (w/v) SDS, 4.8% (w/v) iodoacetamide, bromophenol blue). The second dimension was done on a 10% SDS polyacrylamide gel.

Four gels were run for the wild-type *B. fragilis* GSH18 strain for analysis and protein spot picking. Another gel was run for Far Western analysis, which was performed as described in 3.2.5.

3.3 RESULTS AND DISCUSSION

3.3.1 IDENTIFICATION OF STRAINS

The *16S rRNA* genes of the strains were sequenced in order to confirm the identity of the 24 clinical isolates from Groote Schuur hospital as *B. fragilis*. Additional, sequencing of the *recA* gene of strains GSH 5, 13, 18, and 19 was done for definitive identification of these isolates. The sequencing results, presented in Table 3.3 show that all except one of the strains were, in fact, *B. fragilis*. Strain GSH 24 was, however, identified as *Bacteroides thetaiotaomicron* and was, therefore, not used further in this study.

Table 3.3. Identification of clinical isolate from Groote Schuur Hospital

Strain	Accession	Description	Max. Identity
GSH 2	X83940.1	<i>B.fragilis</i> 12491, 16S rRNA	99%
GSH 3	X83940.1	<i>B.fragilis</i> 12491, 16S rRNA	99%
GSH 5	AP006841.1	<i>B. fragilis</i> YCH46, putative RecA	99%
GSH 6	AP006841.1	<i>B. fragilis</i> 638R, 16S rRNA	99%
GSH 7	X83942.1	<i>B.fragilis</i> B119, 16S rRNA	99%
GSH 8	X83938.1	<i>B.fragilis</i> 2991, 16S rRNA	99%
GSH 9	X83941.1	<i>B.fragilis</i> TAL3636, 16S rRNA	99%
GSH 10	AB542764.1	<i>B. fragilis</i> JCM 11017, 16S r RNA	99%
GSH 12	AP006841.1	<i>B. fragilis</i> 638R, 16S rRNA	99%
GSH 13	AP006841.1	<i>B. fragilis</i> YCH46, putative RecA	99%
GSH 14	X83938.1	<i>B.fragilis</i> 2991, 16S rRNA	99%
GSH 15	X83943.1	<i>B.fragilis</i> Bfr81, 16S rRNA	99%
GSH 16	AP006841.1	<i>B. fragilis</i> 638R, 16S rRNA	99%
GSH 17	AP006841.1	<i>B. fragilis</i> 638R, 16S rRNA	100%
GSH 18	AP006841.1	<i>B. fragilis</i> YCH46, putative RecA	99%
GSH 19	M63029.1	<i>B. fragilis</i> YCH46, putative RecA	99%
GSH 20	AP006841.1	<i>B. fragilis</i> 638R, 16S rRNA	100%
GSH 21	EU728706.1	<i>B. fragilis</i> 638R, 16S rRNA	99%
GSH 22	AB510701.1	<i>B.fragilis</i> JCM 11019, 16S rRNA	99%
GSH 23	AP006841.1	<i>B. fragilis</i> 638R, 16S rRNA	99%
GSH 24	AB510710.1	<i>B. thetaiotaomicron</i> JCM 5827, 16S rRNA	99%
GSH 25	AP006841.1	<i>B. fragilis</i> 638R, 16S rRNA	99%
GSH 28	AB542764.1	<i>B. fragilis</i> JCM 11017, 16S rRNA	99%
GSH 30	AB510701.1	<i>B.fragilis</i> JCM 11019, 16S rRNA	99%

3.3.2 PARTICLE AGGLUTINATION ASSAYS

The 23 *B. fragilis* clinical isolates were first analyzed for their ability to bind to collagen using the particle agglutination method. *S. aureus* and *M. luteus* were used as positive and negative controls respectively, as recommended by Naidu *et al.* (1988). In addition to these controls, a known collagen binding strain, *B. fragilis* P35 (Soki *et al.*, 1999) and a known laminin binder, *B. fragilis* MC2 (Ferreira *et al.*, 2006) were tested for comparison with the South African clinical isolates. *B. fragilis* 638R was also tested to evaluate the hypothesis that recent clinical isolates may be more virulent and, therefore, display greater ability to bind and degrade the ECM protein, collagen.

All, except two of the experimental strains tested (GSH 12 & 25), could bind collagen type I, although none were strong collagen binders (Table 3.4 and Figure. 3.1). The control strains *B. fragilis* 638R and MC2 were both weak collagen binders (binding level 1) as were the majority of the clinical isolates (16/23). As expected *B. fragilis* P35 was a good collagen binder (binding level 2), as were isolates GSH 5, 7, 13, 18 and 19. These GSH isolates were further characterized by collagen far western and collagen zymography and compared to the three control strains.

Table 3.4: Particle agglutination assays (PAA) of *B. fragilis* strains

Strain	Degree of Binding*	Strain	Degree of Binding*
638R	1	GSH 14	1
P35	2	GSH 15	1
MC2	1	GSH 16	1
GSH 2	1	GSH 17	1
GSH 3	1	GSH 18	2
GSH 5	2	GSH 19	2
GSH 6	1	GSH 20	1
GSH 7	2	GSH 21	1
GSH 8	1	GSH 22	1
GSH 9	1	GSH 23	1
GSH 10	1	GSH 25	0
GSH 12	0	GSH 28	1
GSH 13	2	GSH 30	1

*No binding (0), weak binding (1), good binding (2) and strong binding (3).

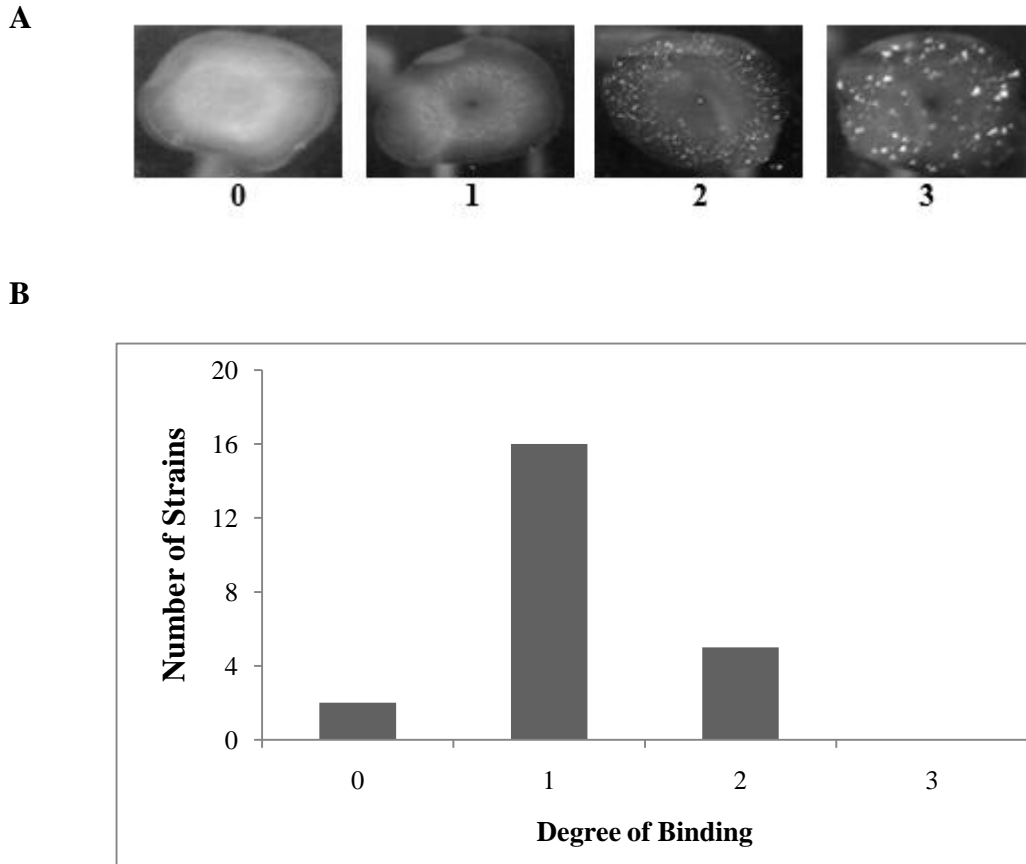


Figure 3.1. Particle agglutination assays. (A) Agglutination scoring system using controls. No binding (0) was represented using *M. luteus*, weak binding (1) using *B. fragilis* MC2, good binding (2) using *B. fragilis* P35, and strong binding (3) using the *S. aureus*. (B) Distribution of binding ability of *B. fragilis* clinical isolates (GSH strains) to collagen type I.

3.3.3 COLLAGEN ADHESINS AND PROTEASES OF CLINICAL STRAINS

In Chapter 2 it was shown that *B. fragilis* 638R produced a number of collagen proteases in different cell fractions. However, in this chapter, five additional clinical strains were identified using the PAA method, which could bind collagen better than *B. fragilis* 638R, and these were taken forward for further studies. The collagen protease and adhesion profiles of these GSH strains were examined by collagen zymography and collagen Far Western analysis.

The examination of the CFEs of the five *B. fragilis* GSH strains, and the control strains, *B. fragilis* 638R, MC2 and P35, revealed that they all produced a protease of about 45 kDa (refer to Figure. 3.2 for representative gel). A less intense protease band, of about 37 to 40 kDa, was also observed (Figure. 3.2 B). Similarly, a collagen adhesin band of approximately 45 kDa was found in all strains (Figure. 3.2 C). This binding was specific as the control blot, in which no collagen binding step was performed, appeared completely blank (data not shown). One of the five best collagen binders, GSH18, was chosen for further analysis.

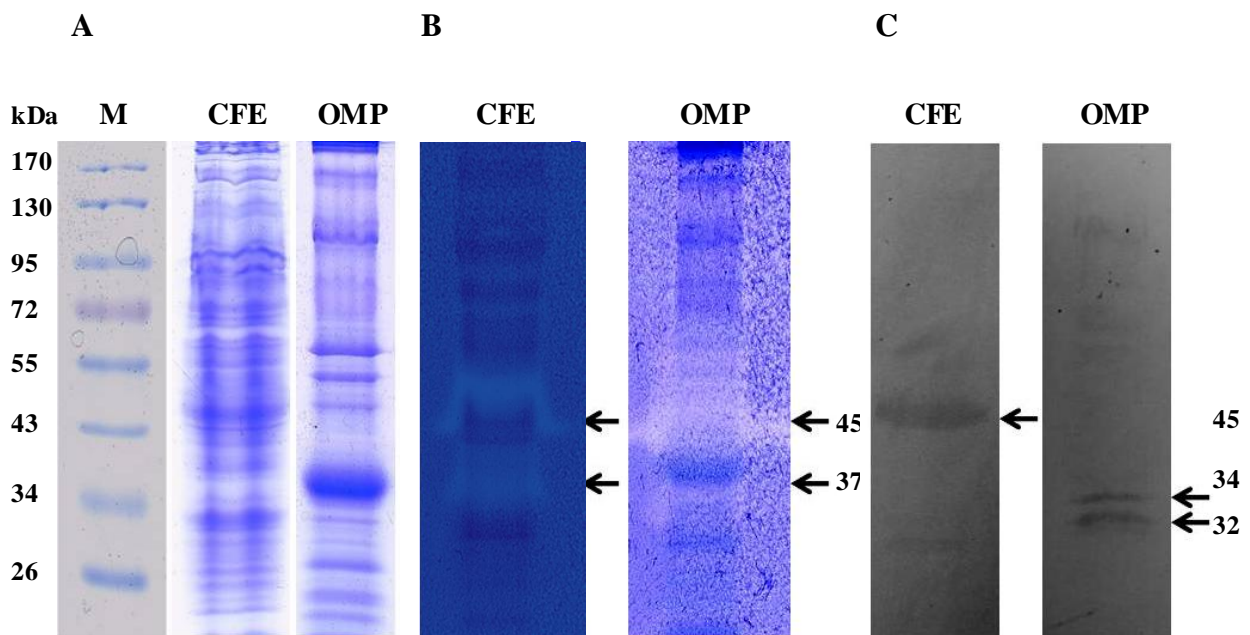


Figure 3.2 Collagen Far Western and zymogram of cell free extracts (CFE) and outer membrane protein (OMP) of *B. fragilis* GSH18. (A) Coomassie stained gel, (B) Zymogram, (C) Far Western. M: PageRuler™ #SM0671

Interestingly, the OMP fraction was not enriched for the adhesin of ~45 kDa, but did reveal the presence of two other collagen adhesins at about 32 and 34 kDa (Figure.3.2 C). The absence of this ~45 kDa adhesin in the OMP implied that it is present either in the cytoplasmic or the inner membrane fraction. The 45 kDa protease band was, however, active in both the fractions, as was the smaller 37 kDa band.

For a bacterium to be able to bind extracellular components, like collagen, for the purposes of attachment, it is necessary that they should express the relevant adhesins on their cell surface. In fact, many of the ECM adhesins described for *B. fragilis* to date have all been found in the outer membranes (Ferreira *et al.*, 2009; Pauer *et al.*, 2009). The presence of an adhesin and protease band of apparently similar size in the initial CFE Far Western and zymogram was interesting since it suggested that the same protein might be responsible for binding and degrading the collagen. However, the lack of a detectable ~45 kDa collagen binding band in the OMP fraction was somewhat disappointing as it indicated that the phenotypes might not be attributed to the same protein after all.

However, another scenario is also possible. The putative collagen-binding adhesin could be present in the OMP at concentrations that were too low for a detectable Far Western band to be seen. The adhesion to collagen is detected as a “snap shot” moment where a limited number of collagen molecules bind to the limited number of adhesin molecules immobilised on the nitrocellulose membrane. The hydrolytic action of the protease component, however, is seen as the cumulative degradation of several collagen molecules over the 48 h activation period meaning that even a limited amount of enzyme could account for the zone of clearing observed in the zymogram. With this possibility in mind, affinity purification was performed using both the CFE and the OMP fractions.

3.3.4 AFFINITY PURIFICATION OF COLLAGEN ADHESINS

The affinity purification method was found to be effective in isolating the collagen adhesins in the CFE and OMP (Figure. 3.3 and 3.4).

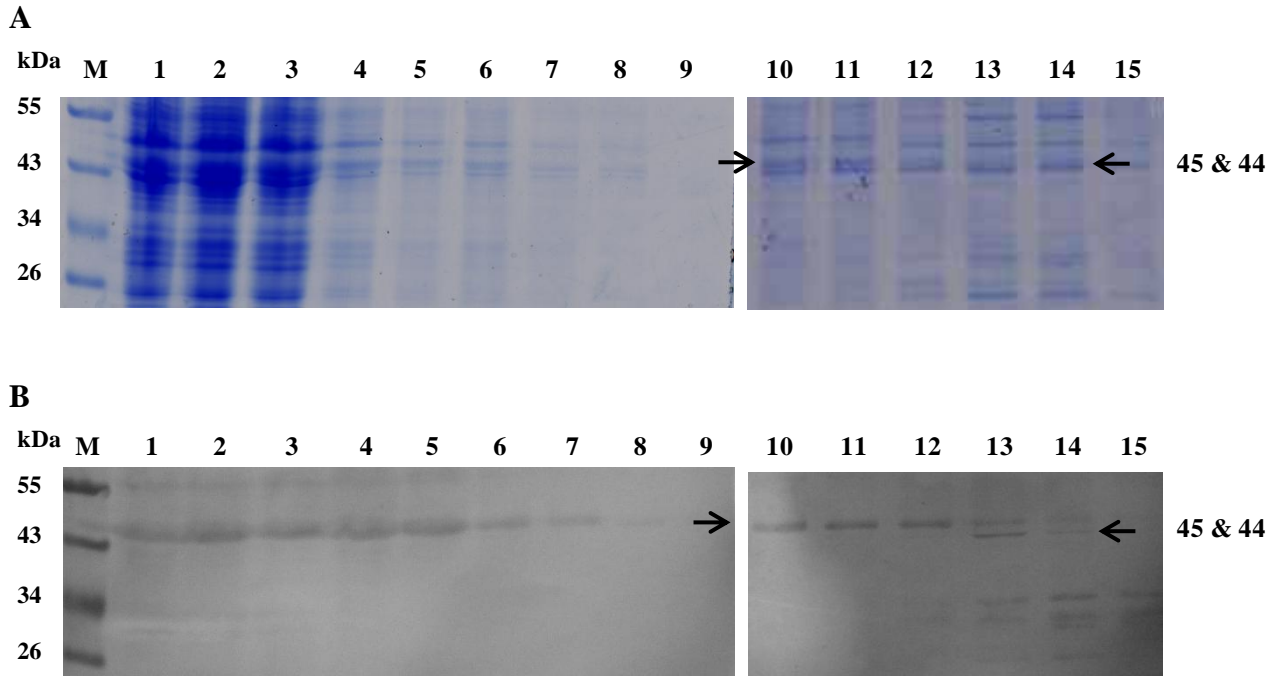


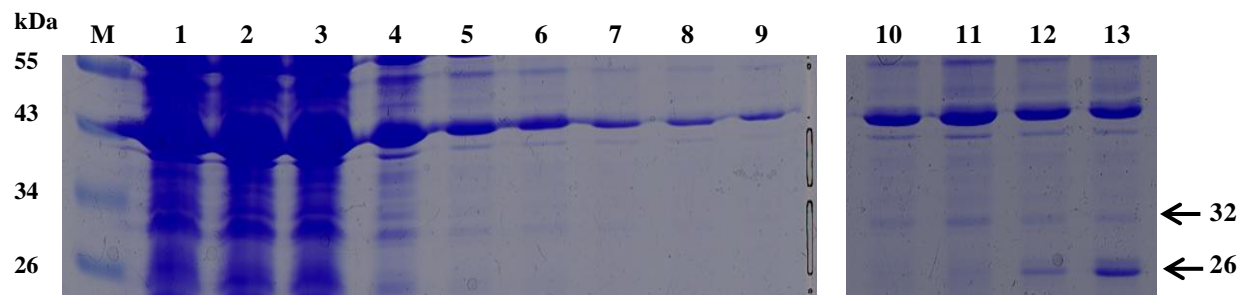
Figure 3.3 Affinity purification of collagen adhesins of the cell free extract (CFE) of *B. fragilis* GSH18: (A) Coomassie stained gel, (B) Far Western. **Lanes:** M: PageRuler™ #SM0671, **1:** CFE, **2:** Non-binding fraction (NBF), **3 – 9:** Washes 1 to 7 (W1 to W7), **10 -15:** Elutions with 0.15, 0.3, 0.6, 1, 1.5 and 2 M NaCl respectively.

The purification process eliminated a large number of proteins, allowing for the visualization of a distinct 45 kDa band. This protein band was excised from lane 10 (0.15 M NaCl). A slightly smaller 44 kDa band just below the above-mentioned band became more predominant from lane 13 (1 M NaCl), as the 45 kDa band concentration decreased. This was clearly seen in the Far Western as well (Figure. 3.3 B) and for this reason both the top (45 kDa) band, in lane 10, and the bottom (44 kDa) band, from lane 13, were excised and subjected to LC MS/MS analysis.

