Characterisation of Campylobacter jejuni glycoprotease and its Role in Bacteria - Host Interactions

Author

Klipic, Zoran

Published

2011

Thesis Type

Thesis (PhD Doctorate)

School

Institute for Glycomics

DOI

10.25904/1912/1035

Characterisation of *Campylobacter jejuni* glycoprotease and its role in bacteria – host interactions

Zoran Klipic
B. Biomed Sci, B. Biomed Sci (Hon)

Department of Microbial Glycobiology, Institute for Glycomics
Griffith University, Gold Coast, Australia

This thesis is submitted in fulfilment of the requirements for the degree of Doctor of Philosophy.

Supervisor: Assoc. Prof. Victoria Korolik
Associate supervisor: Dr Jennifer Wilson

Submitted November 2010

Abstract

The bacterial enzyme glycoprotease (gcp) was first discovered in the culture supernatant of *Mannheimia (Pasteurella) haemolytica* A1 associated with bovine pneumonic pasteurellosis. The enzyme is highly specific for *O*-sialoglycoproteins, but the function of the enzyme in bacterial homeostasis was not fully elucidated.

The *in silico* analysis of *C. jejuni* genome strain NCTC11168 has revealed Cj1344c, a protein with predicted amino acid sequence showing 55% similarity to the *M. haemolytica O*-sialoglycoprotease. The *C. jejuni* Cj1344c homologue was present in all to-date sequenced strains of *C. jejuni* with higher than 97% amino acid identity and an orthologue of this enzyme was present in other *Campylobacter* species, with greater than 70% amino acid similarity. The glycoprotease was also present in the genomes of *Campylobacter* related species, such as *Helicobacter* and *Wollinella* with 65-70% similarity to *C. jejuni* Cj1344c. Comparative analysis also identified presence of orthologues in bacterial species such as *Bacillus anthracis*, *Staphylococcus aureus* and *Haemophilus influenzae*, with more than 49% amino acid similarity to the predicted Cj1344c amino acid sequence.

In this study, the utilisation of gene mutagenesis approach demonstrated that a putative glycoprotease (Cj1344c) is required for *C. jejuni* survival and growth, as the inactivation of the gene by insertion of an antibiotic resistance gene cassette resulted in bacterial death. The gene products essential for bacterial growth *in vitro* and survival during infection constitute an initial set of protein targets for the development of antibacterial vaccines. The results of this study indicate that Cj1344c is a potential novel target for the development of antimicrobials against *C. jejuni* or a target for the development of a vaccine.

In order to utilise the protein in the immunisation and protection studies and to determine its potential as a vaccine candidate, the function of the protein needed to be determined. The protein was expressed and purified utilising a pET-19b system which enabled the overexpression of the protein in *E. coli* and its subsequent purification as a fusion protein with an N-terminal polyhistidine tag.

The role of the Cj1344c in the bacterium could not be elucidated due to the inability to generate the isogenic mutant. In order to gain an insight into the activity of the protein, the Cj1344c His-tagged protein was used. Analysis of the purified His-Cj1344c binding capability by glycan and small molecule array determined that it recognises methionine, lysine and arginine, suggesting that these amino acids are present in the sequences of glycoproteins that are recognised by Cj1344c. The enzyme was also shown to possess specificity to glycosylated structures as it was recognising bovine lactoferrin, but not recombinant lactoferrin which lacks sialic acid. In addition, the specificity of Cj1344c to MUC2 through the use of glycan array methodology was identified, which suggests a putative role for Cj1344c in the degradation of this molecule which was reported to be very important in the *C. jejuni* pathogenesis. Modification or degradation of the mucous layer of the gastrointestinal tract may play a role during the initial stages of *C. jejuni* adherence and invasion of epithelial cells. Enzymatic digestion of MUC2 with His-Cj1344c could not confirm the biological activity of the enzyme. It was speculated that the misfolding of the His-Cj1344c or absence of the enzyme co-factor was probable reason for the reduced enzymatic activity of His-Cj1344c observed in the study. The lack of the enzymatic activity made this protein a good antigen candidate for immunisation trial as it was speculated that its toxicity, due to the reduced activity, would be minimal.

The crucial role of the Cj1344c in the cell survival, its high degree of similarity between campylobacter species as well as its putatively important role in the bacterial pathogenesis through degradation of the mucous layer; makes this protein a potentially very good vaccine candidate. High antibody titres (1:65,000) obtained in the rabbit immunisation with His-Cj1344c provided encouraging preliminary results for the investigation of possible protective role of the Cj1344c against *C. jejuni* infection. The preliminary mouse immunisation trials, assessing different routes of antigen administration, have identified the subcutaneous immunisation to provide the best immune response to the His-Cj1344c. The minimal dose of 5 µg His-Cj1344c during immunisation did not produce any adverse effects in the mice; and produced high IgG antibody titres (1:65,000). The protection studies against *C. jejuni* infection have determined that the mice immunised with His-Cj1344c show lower number of *C. jejuni* cells in their faeces and the small and large intestines, which was indicative of lower colonisation even though the results did not show statistical significance. The immunisation study has also identified that the non-vaccinated mice had His-Cj1344c specific antibodies, which would suggest that the His-Cj1344c specific antibodies have been produced in the non-vaccinated group of animals through a transient infection with *C. jejuni* or related species against the native protein. More importantly, these studies confirm that the enzyme is expressed by *C. jejuni in vivo* and is likely involved in the bacterial pathogenesis.

Statement of Originality

I declare that the work presented in this thesis was performed within the Institute for Glycomics, under the supervision of Associate Professor Victoria Korolik. This work has not been previously been submitted for a degree or a diploma in any university. To the best of my knowledge and belief the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Zoran Klipic

Table of contents

List of Figures

List of Tables

Abbreviations

A	Adenosine
Å	Angstrom. Unit of distance 10^{-12} meters.
A549	Human lung adenocarcinoma epithelial cell line
bp	base pairs
C-	carboxy-
CaCo-2	Human Caucasian colon adenocarcinoma epithelial cell line
Cat	chloramphenicol resistance cassette
CcaA	Campylobacter chemotaxis aspartate receptor A
cDNA	Complementary DNA
cdt	cytolethal distending toxin
cfu	colony forming units
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleotides
DPP	dipeptide binding-protein
E	Glutamic acid, methylation site of cytoplasmic domain
EDTA	ethylenediaminetetraacetic acid
G	Guanine
GAGs	Glycoaminoglycans
GBP	Galactose binding-protein
GBS	Guillain-Barré Syndrome
Hep-G2	Human hepatocellular liver carcinoma cell line
HIV	Human Immunodeficiency Virus
HCR	Highly conserved region (or domain)

HK	Histidine kinase
HRP	Horse radish peroxidase
IMS	Immunomagnetic separation
KDa	Kilodaltons
Km	Kanamycin
μ	micro, 10^{-6}
Mb	Megabase pairs
MBP	Maltose binding-protien
MEM	minimal essential media
MS	Methylation site
Neu5Ac	N-Acetylneuraminic acid
NMR	Nuclear magnetic resonance
ng	nanograms, 10^{-9}
PAGE	Polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PD	periplasmic domain
ppm	parts per million
PVDF	Polyvinylidene Fluoride
Q	Glutamine, methylation site of cytoplasmic domain
Q PCR	Quantitative PCR
RBP	Ribose binding-protein
RPMI	Media for tissue culture developed at Roswell Park Memorial Institute.
RNA	Ribonucleic acid

SDS	Sodium dodecylsulphate
STD	Saturation transfer difference
T	thymine
TAE	Tris/acetate/ethylene diamine tetracetic acid
TBS	tris buffered saline
TBS-T	Tris buffered saline-tween 20
TE	tris/ethylene diamine tetracetic acid
TEMED	N,N,N`,N`-tetramethylethylenediamine
Tris	Tris[hydroxymethyl]aminomethane
U	units

Acknowledgements

I would like to thank Griffith University, Institute for Glycomics for giving me the opportunity to work towards my Doctorate of Philosophy. I would like to express my deep and sincere gratitude to my supervisor, Associate Professor Victoria Korolik, Institute for Glycomics, Griffith University. Her wide knowledge and her logical way of thinking have been of great value. Her understanding, encouraging and personal guidance have been priceless through the years it took me to complete the work presented in these pages. I would also like to thank Dr Jennifer Wilson for her help and guidance with STD-NMR spectroscopy and constructive criticism in writing these pages.

I would like to thank Associate Professor Michael McGuckin at the Mater Medical Research Institute, Brisbane and late Professor Anthony Moran at National University of Ireland, Galway who provided invaluable advice on planning animal immunisation experiments.

I wish to express my warm and sincere thanks to the members of Korolik team, especially Lirio and Lauren for their invaluable help, support and understanding. My gratitude goes to all of them for giving up their valuable time to help me with the animal experiments which would've been difficult for me to complete by myself.

I owe my loving thanks to the people most important in my life; my brother Goran, my parents Damjan and Marjana, and last but not least my partner Shane. You have my deepest gratitude for your continued and unquestioning love, support and understanding. Without your encouragement and understanding it would have been impossible for me to finish this work. Finally, I dedicate this thesis to the loving memory of my grandma "Maja". Your love and support will forever guide me through my adventures.

CHAPTER 1

Introduction to Campylobacter jejuni

1.1 Historical perspectives and *Campylobacter* genus

Campylobacter spp. have been the focus of growing attention for the past 40 years because of the increasing frequency with which they have been isolated from man, animals, food and water. Recognised as a human pathogen in 1970s, campylobacters have probably caused illness in mankind for centuries. The first documented case of campylobacteriosis was published by Thomas Escherisch in 1886, describing spiral bacteria in the colons of children who had died of what he called "cholera infantum" (as reviewed in Butzler, 2004). The organisms were originally assigned to the *Vibrio* genus, due to their spiral appearance and were named by Smith in 1918 as *Vibrio fetus* as they were isolated from aborted bovine foetuses (as reviewed in Moore *et al.*, 2005).