The purification of the adhesins from the OMP fraction gave interesting results. Collagen adhesins were seen from the first elution fraction (Figure. 3.4 B). Only a 32 kDa protein band appeared in the Coomassie stained gel (Figure. 3.4 A), which seemed to be eluted predominantly at 1M NaCl, and no distinct 34 kDa band was present. The Far western

did not reflect an increase in intensity of the 32 kDa band at 1 M NaCl and, surprisingly, it was the 34 kDa adhesin band that was more distinct throughout the elution fraction (Figure. 3.4 B). A novel 26 kDa band, with corresponding collagen binding activity, was eluted at about 2M NaCl. Both the 32 kDa band eluted at 1 M NaCl and the 26 kDa band eluted at 2 M NaCl were excised from the protein gel and subjected to LC MS/MS analysis.

A



B

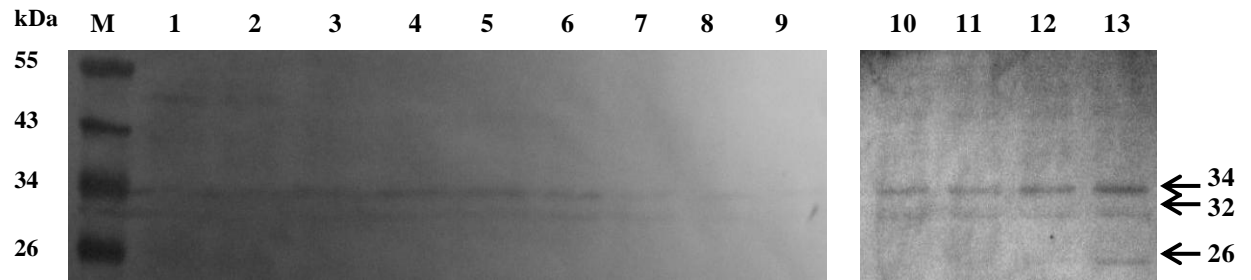


Figure 3.4 Affinity purification of collagen adhesins of the outer membrane protein fraction (OMP) of *B. fragilis* GSH18: (A) Coomassie stained gel, (B) Far Western. **Lanes:** M: PageRuler™ #SM0671, 1: CFE, 2: Non-binding fraction (NBF), 3 – 9: Washes 1 to 7 (W1 to W7), 10 -13: Elutions with 0.6, 1, 1.5 and 2 M respectively.

In each purification profile (Figure. 3.2 and 3.3) the adhesin bands could be seen in the CFE (or OMP), as well as in the non-binding (NBF) and in the washes, suggesting that there was a limit to the amount of adhesins that could bind to this column. The problem of scaling up this experiment was, therefore, a limiting factor in the success of the purification of the adhesins. In order to be able to visualize the proteins on a Coomassie stained SDS-PAGE, the fractions had to be TCA precipitated. This meant that there was insufficient purified product for proper zymographic analysis of the protease activity, as the concentrated samples lost proteolytic activity in the TCA treatment. Confirmation of the collagen adhesive properties of the purified proteins by Far Western analysis was, however, unaffected by the TCA treatment.

It is interesting to note that the collagen binding ability, of the putative collagen adhesins observed here, is not dependant on the proteins' structural conformation. The proteins undergo at times both the denaturing affects of SDS treatment and of TCA exposure and still retain their collagen binding ability without the need for a renaturing step. This phenomenon is observed for other adhesins including the *B. fragilis* fibrinogen binding protein (BF-FBP) (Houston *et al.*, 2010). This property is useful for detection of adhesins which have been separated under highly denaturing conditions such as those used in 2-dimensional gel electrophoresis.

3.3.5 TWO-DIMENSIONAL SEPERATION OF COLLAGEN ADHESINS

Two-dimensional gel electrophoresis is a very powerful separation tool. It has been previously used, in combination with Far Western hybridization, for detection of MSCRAMMs (Microbial Surface Components Recognizing Adhesive Matrix Molecules) such as fibronectin adhesins (Yu *et al.*, 2006). To our knowledge this technique has not been used previously for the detection of collagen adhesins.

Only the OMP fraction was analysed using 2-dimensional gel electrophoresis. Since this fraction contained a smaller number of proteins, the individual protein spots would be

seen more distinctly and, therefore, enable them to be excised more accurately for further analysis. In order to observe the full spectrum of possible OMP adhesins, a pI range of 3 to 10 was used. Figure 3.5A shows an example of one of the 2D gels. The various Coomassie stained gels were analyzed, using the PDQuest Software version 7.4.0 (Bio-Rad), in order to match particular protein spots across the various gels. Matching the spots seen on the corresponding Far Western blot (Figure 3.5 B), however, was done manually due to the lack of coordinating proteins spots on the blot, which are necessary for PD Quest analysis. The specificity control, in which the collagen binding step was not performed, demonstrated that the adhesin spots seen in the experimental blots were not present and therefore the collagen binding is specific (data not shown).

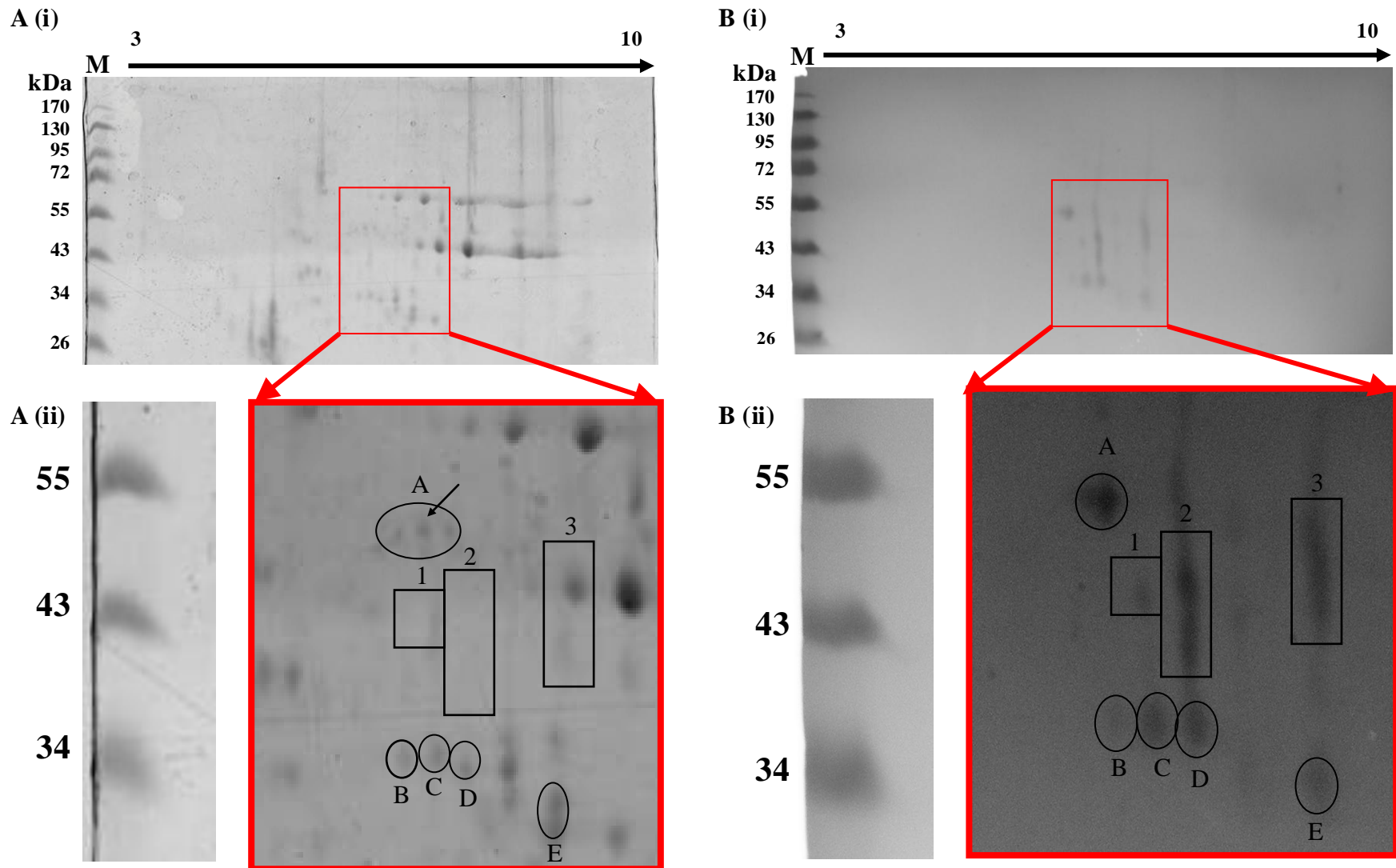


Figure 3.5. Identification of collagen-binding proteins from the OMP of *B. fragilis* GSH18 by two-dimensional gel electrophoresis. OMPs were separated in the first dimension on precast IPG strips (pH 3–10, indicated by the arrow above the figures) followed by separation in the second dimension on 10% SDS-PAGE gels. (A) Coomassie stained gels, (B) Collagen Far Western of proteins from 2D-gels. (i) Full gel images; (ii) Enlargement of the areas of interest encased in red. Black circles and boxes indicate the position of the collagen-binding proteins. M: PageRuler™ #SM0671. Further details discussed in text.

The collagen binding activity was restricted to proteins in an area between pI 5.7 and 7.3, with a size range between 30 and 54 kDa (Figure. 3.5 B ii). The areas that are circled represent distinct spots of adhesion. Adhesion areas B to E were related to distinct protein spots, each of which was excised from the protein gels and subjected to LC MS/MS analysis (see details Chapter 4).

Area A, at about 54 kDa was quite large. The corresponding area in the gel (Figure. 3.5A ii) had various spots in it. Only one of these was successfully isolated and analysed by LC MS/MS. This spot is identified in area A with an arrow. In the previous Far Western analyses performed (Figure. 3.3 and 3.4) only a very faint band was seen at this approximate size in first few lanes, between 54 and 55 kDa. The exact reason for this spot being so distinct in the 2D-SDS PAGE analysis is unclear.

The areas boxed in black in figure 3.5 B (ii) were mainly streaks of adhesion rather than distinct spots and it was, therefore, not possible to relate them to spots in the protein gels. Box 1 was less streaky but the corresponding area of the protein gels was void of distinct protein spots. It should be noted that boxed areas 1 to 3 lie directly above the circled areas C, D and E respectively and could be the same adhesins which have been retained in the gel areas above.

The questions that remain here are whether the boxed areas are distinct collagen adhesins or remnants of the spots below them. The size range of the boxed areas is of about 44kDa for box 1, 36 to 45 kDa for box 2 and 43 to 50 kDa for box 3. These size ranges, and especially that of box 1, are the most similar to the 45kDa adhesin band. It would appear that two-dimensional Far Western analysis of the CFE proteins would also be necessary in future work for elucidating the position of this adhesin.

3.4 CONCLUSION

In this chapter the aims were to use the best collagen binding strain available to us to determine the presence of specific collagen adhesins and to develop methods of isolating both these, and the collagen degrading proteins, for the purpose of identification by mass spectrometry. Various

methods for the separation of *B. fragilis* proteins were successfully used to isolate collagen adhesins, but they could not be used directly for the isolation of collagen proteases. An attempt was made to do 2D gel zymography of the OMPs, in which the second dimension was done using a polyacrylamide gel co-polymerised with collagen and then processed as described for normal zymograms (data not shown). No zones of collagen degradation were observed. This was not surprising as the protein samples in this method are exposed not only to trichloroacetic acid (TCA) treatment but also to harsh denaturants and reducing agents such as urea and dithiothreitol (DTT) with a resulting loss of activity.

This inhibition of proteolytic activity was also problematic in the analysis of the affinity purified collagen adhesin fractions. The elution fractions obtained had a very low protein concentration and, in order to obtain samples that could be visualized on Coomassie stained SDS-PAGE, each elution fraction had to be precipitated using TCA. This treatment did not prevent the ability of the proteins to bind collagen but it completely abolished the collagenase activity of the samples.

The various putative adhesin proteins excised from the gels were analyzed by mass spectrometry analysis in order to determine their identity. This analysis is discussed in detail in the following chapter.

CHAPTER 4

Bioinformatic analysis of the LC MS/MS results of the putative collagen Type I adhesins

4.1 INTRODUCTION.....	80
4.2 METHODS	81
4.2.1 LC MS/MS ANALYSIS OF PUTATIVE ADHESINS	81
4.2.2 DATA ANALYSIS.....	81
4.3 RESULTS	82
4.3.1 AFFINITY PURIFIED ADHESINS OF THE CFES.....	83
4.3.1.1 44 kDa BAND.....	84
4.3.1.2 45kDa BAND.....	85
4.3.2 AFFINITY PURIFIED ADHESINS OF THE OMP FRACTION	85
4.3.3 2D SDS-PAGE SEPARATED ADHESINS OF THE OMP FRACTION	88
4.3.3.1 SPOT A (54 kDa).....	88
4.3.3.2 33-34kDa cluster – SPOTS B to E.....	91
4.4. CONCLUSION	91

4.1 INTRODUCTION

In this study, two techniques for the purification and isolation of collagen adhesins were employed (Chapter 3). These were affinity chromatography, using collagen as a ligand for adhesins found in the CFEs and OMP fractions, and 2D SDS-PAGE. Recognition of collagen binding proteins was done by subsequent collagen Far Western analysis. These methods were successful in isolating several proteins of interest. Table 4.1 below lists the bands and spots that were excised for further analysis and their origins.

Table 4.1 Protein bands and spots excised for protein identification by LC MS/MS

Figure number	Protein fraction	Purification/isolation method	Band/spot
Figure. 3.3 A	CFE	Affinity chromatography	~44 kDa ~45 kDa
Figure. 3.4 A	OMP	Affinity chromatography	~26 kDa ~32 kDa
Figure. 3.5 A (ii)	OMP	2D SDS-PAGE	~54 kDa (spot A) 33 – 34 kDa cluster (spots B to E)

Liquid chromatography followed by tandem mass spectrometry (LC MS/MS) was the technique chosen for protein identification. This method is particularly useful for identifying proteins within protein mixtures (McCormack *et al.*, 1997) and reduces the requirement for performing extensive protein purification. The analysis of the results is also greatly facilitated by the availability of the complete genomes of three *B. fragilis* strains, namely 638R, NCTC 9343 and YCH46 mentioned previously. This allows for the resulting peptide sequences to be directly compared to a species specific database of predicted proteins and the genes encoding them.

This chapter describes the analysis of the LC MS/MS protein identification results and the choice of a putative collagen adhesin for further research.

4.2 METHODS

4.2.1 LC MS/MS ANALYSIS OF PUTATIVE ADHESINS

Protein bands excised from the affinity purification gels and protein spots excised from the 2-D PAGE gels (Chapter 3) were dried in eppendorf tubes for storage and transportation to the mass spectrometry service provider. The samples were sent to the Yale Cancer Centre Mass Spectrometry Resource & W.M. Keck Foundation Biotechnology Resource Laboratory in New Haven, Connecticut, United States of America. They were subjected to tryptic digestion followed by liquid chromatography and tandem mass spectrometry (LC MS/MS) analysis on a LTQ Orbitrap mass spectrometer (McCormack *et al.*, 1997; Stone and Williams, 2009).

4.2.2 DATA ANALYSIS

The spectra from the LC MS/MS were analyzed by the service provider using the Mascot algorithm (Hirosawa *et al.*, 1993). The genome sequences of *B. fragilis* available on NCBI were searched using the peak data generated, and the Mascot results were made available for the user on the YPED (Yale protein expression database) online viewing system.

The criteria used by the service provider for protein identification was that 2 or more MS/MS spectra match the same protein entry in the database searched and that the matched peptides derive from the type of enzymatic digestion performed on the protein, in this case tryptic digestion. Particular attention was also drawn to the fact that since Ile / Leu are isomeric and Phe/oxMet and Gln / Lys are isobaric, differentiation between these amino acids was not possible, although a Gln/Lys in the middle of a peptide is most likely the Gln since a tryptic digest was done. Proteins listed as matching with only 1 significant peptide match were not considered positive identifications and may simply be present in the sample

NCBI and/or the MEROPS peptidase database (Rawlings *et al.*, 2010) was used for BLAST analysis of the results. The PSI-BLAST algorithm (Altschul *et al.*, 1997) was employed.

4.3 RESULTS

A large amount of data was obtained from the LC MS/MS of the various protein gel bands and spots. Most of the results for each sample consisted of a mixture of various proteins which were represented to different extents in as much as the coverage of each is concerned. The selection of the most likely candidates for each sample is not a simple task. The proteins seen to have the highest score may not be the ones of interest but simply those found in higher amounts at the particular molecular weight, and pI, in the case of the 2D-separated proteins. On the other hand, if the protein which is responsible for the phenotype of interest is very abundant, it is likely to be found as a contaminant in another sample and even in the blank gel control.

The results of the LC MS/MS are shown in Tables 4.2, 4.3 and 4.4. These exclude any results that were classified as “enzyme used for digestion”, “probable contaminant” or results that were only tentative due to only one peptide being used for the identification. Peptide matches below the homology or identity threshold of each sample were also excluded.

Some the proteins represented in these tables were disregarded based on annotation that specifically implied a function that clearly had no link to the study, for example, the protein annotated as “DNA-directed RNA polymerase subunit alpha” of *B. fragilis* YCH46. Elimination based on apparent molecular weight discrepancies was, however, not possible due to the lack of knowledge about possible post-translational modifications of the proteins of interest. This is particularly true of the results obtained from the one dimensional gels as these were non-reducing gels (i.e. no β -mercaptoethanol was used) and there was a possibility of unresolved disulphide bonds being present.

Finally a large number of proteins annotated as hypothetical proteins were identified, each of which required more in-depth bioinformatic analysis before they could be either eliminated from the study or selected for further analysis.

4.3.1 AFFINITY PURIFIED ADHESINS OF THE CFEs

Table 4.2 shows the results of the putative collagen binding proteins of 44 and 45 kDa that were purified from the CFE of *B. fragilis* GSH18 using collagen affinity chromatography (from Figure.3.3 A, Chapter 3).

Table 4.2: Mass Spectroscopy Results of Affinity Purified adhesins of CFEs

Band	Score	Expectation	Protein ID	Protein Name	MW (kDa)	% Coverage	
44 kDa	1488	0	gi53712081	hypothetical protein BF0788 [<i>B. fragilis</i> YCH46]	39.9	52.8	
	1244	4.6E-118	gi53715321	major outer membrane protein OmpA BF4037 [<i>B. fragilis</i> YCH46]	43.3	54.9	
	1161	8.8E-110	gi253566135	citrate synthase [<i>Bacteroides</i> sp. 3_2_5]	50.9	37.6	
	782	7.1E-72	gi53712366	hypothetical protein BF1074 [<i>B. fragilis</i> YCH46]	48.7	24.9	
	704	4.4E-64	gi53712058	membrane-bound lytic murein transglycosylase D precursor [<i>B. fragilis</i> YCH46]	49.0	25	
	651	8.9E-59	gi53715439	DNA-directed RNA polymerase subunit alpha [<i>B. fragilis</i> YCH46]	37.4	30.6	
	651	9.2E-59	gi53714500	transcription termination factor Rho [<i>B. fragilis</i> YCH46]	76.6	21.3	
	589	1.4E-52	gi53714920	glutamate dehydrogenase [<i>B. fragilis</i> YCH46]	48.3	36.4	
	569	1.3E-50	gi60679959	hypothetical protein BF0371 [<i>B. fragilis</i> NCTC 9343]	48.4	27.1	
	568	1.8E-50	gi53713031	chaperone protein DnaJ [<i>B. fragilis</i> YCH46]	42.2	23.9	
	434	4E-37	gi53714555	phosphoglycerate kinase [<i>B. fragilis</i> YCH46]	45.1	19.1	
	336	2.5E-27	gi53714597	putative multidrug resistance protein [<i>B. fragilis</i> YCH46]	39.0	24	
	303	5E-24	gi53713703	TPR repeat-containing protein [<i>B. fragilis</i> YCH46]	46.5	16.1	
	45 kDa	573	5.4E-51	gi53715321	major outer membrane protein OmpA BF4037 [<i>B. fragilis</i> YCH46]	43.3	32.9
409		1.3E-34	gi29348149	elongation factor Tu [<i>B. thetaiotaomicron</i> VPI-5482]	43.6	16.8	
280		9.9E-22	gi53712480	phosphopyruvate hydratase [<i>B. fragilis</i> YCH46]	46.4	10	
269		1.5E-20	gi53715832	succinate dehydrogenase/fumarate reductase iron-sulfur subunit [<i>B. fragilis</i> YCH46]	27.2	27.5	
119		0.000013	gi53714650	aminopeptidase C [<i>B. fragilis</i> YCH46]	44.1	5.6	
248		1.6E-18	gi53715707	bifunctional preprotein translocase subunit SecD/SecE [<i>B. fragilis</i> YCH46]	110.0	4.2	
Blank		310	9.9E-25	gi53715321	major outer membrane protein OmpA BF4037 [<i>B. fragilis</i> YCH46]	43.3	17.2
		190	1E-12	gi53712081	hypothetical protein BF0788 [<i>B. fragilis</i> YCH46]	39.9	10.8

Identity or homology score thresholds: 44 kDa (264); 45 kDa (81); Blank (38)

4.3.1.1 44 kDa BAND

In this sample the best hit was a hypothetical protein BF0788 of *B. fragilis* YCH46. PSI-BLAST analysis of this protein sequence, however, yielded no putative function. This protein was also identified in the blank (10.8 % coverage) and for these reasons combined it was not analysed further in this study. The other results included the “major outer membrane protein OmpA” of *B. fragilis* YCH46 as well as two other hypothetical proteins, BF1074 of *B. fragilis* YCH46 and BF0371 of NTCT 9343. PSI-BLAST of the hypothetical proteins did not yield a putative function.

The OmpA had the second best score and showed almost 55% coverage. The percentage coverage relates to the proportion of sequence of the particular protein that was covered by the sequence of the peptides generated from the LC M/MS analysis. Unfortunately, even though OmpA had quite a high coverage it was also the main protein found in the blank control (at 17% coverage). However, eliminating this candidate based on this information alone may be premature. It is not surprising that this protein was found in the blank control as OmpAs are the most abundant outer membrane proteins of *B. fragilis* (Pumbwe *et al.*, 2007). Furthermore, OmpA has been suggested to be involved in the ability of various bacterial strains, including *Bacteroides ovatus* and *B. fragilis*, to bind human cells (Pumbwe *et al.*, 2007; Sato *et al.*, 2010). *B. ovatus* derived *ompA* genes, expressed in *E. coli*, increase the latter's ability to adhere to human cells (Sato *et al.*, 2010).

The OmpA homologue identified in the LC MS/MS was the *B. fragilis* YCH46 homologue of OmpA1, BF4037 (BF3810 of *B. fragilis* NCTC4343). This is one of seven *ompA* family gene homologues (Wexler *et al.*, 2009). *B. fragilis* strains have shown increased *ompA1* expression during exposure to bile-salt treatment, which coincided with an increase in adhesion to human epithelial cells (Pumbwe *et al.*, 2007). The OmpA1 may be a good candidate for the ~45 kDa collagen adhesin. However, without being able to purify the collagen binding protein to greater homogeneity, for more accurate protein identification and mass determination, it is difficult to validate this possibility. Furthermore, the only study where the involvement of OmpA-like proteins in collagen adhesion was investigated, reported that the homologue of *Vibrio cholerae*,

OmpU, could adhere to fibronectin but was unable to bind other matrix proteins such as collagen or laminin (Sperandio *et al.*, 1995). The *B. fragilis* OmpA homologues have previously been the subject of functional studies via gene-specific mutagenesis (Wexler *et al.*, 2009) and it would be interesting, in future studies, to investigate whether these mutants show any differences in collagen adhesion characteristics relative to the wildtype. However, in the current study, it was decided that further novel proteins would be investigated as described below.

4.3.1.2 45kDa BAND

The quality of the results for the 45 kDa band was relatively poor, in that after the elimination process only two putative candidates remained for consideration. The one with the better score was the OmpA1 of *B. fragilis* YCH46 and, for the reasons discussed previously, was not considered further. The other candidate was a putative “aminopeptidase C” of a predicted molecular weight of about 44 kDa. In *B. fragilis* YCH46 this gene (BF3364) is shown to have a conserved C1 peptidase domain. When its predicted protein sequence was blasted on the MEROPS peptidase database (Rawlings *et al.*, 2010), it was identified as being part of subfamily C1A peptidase. Even though the size of the protein is close to that of the larger collagenase band observed in the zymograms, it is unlikely to be responsible for the collagen proteolysis observed as collagenases are generally endopeptidases and not exopeptidases as is the case with aminopeptidases.

4.3.2 AFFINITY PURIFIED ADHESINS OF THE OMP FRACTION

The collagen affinity purification of the OMP fraction resulted in the isolation of two protein bands of ~26 and ~32 kDa with the ability to adhere to collagen (Figure. 3.4, Chapter 3).

Table 4.3: Mass Spectroscopy Results of Affinity Purified adhesins of the OMP

Band	Score	Expectation	Protein ID	Protein Name	MW (kDa)	% Coverage	
26kDa	1567	0	gi 53711952	hypothetical protein BF0662 [<i>B. fragilis</i> YCH46]	31.6	66	
	554	3.7E-49	gi 53715026	putative biopolymer transport protein [<i>B. fragilis</i> YCH46]	28.9	36.6	
	313	5E-25	gi 60682546	hypothetical protein BF3077 [<i>B. fragilis</i> NCTC 9343]	22.0	39.9	
	287	2.2E-22	gi 60680892	hypothetical protein BF1378 [<i>B. fragilis</i> NCTC 9343]	21.8	33.3	
	152	5.9E-9	gi 53712509	hypothetical protein BF1217 [<i>B. fragilis</i> YCH46]	28.1	38.4	
	124	0.0000041	gi 53714054	hypothetical protein BF2762 [<i>B. fragilis</i> YCH46]	22.0	21	
	122	0.0000069	gi 253563053	conserved hypothetical protein [<i>Bacteroides</i> sp. 3_2_5]	37.6	7.3	
	110	0.000093	gi 53713187	OmpA family outer membrane protein BF1897 [<i>B. fragilis</i> YCH46]	23.1	9.8	
32kDa	1420	0	gi 53711952	hypothetical protein BF0662 [<i>B. fragilis</i> YCH46]	31.6	73.3	
	863	5.3E-80	gi 53712590	putative outer membrane protein BF1298 [<i>B. fragilis</i> YCH46]	42.3	29.4	
	625	3.2E-56	gi 53715321	major outer membrane protein OmpA [<i>B. fragilis</i> YCH46]	43.3	33.9	
	466	2.7E-40	gi 254884606	predicted protein [<i>Bacteroides</i> sp. 4_3_47FAA]	27.8	41.6	
	390	9.5E-33	gi 53711673	hypothetical protein BF0382 [<i>B. fragilis</i> YCH46]	33.8	29.7	
	329	1.2E-26	gi 53712065	hypothetical protein BF0772 [<i>B. fragilis</i> YCH46]	27.8	36	
	323	5.2E-26	gi 53715463	50S ribosomal protein L2 [<i>B. fragilis</i> YCH46]	29.7	32.5	
	307	1.8E-24	gi 53713893	hypothetical protein BF2601 [<i>B. fragilis</i> YCH46]	54.8	13.8	
	259	1.1E-19	gi 53715026	putative biopolymer transport protein [<i>B. fragilis</i> YCH46]	30.0	22	
	239	1.2E-17	gi 53714658	electron transfer flavoprotein beta-subunit [<i>B. fragilis</i> YCH46]	31.1	14.1	
	212	5.6E-15	gi 53714795	hypothetical protein BF3510 [<i>B. fragilis</i> YCH46]	35.0	20.8	
	208	1.4E-14	gi 53711881	hypothetical protein BF0590 [<i>B. fragilis</i> YCH46]	27.5	14.7	
	183	4.6E-12	gi 33468445	hypothetical protein [<i>B. ovatus</i>]	40.5	13.7	
	183	4.9E-12	gi 53714576	hypothetical protein BF3290 [<i>B. fragilis</i> YCH46]	33.8	21.1	
	Blank	842	5.9E-78	gi 50787940	putative plasminogen binding protein precursor [<i>B. fragilis</i>]	61.4	39.3
		299	1.3E-23	gi 53715321	major outer membrane protein OmpA BF4037 [<i>B. fragilis</i> YCH46]	43.3	24.1
234		3.9E-17	gi 53711952	hypothetical protein BF0662 [<i>B. fragilis</i> YCH46]	31.6	18.9	
166		2.3E-10	gi 53713185	hypothetical protein BF1895 [<i>B. fragilis</i> YCH46]	71.3	10.2	
150		1.1E-8	gi 53711959	hypothetical protein BF0669 [<i>B. fragilis</i> YCH46]	56.8	9.9	
128		0.0000017	gi 53713013	hypothetical protein BF1724 [<i>B. fragilis</i> YCH46]	59.5	4.4	
121		0.0000079	gi 53711863	hypothetical protein BF0572 [<i>B. fragilis</i> YCH46]	24.3	13.3	

Identity or homology score thresholds: 26 kDa (106); 32 kDa (181); Blank (109)

For both the 26 kDa and the 32 kDa samples the highest hit obtained was a hypothetical protein, BF0662, of *B. fragilis* YCH46. The Mascot coverage maps shown in Figure 4.1 revealed that the peptide sequences obtained for sample 33 kDa covered about 73% of the sequence (seen in red) of this hypothetical protein, whereas those for 26 kDa only covered about 66%. Most significantly, this 66% coverage seen for the smaller protein was restricted to the C-terminal end of the protein sequence after about the 70th amino acid, whereas the peptides obtained from the larger band extended to almost the 30th amino acid. This suggested that the 26kDa band might be

a processed form of the BF0662 homologue, or even a degradation product of the full length protein. The fact that this smaller version was still able to adhere to collagen hints at the presence of a binding domain towards the C-terminal side of the protein.

A

```

1      MKKLAILLVC AAMLASCNGL GGGSKDLKAE NDSLIMELTQ RNAELDEMMG TFNEVQEGFR
61     KINAAESRVD LTRGTISENS ATAKQQIASD IEFITKQMEE NKAQIAKLQA MLKSSKNNSA
121    QLKKAVESLT QELVAKTQRI EELQAELASK NIRIQELDAA VTGLTADKES LAENEAKAK
181    TVAEQDKAIN SAWFVFGTKS ELKTQKILEK GDVLKSADFN KDYFTQIDIR TTKEIKLYSK
241    RAELLTTHPA KSYELVKDDK GQLTLKITNP KEFWSVSKYL VIQVK

```

B

```

1      MKKLAILLVC AAMLASCNGL GGGSKDLKAE NDSLIMELTQ RNAELDEMMG TFNEVQEGFR
61     KINAAESRVD LTRGTISENS ATAKQQIASD IEFITKQMEE NKAQIAKLQA MLKSSKNNSA
121    QLKKAVESLT QELVAKTQRI EELQAELASK NIRIQELDAA VTGLTADKES LAENEAKAK
181    TVAEQDKAIN SAWFVFGTKS ELKTQKILEK GDVLKSADFN KDYFTQIDIR TTKEIKLYSK
241    RAELLTTHPA KSYELVKDDK GQLTLKITNP KEFWSVSKYL VIQVK

```

Figure 4.1 BF0662 MASCOT Coverage maps for affinity purified OMP samples (A) 26 kDa and (B) 33 kDa protein bands. The sequences in red denote the peptide sequence coverage.

In spite of encoding an unknown hypothetical protein, the BF0662 homologue appeared to be the best candidate based on percentage coverage as well as its predicted molecular mass. However, several other hypothetical proteins were also available for consideration. The OmpA1 was only found in the blank control, but two other OmpAs were observed, one in each of the experimental samples. In the 26 kDa sample, the OmpA6 homologue of *B. fragilis* YCH46, BF1897 (BF1959 of *B. fragilis* NCTC4343), was present but it had a very low score and a coverage of only 9.8%. The *B. fragilis* YCH46 OmpA3 homologue, BF1298 (BF1285 of *B. fragilis* NCTC4343), present in the 32 kDa sample, showed a much higher coverage (29.4 %), however this protein has an predicted size of about 43 kDa which is significantly higher than that estimated for the adhesin. None of these other candidates, however, had nearly as high a score and coverage as BF0662 and none were present in both samples. It should be noted that, although, BF0662 was also present in the blank, the coverage was only 19% suggesting that in this sample it was a low level contaminant.

4.3.3 2D SDS-PAGE SEPARATED ADHESINS OF THE OMP FRACTION

The OMP fraction was also separated by 2D SDS-PAGE (Figure. 3.5, Chapter 3). This technique, combined with collagen Far Western analysis, allowed for the identification five adhesin spots which were excised and analyzed by LC MS/MS. These were a protein spot of about 54 kDa, designated spot A, and four spots between 33 and 34 kDa, designated B to E (Table 4.4).

4.3.3.1 SPOT A (54 kDa)

The BF0662 homologue was present in every sample analysed from the 2D SDS-PAGE but this time not in the blank. This indicates that it is probably a very ubiquitous protein. However, the size discrepancy between spot A (~54 kDa) and the BF0662 homologue (~32 kDa) was quite significant and, therefore, it was unlikely that it was responsible for the phenotype of this spot (A). The highest scoring result for this spot had a more convincing predicted size of 52kDa but was a hypothetical protein of unknown function, BF1955 of *B. fragilis* YCH46. BF0662 was the second highest scoring and the third was a serine protease precursor with a predicted size of about 55 kDa. Various other hypothetical protein candidates of unknown function and similar size were also present, as well as the putative plasminogen binding protein.

The most interesting annotated proteins were the plasminogen binding protein (Pbp) and the serine protease precursor. The first has been studied in *B. fragilis*, and was shown to be a plasminogen and laminin binding protein that is also responsible for the conversion of plasminogen to plasmin (Ferreira *et al.*, 2009; Sijbrandi *et al.*, 2008). This proteolytic activation process is discussed in detail in Chapter 1. The Pbp has been shown to be processed to an active form of 59.4 kDa from its original 60 kDa propeptide form (Sijbrandi *et al.*, 2008). Spot A is, therefore, too small to correspond to this protein. Furthermore the putative Pbp was also found in spots A to D and has been seen in the blanks of the other gels (Tables 4.2 and 4.3). Although it was not apparent in the 2D SDS-PAGE blank, it still did not appear to be an ideal candidate in this case.

Table 4.4: Mass Spectroscopy Results of 2D-separated adhesins of OMP fraction

Spot	Score	Expectation	Protein ID	Protein Name	MW (kDa)	% Coverage	
A	1948	0	gi 53713245	hypothetical protein BF1955 [<i>B. fragilis</i> YCH46]	52.1	55.6	
	1026	1.5E-96	gi 53711952	hypothetical protein BF0662 [<i>B. fragilis</i> YCH46]	31.6	49.1	
	806	1.5E-74	gi 53715426	outer membrane efflux protein [<i>B. fragilis</i> YCH46]	50.3	29.8	
	790	6.2E-73	gi 53714036	serine protease precursor [<i>B. fragilis</i> YCH46]	54.7	31.8	
	784	2.6E-72	gi 53711771	alkaline phosphatase III precursor [<i>B. fragilis</i> YCH46]	50.9	25.1	
	738	9.6E-68	gi 255010805	hypothetical protein Bfra3_16818 [<i>B. fragilis</i> 3_1_12]	47.7	30.4	
	698	9E-64	gi 50787940	putative plasminogen binding protein precursor [<i>B. fragilis</i>]	61.4	22.9	
	637	1.3E-57	gi 53713428	hypothetical protein BF2139 [<i>B. fragilis</i> YCH46]	50.1	30	
	633	3.2E-57	gi 53711971	putative outer membrane protein TolC [<i>B. fragilis</i> YCH46]	50.4	23.6	
	469	8.1E-41	gi 29349753	30S ribosomal protein S1 [<i>B. thetaiotaomicron</i> VPI-5482]	67.0	11	
	379	8E-32	gi 53715071	putative chitinase [<i>B. fragilis</i> YCH46]	48.2	25	
	314	2.2E-25	gi 189468231	hypothetical protein BACINT_04627 [<i>B. intestinalis</i> DSM 17393]	50.6	14.5	
	237	1.3E-17	gi 53713300	Na(+)-translocating NADH-quinone reductase subunit A [<i>B. fragilis</i> YCH46]	49.3	14.2	
	B	1739	0	gi 53711952	hypothetical protein BF0662 [<i>B. fragilis</i> YCH46]	31.6	64.9
		646	1.4E-58	gi 50787940	putative plasminogen binding protein precursor [<i>B. fragilis</i>]	61.4	23.4
405		2E-34	gi 53715620	leucine aminopeptidase precursor [<i>B. fragilis</i> YCH46]	37.1	29.3	
227		1.2E-16	gi 53715740	hypothetical protein BF4461 [<i>B. fragilis</i> YCH46]	77.1	5.3	
212		4E-15	gi 53713245	hypothetical protein BF1955 [<i>B. fragilis</i> YCH46]	52.1	9.4	
203		2.8E-14	gi 53713922	hypothetical protein BF2630 [<i>B. fragilis</i> YCH46]	32.7	12.4	
593		3.2E-53	gi 53711952	hypothetical protein BF0662 [<i>B. fragilis</i> YCH46]	31.6	42.8	
D	1230	5.5E-117	gi 53711952	hypothetical protein BF0662 [<i>B. fragilis</i> YCH46]	31.6	55.4	
	945	1.8E-88	gi 53712468	hypothetical protein BF1176 [<i>B. fragilis</i> YCH46]	35.9	48.3	
	848	9.1E-79	gi 53715321	major outer membrane protein OmpA BF4037 [<i>B. fragilis</i> YCH46]	43.3	32.4	
	559	8.2E-50	gi 50787940	putative plasminogen binding protein precursor [<i>B. fragilis</i>]	61.4	19.4	
	353	3.2E-29	gi 53714795	hypothetical protein BF3510 [<i>B. fragilis</i> YCH46]	35.0	33	
	295	1.8E-23	gi 53712259	glyceraldehyde 3-phosphate dehydrogenase [<i>B. fragilis</i> YCH46]	35.6	23.7	
	279	7.4E-22	gi 53712060	putative chromosome partitioning protein ParB [<i>B. fragilis</i> YCH46]	33.6	17.6	
	227	1.1E-16	gi 53715425	cation efflux system protein [<i>B. fragilis</i> YCH46]	37.1	10.7	
	E	877	1.4E-81	gi 53715832	succinate dehydrogenase/fumarate reductase iron-sulfur subunit [<i>B. fragilis</i> YCH46]	27.2	59.8
		804	2.4E-74	gi 53715321	major outer membrane protein OmpA [<i>B. fragilis</i> YCH46]	43.3	42
710		5.9E-65	gi 53711952	hypothetical protein BF0662 [<i>B. fragilis</i> YCH46]	31.6	37.2	
287		1.1E-22	gi 53711479	UDP-N-acetylglucosamine acyltransferase [<i>B. fragilis</i> YCH46]	27.9	27.7	
274		2.5E-21	gi 53713932	putative glycerophosphodiester phosphodiesterase [<i>B. fragilis</i> YCH46]	28.5	28.5	
268		9.9E-21	gi 53712899	hypothetical protein BF1609 [<i>B. fragilis</i> YCH46]	54.3	15.6	
244		2.6E-18	gi 53714795	hypothetical protein BF3510 [<i>B. fragilis</i> YCH46]	35.0	19.9	
185		1.9E-12	gi 60682598	hypothetical protein BF3129 [<i>B. fragilis</i> NCTC 9343]	33.9	19.4	
Blank	155	3.6E-9	gi 53713184	outer membrane protein Omp121 [<i>B. fragilis</i> YCH46]	124.6	3.3	
	95	0.0035	gi 53715321	major outer membrane protein OmpA BF4037 [<i>B. fragilis</i> YCH46]	43.3	9.9	

Identity or homology score thresholds: A (211); B (184); C (22); D (200); E (139); Blank (26)

The serine protease precursor candidate is a homologue of the BF2744 gene of *B. fragilis* YCH46, which corresponds to BF2761 of strain NCTC 9343 and BF638R2772 of strain 638R. This open reading frame (ORF) has an expected size of about 55 kDa and is annotated as a

trypsin-like serine protease of the DegP/HtrA family. This family has been extensively studied in various bacterial and eukaryotic species and is characterized by trypsin-like catalytic domain with one or more PDZ domains, responsible for protein-protein interactions (Clausen *et al.*, 2002). These are highly conserved proteins with protease and chaperone activity, located in the periplasm of gram negative bacteria. They are responsible for ensuring misfolded proteins are degraded and that outer membrane proteins are carried through the periplasm (Ingmer and Brondsted, 2009).