These observations, however, failed to attract worldwide recognition until 1970s. The breakthrough in the identification and classification of the organism was accomplished by the isolation of *Campylobacter* spp. from faeces in 1968 (Dekeyser *et al.*, 1972). A few years later, the discovery of selective media for the growth of *Campylobacter* spp. by Skirrow, brought the research of campylobacters into a new era (Skirrow, 1977).

Campylobacter jejuni belongs to the epsilon class of proteobacteria, in the order *Campylobacteriales*; this order includes the other two genera, *Helicobacter* and *Wolinella* (Vandamme, 2000). Although the genus *Campylobacter* is composed of 16 described species (Vandamme, 2000), human illness is associated primarily with *C. jejuni* and *C. coli* and infrequently with *C. upsaliensis, C. lari* and *C. fetus* (Vandamme, 2000).

1.2 Microbiology and genetics

Campylobacter jejuni cells are small (1.5–6 µm long and 0.2–0.5 µm wide), spirally curved, Gram-negative bacilli (Figure 1.1) that exhibit rapid spinning motions by means of a single polar flagellum at one or both poles (Park, 2002). The organisms are generally considered to be microaerophilic, that is they are unable to grow in the presence of air and grow optimally in atmospheres containing 5% oxygen (Thompson *et al.*, 1990). In addition, campylobacters have a restricted temperature growth range and whilst they grow optimally at 42˚C, the organisms do not grow at temperatures below 30˚C (Nachamkin *et al.*, 2000b).

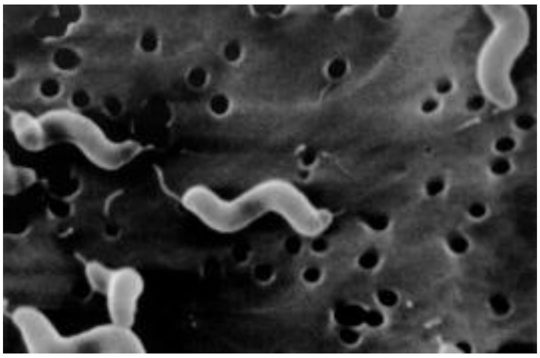


Figure 1.1 Scanning electron microscope image of *C. jejuni*, illustrating its corkscrew appearance and bipolar flagella (Altekruse *et al.*, 1999)

As a consequence of environmental stress (temperature extremes, starvation, oxidative stress and pH extremes) the bacteria are believed to be able to enter a viable, non-culturable state which means that they are still infectious but cannot be cultured in the laboratory (Bovill & Mackey, 1997).

C. jejuni has a small genome of 1.64 Mbp that is AT-rich with a GC ratio of 30.6% (Parkhill *et al.*, 2000b). The *C. jejuni* genome is one of the densest bacterial genomes sequenced to date, with 94.3% of the genome encoding for proteins (Parkhill *et al.*, 2000b). The small genome may reflect the bacteria's habitat of the animal gastrointestinal tract, its requirements for complex-media and its inability to ferment carbohydrates and degrade complex compounds (Altekruse *et al.*, 1999).

1.3 Reservoirs and Transmission

Campylobacter enteritis is considered to be a zoonosis, with many animals serving as possible reservoirs for human disease. Enteric campylobacters are frequently isolated from the faeces of many mammals and domestic and wild birds (Altekruse *et al.*, 1999). The prevalence of *C. jejuni* in the majority of domestic animal sources (including cattle, pigs, sheep and poultry) ranged from 22% to 28% with poultry being highest at 41% (Ogden *et al.*, 2009). Domesticated pets are known to harbour *Campylobacter* spp. in their digestive tracts (Horrocks *et al.*) and as such are a potential infection source.

Campylobacter appears to permanently colonise the gastrointestinal tract of birds with few noticeable effects and only occasionally is diarrhoea observed with *Campylobacter* infection in birds (Newell, 2001). In case of domestic birds, especially chickens, colonisation can occur with as few as 35 organisms (Kaino *et al.*, 1988) and by four weeks most chickens in commercial operations are colonised (Humphrey *et al.*, 2007). *C. jejuni* is often carried by migratory birds – cranes, ducks and geese (Luechtefeld *et al.*, 1980). Shedding of campylobacter by wild birds causes contamination of waterways, and, as campylobacters can survive in water for weeks (Bolton *et al.*, 1987), open waters may act as a source of infection in domestic animals.

The routes of transmission of *C. jejuni* in humans is most often by ingestion of contaminated poultry, raw milk (Crushell *et al.*, 2004, Altekruse *et al.*, 1999) and drinking contaminated water (Ashbolt, 2004). Poultry, however, is considered to be a major source of transmission with epidemiological studies suggesting a significant link between infection with *C. jejuni* and the consumption of raw or undercooked chicken (Park, 2002). Studies have shown very high rates of *Campylobacter* contamination among supermarket chicken and meat products (as reviewed in Humphrey *et al.*, 2007. Although traditional cooking methods kill *Campylobacter*, these organisms may survive the cooking process if the meat is not cooked sufficiently and pose a risk of infection (Allerberger *et al.*, 2003).

1.4 Epidemiology

Since the discovery of *C. jejuni* as the causative agent of campylobacter enteritis, or campylobacteriosis, in the 1970s, the bacterium emerged as the most frequent cause of infectious diarrhoea (Skirrow, 1991, Ketley, 1997, MMWR, 2005). In the study of patients suffering from diarrhoea, *Campylobacter* spp. are frequently isolated from the faeces along with other causative agents of diarrhoea (Rotavirus, ETEC, EPEC, *Aeromonas* spp, *Shigella* sp. and *Vibrio cholerae*) (Albert *et al.*, 1999). *C. jejuni* is, however, documented to be the major cause of gastroenteritis in humans (Ketley, 1997). Over 99% of reported *Campylobacter* isolates are *C. jejuni* which affects over 450 million people every year globally; and is responsible for a large economic burden (Friedman *et al.*, 2000). However, the true number of infections may be much higher than the figures currently reported. The main reason for this is the fact that many of these infections go undiagnosed, and in passive surveillance, most diagnosed infections are not reported.

In the USA alone, the estimated number of cases of campylobacteriosis reaches 2.4 million cases annually (Crushell *et al.*, 2004). In Australia, approximately 16,000 cases of gastroenteritis caused by *C. jejuni* are reported yearly, which accounts for

73-80% of all reported gastrointestinal infections in this country (Liu *et al.*, 2009). Considerably high numbers of *C. jejuni*-related gastrointestinal disease can be attributed to the combination of the wide spread of the bacteria in many countries and their commensal presence in poultry and other animal products; and a relatively low infectious dose of this microorganism in humans. The human infectious doses of *C. jejuni* have been reported to be as low as 500-800 organisms (as reviewed (Skirow & Blaser, 2000), however, the incidence and the clinical manifestations of the disease vary in different parts of the world. Variations in incidence rate have been observed between countries and even between different regions within countries (Brieseman, 1990, NNDSS, 2003). In addition, the age-specific incidence of *C. jejuni* has also been repeatedly observed (Friedman *et al.*, 2000, Coker *et al.*, 2002), with a marked difference between developed and developing countries as explained below.

1.4.1 Developed Countries

The isolation rates of *C. jejuni* in developed countries vary, but most show an increased in number of *C. jejuni* cases over the last 20 years period. These differences in incidence rates between countries may be attributed to different carriage rates in food animals, differences in food preparation practices and patterns of food consumption (as reviewed in Friedman *et al.*, 2000. Figure 1.2 shows that *C. jejuni* infections affect all age groups in different developed countries, but the main age groups affected are children (less than 4 years) and young adults (15 to 44 years). It has been speculated that the first peak could be due to increased sampling in this age group as parents are more likely to seek medical attention for children. The second peak could be attributed to the increased foreign travel in this age group (Friedman *et al.*, 2000).

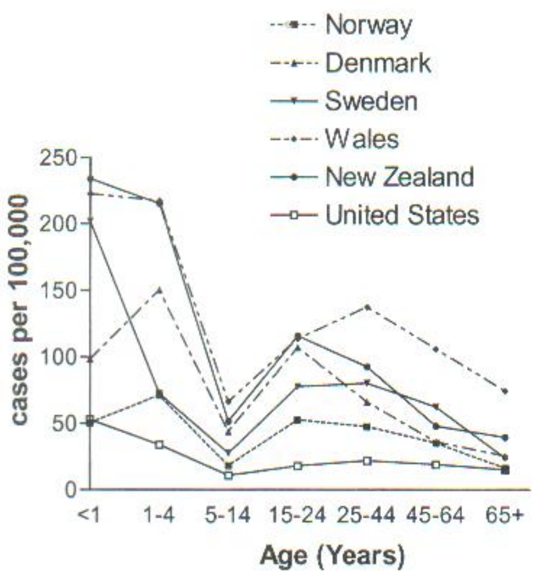


Figure 1.2 Incidence of *C. jejuni* infections by age group in developed countries (Friedman *et al.*, 2000)

1.4.2 Developing Countries

The epidemiology of *Campylobacter* infections in the developing world differs markedly from that in the developed world. Beside the difference in climate, population density and ethnic background, the epidemiological differences in the developed and developing world are best explained by the increased exposure and high infection rates early in life in the developing-world populations (Oberhelman & Taylor, 2000a). The main age group affected in the developing countries are children younger than five, with the incidence of *Campylobacter* infections as high as 40,000 per 100,000 (Oberhelman *et al.*, 1999). At the same time, the disease does not appear to be important in adults. Poor hygiene and sanitation; and the close proximity to animals, causing drinking water contamination, in developing countries all contribute to ease and frequent acquisition of *Campylobacter* (Coker *et al.*, 2002).