A



B

YCH46 ORF	NCTC 9343 ORF	638R ORF	GENE DESCRIPTION
BF2747	BF2764	BF638R2772	Dipeptidyl peptidase VI
BF2746	BF2763	BF638R2771	Muconate cycloisomerase
BF2745	BF2762	BF638R2770	Putative cystein protease
BF2744	BF2761	BF638R2769	Serine protease precursor
BF2743	BF2760	BF638R2768	Hypothetical protein
BF2742	BF2759	BF638R2767	RNA polymerase factor RpoD
BF2741	BF2758	BF638R2766	Hypothetical protein
BF2740	BF2757	BF638R2765	Clostripain-related protein

Figure 4.2 Genomic context of the putative serine protease precursor of *B. fragilis* YCH46. (A) Arrangement of genes in *B. fragilis* YCH46; (B) ORF numbers and gene annotation in *B. fragilis* strains YCH46, NCTC9343 and 638R

HtrA serine proteases are known to contribute to pathogenesis in various ways. In *E. coli* they are responsible for resistance to temperature and oxidative shock (Pallen and Wren, 1997). HtrA was found to be needed for systemic infection of mice by *Salmonella typhimurium* and similarly for cell-to-cell translocation by *Shigella flexneri*. In *Campylobacter jejuni*, HtrA appears to stimulate adherence to gastrointestinal epithelial cells (Ingmer and Brondsted, 2009). The only published report of this protein interacting with collagen involved the interaction of the mammalian HtrA1 protease with C-propeptides of fibrillar collagens. Here it was found to bind various C-propeptides of fibrillar collagens, including type I, and to degrade type III procollagen $\alpha 1$ C-propeptide (Murwantoko *et al.*, 2004). No such interaction has been reported for bacterial HtrA type proteins.

It seems unlikely that this protein is responsible for the type of collagen adhesin and protease properties observed in *B. fragilis* in this study, but it may play a role in other proteolytic pathways in light of its genomic context in *B. fragilis*. Figure 4.2 shows the arrangement of the genes (Figure 4.2.A) and how they are conserved between three *B. fragilis* strains (Figure 4.2.B). In this gene cluster there are three more proteins annotated as putative peptidases, including the clostripain protein candidate (BF638R2765) mentioned in Chapter 2 (Table 2.3).

4.3.3.2 33-34kDa cluster – SPOTS B to E

Spots B to E each have the BF0662 hypothetical protein of *B. fragilis* YCH46 as their highest hit, with spot B having a coverage of about 65%. These results correlate well with those of the affinity purified adhesins of the OMP fraction (Section 4.3) where the 33 kDa band also shows the presence of the BF0662 homologue.

Other proteins identified in these samples again include OmpA1 and the putative plasminogen binding protein. Various hypothetical proteins of unknown function were also present, as well as a putative leucine aminopeptidase precursor of *B. fragilis* YCH46. This putative metallopeptidase has a predicted molecular weight of about 37kDa and has a conserved M28 peptidase domain as identified using the MEROPS peptidase database (<http://merops.sanger.ac.uk/index.shtml>) (Rawlings *et al.*, 2010). Even though the size of the protein is close to that of the smaller 37 kDa, collagenase band observed in the zymograms in Chapter 3 (Figure 3.2 B), it is again not an endopeptidase and thus unlikely to be responsible for the collagen proteolysis observed.

4.4. CONCLUSION

Having successfully separated the a number of collagen binding proteins of *B. fragilis* for the purpose of mass spectrometric analysis, a through analysis of these results had to be performed in order to determine whether there were any putative collagen adhesin and/or protease gene candidates which could be responsible for the production of the collagen adhesins and proteases observed in previous chapters. The mass spectroscopy results discussed in this chapter present a wealth of information about a subset of outer membrane proteins being expressed in *B. fragilis*.

Several candidates which may function as collagen adhesins were identified, namely the major outer membrane protein, OmpA1, and the putative plasminogen binding protein, Pbp. As discussed above, there were various considerations which excluded them from being investigated further in this study. These included size discrepancies and their presence in the blank control and in samples obtained from areas of very different molecular weight range. This was particularly true in the case of OmpA1 which was always present in the blanks of all the gels analysed. Pbp was found only once in a blank but was also found in areas with much smaller molecular weight proteins, which could be the result of the presence of degradation products of this protein. In spite of this, both of these have been implicated in the ability of *B. fragilis* to adhere to human cells or components of the ECM and may be worth investigating further in a follow up study in terms of their relationship to collagen.

The HtrA-like serine protease, identified in the 2D SDS-PAGE separated spot A, of ~54kDa, showed that at least one protein of a putative protease cluster was being expressed in *B. fragilis*. The last gene of the cluster, the clostripain-like protein coding gene, appears to be a good candidate for a collagenase. However, an insertional mutant of the putative clostripain-like gene, produced when attempts at identifying putative collagenase genes from genome annotation were made (Chapter 2, data not shown), did not show the loss of any of the protease activity bands seen in this study. This gene might, however, still encode an active protease with alternative substrate specificity and different activation buffer requirements. Further analysis of this gene cluster would be of interest especially in light of this mass spectroscopy information.

The hypothetical protein BF0662 of *B. fragilis* YCH46 was repeatedly found as the principal component of the samples analysed, especially those of the size range of the 32 – 34 kDa doublet. This protein had a predicted molecular weight (31.6 kDa) which was in the correct range for the adhesin as determined by one dimensional SDS-PAGE. The observation that the 26 kDa band, obtained during affinity purification of OMP, showed peptides that covered only the C-terminal side of the protein was an interesting fact that could hint at the position of the collagen binding domain of this adhesin. For these reasons the BF0662 homologue was chosen as the candidate gene for further genetic and functional investigations in this study. This work is described in Chapter 5.

CHAPTER 5

Molecular characterisation of the putative collagen adhesin gene *cbp* of *B. fragilis* GSH18

5.1 INTRODUCTION.....	94
5.2 MATERIALS AND METHODS	95
5.2.1 BACTERIAL STRAINS AND GROWTH CONDITONS	95
5.2.2 INSERTIONAL MUTA GENESIS OF HOMOLOGUE OF <i>CBP</i>	96
5.2.3 CONFIRMATION OF INSERTION	97
5.2.4 COLLA GEN FAR WESTERN ANALYSIS AND ZYMOGRAPHY	97
5.2.5 TWO-DIMENSIONAL GEL ELECTROPHORESIS	97
5.2.6 OVEREXPRESSION AND EXTRACTION OF CbP ^{His}	98
5.2.7 HIS-TAG WESTERN ANALYSIS	99
5.3. RESULTS AND DISCUSSION.....	99
5.3.1. INSERTIONAL MUTA GENESIS	99
5.3.2 ADHESIN & PROTEA SE PROFILE OF MUTANT AND WILD-TYPE.....	101
5.3.3 TWO-DIMENSIONAL ANALYSIS OF THE <i>B. FRAGILIS</i> GSH18 <i>CBP</i> MUTANT	105
5.3.4 PET OVER-EXPRESSION OF THE CbP ^{His} COLLA GEN ADHESIN.....	106
5.4 CONCLUSION	109

5.1 INTRODUCTION

The discovery of *B. fragilis* collagen adhesins by Far Western analysis of the cell free extracts and outer membrane protein fractions was described in previous chapters. Three distinct adhesin bands were seen, of approximately 32, 34 and 45 kDa respectively. These were isolated by affinity chromatography and 2D SDS-PAGE and then analyzed by LC-MS/MS. The ORF BF0662 of *B. fragilis* YCH46, was found to be the most likely candidate gene for the production of the ~32 kDa adhesin. For the purpose of this study, the BF0662 homologue in strain GSH18 has been named *cbp* (collagen binding protein).

In *B. fragilis* YCH46, the Cbp coding gene was annotated as a hypothetical protein and further comparison of its predicted protein sequence with proteins lodged in the database of the National Centre for Biotechnology Information (NCBI) did not yield any further information of the putative function of this gene.

Gene specific mutagenesis is a useful technique for identifying the function of a gene of interest in the bacterium under investigation. In *B. fragilis*, insertional mutagenesis (Shoemaker *et al.*, 2000), has been established as a reliable method for this purpose. The resulting mutant can be compared to the wild-type strain using the two methods used previously to identify the collagen adhesin, namely one dimensional SDS-PAGE and 2D SDS-PAGE followed by Far Western analysis in order to determine whether the genetic mutation has lead to the loss of ~32 kDa adhesin band.

In order to confirm that a specific gene encodes a protein with a particular function, the gene can be cloned into an over-expression vector for the heterologous production of the protein in *E. coli*. The appearance of a novel collagen adhesin in the recombinant strain would confirm *cbp* as the gene responsible for the production of the collagen binding protein of interest.

This chapter, therefore, describes the functional characterization of the *cbp* gene, and the expressed Cbp protein, in *B. fragilis* GSH18 and *E. coli* respectively.

5.2 MATERIALS AND METHODS

5.2.1 BACTERIAL STRAINS AND GROWTH CONDITONS

The bacterial strains and plasmids used in this chapter are described in Table 5.1. below. *B. fragilis* strains were maintained on BHIS agar and broth as described in Chapter 2, except during transconjugant selection following mutagenesis (section 5.2.2) when the BHIS agar was supplemented with gentamycin (200 µg/ml) and erythromycin (10 µg/ml). *E. coli* strains were grown in Luria-Bertani (LB) broth or agar, aerobically, at 37°C (Maniatis *et al.*, 1982) supplemented with ampicillin (100 µg/µl) or kanamycin (30 µg/µl) when necessary.

Table 5.1. Description of bacterial strains and plasmids

Strain/Plasmid	Genotype/phenotype	Source/reference
Plasmids		
pGERM	pUC19-based suicide vector	(Shoemaker <i>et al.</i> , 2000)
pGERMcbp-int	pGERM containing <i>cbp</i> internal fragment	This study
pTZ57R/T	Poly-T tailed cloning vector with 3'-ddT overhangs. <i>lacZa ori_{E.coli}</i> . Amp ^R	Fermentas
pTZ57cbp	pTZ57R/T containing the <i>cbp</i> gene. Amp ^R	This study
pET29b(+)	IPTG-inducible <i>E. coli</i> expression vector containing a N-terminal S-Tag and a C-terminal hexahistadyl tag. Kan ^R	Novagen
pETcbp	pET29b(+) containing the <i>cbp</i> gene in-frame to sequences encoding the S-Tag and the hexahistadyl tag. Kan ^R	This study
Strains		
<i>Bacteroides fragilis</i>		
GSH18	Clinical isolate, Gent ^R	This study
GSH18 <i>cbp</i> ⁻	GSH18 derivative, <i>cbp</i> ⁻ , Gent ^R Erm ^R	This study
<i>Escherichia coli</i>		
S17-1	RP4-2-Tc::Mu <i>aph</i> ::Tn7 <i>recA</i> ⁻ , Strep ^R	(Simon <i>et al.</i> , 1983)
S17-1 pGERMcbp-int	RP4-2-Tc::Mu <i>aph</i> ::Tn7 <i>recA</i> ⁻ , Strep ^R Amp ^R	This study
BL21(DE3)	F ⁻ <i>ompT hsdS_B (r_B⁻ m_B⁻) gal dcm</i> (DE3)	(Studier and Moffatt, 1986)
BL21(DE3) pET29b(+)	F ⁻ <i>ompT hsdS_B (r_B⁻ m_B⁻) gal dcm</i> (DE3). Kan ^R	This study
BL21(DE3) pETcbp	F ⁻ <i>ompT hsdS_B (r_B⁻ m_B⁻) gal dcm</i> (DE3). Kan ^R	This study

Gent^R, Erm^R, Strep^R, Amp^R and Rif^R denote resistance to gentamycin, erythromycin, streptomycin, ampicillin and kanamycin respectively

5.2.2 INSERTIONAL MUTAGENESIS OF HOMOLOGUE OF *cbp*

A 0.339 kb fragment corresponding to an internal region of the *cbp* gene, beginning at base 43 and terminating at base 381, was PCR amplified using the primers BF0662-Fstop and BF0662-Rstop (Table 5.2). Translation stop codons were designed into these primers to ensure that no truncated version of the Cbp protein was produced by the recognition of a *Bacteroides* type promoter sequence in the pGERM insert by a *B. fragilis* RNA polymerase. The product was then cloned into the *Sma*I site of the suicide vector pGERM by blunt cloning according to standard protocols (Maniatis *et al.*, 1982). The insert was sequenced from the recombinant plasmid pGERMcbp-int using M13 primers as described in Chapter 2.

The recombinant plasmid was transformed into *E. coli* S17-1 CaCl₂ competent cells (Dagert and Ehrlich, 1979). The plasmid pGERMcbp-int was then transferred into *B. fragilis* GSH18 cells by conjugation as described previously in Chapter 2.

Table 5.2. Description of PCR primers

Name	Primer	*Hyb T°	Fragment size (kb)	Description
BF0662-Fstop	5'-GCTT <u>A</u> GTGTAACGGCCTGG-3'	52.1°C	0.339 kb	Primer pair amplifies an internal fragment of BF0662 of <i>B. fragilis</i> YCH46 gene homologue; <i>cbp</i> . Stop codons are bolded with altered bp underlined.
BF0662-Rstop	5'-ACTCAACAGCCTTCT <u>A</u> CAGC-3'	50.4°C		
BF0662-F2PET	5'-AAATGTAAT <u>CCC</u> <u>A</u> T <u>G</u> GAAAAGTTAGC-3' (<i>Nco</i> I)	51.0°C	0.87 kb	Primer pair amplifies the full length BF0662 of <i>B. fragilis</i> YCH46 gene homologue; <i>cbp</i> . Restriction enzyme sites (in brackets) are bolded with altered bp underlined.
BF0662-R2PET	5'-TTACT <u>CGA</u> GTTGGATAACCAAATAC-3' (<i>Xho</i> I)	50.7°C		

*Hyb T°. Hybridization temperature

5.2.3 CONFIRMATION OF INSERTION

Confirmation of mutagenesis was done by PCR of erythromycin and gentamycin resistant *B. fragilis* transconjugant colonies, using M13 forward and reverse primers (Table 2.2), internal to the pGERM vector, and primers for the full length gene, BF0662-F2PET and BF0662-R2PET (Table 5.2). The PCR products obtained from the mutant colonies, using M13R with BF0662-F2PET, and M13F with BF0662-R2PET, and from wild-type GSH18 colonies using BF0662-F2PET with BF0662-R2PET, were sequenced by Macrogen Inc., Seoul, Korea. Sequencing was done using M13R and M13F primers for the respective mutant PCR products and BF0662-F2PET and BF0662-R2PET for the wild-type PCR product. The resulting sequences were analyzed using DNAMAN (version 4.13).

5.2.4 COLLAGEN FAR WESTERN ANALYSIS AND ZYMOGRAPHY

The CFE and OMP fractions of the wild-type *B. fragilis* GSH18 and the GSH18 *cbp*⁻ mutant were extracted using the methods described in Section 3.2.4 and compared by Far Western analysis as per Section 3.2.5. Collagen zymography was performed on the same samples as described in Section 2.2.5.

Collagen Far Western analysis of the *E. coli* BL21(DE3) (pETcbp) and *E. coli* BL21(DE3) (pET29b(+)) protein extracts (Section 5.2.6) was performed to assesses the collagen binding ability of Cbp^{His}.

5.2.5 TWO-DIMENSIONAL GEL ELECTROPHORESIS

The outer membrane protein fractions of the wild-type *B. fragilis* GSH18 and the GSH18*cbp*⁻ mutant were separated by 2D SDS-PAGE as described in section 3.2.8. Collagen Far Western analysis of the wild-type and mutant proteins on the gel was performed and the presence or absence of collagen binding proteins was compared.

5.2.6 OVEREXPRESSION AND EXTRACTION OF Cbp^{His}

The *B. fragilis* GSH18 *cbp* gene was PCR amplified using primers BF0662-F2PET and BF0662-R2PET (Table 5.2). PCR reactions were done as per Section 2.2.10. The PCR product was purified using the BioSpin PCR Purification Kit (BioFlux, Cat# BSC03S1), as per the manufacturer's instructions, and ligated into the TA-tailed vector pTZ57R/T using the InsTAclone PCR Cloning Kit (Fermentas) to produce the plasmid pTZ57cbp. The *cbp* gene was then subcloned into the pET29b(+) protein expression vector by digestion of the pTZ57cbp plasmid with *NcoI* and *XhoI* restriction enzymes (Fermentas) and subsequent ligation into these sites in pET29b(+). Confirmation that the gene was in frame with the pET29b(+) encoded His and S-tags was obtained by sequencing of the resulting clone, pETcbp.

The overexpression of Cbp^{His} was done based on the methodology outlined in the pET System Manual (Novagen) with various modifications. Briefly, 50 ml of LB with kanamycin was inoculated with *E. coli* BL21(DE3) pETcbp (Cbp^{His} expressor) and grown to an OD_{600nm} of between 0.4 and 0.6. Expression of the protein was then induced for 2 h by addition of IPTG (0.5 mM). The same procedure was followed for *E. coli* BL21(DE3) containing pET29b(+) which was used as a negative control.

Four ml of induced culture were collected by centrifugation in an Eppendorf 5415C bench top centrifuge (14 000 rpm or 16 000 g, RT, 10 min), the cell pellet resuspended in 200 µl PBS, and the cells disrupted by sonication using a Misonic sonicator 3000 at a power output of 3 W for 5 rounds of 30 s. The cell lysate was centrifuged in an Eppendorf 5415C bench top centrifuge (14 000 rpm or 16 000 g, 4°C, 10 min) and the cell debris and supernatant were collected. The supernatant, or CFE, was designated as the soluble fraction, and the cell debris as the insoluble fraction. The insoluble fraction was resuspended in 200 µl PBS.

5.2.7 HIS-TAG WESTERN ANALYSIS

To confirm the expression of Cbp^{His}, the *E. coli* BL21(DE3) (pETcbp) and *E. coli* BL21(DE3) (pET29b(+)) protein extracts (Section 5.2.6) were tested by Western blot analysis using anti-6-His antibodies (Sambrook *et al.*, 2001). Fifteen microliters of each protein sample was mixed in a 1:1 ratio with 2X SDS gel loading buffer (without β -mercaptoethanol) (0.125M Tris-HCl [pH 6.8], 4% SDS, 20% glycerol, 0.002% bromophenol blue) and incubated at room temperature for 10 min prior to electrophoresis in a 12% SDS (w/v) polyacrylamide gel (without β -mercaptoethanol) (Laemmli, 1970). PageRulerTM Prestained Protein Ladder (Fermentas, #SM0671) was used as the molecular weight standard for all protein gels. Duplicate protein gels were prepared, where one was stained with Coomassie brilliant blue, and the other was transferred on to a nitrocellulose membrane using the Trans-Blot® SD DNA/RNA Blotting Kit (Bio-Rad) as per the manufacturer's instructions.

The nitrocellulose blots were blocked overnight by incubating at room temperature (RT) in blocking buffer (1X PBS buffer, pH7.4, 0.1% Tween 20, 5% skim milk). The membranes were then incubated with affinity purified rabbit anti-6xHis (1:5000) (GeneTex. GTX77352) for 1 h at RT and washed four times with blocking buffer. This was followed by incubation with peroxidase conjugated anti-rabbit IgG (Goat) antibody (1:2500) (Gene Tex, GTX77060) for 1 h at RT and four washes with blocking buffer without skim milk. The membrane was then developed with TMB Membrane Peroxidase substrate (KPL, 50-77-18).

5.3. RESULTS AND DISCUSSION

5.3.1. INSERTIONAL MUTAGENESIS

Insertional mutation of the putative collagen adhesin gene, named here *cbp*, was the first step in determining whether it was responsible for the production of the ~32 kDa collagen adhesin of *B.fragilis* GSH18. A fragment of approximately 0.3 kb was used as the gene specific target region for disruption (Figure. 5.1). To confirm the insertion of the suicide vector into the gene of

interest, PCR of the putative transconjugants was performed using primers for the full length gene, in combination with M13 primers which bind to the pGERM vector sequence.

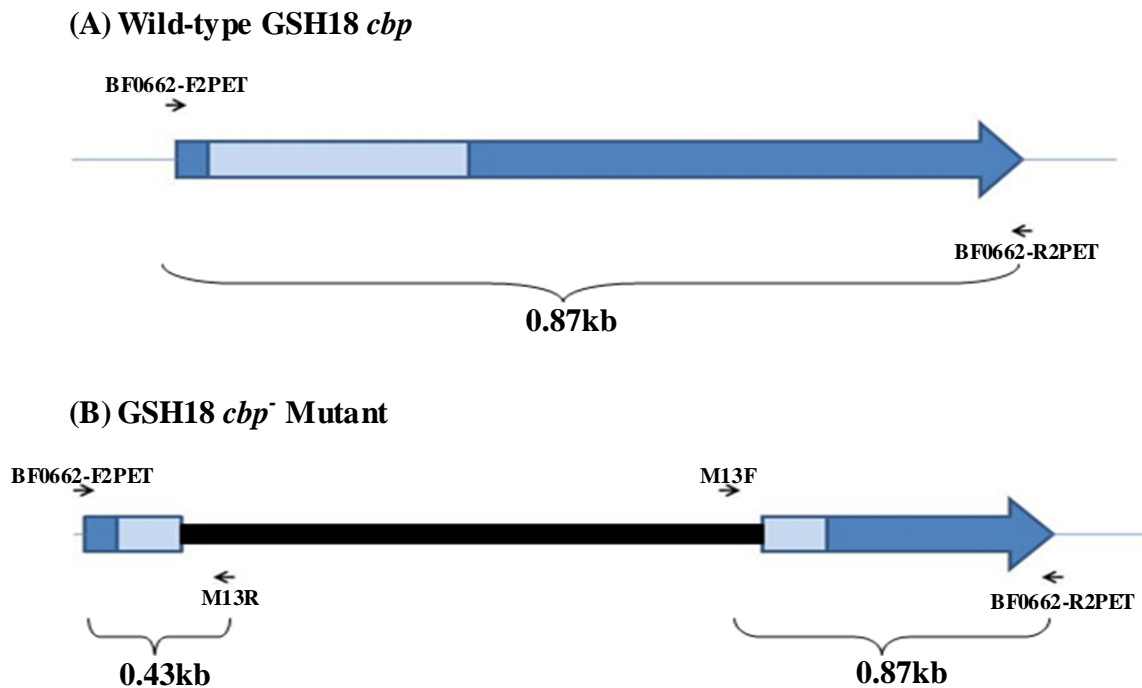


Figure 5.1 Schematic diagram of the targeted gene disruption. (A) Wild-type *B. fragilis* GSH18 *cbp* expected PCR products; (B) *B. fragilis* GSH18 *cbp*⁻ mutant expected PCR products. The light blue area indicates the internal fragment of the *cbp* gene used for the homologous recombination. The dark blue area is the rest of the *cbp* gene. The black area represents the pGERM plasmid DNA and the arrows indicate the primers used and their annealing sites.

The *cbp* gene was indeed interrupted, as confirmed by the production of 0.43 kb and 0.87 kb PCR fragments with primers BF0662-F2PET and M13R, and BF0662-R2PET and M13F respectively (Figure. 5.2, lanes 6 and 8 respectively). These fragments were not produced when wild-type colonies were used (lanes 7 and 9). PCR with BF0662-F2PET and BF0662-R2PET amplified the full length *cbp* gene from wild-type *B. fragilis* GSH18 (lane 3) and not from the mutant (lane 2).

Sequencing of the mutant PCR products (Figure. 5.2. Lanes 6 and 8) with primers M13F and M13R confirmed the junction between the pGERM inserted vector and the interrupted *cbp* gene. The presence of the stop codons, designed into the BF0662-F2stop and BF0662-R2stop, was also confirmed.

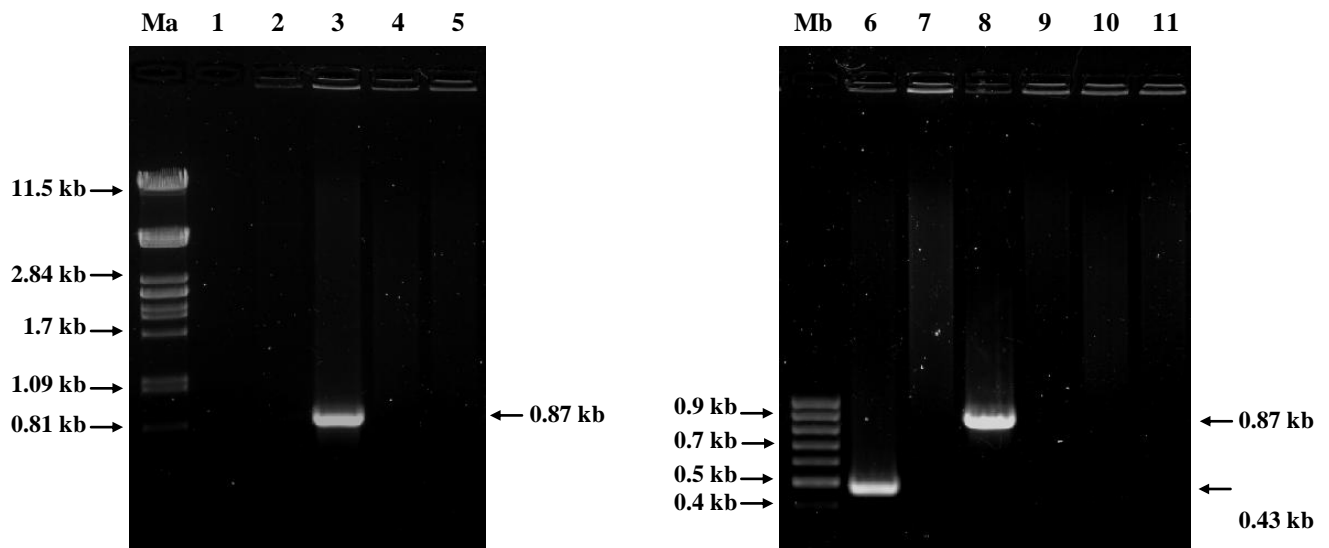


Figure 5.2 PCR of DNA from wild-type *B. fragilis* GSH18 and *B. fragilis* GSH18 *cbp*⁻ mutant for confirmation of insertion. **Lanes:** (Ma) molecular weight marker of λ -phage DNA digested with *Pst*I; (Mb) molecular weight marker O'GeneRuler™ 100bp DNA Ladder (Fermentas #SM1143); Primers BF0662-F2PET and BF0662-R2PET with (1) no DNA template; (2) mutant DNA; (3) wild-type DNA; Primers BF0662-F2PET and M13F with (4) mutant DNA; (5) wild-type DNA; Primers BF0662-F2PET and M13R with (6) mutant DNA; (7) wild-type DNA; Primers BF0662-R2PET and M13F with (8) mutant DNA, (9) wild-type DNA; Primers BF0662-R2PET and M13R with (10) mutant, (11) wild-type.

5.3.2 ADHESIN & PROTEASE PROFILE OF MUTANT AND WILD-TYPE

A comparison of the collagen adhesin and protease profiles of the *B. fragilis* GSH18 *cbp*⁻ mutant with that of the wild-type *B. fragilis* GSH18 was performed to establish if the *cbp* gene coded for the production of the 32 kDa adhesin protein. A complete analysis of the *cbp*⁻ mutant's ability to bind to or degrade collagen, was done using the previously established methods of collagen Far Western analysis and collagen zymography. Both the CFEs (Figure. 5.3) and OMP fractions (Figure. 5.4) of the *B. fragilis* GSH18 wild-type and *cbp*⁻ mutant strains were investigated.

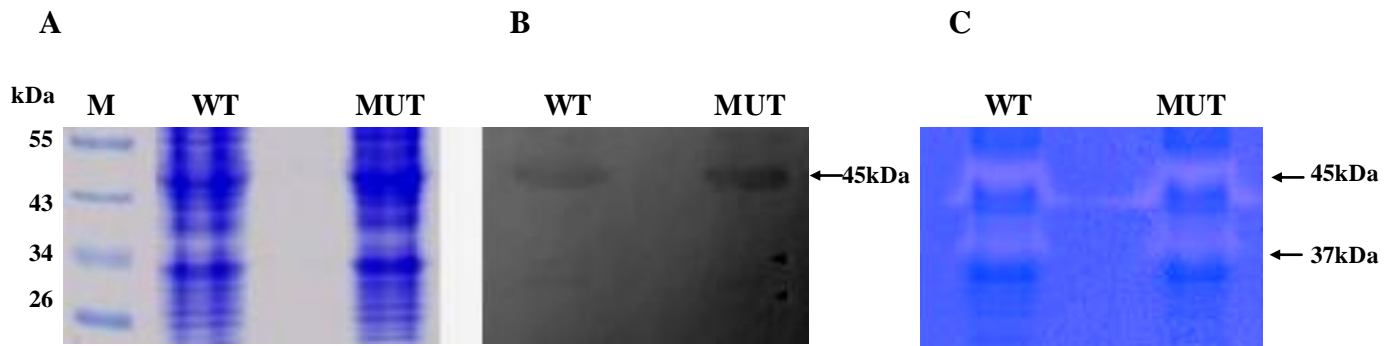


Figure 5.3 Collagen Far Western blot and zymogram comparisons of *B. fragilis* GSH18 wild-type (WT) and *B. fragilis* GSH18 *cbp*⁻ mutant (MUT) CFEs. (A) Coomassie stained SDS-PAGE, (B) Collagen Far Western, (C) Collagen zymogram. **M**: PageRuler™ #SM0671

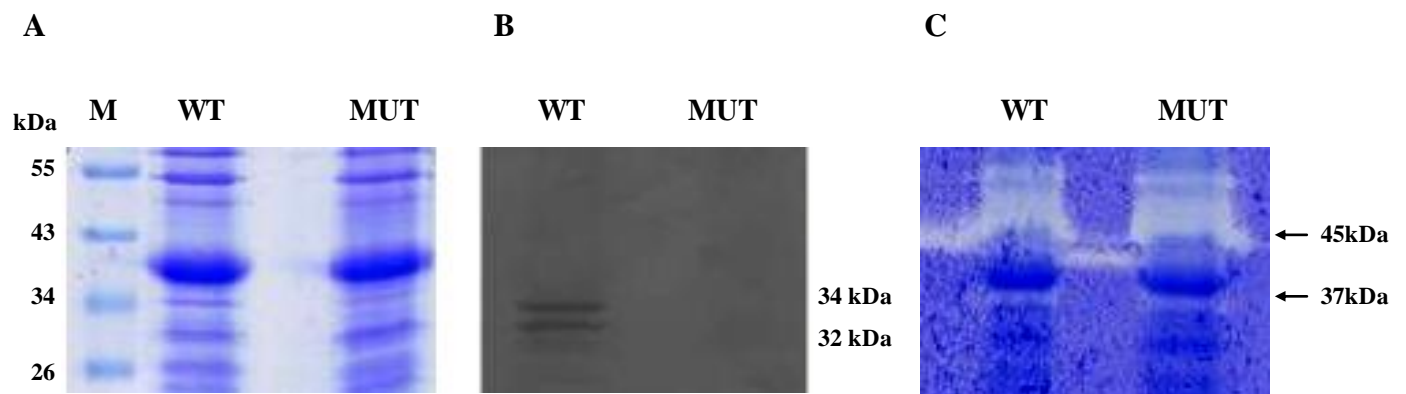


Figure 5.4 Collagen Far Western blot and zymogram comparison of wild-type *B. fragilis* GSH18 (WT) and *B. fragilis* GSH18 *cbp*⁻ mutant (MUT) outer-membrane protein fraction. (A) Coomassie stained gel, (B) Far Western, (C) Zymogram. **M**: PageRuler™ #SM0671

This analysis revealed that the mutant strain had an identical CFE profile as compared to the wild-type, producing the ~45 kDa collagen adhesin (Figure. 5.3 B) as well as the protease of the same size as well as the ~37 kDa protease (Figure. 5.3 C). The OMP collagen adhesin profile (Figure. 5.4 B) was, however, different in the mutant strain as compared to the wild-type. The *cbp*⁻ mutant did not produce either the 32 or the 34 kDa collagen adhesin band, indicating that the disruption of the *cbp* gene had resulted in the loss of the putative collagen adhesin. There was no change in the collagen zymography pattern in the OMP fraction of the *cbp*⁻ mutant as compared to the wild-type strain (Figure. 5.4 C).

Interestingly, by insertionally interrupting the *cbp* gene, the 34 kDa adhesin band also disappeared. There are two possibilities for this phenomenon. The first is that the gene encoding the 34 kDa adhesin protein band might be located downstream of the *cbp* gene and that both are transcribed from the same promoter. In this scenario, by interrupting the transcription of *cbp*, the subsequent genes and their product are also not produced.

This possibility was investigated by determining the genomic context of the *B. fragilis* YCH46 Cbp is shown in Figure 5.5 A. The *cbp* gene and the genes surrounding it are conserved throughout the three published genomes of *B. fragilis* strains YCH46, NCTC 9343 and 638R (Figure. 5.5 B). It is, of course, important to keep in mind that the mutation of the *cbp* gene was done in the clinical isolate, *B. fragilis* GSH18, and that the conservation of the genes discussed would have to be confirmed in this strain.

A



B

YCH46 ORF	NCTC 9343 ORF	638R ORF	GENE DESCRIPTION	PREDICTED MW
BF0660	BF0584	BF638R0631	Hypothetical protein	27 kDa
BF0661	BF0585	BF638R0632	Putative tetrapyrrole (corrin/porphyrin) methylase	25 kDa
BF0662	BF0586	BF638R0633	Hypothetical protein (<i>cbp</i>)	32 kDa
BF0663	BF0587	BF638R0634	Putative haloacid dehalogenase-like hydrolase protein	27 kDa
BF0664	BF0588	BF638R0635	MarC family integral membrane protein (putative amino acid transporter)	23.5 kDa

Figure 5.5 Genomic context of the *cbp*. (A) Arrangement of genes in *B. fragilis* YCH46 (The *cbp* gene is highlighted in black); (B) ORF numbers and gene annotation in *B. fragilis* strains YCH46, NCTC9343 and 638R.

The *cbp* gene (BF0662 of strain YCH46) was found to be the middle gene in a cluster of five closely positioned genes which could form part of an operon. If this is the case then the interruption of the *cbp* gene could prevent the transcription of the two downstream genes.

However, these genes are annotated as a putative haloacid dehalogenase-like hydrolase protein (BF0663 – YCH46) and a MarC family integral membrane protein (BF0664 – YCH46) with predicted protein sizes of 27 and 23.5 kDa respectively. Neither of these encodes a protein of the appropriate size, nor can their annotated function be linked to collagen adhesion, so the hypothesis that the *cbp* mutation causes a polar effect on another collagen binding protein seems unlikely.

The second possibility is that the 34 kDa protein band is a post-translationally modified version of the Cbp protein. Post-translational modifications include processes such as phosphorylation and methylation. Using the NetPhosBac 1.0 server (Blom *et al.*, 1999), which predicts serine and threonine phosphorylation sites in bacterial proteins, five potential serine phosphorylation sites (S 24, S 119, S128, S239 and S 277) were identified.