1.4.3 *Campylobacter* infections in HIV-infected patients

Patients with AIDS are usually more prone to *Campylobacter* spp. infections and in these patients the infection is usually more severe (Molina *et al.*, 1995). Chronic carriage and recurrent enteritis, often with bacteraemia, is a typical problem. A study of *Campylobacter* infections in human immunodeficiency virus-infected patients showed that 10% had bacteraemia (Molina *et al.*, 1995). Already immunocompromised, patients usually developed debilitating, febrile illness requiring multiple and prolonged courses of antimicrobial therapy (Tee & Mijch, 1998). Overall, it has been estimated that the incidence of *Campylobacter* infections in patients with AIDS is 40-fold higher than in immunocompetent patients (Sorvillo *et al.*, 1991).

1.4.4 Epidemiology of Traveller's diarrhoea

Traveller's diarrhoea is the most common illness acquired by visitors to the developing countries, affecting 20-30% of the 35 million people who travel from industrialised countries each year (Adachi *et al.*, 2000, Castelli & Carosi, 1995). *Campylobacters* are reported to be the main causative agent of diarrhoea in travellers to the developing countries (Black, 1990). It is estimated that 3-50% of all campylobacteriosis cases are associated with foreign travel, and usually result from the consumption of contaminated food or water (Butzler, 2004). The use of antimicrobial agents reduces the duration of the symptoms, but the success of treatment with these agents is becoming limited because of increasing bacterial resistance (Murphy *et al.*, 1996). Education on hygiene and safe food practices remain an effective way in preventing many diarrhoeal diseases, including *Campylobacter* gastroenteritis.

1.5 Clinical features of infection

Infection with *C. jejuni* can induce a spectrum of disease symptoms and variable severity of disease in humans. The clinical presentation of patients with *C. jejuni* infections differs between developing and industrialised countries. Variations in bacterial virulence and host immune response each may play a role in these different phenotypic expressions of disease (Ketley, 1997). In the developing world, infections are usually asymptomatic or there may be mild non-inflammatory diarrhoea, predominantly effecting young children (Oberhelman & Taylor, 2000b). In the industrialised world, acute self-limiting gastrointestinal illness, characterised by diarrhoea (ranging from a watery, non-bloody, non-inflammatory diarrhoea to a severe inflammatory diarrhoea), fever and abdominal cramps, is the most common presentation of *C. jejuni* infection (Butzler *et al.*, 1992, Butzler & Oosterom, 1991, Coker *et al.*, 2002). The signs and symptoms of the disease are not very characteristic and cannot determine the causative agent of the illness, as they are very similar in presentation to those caused by *Salmonella* and *Shigella* (Butzler *et al.*, 1992).

The mean incubation period of *C. jejuni* is 3.2 days, with a range of 18 hours to 8 days, but can last up to 10 days (Blaser & Engberg, 2008). Further progress of infection is characterised by the onset of diarrhoea; and is estimated that 50% of the patients attending emergency rooms have 10 or more bowel motions per day (Skirrow & Blaser, 2000). About 15% of patients have reported blood in their stools, 1 to 2 days after the onset of diarrhoea (Skirrow & Blaser, 2000). Once the diarrhoea stops (usually about 3 to 4 days) the discomfort and abdominal pain may persist for several more days. Examination of faecal samples usually shows numerous *Campylobacter* organisms and leukocytes (Butzler, 2004). Patients continue to excrete *Campylobacters* in their faeces for several weeks after they have clinically recovered (Ketley, 1997). The average duration of illness is difficult to measure due to many variables, such as the immune status of the host, the virulence of the strain and criteria used to define illness, but usually lasts about 5 days, and sometimes extends to 7 days (Blaser & Engberg, 2008).

Apart from campylobacteriosis, *C. jejuni* can also be responsible for rare cases of pseudoappendicitis, colitis, hepatitis, pancreatitis, renal and urinary tract infections (Skirow & Blaser, 2000).

1.5.1 Guillain-Barré Syndrome

The most serious complication caused by *C. jejuni* infection is Guillain-Barré syndrome (GBS) with indications that *C. jejuni* infections precede GBS in 20 to 50% of cases reported in Europe, North and South America, Japan and Australia; while in countries such as China and Bangladesh the frequency can be even higher (Jacobs *et al.*, 2008). In general, one in three GBS patients have suffered from a preceding *C. jejuni* infection (Hughes *et al.*, 1999, Nachamkin *et al.*, 2000a). GBS is the most common cause of acute neuromuscular paralysis, whereby the body's own immune system causes acute demyelination of the nerves in the peripheral nervous system (Nachamkin *et al.*, 1998). Structural mimicry between the microbial pathogen surface molecules and particular host antigens (Figure 1.3) is the main factor in the development of this disease (Ang *et al.*, 2003, Godschalk *et al.*, 2004, Hughes *et al.*, 1999, Tsang, 2002).

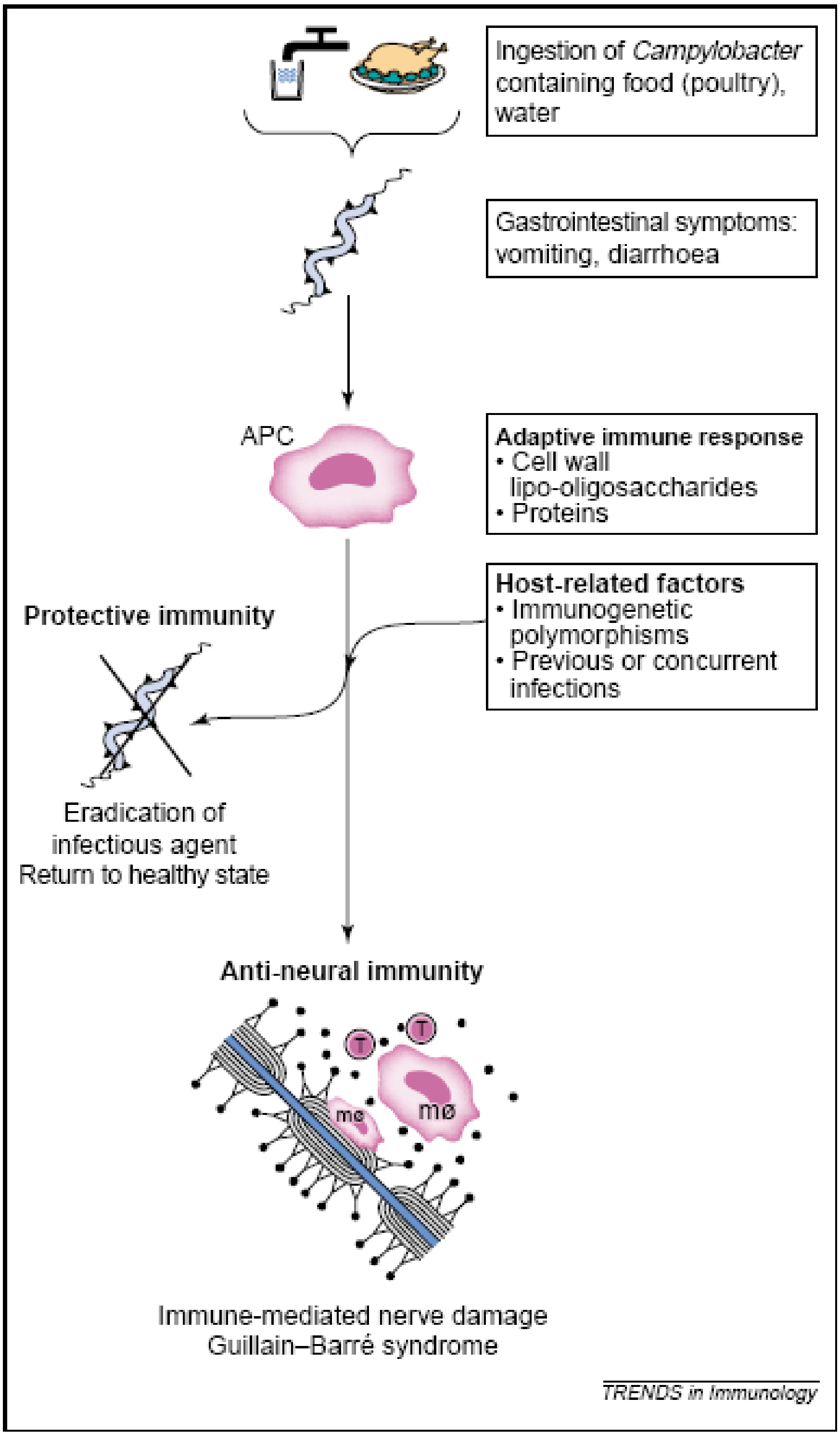


Figure 1.3 Presumed role of molecular mimicry in the Gillain-Barre syndrome (Ang *et al.*, 2003)

Analysis of serum taken from GBS patients in the acute phase of the disease shows a notable presence of antibodies against gangliosides, major constituents of the nerve cell membrane (Walsh et al., 1991), which are sialic acid-containing glycolipids implicated in cell growth and differentiation and signal transduction (Hughes *et al.*, 1999). The main reason for molecular mimicry is the presence of sialic acid in the lipo-oligosaccharides (LOS) structures of some C. jejuni isolates, which mimic the human gangliosides (Figure 1.4) (Yuki *et al.*, 1993, Ang *et al.*, 2003). Further evidence in support of molecular mimicry comes from studies showing that anti-ganglioside antibodies from GBS patients recognise C. jejuni LOS (Ang *et al.*, 2003).

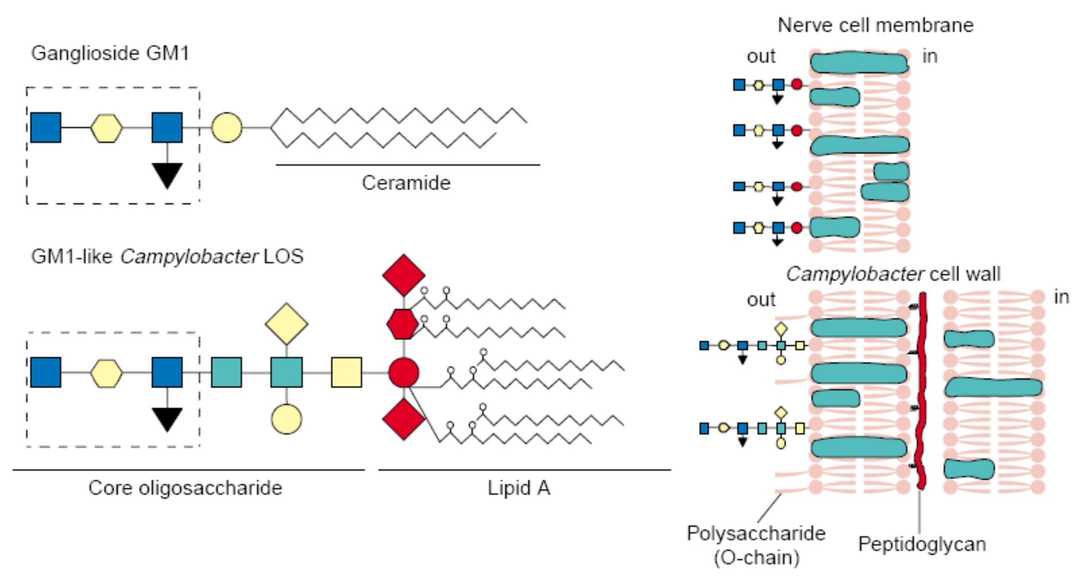


Figure 1.4 Molecular mimicry of gangliosides and *Campylobacter jejuni* lipooligosaccharides (LOS) (Ang *et al.*, 2003)

Persons affected by GBS rapidly develop weakness of the limbs and the respiratory muscles and areflexia. Although most people do not require any treatment, a significant proportion of patients may require mechanical ventilation and an estimated 15 to 20% may exhibit a severe neurologic deficit (Nachamkin *et al.*, 2000a). Mortality rates of GBS have been reduced to 2 to 3% in the developed world due to adequate medical support, but remain higher in the developing world (Nachamkin *et al.*, 2000a).

1.6 Treatment and Antibiotic resistance

The gastroenteritis caused by *C. jejuni* is a self-limiting disease and most patients recover without any specific treatment (Skirrow & Blaser, 2000). The most common form of treatment is replacing fluid and electrolytes (Piddock *et al.*, 2000). In particularly severe cases, fluids are delivered intravenously for rapid volume expansion, but for mild cases re-hydration is performed orally (Altekruse *et al.*, 1999). Antibiotic therapy is recommended for patients acutely ill with enteritis, having persistent fever and bloody diarrhoea or history of diarrhoea for more than 7 days (Butzler, 2004). Immunocompromised individuals are also encouraged to have antibiotic treatment due to their already compromised immune system (Butzler, 2004).

Antimicrobial chemotherapy has traditionally involved treatment with erythromycin and ciprofloxacin. Resistance of *C. jejuni* to a number of antibiotics such as tetracycline, erythromycin, ciprofloxacin, kanamycin and chloramphenicol has been reported (Piddock *et al.*, 2000, Pratt & Korolik, 2005, Gaudreau & Gilbert, 2003, Taylor *et al.*, 1987). The rate of resistance is rapidly increasing in both developed and developing countries; and world travel has increased the spread of drug resistant strains of the bacteria. The use of antimicrobial drugs in poultry farming has also greatly increased resistance to antimicrobial drugs (Butzler, 2004). The increase in multi-antibiotic resistant strains of *C. jejuni* places an increased pressure on the medical system and challenges the current treatment regimes. The best treatment still remains the prevention of infection through education and better hygiene which have far greater roles in reducing infections than active treatment with antibiotics.

1.7 Pathogenesis

The association of *C. jejuni* with human enteric disease has increased interest in understanding the general clinical, microbiological and epidemiological aspects of infection. The molecular mechanisms involved in pathogenesis, however, are still poorly understood. The factors or virulence determinants required to establish an infection are multifactorial in nature and campylobacters are no exception to this. Few of the determinants involved in *Campylobacter* pathogenesis are known or have a proven role. Virulence determinants involved in pathogenesis include motility and chemotaxis; adherence and invasion; toxins, surface polysaccharide structures and flagella, all of which appear to be required for colonisation of the gut mucosa (Ketley, 1997).

1.7.1 Motility and Chemotaxis

Campylobacter species are highly motile by means of a single polar flagellum. The direction of the flagellum is controlled by the response regulatory transduction pathway called chemotaxis. Chemotaxis is the ability of an organism to detect and move up and down chemical gradients. Both motility and chemotaxis are essential for *C. jejuni* colonisation (Korolik & Ketley, 2008). This was shown by the creation of non-chemotactic mutants which were unable to colonise the intestine in animal models (Takata *et al.*, 1992), while aflagellated mutants showed decreased colonisation when compared to the wild type strain (Nachamkin *et al.*, 1993b)

The flagellum of *C. jejuni*, involved in chemotaxis process, consists of an unsheathed polymer of flagellin subunits, which are encoded by the adjacent *fla*A and *fla*B genes (Nuijten *et al.*, 1995). The *fla*A and *fla*B genes show a very high degree of sequence identity (95%), however they exhibit both antigenic and phase variation (Nuijten *et al.*, 1995). Flagellin amino acid sequences show considerable interstrain diversity, especially in the central region of the FlaA flagellin proteins (Penn, 2001), which may account for differences in the colonisation potential of different strains.

1.7.2 Bacteria-Host Interactions

Adherence and invasion are key determinants in bacterial pathogenesis and are crucial for bacterial survival and subsequent development of disease. Adherence ensures initial contact between bacteria and the host surface. Invasion provides a mechanism of survival and protection from the host immune system. Upon infection, *C. jejuni* crosses the mucus layer covering the epithelial cells and adheres to these cells, with a subsequent co-population of these bacteria invading the epithelial cells (Vliet & Ketley, 2001)

1.7.2.1 Adherence

The ability of *C. jejuni* to bind to the cells lining the gastrointestinal tract is essential for the development of *C. jejuni* enteritis since it prevents the organism from being swept away by mechanical cleansing forces. Upon infection, *C. jejuni* crosses the mucous layer covering the epithelial cells and attaches to these cells (Konkel *et al.*, 2000). *In vitro* adherence assays have been extensively used to determine the factors that mediate bacterial binding to these cells (Ketley, 1997). It is evident that *C. jejuni* strains isolated from patients with fever and diarrhoea adhere to cultured cells at higher levels than strains isolated from individuals without these symptoms (Konkel *et al.*, 2000). These finding may suggest that different strains utilise various adhesion molecules in the adherence process.

One of the first structures implicated in *C. jejuni* adherence is the flagellum. Wassenaar *et al.* reported that genetically defined *C. jejuni flaA* and *flaB* mutants adhered to host cells at levels lower to those of their isogenic wild-type counterparts (Wassenaar *et al.*, 1991). These results were later confirmed by Yao *et al*. with a 50-fold reduction of adherence of isogenic mutants (Yao *et al.*, 1994). Furthermore, the tips of the flagellin were observed in contact with cells in scanning electron microscopy examinations of *C. jejuni* infected cells (Konkel *et al.*, 2000) confirming a definite interaction between flagella and the host cells. Reduced adherence, rather than complete absence of adherence in these experiments suggests the presence of other adhesion molecules involved in the initial interactions between bacteria and host cells.

The early work of McSweegan and Walker (Konkel *et al.*, 2000) proposed the role of Lipopolysaccharides (LPS) as the mediator of adherence to host cells. LPS are the major components of the outer membrane of many Gram-negative bacteria including *C. jejuni*. Identification of the genes involved in LPS biosynthesis has opened new perspectives in examining the potential role of these structures in adherence by means of defined LPS mutants (Fry *et al.*, 1998). The inactivation of the *galE* gene, encoding UDP-glucose 4-epimerase, an enzyme involved in LPS biosynthesis, significantly reduced *C. jejuni* adherence (Fry *et al.*, 2000) confirming that the LPS play an important role in bacterial adherence.

Additionally, significant advances have been made in the characterisation of the outer membrane proteins that mediate the binding of *C. jejuni* to host cells. A vast number of candidate proteins have been identified. However, the roles and the mechanisms of action of these factors need to be further characterised before making any conclusive statements. One protein, suggested important in adherence was PEB1, identified by Pei *et al*., which shows homology to the binding components of other Gram-negative bacteria (Pei & Blaser, 1993). Subsequent mutation of the gene caused 50 to 100-fold reduction in adherence compared to wild type bacteria (Pei *et al.*, 1998).

In addition to investigations of *C. jejuni* adhesins, a great deal of attention has focused over the past few years in the search for host cell structures involved in the initial contact with bacterial adhesins. Fibronectin, an extracellular matrix protein was found to be an attachment site for *C. jejuni*. Binding of *C. jejuni* cells to fibronectin was characterised by Konkel *et al*. who found that fibronectin specifically binds to *C. jejuni* 37 kDa outer membrane protein (Konkel *et al.*, 1997). Subsequent cloning and mutation of the gene encoding this protein (termed cadF) confirmed the role of the protein in adherence, as the isogenic mutant bacteria of this gene could not any longer adhere to fibronectin at wild type levels. In addition to fibronectin, *C. jejuni* also binds to wide variety of lipids including phosphatidylethanolamine, phosphatidylcholine, phospatidylinositol, phosphatidylserine, phosphatidylglycerol and sphingomyelin (Konkel *et al.*, 2000).

1.7.2.2 Invasion

Crossing of the epithelial mucosa is considered to be an essential virulence mechanism of several pathogenic bacteria, including *Salmonella, Shigella, Yersinia* and *Listeria* (Hu & Kopecko, 2000). The results of studies of *Campylobacter* infection in infant animals suggest that invasion is also a key component of *Campylobacter* pathogenesis (as reviewed in Hu & Kopecko, 2000. A variety of bacterial and host factors are involved in the process of invasion, none of which appear to be more important than the other. The pathways, the mechanisms and the functions of many of these factors are currently being investigated by different research groups around the world.