Interestingly, the gene upstream of *cbp* is annotated as a putative tetrapyrrole methylase. According to the protein family database, Pfam (Finn *et al.*, 2010), these methylases are involved in the methylation of substrates using S-adenosyl-L-methionine as a methyl source. Methylation is known to alter the function of target proteins and has been implicated in processes like signal transduction and transcriptional regulation (Grillo and Colombatto, 2005). If methylation of the *B. fragilis* Cbp is responsible for the adhesin band of increased molecular weight, the mutation of the *cbp* gene would account for the loss of the collagen adhesin doublet and not just the ~32 kDa band. Using the BPB-PPMS server (Shao *et al.*, 2009), which performs *in silico* identification of protein methylation sites, two putative lysine methylation sites (K 102 and K 123) were identified.

In *Mycobacterium tuberculosis* and *Mycobacterium bovis* a heparin-binding hemagglutinin adhesin (HBHA) and a laminin-binding protein (LBP), involved in extrapulmonary dissemination and cytoadherence respectively, were found to be methylated *in vivo* (Pethe *et al.*, 2002). This modification resulted in the apparent increase in the molecular weights of the adhesins. Methylation did not however appear to alter the affinity of HBHA for heparin-sulfate receptors but it was shown to protect HBHA and LBP from proteolytic degradation.

If a similar situation occurs in the case of the Cbp this could be determined in a similar manner to that used in Pethe *et al.* (2002), which is by mass spectrometry. Although determination of methylation by this method could have been performed as part of the LC-MS/MS analysis, done to determine the presence of Cbp (the BF0662 homologue) in the samples analysed, it was not known at the time that such an analysis would be of interest. Unfortunately due to the time and financial constraints associated in the sample preparation and analysis at the Yale Cancer Centre Mass Spectrometry Resource & W.M. Keck Foundation Biotechnology Resource Laboratory the United States of America, it was not possible in this thesis to accomplish this. The determination of methylation of the wild type Cbp and the mechanisms by which this modification occurs may be best analysed as part of a new study in which the methylation states of the wild type and recombinant proteins can be compared as was describe by Pethe *et al.* (2002).

5.3.3 TWO-DIMENSIONAL ANALYSIS OF THE *B. FRAGILIS* GSH18 *cbp*⁻ MUTANT

Two-dimensional gel electrophoresis proved to be a powerful tool for the separation of the OMP fraction with the aim of identifying collagen adhesin protein spots (Chapters 3 and 4). Computer programs such as PDQuest software version 7.4.0 (Bio-Rad), used in this study, allow for detailed comparison of protein profiles under different sets of experimental conditions or in this case between a wild-type and mutant strain.

The OMPs of *B. fragilis* GSH18 wild-type and *B. fragilis* GSH18 *cbp*⁻ mutant strains were separated by 2D SDS-PAGE and examined by subsequent collagen Far Western analysis (Figure. 5.6). It was observed that the mutant strain only produced the collagen adhesin spot of ~54 kDa (Figure. 5.6 C – spot A) and none of the others seen in the wild-type (Figure. 5.6 C – spots B to E). PDQuest was used to analyse the wild-type and mutant gels. Spots B to E were those that were excised and analysed by LC MS/MS. Use of the spot matching tool revealed that protein spots C to E were still present in the *cbp*⁻ mutant, but that spot B was missing from all mutant gels. The mass spectrometry results for the 2D SDS-PAGE gels (Table 4.3) showed that spot B has the greatest identity to BF0662 of *B. fragilis* YCH46. BF0662 was also detected in the case for spots C and D but these results had a much lower identity than that of spot B.

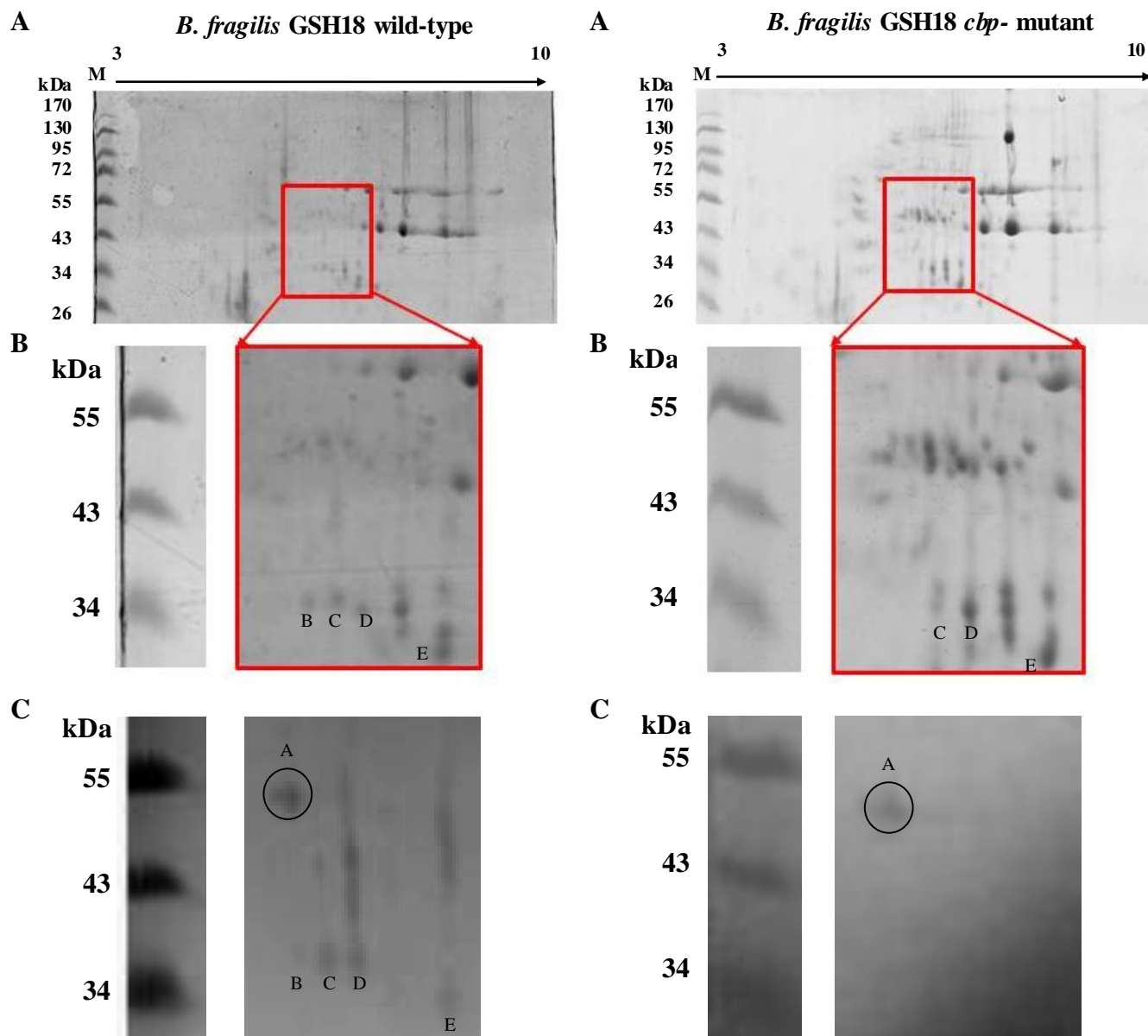


Figure 5.6. Comparison of collagen-binding proteins from the outer membrane of *B. fragilis* GSH18 wild-type and *B. fragilis* GSH18 *cbp*⁻ mutant by two-dimensional gel electrophoresis. OMPs were separated in the first dimension on precast IPG strips (pH 3–10) followed by separation in the second dimension on 10% SDS-PAGE gels: (A) Coomassie stained gels; (B) Magnified area of Coomassie stained gels shown in red box of (A); (C) Collagen Far Western of proteins from 2D-gels in the same magnified area indicated in the red box of (A);. M: PageRuler™ #SM0671.

The results of this work, therefore, confirm that Cbp is the collagen adhesin responsible for the collagen binding activity seen between about 32 and 35 kDa, and suggest that the collagen binding seen in spots C to E (Figure. 5.6 B – wild-type) was due to the presence of contaminating Cbp protein.

5.3.4 pET OVER-EXPRESSION OF THE Cbp^{His} COLLAGEN ADHESIN

To further prove that the *cbp* gene of *B. fragilis* GSH18 was, in fact, responsible for the production of a collagen adhesin of approximately 32 kDa, this gene was cloned into pET29b⁺ and expressed in *E. coli*. The fusion protein of Cbp with an S-Tag at the N-terminus and a C-terminal hexahistadyl had a calculated molecular weight of approximately 35 kDa.

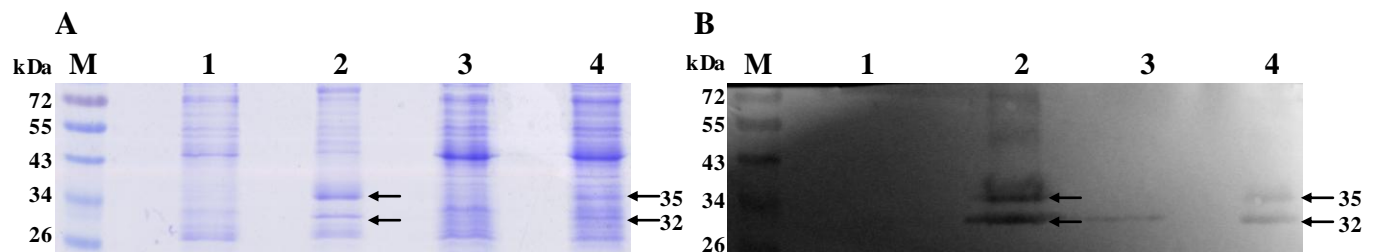


Figure 5.7 Expression of Cbp^{His}. (A) Coomassie stained gel, (B) Anti-6-His Western. Lanes: (M) PageRuler™ #SM0671; (1) insoluble fraction of *E. coli* BL21(DE3) pET29b⁺; (2) insoluble fraction of *E. coli* BL21(DE3) pETcbp; (3) soluble fraction of *E. coli* BL21(DE3) pET29b⁺; (4) soluble fraction of *E. coli* BL21(DE3) pETcbp.

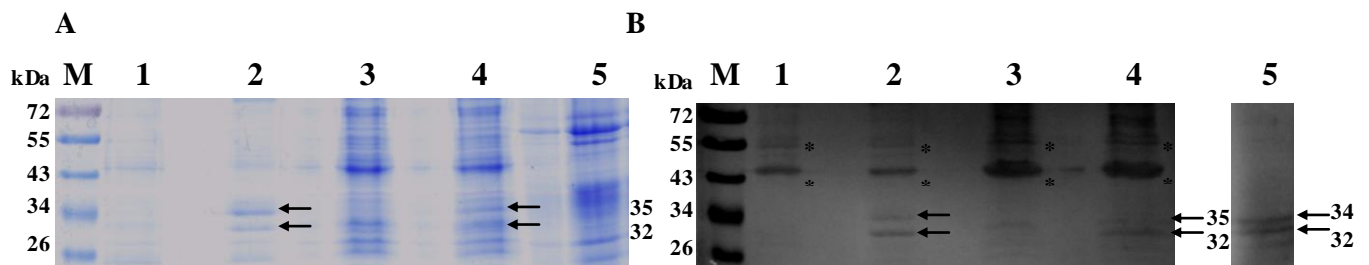


Figure 5.8 Collagen far western analysis of Cbp^{His}. (A) Coomassie stained gel, (B) Collagen Far Western. Lanes: (M) PageRuler™ #SM0671; (1) insoluble fraction of *E. coli* BL21(DE3) pET29b⁺; (2) insoluble fraction of *E. coli* BL21(DE3) pETcbp; (3) soluble fraction of *E. coli* BL21(DE3) pET29b⁺; (4) soluble fraction of *E. coli* BL21(DE3) pETcbp; (5) 90µg OMP of *B. fragilis* GSH18 as a positive control. (*) indicate collagen adhesins in the *E. coli* protein background.

After expression for 2 h at 30°C, the fusion protein, Cbp^{His} was expressed in the recombinant *E. coli* BL21(DE3) pETcbp strain (Figure. 5.7 A, lanes 2 and 4). Cbp^{His} was found predominantly in the insoluble fraction as can be seen by the presence of a distinct 35 kDa band (lane 2) not present in the control strain harboring only the parent vector (lane 1). Interestingly a

novel band of ~32 kDa is also seen in the pETcbp strain (lane 2). Confirmation of the presence of the Cbp^{His} was done by western blot detection of the hexahistadyl using anti-6-His antibodies (Figure. 5.7 B). This showed that the two bands in *E.coli* BL21(DE3) pETcbp and possessed a hexahistadyl. Although the reaction to the His-tagged proteins in the soluble fraction (lane 4) was weaker than that of the insoluble fraction (lane 2) it demonstrated that at least some of the Cbp^{His} is soluble during the expression protocol used which will assist in the purification of the recombinant protein for use in future studies. The *E.coli* BL21(DE3) pET29b⁺ negative control strain only displayed a faint band at ~32 kDa in the soluble fraction (lane 3, Figure. 5.7 B) but this was attributed to non-specific binding of the anti-6-His antibodies to *E. coli* proteins.

To analyse for the ability of the recombinant Cbp^{His} to bind to collagen, SDS-PAGE and Far Western analysis was performed (Figure. 5.8) using *B. fragilis* GSH18 OMP as a positive control (lane 5). Both the ~35 and ~32 kDa bands were able to bind collagen as compared to the positive control (lane 5). It should also be noted that other collagen binding adhesins were detected in the *E. coli* background in all fractions, as seen at about 44 kDa And 54 kDa, indicated with asterisks (*) (Figure. 5.8 B), but not at the ~32 kDa position where the non-specific binding of anti-6-His antibodies had been observed (lane 3, Figure. 5.7 B). There is very little reported information on collagen type I adhesins of *E. coli* that could be used to interpret this data. An EhaB autotransporter protein of *E. coli* O157:H7 has been shown to bind to laminin and collagen Type I, but this has a size of 107 kDa (Wells *et al.*, 2009) and is, therefore, clearly not among the proteins visualized in Figure 5.8.

The 32 kDa band, having a smaller molecular weight than the one predicted for the Cbp^{His} is, therefore, not modified in the same manner as the wild-type Cbp in *B. fragilis*. This was also the case with the recombinant heparin-binding hemagglutinin adhesin (HBHA) of *M. tuberculosis* and *M. bovis*, mentioned in section 5.3.2., which was not methylated when produced in *E. coli* (Pethe *et al.*, 2002). This smaller sized band is likely to be the Cbp^{His} without the S-Tag which is of about 2.8 kDa. Proteolytic cleavage of this tag can occur at a thrombin site found between the S-tag and the cloned gene. For future work requiring pure Cbp^{His} found in one conformation, protease inhibitors will have to be included in the protein preparation to avoid removal of the S-Tag or the S-Tag will have to be removed completely in a controlled fashion.

5.4 CONCLUSION

In this chapter, the aim was to determine whether the putative collagen binding protein gene, *cbp*, was responsible for the production of the collagen adhesin of approximately 32 kDa. This was successfully achieved by using two methods to functionally characterize this gene. The first was insertional inactivation of the *cbp* gene to prevent the production of the Cbp in *B. fragilis* GSH18. The mutation was confirmed by PCR with a combination of primers internal to the suicide vector, pGREM and primers for the full length gene, followed by sequencing of the PCR products.

The wild-type and mutant strains were compared by one dimensional gel electrophoresis in combination with collagen zymography and far western analysis, to reveal the loss of the ~32 and ~34 kDa adhesin bands, but no loss of the ~45 kDa adhesin or the ~45 kDa and ~37 kDa collagen protease bands. Two-dimensional PAGE of the OMPs of the two strains also revealed the loss of collagen binding activity at the 32 – 35 kDa area, but again retained the collagen binding spot seen at about 54 kDa. In the second approach the *cbp* gene was cloned and expressed in *E. coli* and the recombinant Cbp^{His} was shown to bind collagen as well.

This is, to our knowledge, the first report of the identification of a collagen binding protein in *B. fragilis*, which is now available for further characterisation. Future work will include resolving any possible mechanisms of post-translational modification of the protein, which could result in the production of an adhesin with a greater apparent molecular mass and similarly if these modifications have any influence on the affinity of Cbp for collagen. It would also be of interest to determine the involvement, if any, of the putative methylase gene found directly upstream of *cbp*, and whether this alleged modification on the Cbp has any effect on the protein's function.

The experimental approach undertaken in this dissertation has allowed for the assignment of function to the *cbp* gene, annotated as a hypothetical protein encoded by the ORF BF0662 in *B. fragilis* YCH46. This gene sequence had no detectable conserved domains and without the functional identification used here, its role in *B. fragilis* would not have been easily elucidated.

CHAPTER 6

General conclusions and future research

The increasing incidence of antibiotic resistance in the important human opportunistic pathogen, *B. fragilis*, has prompted the search for new therapeutic targets such as virulence factors (Barczak and Hung, 2009). In this dissertation, the emphasis has been on a particular subset of putative virulence factors in *B. fragilis*, namely adhesins and proteases that target the main component of the human extracellular matrix, collagen.

At the beginning of this study, little was known about the relationship of *B. fragilis* to this substrate. Previous studies had not been able to show any evidence of its ability to degrade collagen (Gibson and Macfarlane, 1988b; Patrick, 2002), although, a few instances of activity against collagen-like substrates, such as Azocoll and gelatin, had been reported (Chen *et al.*, 1995; Moncrief *et al.*, 1998). The ability of various *B. fragilis* strains to adhere to collagen type I had, however, been established (Szoke *et al.*, 1996; Szoke *et al.*, 1997). It is important to note that, even though collagen type I is the most abundant type found in the human body, it is not the collagen type that this bacterium is most likely to encounter. The basal lamina, which underlies endothelial and epithelial cells, in fact contains the non-fibrillar, or net-work forming, type IV collagen. In this study, however, we chose to begin investigating the relationship with this ECM component using the only collagen type that *B. fragilis* had been, thus far, confirmed to interact with, namely collagen type I.

With the ability of *B. fragilis* to bind this substrate already established, the first step in this study was to assess whether or not it could degrade it. Using zymography, it was determined that it did, in fact, hydrolyse collagen type I and that a distinct protease band of about 45 kDa and several other zones of clearing were present. This prominent collagenolytic band was observed in the extracellular fraction as well as in the CFEs and in the OMP fraction. Initial attempts to use bioinformatic searches of the annotated genomes of *B. fragilis* to look for candidate collagenase genes proved to be a somewhat narrow approach. The best candidate gene from *in silico*

analyses, annotated as a *prtC*-like protease, was shown not to be responsible for the degradation of collagen seen here.

This candidate gene was, however, found in a putative gene cluster with very high similarity to its *P. gingivalis* counterpart. Although interruption of this gene showed that it was not responsible for the production of this particular collagen type I collagenase, it may have alternative substrates or require specific gene induction conditions, which were not investigated here. Further studies of this gene cluster would be, therefore, of great interest.