The current knowledge in the field recognises the ability of *C. jejuni* to invade different host cells. The level of invasion, however, differs between different strains and also depends on the host cells used in the experiments (Everest *et al.*, 1992, Biswas *et al.*, 2003, Monteville & Konkel, 2002). It appears that different strains require different host cell cytoskeleton structures. Different groups of authors stress the importance of flagella and motility of *C. jejuni* for successful invasion (Grant *et al.*, 1993). In addition, current knowledge suggests the importance of temperature and iron availability (Hu & Kopecko, 2000). The majority of work done on invasion in *C. jejuni* has been done with 81-176 strain, strain originally isolated from patients with inflammatory diarrhoea. However, the difference in invasion between strains and their preference to certain cell lines that is observed in different studies suggests the complexity of *C. jejuni* and multiple factors associated in invasion process. The current model of invasion (Figure 1.5) suggests a microtubule-dependent entry into the intestinal epithelium with subsequent translocation to the basolateral surface of the cells and important role of CDT toxin (Kopecko *et al.,* 2001).

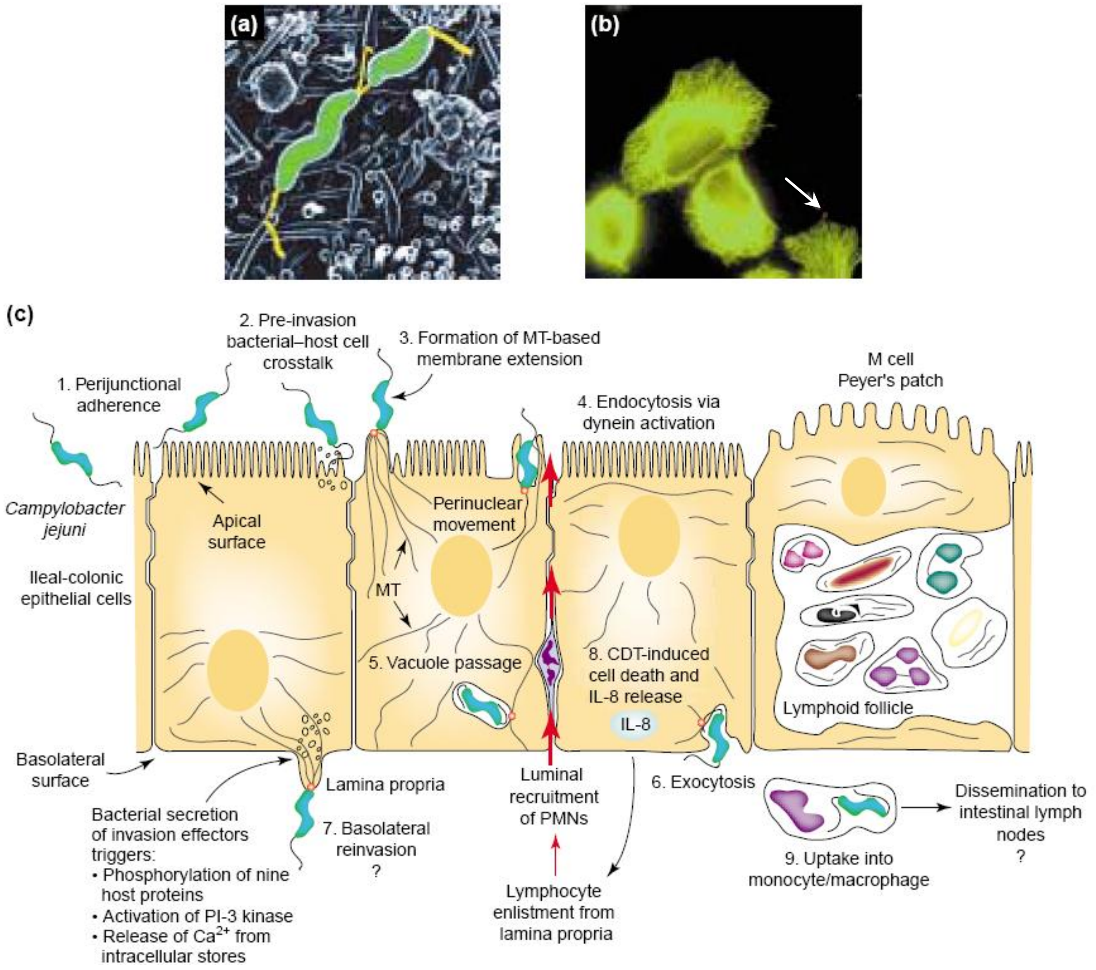


Figure 1.5 Combined microscopic images and schematic illustration demonstrating the current working model of *Campylobacter jejuni* pathogenesis (Kopecko *et al.*, 2001);

Legend: **(a)** and **(b)** have been enhanced and coloured. **(a)** Scanning electron micrograph looking down on apical surface of infected Caco-2 cells containing two *C. jejuni* (green) bound via flagella (yellow) to the tips of microvilli. **(b)** Immunofluorescence micrograph of differentially labelled, infected INT407 cells with microtubules (MTs) labelled with FITC (green) and *C. jejuni* labelled with Fast Red. During the early stages of infection, the host cell forms finger-like membrane protrusions that meet the adjacent bacteria (see arrows). **(c)** Schematic diagram illustrating the overall steps in *C. jejuni* gut invasion and pathogenesis.

1.7.3 Toxins

One important mechanism by which bacterial enteropathogens induce diarrhoea is through the production of potent toxins, which either damage membranes or act intracellularly (Wallis, 1994). *Campylobacters* reportedly produce a variety of toxins including a cholera-like toxin (CLT) as the genes encoding CLT have been identified; however, *in vitro* production has not been reported. In addition, several other cytotoxins were reported (Wassenaar, 1997), however the mechanism of action of these toxins is still not fully understood and their proposed mechanisms of action is based on comparison with other similar toxins.

Knowledge regarding one *Campylobacter* toxin has dramatically increased over the past few years; cytolethal distending toxin (CDT) was first described by Johnson and Lior as CLDT (as reviewed (Wassenaar, 1997). The toxin caused elevated intracellular cAMP levels and major elongation of cells followed by cell death. In 1996 Picket *et al*. (Pickett *et al.*, 1996) reported the isolation and characterisation of the *cdt* genes from *C. jejuni*, the first and until now only *C. jejuni* toxin-encoding genes. CDT activity is encoded by three genes named *cdt*A, *cdt*B, and *cdt*C that show similarity with *E. coli cdt* genes. *cdt* genes were also analysed in three other *Campylobacter* species; *C. coli* strain RM2228, *C. lari* strain RM2100 and *C. upsaliensis* strain RM3195. The study shows that the *cdt*A, *cdt*B and *cdt*C genes are conserved across the four *Campylobacter* species examined (Fouts *et al.*, 2005).

Although all *C. jejuni* and *C. coli* strains tested contain the *cdt* genes, there is a profound variation in CDT titres (Pickett *et al.*, 1996), and this variation is yet to be explained. The role of CDT in *C. jejuni* pathogenesis is still not fully elucidated. The most suitable method of determining a contribution of CDT to diarrhoeal disease would be to use an established animal model for *C. jejuni* diseases. So far the only study comparing CDT positive and CDT negative *C. jejuni* strains demonstrated that CDT(+) *C. jejuni* strains adhere to and invade epithelial cells more efficiently than CDT(-) strains. In addition these studies also found that CDT is responsible for the typical intestinal pathology of the colon in a murine model of *C. jejuni* infection (Jain *et al.*, 2008).

1.8 Mucins and their role in pathogenesis of gastrointestinal bacteria

Intestinal mucins, the major protein component of the mucous covering the epithelium of the gastrointestinal tract, are highly glycosylated macromolecules distinguished by the presence of dense *O*-glycosylation on the amino acids serine and threonine (Robbe *et al.*, 2004). Some of the functions of these structures include lubrication and modulation of water and electrolyte absorption. In addition, mucins play an important role in protecting the underlying epithelium from mechanical and chemical stress; and may also provide attachment sites for commensal and pathogenic microbes (Robbe *et al.*, 2004). Constant regeneration of mucous layer also presents a challenge for pathogenic bacteria.

One of the ways pathogenic bacteria as well as bacteria of normal flora, have solved the problem of mucus lining is degradation of the mucin layer. Namely, production of mucinases, enzymes capable of degrading mucins enables some of bacteria to cross this barrier and invade host cells beneath this layer (Robbe *et al.*, 2004). Variety of different proteases, sulphatases, sialidases, glycoproteases, neuraminidases all exert some effect on mucins. Unmasking of these glycoprotein structures by removal of terminal sugars may increase the binding opportunities and invasion of bacteria.

Intestinal mucins have long been recognised as a chemoattractant to *C. jejuni* (Hugdahl *et al.*, 1988). In addition, they have been implicated in *C. jejuni* (De Melo & Pechère, 1988) and *C. upsaliensis* (Sylvester *et al.*, 1996) adhesion and internalisation in tissue culture cells. The transmembrane mucin-1 (MUC1) and secretory mucin-2 (MUC2) are two well characterised constituents of intestinal mucus (Dekker *et al.*, 2002). MUC1 is speculated to have a protective role in *C. jejuni* colonisation as the levels of expression increased in mice after oral challenge with *C. jejuni* (McAuley *et al.*, 2007). Bacteria were found in the spleen, lung and liver of most Muc1$^{-/-}$ mice (strain 129/SvJ), but not in the wild-type mice, indicating that MUC1 contributes to innate defence against *C. jejuni* in mice (McAuley *et al.*, 2007). MUC2, on the other hand, is the major secretory mucin in the intestine and may account for the chemoattractant property of intestinal mucus (Tu *et al.*, 2008). In addition, it may also trigger the upregulation of the genes associated with virulence and invasion, as well as putative mucin-degrading enzymes (Tu *et al.*, 2008).

One of relatively recent discoveries of such an enzyme is a sialoglycoprotease from *Pasteurella (Mannheimia) haemolytica*, the bacteria that cause bovine pneumonic pasteurellosis (Abdullah *et al.*, 1991).

1.9 *O*-sialoglycoprotease

The *O*-sialoglycoprotease was first discovered in the culture supernatant of *M. haemolytica* A1 (Otulakowski *et al.*, 1983), the principal microorganism associated with bovine pneumonic pasteurellosis and the major cause of sickness and death in cattle in North America (Whiteley *et al.*, 1992). The glycoprotease of *M. haemolytica* is highly specific for O-linked, but not N-linked sialoterminal oligosaccharides of glycoproteins such as the sialylated membrane glycoprotein, glycophorin A of human red blood cells (Abdullah *et al.*, 1992, Mellors & Jiang, 1998). The enzyme has a marked specificity for human CD34 (an antigen expressed on panhematopoietic stem cells in the bone marrow), human CD43 (a sialomucin that has been implicated in immune and human cell function and cell-signalling phenomena) and human CD44 (Sutherland *et al.*, 1992). In addition, Hu *et al*. have demonstrated that the enzyme degrades epitectin and other mucin-type sialoglycoproteins (Hu *et al.*, 1994).

It has been shown that a number of *P. haemolytica* glycoprotease-susceptible substrates become resistant to cleavage by the glycoprotease if they are first treated with sialidase form *Clostridium perfringens* or *Vibrio cholerae* (Sutherland *et al.*, 1992). Thus, the proteolysis seems to be dependent on the presence of terminal sialyl residues. Abolition or reduction of the glycoprotease activity by desialylation has been shown for glycophorin A (Abdullah *et al.*, 1992), CD34 and CD44 (Sutherland *et al.*, 1992).

The predicted amino acid sequence for the *O*-sialoglycoprotein endopeptidase was thought to contain a putative metal-binding site, namely His^{110}-His-Met-Glu-Gly-His. This site is conserved in the homologous open reading frames found in *Haemophilus influenzae* (Lo *et al.*, 1994) and *E. coli* (Nesin *et al.*, 1987). This site was originally proposed to be a zinc ion binding domain (Abdullah *et al.*, 1991), but subsequent analysis of 12 types of zinc ion binding motifs (Hooper, 1994) has shown that the *O*-sialoglycoprotein endopeptidase putative metal-ion binding site does not fall into any of the known categories, since it shows a HHMEXH pattern rather than the general HEXXH sequence for Zn^{2+} binding. Although many metal ions have been tested for their ability to reactivate EDTA-inactivated and dialyzed *O*-sialoglycoprotein endopeptidase, no metal ion activator has yet been found. The glycoprotease may contain a tightly-bound Zn^{2+}, Ni^{2+}, Co^{2+} ions, or the transition metals such as Hg^{2+} and Cu^{2+} which commonly inactivate enzymes, may displace an essential metal ion activator from its binding site. The predicted amino acid sequence for the glycoprotease includes four cysteine residues, and the sensitivity of the glycoprotease to metal ions could be due to interactions with the cysteine thiols, though no effects of thiol-protecting reagents such as dithiothreitol has been detected (Otulakowski *et al.*, 1983).

A homologue of the enzyme has been identified in every whole genome sequenced so far; and in all kingdoms of life. Enzymatic activity has, however, only been demonstrated *in vitro* for the protein from *M. haemolytica*, but the biological function of the enzyme is still unknown. Glycoproteases appear to have a variety of functions. Calves vaccinated with *M. haemolytica* sialoglycoprotease expressed in *E. coli* had a significantly lower percentage of pneumonic tissue necropsy than unvaccinated controls (Shewen *et al.,* 2003), suggesting that the enzyme has a role in the induction of protective immunity. Furthermore, analysis of the bovine sera from calves vaccinated with the live *M. haemolytica*, show the presence of anti-Gcp antibodies, suggesting that the sialoglycoprotease is immunogenic and that the bacterium produces the enzyme *in vivo* (Lee *et al.,* 1994). On the other hand, the Gcp homolog in *E. coli,* ygjY (56% similarity) has been shown, both by promoter analysis and Northern (RNA) blot analysis, to be expressed as an intracellular protein possibly involved in the regulation of the *rpsUdnaG-rpsD* macromolecular synthesis operon (Nesin *et al.*, 1987).

A recent publication by Katz *et al*. has provided an insight into the role and function of the glycoprotease encoded by the *ygjY* gene in *E. coli.* Using the gcp depletion approach, the study performed by Katz *et al*. showed the importance of the enzyme in the metabolism of glycated proteins (Amadori-modified proteins (AMPs) and advanced glycated end products (AGEs)). AMPs are the products of non-enzymatic glycation formed by the reaction of reducing sugars with primary amine-containing amino acids in proteins (Horvat & Jakas, 2004). These glycated proteins formed via a multistep reaction called an Amadori rearrangement can further develop into irreversible, highly stable compounds known as AGEs (Singh *et al.*, 2001). The experiments performed by Katz *et al.* show that the depletion of gcp results in accumulation of AMPs, which serve as the potential for the development of AGEs. This finding suggests that Gcp is involved in Amadori product neutralisation by protein degradation (Katz *et al.*, 2010). Though the enzymatic activity of the *E. coli* enzyme has not been examined, these findings suggest that the enzyme in *E. coli* does not have a specificity for the *O*-sialoglycoproteins observed in *M. haemolytica* gcp (Abdullah *et al.*, 1992).

Another homologous open reading frame of unknown function has been found on chromosome IV of *Saccharomyces cerevisiae* where it is associated with genes essential for cellular function (Simon *et al.*, 1994). However, in the cyanobacterium *Synechocystis* sp., mutation of the glycoprotease gene results in a reduction of salt tolerance and alters the pigmentation and cyanophycin accumulation (Zuther *et al.*, 1998). This study revealed an increase in the amount of inclusion bodies containing the high-molecular-weight nitrogen storage polymer cyanophycin (polyaspartate and arginine). Cyanophycin accumulation was hypothesised to be caused by inactivation of the putative protease encoded by the *gcp* gene which was responsible for cyanophycin degradation in the *Synechocystis* (Zuther *et al.*, 1998).

In *Staphylococcus aureus*, a bacterium generating enormous public health concerns, a regulated gene expression approach demonstrated that the putative glycoprotease is essential for staphylococcal growth in culture (Zheng *et al.*, 2005), though the biological function of the enzyme could not be assigned. Gcp was demonstrated to be a critical mediator involved in the modification of cell wall biosynthesis through modification of cell wall peptidoglycans (Zheng *et al.*, 2007). The gcp was also suggested to be involved in the regulation of expression and/or activity of some murein hydrolases associated with the modification of cell wall peptidoglycan synthesis. The consequences of down-regulating *gcp* expression include increased bacterial tolerance to detergent-, penicillin-, and vancomycin-induced lysis (Zheng *et al.*, 2007), which, in turn, may have played an important role in bacterial viability (Zheng *et al.*, 2005). These data indicate that Gcp is likely associated with extracellular hydrolase activity and possibly the posttranslational regulation of these hydrolases, confirming the findings by Otulakowski *et al*. that Gcp is a secreted enzyme (Otulakowski *et al.*, 1983).

The sequence search of the *C. jejuni* genome strain 11168 has revealed a gene (*Cj1344c*) with similarity to the *M. haemolytica* sialoglycoprotease gene (this study). In the study of the *C. jejuni* response to human mucin MUC2, the expression of *Cj1344c* was upregulated suggesting a putative role for the enzyme in MUC2 degradation and its biological function in the gastrointestinal tract (Tu *et al.*, 2008). As complex mucin structures made of high molecular weight glycoproteins are the main constituent of this mucous layer of gastrointestinal tract (Robbe *et al.*, 2004), it can be speculated that the ability of *C. jejuni* to attach to these structures is beneficial for their survival in this hostile environment. Sylvester *et al*. showed that *C. upsaliensis*, a bacterium closely related to *C. jejuni,* is capable of binding to gangliotetraosylceramide (Gg_4), a structure derived from the cleavage of sialic acid residues of membrane GM_1 gangliosides by enzymes produced by normal bacterial flora (Sylvester *et al.*, 1996). Hypothetically, an enzyme able to remove or cleave glycoproteins, the main constituents of mucous layer, may expose different glycoprotein structures and thus enhance attachment and subsequent internalisation of the pathogen. Grys *et al.* have demonstrated that another metalloprotease from enterohemorrhagic *E. coli*, StcE, cleaves glycoprotein 340 and mucin 7. In addition, adherence studies performed with the *stcE* mutant showed a difference in intimate adherence, allowing a closer interaction between the bacterium and host cell (Grys *et al.*, 2005).

1.10 Development of a *C. jejuni* vaccine

Natural immunity in children and data obtained from volunteer challenge studies, suggest development of an effective *Campylobacter* vaccine is feasible. In the developing world, acquisition of immunity during the first 2 years of life has been shown to be accompanied by raising titres of *Campylobacter*-specific antibodies (Blaser *et al.*, 1985). Similarly adult volunteers challenged with *C. jejuni* developed serum and intestinal antibodies and were protected from the subsequent illness, but not against colonisation, following challenge (Black *et al.*, 1988).