Having discovered the presence of specific collagen degrading enzymes, the next objective was to investigate the mechanisms of collagen type I adhesion. The first step was to determine if the previously reported differences in the ability of *B. fragilis* strains to bind collagen were also seen within the subset of clinical strains available for this study. The initial discovery of the collagen proteases had been performed using the *B. fragilis* 638R strain. Although various genetic tools, such as insertional mutagenesis, had been well documented for use with this strain, it was important to make sure that it was in fact the best candidate for the study. Using particle agglutination assays, both as a measure of affinity for collagen and as an indicator of virulence, it was seen that *B. fragilis* 638R was in fact not the best candidate strain, and so studies were continued using the more adherent clinical isolate, *B. fragilis* GSH 18.

Using this strain, the presence of four collagen adhesin proteins was established. A ~45 kDa adhesin was observed in the CFE and a ~54 kDa collagen binding protein as well as an adhesin doublet, containing ~32 and a ~34 kDa bands, were found in the OMP fraction. The presence of an adhesin of a similar size to the predominant collagenase band (45 kDa) suggested this could be a collagenase with a collagen binding domain and, therefore, it was hoped that collagen affinity purification would also allow for the isolation of the respective collagenase.

Collagen affinity purified and 2D PAGE separated putative collagen adhesin proteins were analysed by mass spectrometry and the predicted protein sequences of the published genomes were used for comparison of the resulting protein sequence information. The LC MS/MS method used allows for analysis of mixed protein samples which meant that good quality results were

obtained after relatively limited purification procedures. Unfortunately, the wealth of information that resulted from this analysis did not allow for easy determination of the identity of all the proteases and adhesins of interest. No distinct putative collagenase genes were identified. In fact very few protease genes were seen at all. A similar experience was reported by Houston *et al.* this year (2010) where their mass spectrometry analysis of *B. fragilis* extracellular proteins between 35 and 60 kDa in size yielded no putative protease gene candidates.

The identity of the genes encoding the collagenase activity bands observed in this study was unfortunately not established. Fresh approaches to this task will have to be looked at in future studies. In spite of the difficulties associated with expressing *B. fragilis* promoters in *E. coli*, the production of a gene bank could be a viable alternative, since a number of *B. fragilis* genes have been identified in this way, an example being the glutamine synthetase gene *glnA* (Southern *et al.*, 1986). Alternatively, every putative protease gene identified *in silico* could be expressed in *E. coli* by cloning the genes into vectors, such as pET, in which expression is mediated by *E. coli* compatible promoters, such as the *lacZ* promoter. More robust screening methods, including the use of gelatin or casein agar, could be used to select for collagenase positive clones and, similarly, for collagenase deficient transposon mutants of *B. fragilis*. The recent publication mentioned above (Houston *et al.*, 2010) already hints at the substrate promiscuity of the ~45 kDa collagenase observed in this study. In it two fibrinogen degrading protease bands, of ~45 and 50 kDa were observed using fibrinogen zymography. It would be very interesting to see if the same enzyme is responsible for both the proteolytic activities. Failing this, however, the use of extensive protein purification methods will have to be employed.

The LC MS/MS analysis also didn't identify any genes annotated as putative collagen MSCRAMMs. The only proteins identified which have been at all linked to the adhesion of *B. fragilis* to human cells or ECM components were the major outer membrane protein, OmpA (Pumbwe *et al.*, 2007), and the plasminogen and laminin binding protein Bfp60 (Ferreira *et al.*, 2009; Sijbrandi *et al.*, 2008). Although these proteins did not appear to be ideal candidates for the particular adhesins of interest in this study, it may still be of interest to follow up on their relationship with collagen. Deletions mutants of the *ompA1* to 7 genes have already been created and tested for alterations in their ability to deal with several stress conditions, such as oxygen

exposure and osmotic stress (Wexler *et al.*, 2009). These could similarly be tested for alterations in their human cell binding ability and for the specific loss of any of the collagen adhesins. No reports of *bfp60* have been published although the gene has been successfully expressed in *E. coli* (Ferreira *et al.*, 2009; Sijbrandi *et al.*, 2008). Interestingly the recent study by Houston *et al.* (2010) discussed above, reports on the identification of a fibrinogen binding protein (BF-FBP) of 54 kDa, which is the same size as one of the collagen adhesins seen in this study. However, BF-FBP is encoded by the BF1705 ORF of *B. fragilis* NCTC 9343 (BF1698 of YCH46), which was not identified in the LC MS/MS results described here. It remains of interest to follow up on whether or not this protein has any affinity for collagen type I as well.

A large number of proteins, identified as hypothetical proteins of unknown function were seen in the mass spectrometry results in this study. This was a very interesting observation as it indicated that a significant number of proteins being expressed by *B. fragilis* have little to no genetic and protein sequence similarity to known functional proteins. This confirms that the difficulty experienced in finding putative protease gene candidates at the beginning of the study is not surprising.

One such hypothetical protein encoding gene proved to be responsible for the production of not one but two of the collagen adhesin bands observed by collagen Far Western analysis. The homologue of the BF0662 ORF of *B. fragilis* YCH46, referred to in this study as *cbp*, was mutated in the clinical strain *B. fragilis* GSH18, demonstrating the loss of the collagen adhesin doublet of ~32 and a ~34 kDa. The *cbp* gene expression in *E. coli* confirmed the collagen binding phenotype of the protein expressed by the gene. Possible reasons for the production of the adhesin doublet by the same protein coding gene have been discussed. The possibility of post-translational modifications of the Cbp protein and the involvement of the putative methylase gene, found upstream of *cbp* in the three published *B. fragilis* genomes, is to be the one of the focuses of the follow up study of this collagen adhesin. The other focus will be on establishing the contribution of this gene to the ability of *B. fragilis* to adhere to collagen, and to human cells as a possible indicator of its participation in the virulence potential of *B. fragilis*. The latter can now easily be done by comparing the wild-type *B. fragilis* GSH 18 to the *cbp*⁻ mutant of the same strain, produced in this study, using methods previously described for *B. fragilis* in the context of

the interaction with other ECM components like laminin (Ferreira *et al.*, 2002; Ferreira *et al.*, 2006).

The work presented in this dissertation, therefore, describes for the first time the production of proteases and adhesins with the ability to, respectively, bind to and degrade fibrillar collagen (type I). Although the identification of the collagenase proteins was not successful, the development of the collagen zymography technique for use in the context of *B. fragilis* is an important development in the study of these proteases. Similarly, the use of the collagen type I Far Western analysis with *B. fragilis* has allowed for the detection of specific adhesins to this substrate. Although Far Western analysis of 2D SDS-PAGE separation of proteins has been previously described, this is, to our knowledge, the first report of the technique being used with collagen Far Western analysis.

This study, especially with the identification of the first collagen binding protein of *B. fragilis*, Cbp, has opened the doors to the investigation of the relationship of this important human opportunistic pathogen and the main component of the human ECM, collagen. The identification of molecules which mediate the interaction of *B. fragilis* to any ECM component may eventually lead to the development of inhibitors of these adhesive and proteolytic properties to be used as potential therapeutic methods.

Reference List

- Abrahams,G.L. and Abratt,V.R. (1998) The NADH-dependent glutamate dehydrogenase enzyme of *Bacteroides fragilis* Bf1 is induced by peptides in the growth medium. *Microbiology* **144** (Pt **6**): 1659-1667.
- Alberts,B., Johnson,A., Lewis,J., Raff,M., Roberts,K., and Walter,P. (2008a) Cell Junctions, cell Adhesion, and the Extracellular Matrix. In *Molecular Biology of the Cell*. Anderson,M. and Granum,S. (eds). New York: Garland Science, pp. 1164-1194.
- Alberts,B., Johnson,A., Lewis,J., Raff,M., Roberts,K., and Walter,P. (2008b) Membrane Structure. In *Molecular Biology of the Cell*. Anderson,M. and Granum,S. (eds). New York: Garland Science, pp. 617-648.
- Alexander,C.M., Birk,D.E., Farquhar,M.G., Hascall,V.C., Hay,E.D., Heinegard,D.K. *et al.* (1991) *Cell Biology of Extracellular Matrix*. New York: Plenum Press.
- Alhawi, M., Stewart, J., Erridge, C., Patrick, S., and Ian R. Poxton, I.R. (2009) *Bacteroides fragilis* signals through Toll-like receptor (TLR) 2 and not through TLR4. *J Med Microbiol* , **58**: 1015 – 1022.
- Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W., and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402.
- Barczak,A.K. and Hung,D.T. (2009) Productive steps toward an antimicrobial targeting virulence. *Curr Opin Microbiol* **12**: 490-496.
- Barrett,A.J. (2001) Proteases. *Curr Protoc Protein Sci* **Chapter 21**: Unit.
- Baughn,A.D. and Malamy,M.H. (2002) A mitochondrial-like aconitase in the bacterium *Bacteroides fragilis*: implications for the evolution of the mitochondrial Krebs cycle. *Proc Natl Acad Sci U S A* **99**: 4662-4667.
- Bayley,D.P., Rocha,E.R., and Smith,C.J. (2000) Analysis of *cepA* and other *Bacteroides fragilis* genes reveals a unique promoter structure. *FEMS Microbiol Lett* **193**: 149-154.
- Bergmann,S. and Hammerschmidt,S. (2007) Fibrinolysis and host response in bacterial infections. *Thromb Haemost* **98**: 512-520.
- Blom,N., Gammeltoft,S., and Brunak,S. (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* **294**: 1351-1362.
- Bonifait,L., de la Cruz Dominguez-Punaro, Vaillancourt,K., Bart,C., Slater,J., Frenette,M. et al. (2010) The cell envelope subtilisin-like proteinase is a virulence determinant for *Streptococcus suis*. *BMC Microbiol* **10**: 42.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248-254.

Brigham, C.J. and Malamy, M.H. (2005) Characterization of the RokA and HexA broad-substrate-specificity hexokinases from *Bacteroides fragilis* and their role in hexose and N-acetylglucosamine utilization. *J Bacteriol* **187**: 890-901.

Brook, I. (2002) Clinical review: bacteremia caused by anaerobic bacteria in children. *Crit Care* **6**: 205-211.

Buckwold, S.L., Shoemaker, N.B., Sears, C.L., and Franco, A.A. (2007) Identification and Characterization of Conjugative Transposons CTn86 and CTn9343 in *Bacteroides fragilis* Strains. *Appl Environ Microbiol* **73**: 53 - 63.

Caceres, M., Zhang, G., Weintraub, A., and Nord, C.E. (2000) Prevalance and antimicrobial susceptibility of enterotoxigenic *Bacteroides fragilis* in children with diarrhoea in Nicaragua. *Anaerobe* **6**: 143-148.

Chatzidaki-Livanis, M., Weinacht, K.G., and Comstock, L.E. (2010) *Trans* locus inhibitors limit concomitant polysaccharide synthesis in the human gut symbiont *Bacteroides fragilis*. *PNAS* **107**: 11976 - 11980.

Chan, K.W., Lee, K.H., Mou, J.W., Cheung, S.T., Sihoe, J.D., and Tam, Y.H. (2010) Evidence-based adjustment of antibiotic in pediatric complicated appendicitis in the era of antibiotic resistance. *Pediatr Surg Int* **26**: 157-160.

Chen, T. and Duncan, M.J. (2004) Gingipain adhesin domains mediate *Porphyromonas gingivalis* adherence to epithelial cells. *Microb Pathog* **36**: 205-209.

Chen, Y., Kinouchi, T., Kataoka, K., Akimoto, S., and Ohnishi, Y. (1995) Purification and characterization of a fibrinogen-degrading protease in *Bacteroides fragilis* strain YCH46. *Microbiol Immunol* **39**: 967-977.

Cheneby, D., Philippot, L., Hartmann, A., Henault, C., and Germon, J. (2000) 16S rDNA analysis for characterization of denitrifying bacteria isolated from three agricultural soils. *FEMS Microbiol Ecol* **34**: 121-128.

Clausen, T., Southan, C., and Ehrmann, M. (2002) The HtrA family of proteases: implications for protein composition and cell fate. *Mol Cell* **10**: 443-455.

Coyne, M.J., Kalka-Moll, W., Tzianabos, A.O., Kasper, D.L., and Comstock, L.E. (2000) *Bacteroides fragilis* NCTC9343 produces at least three distinct capsular polysaccharides: cloning, characterization, and reassignment of polysaccharide B and C biosynthesis loci. *Infect Immun* **68**: 6176-6181.

- Coyne,M.J., Weinacht,K.G., Krinos,C.M., and Comstock,L.E. (2003) Mpi recombinase globally modulates the surface architecture of a human commensal bacterium. *Proc Natl Acad Sci U S A* **100**: 10446-10451.
- Cunningham,D.F. and O'Connor,B. (1997) Proline specific peptidases. *Biochim Biophys Acta* **1343**: 160-186.
- Cutting,S., Roels,S., and Losick,R. (1991) Sporulation operon spoIVF and the characterization of mutations that uncouple mother-cell from forespore gene expression in *Bacillus subtilis*. *J Mol Biol* **221**: 1237-1256.
- Dagert,M. and Ehrlich,S.D. (1979) Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. *Gene* **6**: 23-28.
- Dargatz,H., Diefenthal,T., Witte,V., Reipen,G., and von,W.D. (1993) The heterodimeric protease clostripain from *Clostridium histolyticum* is encoded by a single gene. *Mol Gen Genet* **240**: 140-145.
- Dobrindt,U. and Hacker,J. (2008) Targeting virulence traits: potential strategies to combat extraintestinal pathogenic *E. coli* infections. *Curr Opin Microbiol* **11**: 409-413.
- Domingues,R.M., Cavalcanti,S.M., Andrade,A.F., and Ferreira,M.C. (1992) Sialic acid as receptor of *Bacteroides fragilis* lectin-like adhesin. *Zentralbl Bakteriologie* **277**: 340-344.
- El,T.Y. and Skurnik,M. (2001) YadA, the multifaceted *Yersinia* adhesin. *Int J Med Microbiol* **291**: 209-218.
- Esmay,P.A., Billington,S.J., Link,M.A., Songer,J.G., and Jost,B.H. (2003) The *Arcanobacterium pyogenes* collagen-binding protein, CbpA, promotes adhesion to host cells. *Infect Immun* **71**: 4368-4374.
- Ferreira,E.O., Falcao,L.S., Vallim,D.C., Santos,F.J., Andrade,J.R., Andrade,A.F. et al. (2002) *Bacteroides fragilis* adherence to Caco-2 cells. *Anaerobe* **8**: 307-314.
- Ferreira,E.O., Araujo,L.L., Barreiros,P.D., dos S Avelar KE, Ferreira,M.C., Silva Filho,F.C., and Domingues,R.M. (2006) A *Bacteroides fragilis* surface glycoprotein mediates the interaction between the bacterium and the extracellular matrix component laminin-1. *Res Microbiol* **157**: 960-966.
- Ferreira,E.O., de Carvalho,J.B., Peixoto,R.J., Lobo,L.A., Zingalli,R.B., Smith,C.J. et al. (2009) The interaction of *Bacteroides fragilis* with components of the human fibrinolytic system. *FEMS Immunol Med Microbiol* **56**: 48-55.
- Ferreira,E.O., Yates,E.A., Goldner,M., Vommaro,R.C., Silva Filho,F.C., Petropolis,D.B., and Domingues,R.M. (2008) The redox potential interferes with the expression of laminin binding molecules in *Bacteroides fragilis*. *Mem Inst Oswaldo Cruz* **103**: 683-689.

- Finn,R.D., Mistry,J., Tate,J., Coggill,P., Heger,A., Pollington,J.E. et al. (2010) The Pfam protein families database. *Nucleic Acids Res* **38**: D211-D222.
- Foster,T.J. and Hook,M. (1998) Surface protein adhesins of *Staphylococcus aureus*. *Trends Microbiol* **6**: 484-488.
- Franco, A. A. (2004) The *Bacteroides fragilis* pathogenicity island is contained in a putative novel conjugative transposon. *J. Bacteriol.* **186**: 6077 – 6092.
- Gannong,W.F. (2001) Circulating body fluids. In *Review of Medical Physiology*. Flotin,J., Nogueira,I., Ransom,J., and Sheinis,L.A. (eds). International Edition: McGraw-Hill, pp. 499-527.
- Gibson,S.A. and Macfarlane,G.T. (1988a) Characterization of proteases formed by *Bacteroides fragilis*. *J Gen Microbiol* **134**: 2231-2240.
- Gibson,S.A. and Macfarlane,G.T. (1988b) Studies on the proteolytic activity of *Bacteroides fragilis*. *J Gen Microbiol* **134**: 19-27.
- Godoy,V.G., Dallas,M.M., Russo,T.A., and Malamy,M.H. (1993) A role for *Bacteroides fragilis* neuraminidase in bacterial growth in two model systems. *Infect Immun* **61**: 4415-4426.
- Grillo,M.A. and Colombatto,S. (2005) S-adenosylmethionine and protein methylation. *Amino Acids* **28**: 357-362.
- Harrington,D.J. (1996) Bacterial collagenases and collagen-degrading enzymes and their potential role in human disease. *Infect Immun* **64**: 1885-1891.
- Hawser,S.P., Hackel,M., and Hoban,D.J. (2010) Antibiotic susceptibility profiles of European *Bacteroides fragilis* with reduced carbapenem susceptibility. *J Antimicrob Chemother* **65**: 803-804.
- Hirosawa,M., Hoshida,M., Ishikawa,M., and Toya,T. (1993) MASCOT: multiple alignment system for protein sequences based on three-way dynamic programming. *Comput Appl Biosci* **9**: 161-167.
- Holdeman,L.V. and Moore,W.E.C. (1972) *Anaerobe Laboratory Manual*. Blackbury, V. A.: Virginia Polytechnic Institute and State University Anaerobe Laboratory.
- Houston,S., Blakely,G.W., McDowell,A., Martin,L., and Patrick,S. (2010) Binding and degradation of fibrinogen by *Bacteroides fragilis* and characterisation of a 54kDa fibrinogen binding protein. *Microbiology*.
- Ingmer,H. and Brondsted,L. (2009) Proteases in bacterial pathogenesis. *Res Microbiol* **160**: 704-710.

- Isaacson,T., Damasceno,C.M., Saravanan,R.S., He,Y., Catala,C., Saladie,M., and Rose,J.K. (2006) Sample extraction techniques for enhanced proteomic analysis of plant tissues. *Nat Protoc* **1**: 769-774.
- Itoi,Y., Horinaka,M., Tsujimoto,Y., Matsui,H., and Watanabe,K. (2006) Characteristic features in the structure and collagen-binding ability of a thermophilic collagenolytic protease from the thermophile *Geobacillus collagenovorans* MO-1. *J Bacteriol* **188**: 6572-6579.
- Janoir,C., Pechine,S., Grosdidier,C., and Collignon,A. (2007) Cwp84, a surface-associated protein of *Clostridium difficile*, is a cysteine protease with degrading activity on extracellular matrix proteins. *J Bacteriol* **189**: 7174-7180.
- Jin,F., Matsushita,O., Katayama,S., Jin,S., Matsushita,C., Minami,J., and Okabe,A. (1996) Purification, characterization, and primary structure of *Clostridium perfringens* lambda-toxin, a thermolysin-like metalloprotease. *Infect Immun* **64**: 230-237.
- Kalka-Moll,W.M., Wang,Y., Comstock,L.E., Gonzalez,S.E., Tzianabos,A.O., and Kasper,D.L. (2001) Immunochemical and biological characterization of three capsular polysaccharides from a single *Bacteroides fragilis* strain. *Infect Immun* **69**: 2339-2344.
- Kato,T., Takahashi,N., and Kuramitsu,H.K. (1992) Sequence analysis and characterization of the *Porphyromonas gingivalis* prtC gene, which expresses a novel collagenase activity. *J Bacteriol* **174**: 3889-3895.
- Keane,F.M., Loughman,A., Valtulina,V., Brennan,M., Speziale,P., and Foster,T.J. (2007) Fibrinogen and elastin bind to the same region within the A domain of fibronectin binding protein A, an MSCRAMM of *Staphylococcus aureus*. *Mol Microbiol* **63**: 711-723.
- Koch,C.L., Derby,P., and Abratt,V.R. (1998) In-vitro antibiotic susceptibility and molecular analysis of anaerobic bacteria isolated in Cape Town, South Africa. *J Antimicrob Chemother* **42**: 245-248.
- Laemmli,U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.
- Macfarlane,G.T., Allison,C., Gibson,S.A., and Cummings,J.H. (1988) Contribution of the microflora to proteolysis in the human large intestine. *J Appl Bacteriol* **64**: 37-46.
- Maeda,H. (1996) Role of microbial proteases in pathogenesis. *Microbiol Immunol* **40**: 685-699.
- Mancuso,G., Midiri,A., Biondo,C., Beninati,C., Gambuzza,M., Macri,D. et al. (2005) *Bacteroides fragilis*-derived lipopolysaccharide produces cell activation and lethal toxicity via toll-like receptor 4. *Infect Immun* **73**: 5620-5627.
- Maniatis,T., Fritsch,E.F., and Sambrook,J. (1982) *Molecular cloning: a laboratory manual*. Cold Spring Harbour, N. Y..

- Matsushita,O., Jung,C.M., Katayama,S., Minami,J., Takahashi,Y., and Okabe,A. (1999) Gene duplication and multiplicity of collagenases in *Clostridium histolyticum*. *J Bacteriol* **181**: 923-933.
- Matsushita,O., Yoshihara,K., Katayama,S., Minami,J., and Okabe,A. (1994) Purification and characterization of *Clostridium perfringens* 120-kilodalton collagenase and nucleotide sequence of the corresponding gene. *J Bacteriol* **176**: 149-156.
- Mazmanian,S.K., Liu,C.H., Tzianabos,A.O., and Kasper,D.L. (2005) An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122**: 107-118.
- McCormack,A.L., Schieltz,D.M., Goode,B., Yang,S., Barnes,G., Drubin,D., and Yates,J.R., III (1997) Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal Chem* **69**: 767-776.
- Merrick,M.J. and Edwards,R.A. (1995) Nitrogen control in bacteria. *Microbiol Rev* **59**: 604-622.
- Meyer,L., Brink,D., and Weldhagen,G.F. (2006) Antibiotic susceptibility patterns of anaerobic bacteria isolated in Pretoria, South Africa, during 2003 - 2004. *The South African Journal of Epidemiology and Infection* **21**: 161-163.
- Mikamo,H., Kawazoe,K., Sato,Y., and Tamaya,T. (1999) Elastase activity of anaerobes isolated from amniotic fluid with preterm premature rupture of membranes. *Am J Obstet Gynecol* **180**: 378-380.
- Moncrief,J.S., Duncan,A.J., Wright,R.L., Barroso,L.A., and Wilkins,T.D. (1998) Molecular characterization of the fragilysin pathogenicity islet of enterotoxigenic *Bacteroides fragilis*. *Infect Immun* **66**: 1735-1739.
- Moncrief,J.S., Obiso,R., Jr., Barroso,L.A., Kling,J.J., Wright,R.L., Van Tassell,R.L. et al. (1995) The enterotoxin of *Bacteroides fragilis* is a metalloprotease. *Infect Immun* **63**: 175-181.
- Murwantoko, Yano,M., Ueta,Y., Murasaki,A., Kanda,H., Oka,C., and Kawaichi,M. (2004) Binding of proteins to the PDZ domain regulates proteolytic activity of HtrA1 serine protease. *Biochem J* **381**: 895-904.
- Naidu,A.S., Paulsson,M., and Wadstrom,T. (1988) Particle agglutination assays for rapid detection of fibronectin, fibrinogen, and collagen receptors on *Staphylococcus aureus*. *J Clin Microbiol* **26**: 1549-1554.
- O'Brien-Simpson,N.M., Veith,P.D., Dashper,S.G., and Reynolds,E.C. (2003) *Porphyromonas gingivalis* gingipains: the molecular teeth of a microbial vampire. *Curr Protein Pept Sci* **4**: 409-426.
- Orgel,J.P., Wess,T.J., and Miller,A. (2000) The in situ conformation and axial location of the intermolecular cross-linked non-helical telopeptides of type I collagen. *Structure* **8**: 137-142.

- Pallen,M.J. and Wren,B.W. (1997) The HtrA family of serine proteases. *Mol Microbiol* **26**: 209-221.
- Park,P.W., Senior,R.M., Griffin,G.L., Broekelmann,T.J., Mudd,M.S., and Mecham,R.P. (1995) Binding and degradation of elastin by the staphylolytic enzyme lysostaphin. *Int J Biochem Cell Biol* **27**: 139-146.
- Patrick,S. (2002) *Bacteroides*. In *Molecular Medical Microbiology*. Sussman,M. (ed). London: Acedemic Press, pp. 1921-1948.
- Patrick, S., Blakely, G.W., Houston, S., Moore, J., Abratt, V.R., Bertalan, M., Cerdeño-Tárraga, A.M., Quail, M.A., Corton, N., Corton, C., Bignell, A., Barron, A., Clark, L., Bentley, S.D., and Parkhill, J. (2010) Twenty-eight divergent polysaccharide loci specifying within- and amongst-strain capsule diversity in three strains of *Bacteroides fragilis*. *Microbiol* **156**: 3255 - 3269.
- Patrick,S., Houston,S., Thacker,Z., and Blakely,G.W. (2009) Mutational analysis of genes implicated in LPS and capsular polysaccharide biosynthesis in the opportunistic pathogen *Bacteroides fragilis*. *Microbiology* **155**: 1039-1049.
- Patrick,S., Parkhill,J., McCoy,L.J., Lennard,N., Larkin,M.J., Collins,M. et al. (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* **149**: 915-924.
- Patti,J.M., Allen,B.L., McGavin,M.J., and Hook,M. (1994) MSCRAMM-mediated adherence of microorganisms to host tissues. *Annu Rev Microbiol* **48**: 585-617.
- Pauer,H., Ferreira,E.O., dos Santos-Filho,J., Portela,M.B., Zingali,R.B., Soares,R.M., and Domingues,R.M. (2009) A TonB-dependent outer membrane protein as a *Bacteroides fragilis* fibronectin-binding molecule. *FEMS Immunol Med Microbiol* **55**: 388-395.
- Pethe,K., Bifani,P., Drobecq,H., Sergheraert,C., Debrie,A.S., Locht,C., and Menozzi,F.D. (2002) Mycobacterial heparin-binding hemagglutinin and laminin-binding protein share antigenic methyllysines that confer resistance to proteolysis. *Proc Natl Acad Sci U S A* **99**: 10759-10764.
- Phillips,A.L., Means,W.J., Kalchayanand,N., McCormick,R.J., and Miller,K.W. (2000) Bovine placental protease specificity toward muscle connective tissue proteins. *J Anim Sci* **78**: 1861-1866.
- Privitera,G., Dublanchet,A., and Sebald,M. (1979) Transfer of multiple antibiotic resistance between subspecies of *Bacteroides fragilis*. *J Infect Dis* **139**: 97-101.
- Pucci-Minafra,I., Minafra,S., La,R.G., Barranca,M., Fontana,S., Alaimo,G., and Okada,Y. (2001) Zymographic analysis of circulating and tissue forms of colon carcinoma gelatinase A (MMP-2) and B (MMP-9) separated by mono- and two-dimensional electrophoresis. *Matrix Biol* **20**: 419-427.

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

Pumbwe,L., Skilbeck,C.A., and Wexler,H.M. (2006) The *Bacteroides fragilis* cell envelope: quarterback, linebacker, coach-or all three? *Anaerobe* **12**: 211-220.

Rao,M.B., Tanksale,A.M., Ghatge,M.S., and Deshpande,V.V. (1998) Molecular and biotechnological aspects of microbial proteases. *Microbiol Mol Biol Rev* **62**: 597-635.

Rawlings,N.D., Barrett,A.J., and Bateman,A. (2010) MEROPS: the peptidase database. *Nucleic Acids Res* **38**: D227-D233.

Reid,J.H. and Patrick,S. (1984) Phagocytic and serum killing of capsulate and non-capsulate *Bacteroides fragilis*. *J Med Microbiol* **17**: 247-257.

Rhee, K.J., Wu, S., Wu, X., Huso, D.L., Karim, B., Franco, A.A., Rabi-zadeh, S., Golub, J., Mathews, L.E., Shin, J., Sartor, R.B., Golenbock, D., Hamad, A., Gan, C. M., Housseau, F., and Sears, C.L. (2009) Induction of persistent colitis by a human commensal, enterotoxigenic *Bacteroides fragilis*, in wild-type C57BL/6 mice. *Infect. Immun.* **77**: 1708 – 1718.

Robertson,K.P., Smith,C.J., Gough,A.M., and Rocha,E.R. (2006) Characterization of *Bacteroides fragilis* hemolysins and regulation and synergistic interactions of HlyA and HlyB. *Infect Immun* **74**: 2304-2316.

Roh,K.H., Kim,S., Kim,C.K., Yum,J.H., Kim,M.S., Yong,D. et al. (2009) Resistance trends of *Bacteroides fragilis* group over an 8-year period, 1997-2004, in Korea. *Korean J Lab Med* **29**: 293-298.

Round,J.L., and Mazmanian, S.K. (2010) Inducible Foxp3 + regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *PNAS* **107**: 12204 - 12209.

Sambrook,J., Fritsch,E.F., and Maniatis,T. (2001) *Molecular cloning: a laboratory manual*. Cold Spring Harbour, N. Y..

Sasaki,T., Fassler,R., and Hohenester,E. (2004) Laminin: the crux of basement membrane assembly. *J Cell Biol* **164**: 959-963.

Sato,K., Kumita,W., Ode,T., Ichinose,S., Ando,A., Fujiyama,Y. et al. (2010) OmpA variants affecting the adherence of ulcerative colitis-derived *Bacteroides vulgatus*. *J Med Dent Sci* **57**: 55-64.

Schaffer,A.A., Aravind,L., Madden,T.L., Shavirin,S., Spouge,J.L., Wolf,Y.I. et al. (2001) Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res* **29**: 2994-3005.

Scozzafava,A., Ilies,M.A., Manole,G., and Supuran,C.T. (2000) Protease inhibitors. Part 12. Synthesis of potent matrix metalloproteinase and bacterial collagenase inhibitors incorporating sulfonylated N-4-nitrobenzyl-beta-alanine hydroxamate moieties. *Eur J Pharm Sci* **11**: 69-79.

Sears,C.L. (2001) The toxins of *Bacteroides fragilis*. *Toxicon* **39**: 1737-1746.

Sears, C.L. (2009) Enterotoxigenic *Bacteroides fragilis*: a Rogue among Symbiotes. *Clin Microbiol Rev* **22**: 349 - 369.

Shah, H.N. , Olsen, I., Bernard, K., Finegold, S.M., Gharbia, S., and Gupta, R.S. (2009) Approaches to the study of the systematics of anaerobic, Gram-negative, non-sporeforming rods: Current status and perspectives. *Anaerobe* **15**: 179 -194.

Shao,J., Xu,D., Tsai,S.N., Wang,Y., and Ngai,S.M. (2009) Computational identification of protein methylation sites through bi-profile Bayes feature extraction. *PLoS One* **4**: e4920.

Shoemaker,N.B., Wang,G.R., and Salyers,A.A. (2000) Multiple gene products and sequences required for excision of the mobilizable integrated *Bacteroides* element NBU1. *J Bacteriol* **182**: 928-936.

Sijbrandi,R., Stork,M., Luirink,J., and Otto,B.R. (2008) Pbp, a cell-surface exposed plasminogen binding protein of *Bacteroides fragilis*. *Microbes Infect* **10**: 514-521.

Simon,R., Prierer,U., and Puhler,A. (1983) A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. *Nat Biotech* **1**: 784-791.

Snydman,D.R., Jacobus,N.V., McDermott,L.A., Golan,Y., Hecht,D.W., Goldstein,E.J. et al. (2010) Lessons learned from the anaerobe survey: historical perspective and review of the most recent data (2005-2007). *Clin Infect Dis* **50 Suppl 1**: S26-S33.

Soki,J., Szoke,I., and Nagy,E. (1999) Characterisation of a 5.5-kb cryptic plasmid present in different isolates of *Bacteroides* spp. originating from Hungary. *J Med Microbiol* **48**: 25-31.

Southern,J.A., Parker,J.R., and Woods,D.R. (1986) Expression and purification of glutamine synthetase cloned from *Bacteroides fragilis*. *J Gen Microbiol* **132**: 2827-2835.

Spence,C., Wells,W.G., and Smith,C.J. (2006) Characterization of the primary starch utilization operon in the obligate anaerobe *Bacteroides fragilis*: Regulation by carbon source and oxygen. *J Bacteriol* **188**: 4663-4672.

Spencer,J., Murphy,L.M., Connors,R., Sessions,R.B., and Gamblin,S.J. (2010) Crystal structure of the LasA virulence factor from *Pseudomonas aeruginosa*: substrate specificity and mechanism of M23 metallopeptidases. *J Mol Biol* **396**: 908-923.

Sperandio,V., Giron,J.A., Silveira,W.D., and Kaper,J.B. (1995) The OmpU outer membrane protein, a potential adherence factor of *Vibrio cholerae*. *Infect Immun* **63**: 4433-4438.

- Steffens,L.S., Nicholson,S., Paul,L.V., Nord,C.E., Patrick,S., and Abratt,V.R. (2010) *Bacteroides fragilis* RecA protein overexpression causes resistance to metronidazole. *Res Microbiol* **161**: 346-354.
- Stone,K.L. and Williams,K.R. (2009) Reverse-Phase HPLC Separation of Enzymatic Digests of Proteins. In *The Protein Protocols Handbook*. Walker,J. (ed). New Haven, CT: pp. 937-946.
- Stubben,C.J., Duffield,M.L., Cooper,I.A., Ford,D.C., Gans,J.D., Karlyshev,A.V. et al. (2009) Steps toward broad-spectrum therapeutics: discovering virulence-associated genes present in diverse human pathogens. *BMC Genomics* **10**: 501.
- Studier,F.W. and Moffatt,B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**: 113-130.
- Supuran,C.T., Scozzafava,A., and Mastrolorenzo,A. (2001) Bacterial proteases: current therapeutic use and future prospects for the development of new antibiotics. *Expert Opinion on Therapeutic Patents* **11**: 221-259.
- Szoke,I., Pascu,C., Nagy,E., Ljung,A., and Wadstrom,T. (1996) Binding of extracellular matrix proteins to the surface of anaerobic bacteria. *J Med Microbiol* **45**: 338-343.
- Szoke,I., Pascu,C., Nagy,E., Ljung,A., and Wadstrom,T. (1997) Binding of extracellular matrix proteins to the surface of *Bacteroides* spp. *Anaerobe* **3**: 91-95.
- Takahashi,N., Kato,T., and Kuramitsu,H.K. (1991) Isolation and preliminary characterization of the *Porphyromonas gingivalis prtC* gene expressing collagenase activity. *FEMS Microbiol Lett* **68**: 135-138.
- Thumm,G. and Gotz,F. (1997) Studies on prolystaphin processing and characterization of the lysostaphin immunity factor (Lif) of *Staphylococcus simulans* biovar staphylolyticus. *Mol Microbiol* **23**: 1251-1265.
- Toprak, N.U., Yagci, A., Gulluoglu, B.M., Akin, M.L., Demirkalem, P., Celenk, T., and Soyletir, G. (2006) A possible role of *Bacteroides fragilis* entero-toxin in the aetiology of colorectal cancer. *Clin. Microbiol. Infect.* **12**: 782 – 786.
- Travis,J. and Potempa,J. (2000) Bacterial proteinases as targets for the development of second-generation antibiotics. *Biochim Biophys Acta* **1477**: 35-50.
- Tzianabos,A.O., Onderdonk,A.B., Rosner,B., Cisneros,R.L., and Kasper,D.L. (1993) Structural features of polysaccharides that induce intra-abdominal abscesses. *Science* **262**: 416-419.
- Van de Westhuyzen,I. and Chalkey,L.J. (1992) Antibiotic susceptibility of anaerobic bacteria isolated in Johannesburg. *South African Medical Journal* **82**: 314-317.
- van Rooyen,J.M., Abratt,V.R., and Sewell,B.T. (2006) Three-dimensional structure of a type III glutamine synthetase by single-particle reconstruction. *J Mol Biol* **361**: 796-810.

- Watanabe,K. (2004) Collagenolytic proteases from bacteria. *Appl Microbiol Biotechnol* **63**: 520-526.
- Wells,T.J., McNeilly,T.N., Totsika,M., Mahajan,A., Gally,D.L., and Schembri,M.A. (2009) The *Escherichia coli* O157:H7 EhaB autotransporter protein binds to laminin and collagen I and induces a serum IgA response in O157:H7 challenged cattle. *Environ Microbiol*.
- Wexler,H.M., Tenorio,E., and Pumbwe,L. (2009) Characteristics of *Bacteroides fragilis* lacking the major outer membrane protein, OmpA. *Microbiology* **155**: 2694-2706.
- Witte,C., Jensen,R.E., Yaffe,M.P., and Schatz,G. (1988) MAS1, a gene essential for yeast mitochondrial assembly, encodes a subunit of the mitochondrial processing protease. *EMBO J* **7**: 1439-1447.
- Wu, S., Dreyfus, L.A., Tzianabos, A.O., Hayashi, C., and Sears, C.L. (2002) Diversity of the Metalloprotease Toxin Produced by Enterotoxigenic *Bacteroides fragilis*. *Infect Immun* **70**: 2463 - 2471.
- Yanisch-Perron,C., Vieira,J., and Messing,J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103-119.
- Yoshihara,K., Matsushita,O., Minami,J., and Okabe,A. (1994) Cloning and nucleotide sequence analysis of the *colH* gene from *Clostridium histolyticum* encoding a collagenase and a gelatinase. *J Bacteriol* **176**: 6489-6496.
- Yu,F., Iyer,D., Anaya,C., and Lewis,J.P. (2006) Identification and characterization of a cell surface protein of *Prevotella intermedia* 17 with broad-spectrum binding activity for extracellular matrix proteins. *Proteomics* **6**: 6023-6032.
- Zhang,G., Svenungsson,B., Karnell,A., and Weintraub,A. (1999) Prevalence of enterotoxigenic *Bacteroides fragilis* in adult patients with diarrhea and healthy controls. *Clin Infect Dis* **29**: 590-594.