Development of vaccine against *C. jeju*ni is both, necessary and desirable but may be complicated due to a number of factors that include the tremendous antigenic diversity of the organism, a lack of understanding of the nature of acquired immunity, a lack of small animal models suitable for vaccine evaluation, as well as the fact that the protective epitopes are not clearly defined (Scot, 1997). Several antigens have been suggested and tested as vaccine candidates. The importance of flagellin in *C. jejuni* pathogenesis and high immunogenicity have suggested its use as a good vaccine candidate (Martin *et al.*, 1989, Hu & Kopecko, 1999). Flagellin-based vaccines have been assessed for their role to prevent *C. jejuni* infections (Lee *et al.*, 1999) however the variability of the flagellin proteins among *C. jejuni* strains due to highly variable regions and posttranslational modifications, as well as a low immunogenic response or recombinant flagellin presented problems and hindered further development of the vaccine. Different flagellar proteins were therefore assessed for their suitability as *C. jejuni* vaccine candidates.

Many Gram-negative pathogens use the type III secretion apparatus (TTSS) for transporting effector proteins to eukaryotic cells (Coburn *et al.*, 2007). However, genes encoding structural proteins of TTSS were not found in the genomes of *Campylobacter* strains. In this bacterium, protein secretion, including secretion of several significant virulence factors, depends on the flagellar apparatus (Konkel *et al.*, 2004). Among proteins secreted by this apparatus are FlaC, FspA1 and FspA2, whose immunogenicity and induced protective effect were tested in a mouse model (Baqar *et al.*, 2008). Proteins purified from *E. coli* cells were intranasally administered to BALB/c mice, alone or with LT $_{R192G}$ as an adjuvant. Although all three proteins were observed to be highly immunogenic, immunization did not provide a high protection level (Baqar *et al.*, 2008).

The new approach to prevent human campylobacteriosis is by the introduction of a glycoconjugated vaccine. The first conjugated vaccine was developed against *Heamophilus influenzae* type b (Hib). Later, conjugated vaccines that offer protection against selected serotypes of *Neisseria meningitidis* and *Streptococcus pneumoniae* were introduced for routine infant immunization (Trotter *et al.*, 2008). The Capsular Polysaccharides (CPS)-conjugated vaccine was tested in mice and in the New World monkey, *Aotus nancymae* and was found to induce significant serum immune response and protection. $CPS_{81\text{-}176}$ -CRM_{197} (*C. jejuni* 81-176 CPS conjugated to diphtheria toxin mutant) immunization resulted in significant specific serum IgM, IgA and IgG responses, and a reduction in the illness symptoms in mice after intranasal challenge with the homologous strain (Monteiro *et al.*, 2009). Although vaccination of *A. nancymae*, does not induce serum IgA, it protects animals against diarrhoea but not against intestinal colonization after challenge with the homologous strain (Monteiro *et al.*, 2009). Although preliminary experiments on CPS-conjugated vaccination provided promising results, many important issues concerning this type of vaccine still await clarification. These include surveillance of serotypes distribution, analysis of the capsular-switching phenomenon, cross-reactivity and safety concerns.

A lack of defined epitopes and antigenic complexity among *C. jejuni* strains may suggest the development of a live attenuated vaccine or killed *Campylobacter* whole cell (CWC) vaccine as an answer to the problem. Challenge studies have shown that infection with a wild-type strain produce solid protective immunity in volunteers, so it is reasonable to expect that a live attenuated vaccine or killed CWC could produce same results. However, the paucity of information regarding the pathogenesis complicates this approach. In addition, the association between *C. jejuni* infection and GBS makes the development of the attenuated or killed CWC vaccine very difficult.

Sequencing the genomes of many *Campylobacter* strains as well as the development of transcriptomic and proteomic techniques have allowed a faster and more efficient identification of antigens that are potential candidates for vaccine development. Identification of *O*-sialoglycoprotease homologue (Cj1344c) within *C. jejuni* genome and the high level of conservation between *Campylobacter* strains initiated studies to determine the role of this enzyme in *C. jejuni* pathogenesis. The successful use of the *M. haemolytica O*-sialoglycoprotease in vaccination trials to prevent bovine pneumonic pasteurellosis (Shewen *et al.*, 2003), also opened possibilities of using the *C. jejuni* glycoprotease in vaccination trials to combat diarrhoeal illness. The aims of this study were to examine the role of putative glycoprotease in *C. jejuni* pathogenesis and determine its action; and secondly to examine the possibility of using the glycoprotease in vaccination trials to eliminate *C. jejuni* colonisation in mice.

CHAPTER 2

Materials and Methods

2.1 General procedures

Media, glassware, and general solutions were sterilised by autoclaving at 121 ˚C for 20 minutes (1,000 kPa), unless otherwise specified. Eppendorf pipettes were used to measure volumes ranging from 0.1 µL to 10 mL. Pipette tips were sterilized at 121 ˚C for 20 minutes (1,000 kPa).

2.2 General Chemicals and Equipment

Chemicals used in this study were of analytical grade and were purchased from Sigma Chemical Co., Chem-Supply, Oxoid, Merck, Amresco, Fluka, Bio-Rad, Lancaster, APS Finechem or Applichem. Unless specified, chemicals were made up to the desired concentration in MilliQ purified water and sterilized by filter sterilization or autoclaving at 121 ˚C for 20 minutes (1,000 kPa). Enzymes used in this study were purchased from New England Biolabs, Promega or Roche. Plastic ware was purchased from Becton Dickson or Eppendorf, with the exception of plastic ware required for RNA which was acquired from Axygen.

Centrifuges used were: Beckman Coulter Allegra 25R with rotors TA-14-50, TA 10.250, Sigma 1-15 bench top microcentrifuge and Sigma 3-16 refrigerated benchtop centrifuge.

2.3 Bacteriological Techniques

2.3.1 Media

Horse Blood Agar: 39 g/L Columbia agar base in deionised water. Autoclaved at 121 ˚C for 15 minutes (1,000 kPa), cooled to 50-55 ˚C, 5% sterile defibrinated horse blood (IMVS) added. Campylobacter selective supplements; Skirrow also added (Oxoid) which comprises of 5 mg vancomycin, 2.5 mg trimethoprim and 1250 IU Polymyxin B. Agar poured into sterile Petri dishes and stored at 4 ˚C. For 2% HBA plates, an additional 1% (w/v) Agar Bacteriological added prior to autoclaving.

Luria-Bertani (LB) agar: 25 g/L Luria Broth Base (Oxoid) and 12 g/L Bacteriological Agar in deionised water. Autoclaved at 121 ˚C for 20 minutes (1,000 kPa), cooled to 50-55 ˚C, antibiotics or other supplements if required were added, mixed and poured into sterile Petri dishes. Stored at 4 ˚C.

Luria-Bertani (LB) broth: 25 g/L Luria Broth Base (Oxoid) in deionised water. Autoclaved at 121 ˚C for 20 minutes (1,000 kPa); cooled to room temperature, antibiotics or other supplements if required were added. Stored at 4 ˚C.

SOC media: 20 g/L bacto-tryptone, 5 g/L yeast extract, 10 mM NaCl and 2.5 mM KCl in distilled water and autoclaved at 121 ˚C for 20 minutes (1,000 kPa), with the addition of 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose (all filter sterilized with a 0.2 micron filter) when cooled.

Storage medium: 10% skim milk powder, 1% bacto-tryptone, 10 mM Tris-Cl pH 7.5. Autoclaved at 109 ˚C for 30 minutes.

X-gal agar: LB agar supplemented with 2% X-Gal and IPTG to a final concentration of 0.5 mM.

2.3.2 Antibiotics

Ampicillin: Stock solution of 50 mg/mL in distilled water, 0.22 µm filter sterilized.

Chloramphenicol: Stock solution of 25 mg/mL in 100% ethanol.

Kanamycin: Stock solution of 50 mg/mL in distilled water, 0.22 µm filter sterilized.

2.3.3 Stock solutions and Buffers

10X TBS: 24.2 g Tris base, 80 g NaCl. The pH adjusted to 7.6 with HCl. Made up to 1 L with deionised water.

1X SDS-PAGE sample buffer: 100 mM Tris-Cl pH 6.8, 2% SDS, 0.1% bromophenol blue, 10% glycerol, 5% β-mercaptoethanol

30% acrylamide: 29.2% acrylamide, 0.8% bis-acrylamide in distilled water. 0.2 µm filter sterilized and stored in the dark at 4 ˚C.

50X TAE: 242 g Tris base, 57.1 mL glacial acetic acid, 100 mL 0.5M EDTA pH 8.0, made up to 1 L with deionised water. Stored at room temperature in glass bottle.

5X TBE: 54 g Tris base, 27.5 g boric acid, 20 mL 0.5 M EDTA pH 8.0, made up to 1 L with deionised water.

6X Gel-loading sample buffer: 40% sucrose, 0.25% bromophenol blue, 0.25% xylene cyanol FF

Blocking buffer: 1X TBS, 0.1% Tween 20, 1% non-fat dry milk powder.

Coomassie brilliant blue stain: 0.1% brilliant blue in 50% methanol, 40% deionised water, 10% acetic acid

Coomassie brilliant blue destaining solution: 50% methanol, 40% water, 10% acetic acid

ELISA blocking buffer: 3% BSA in Tris-buffered saline, 0.05% Tween 20

Plasmid Mini Preparation Solution I: 50 mM glucose, 25 mM Tris, 10 mM EDTA pH 8.0 containing 10μg RNase

Plasmid Mini Preparation Solution II: 0.2 M NaOH, 1% [w/v] SDS

Plasmid Mini Preparation Solution III: 3 M potassium, 5 M acetate

Protein purification binding buffer: 50 mM $NaHPO_4$, 0.3 M NaCl, 0.5% Triton X-100, 10 mM imidazole, pH 7.0

Protein purification elution buffer: 50 mM $NaHPO_4$, 0.3 M NaCl, 500 mM imidazole, pH 7.0

Protein purification wash buffer: 50 mM $NaHPO_4$, 1.0 M NaCl, 20 mM β-mercaptoethanol, 50 mM imidazole, pH 7.0

Tbf1 buffer: 30 mM potassium acetate, 100 mM potassium chloride, 60 mM calcium chloride and 15% glycerol, pH adjusted to 5.8 with 0.2 M acetic acid and 0.2 µm filter sterilized

Tbf2 buffer: 10 mM MOPS, 75 mM Calcium chloride, 10 mM Potassium chloride, 15% glycerol. pH adjusted to 6.5 with 1 M KOH and 0.2 µm filter sterilized

TBS/Tween wash buffer: 1X TBS + 0.1% Tween20

TE buffer: 10 mM Tris-Cl and 1 mM EDTA, pH 8.0

Towbin buffer: 3 g Tris base, 14.4 g glycine, 800 mL MilliQ purified water, 200 mL methanol

Tris-glycine electrophoresis running buffer: 0.025 M Tris, 0.250 M glycine, 0.1% SDS

2.4 Bacterial strains and plasmids

2.4.1 *C. jejuni* strains and growth conditions

The *C. jejuni* strains used in this study are listed in Table 2.1. The *C. jejuni* NCTC11168 genome strain will be referred to as 11168-GS, and the *C. jejuni* NCTC11168 original strain, kindly donated by D.G Newell, London, will be referred to as 11168-O. The bacteria were grown on Columbia agar supplemented with 5 % defibrinated horse blood (HBA) with Skirrow antibiotic supplement (Oxoid) under microaerophilic conditions (5% O_2, 15% CO_2, 80% N_2; BOC gases) for 48 hours at 42 ˚C. *C. jejuni* was harvested from the agar plates in sterile Brucella Broth (Oxoid) and the CFU/mL was determined by measuring OD_{600} and comparing to a standard growth curve.

Table 2.1 *C. jejuni* strains used in this study

***C. jejuni* strain**	**Isolated from**	**Original Source**
11168-GS	Human	D.G. Newell, Centre of Veterinary Laboratories, London, UK
81116	Human	D.G. Newell, Centre of Veterinary Laboratories, London, UK
11168-O	Human	D.G. Newell, Centre of Veterinary Laboratories, London, UK

2.4.2 Preparation of conditioned media

"Conditioned" Brucella Broth to grow mutant *C. jejuni* bacteria was prepared by growing wild-type *C. jejuni* strain 11168 in Brucella Broth for 24 hours, removing the cells by centrifugation, and collecting and filtering the supernatant. The absence of viable cells in the supernatant was verified by plating it and observing no growth on the plates.

In addition, the Brucella Broth medium was supplemented with the *E. coli* expressed and purified His-Cj1344c protein to enhance the chance of mutant recovery.

2.4.3 Bacterial strains and growth conditions

Table 2.2 lists the *E. coli*, *S. aureus* and *H. pylori* strains used in this study. *E. coli* and *S. aureus* strains were cultured from frozen stocks onto LB agar plates (supplemented with ampicillin (100 µg/mL), kanamycin (50 µg/mL) or chloramphenicol (10-20 µg/mL) when required) and incubated at 37 ˚C for 18 hours under aerobic conditions. *H. pylori* and *C. coli* and *C. fetus* strains were grown on Columbia agar supplemented with 5% defibrinated horse blood (HBA) with Skirrow antibiotic supplement (Oxoid) under microaerophilic conditions (5% O_2, 15% CO_2, 80% N_2; BOC gases) at 42 ˚C for 72 and 48 hours respectively.

Table 2.2 Bacterial strains used in this study

Species	Strain	Genotype	Reference/source
E. coli	DH5α	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	Invitrogen
E. coli	XL1-blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac*	Stratagene
E. coli	BL21(DE3)	F^- *ompT* $hsdS_B$ (r_B^- m_B^-) *gal dcm* (DE3)	Novagen
H. pylori	26695		Clinical isolate
C. coli	18		Clinical isolate
C. fetus	82-40		Clinical isolate

2.4.4 Plasmids

Table 2.3 Plasmids used in this study

Plasmid	Features	Reference/source
pGEM-T Easy	amp^R, (100 µg/mL), blue/white screening	Promega
pET-19b	amp^R, (100 µg/mL) His-Tag	Novagen
pMW2	km^R, (50 µg/mL)	(Wosten, M)
pAV110	cat^R	(Ketley, J)
pGU0401	Km^R, *oxa-61*	(Alfredson, D)
pGU0501	amp^R, (100 µg/mL) *Cj1344c*	This study
pGU0501*Bgl*II	amp^R, (100 µg/mL) *Cj1344c::Bgl*II	This study
pGU0509	amp^R , km^R, promoted *amph(3')-III*	This study
pGU0513	amp^R, (100 µg/mL) His-Tag *Cj1344c*	This study
pGU0522	amp^R , km^R, promoted *amph(3')-III*, $P_{oxa\text{-}61}$	This study
pGU0523	amp^R , km^R, *Cj1344*c::promoted *amph(3')-III*, $P_{oxa\text{-}61}$	This study
pGU0706	amp^R , promoter-less *amph(3')-III*	This study
pGU0707	amp^R , *Cj1344c*::promoter-less *amph(3')-III*	This study
pGU0613	amp^R , km^R, *Cj1344c*::promoted *amph(3')-III*	This study
pGU0804	amp^R , promoter-less *cat*	This study
pGU0805	amp^R , *Cj1344c*::promoter-less *cat*	This study

2.4.5 Storage of bacterial strains

For **long term storage of *Campylobacter* and *Helicobacter*** strains, microorganisms were stored using either storage medium or 20% glycerol. When using storage medium, an overnight lawn culture on appropriate media was grown, 2 mL storage media was pipetted onto a lawn culture plate using aseptic technique and harvested using a sterile spreader. 1 mL of liquid culture was pipetted into a cryovial and placed at -80 ˚C. Storage media consisted of 10% (w/v) skim milk powder, 1% (w/v) tryptone and 10 mM Tris-Cl, pH 7.5, sterilised by autoclaving.

For **long term storage of *E. coli***, strains were grown overnight in LB media with the appropriate antibiotic. The cultures were spun down and then resuspended in sterile LB media containing 20% glycerol. 1 mL of the resuspension was added to cryovial and stored at -80 ˚C.

For **short term storage** *C. jejuni* strains were stored up to six weeks in 10 mL bottles containing semisolid (0.4%) agar at 37 ˚C. *E. coli* was stored for up to 4 weeks on LB agar plates at 4 ˚C.

2.5 General Molecular Biological Techniques

2.5.1 Crude DNA isolation

DNA from the bacteria was extracted using the crude boiling method, where a colony or few colonies were selected from the plate and placed into 100 μL of sterile water and boiled for 5 min. The cell debris was removed by centrifugation at 14,000*g* for 5 min, and the supernatant was stored at -20 ˚C or used immediately.

2.5.2 Purification of plasmid DNA from *E. coli*

Plasmid DNA was extracted from overnight cultures containing the appropriate antibiotic of *E. coli* using either the method outlined below or one of the following commercial extraction kits: Fast Plasmid mini prep kit (Eppendorf), PureLink HiPure mega prep kit (Invitrogen) or QIAGEN Plasmid midi prep kit (QIAGEN). Purifications were performed as per manufacturer's instructions with no modifications.

2.5.2.1 Alkaline lysis plasmid mini preparation

A plasmid DNA using the alkaline lysis method was prepared using the modified alkaline lysis plasmid mini preparation method outlined by Sambrook and associates (Sambrook & Russel, 2001). A single colony of *E. coli* containing the recombinant plasmid was inoculated into 1.5 mL of broth containing the appropriate antibiotic. The culture was incubated at 37 ˚C on an orbital shaker at 220 rpm for 18 hours. Cultures were pelleted by centrifugation at 14,000 rpm for 1 minute. Supernatant was discarded and pellet re-suspended in 200 μL **solution I** (50 mM glucose, 25 mM Tris, 10 mM EDTA pH 8.0) containing 10 µg RNase A by vortexing. 200 µL of **solution II** (0.2 M NaOH, 1% [w/v] SDS) was added and the tubes inverted 6-8 times. Samples were incubated on ice for 5 minutes. 150 µL of **solution III** (3 M potassium, 5 M acetate)) was added and the tubes were gently mixed for 10 seconds. Samples were incubated on ice for 10 minutes followed by a centrifugation at 14,000 rpm for 10 minutes to remove unlysed cells. Supernatant was transferred to a new tube and an equal volume of isopropanol was added. Tubes were mixed, incubated at room temperature for 5 minutes and subsequently centrifuged for 10 minutes at 14,000 rpm. Supernatant was removed and the pellet washed with 100 µL 70% ethanol. Tubes were centrifuged for 2 minutes at 14,000 rpm. Supernatant was carefully removed and the pellets air dried for 5-10 minutes. DNA was resuspended in 25-50 µL of sterile distilled water and stored at -20 ˚C.

2.5.3 Standard Ethanol Precipitation

Standard ethanol precipitation was performed without modification as outlined in Sambrook *et al.*, 2001.

2.5.4 Quantitation of DNA

Quantification of DNA in samples was performed by the spectrophotometric method using $A_{260/280}$. These methods are described in Sambrook *et al.* (2001) and used without modification.

2.5.5 Agarose gel electrophoresis

Agarose gel electrophoresis was performed for the size estimation of DNA fragments produced by PCR and restriction endonuclease digest products as well as estimation of DNA purity. Bio-Rad Power Pac 200, sub cell GT electrophoresis tanks and gel trays were utilized. 0.8%-2% (depending on the size of fragments to be discerned) biotechnology grade agarose was added to 1XTAE buffer and boiled in a microwave oven to dissolve the agarose. This was cooled to ~50 ˚C and 10 mg/mL ethidium bromide (Bio-Rad) was added to a final concentration of 0.5 µg/mL. Samples were loaded in 1X sample buffer and then electrophoresed at 100 V for 45 minutes. The DNA fragments were visualized with an UVP white/ultraviolet transilluminator and documented using BioRad Quantity One 1D Analysis Software. 100 bp and 1 kb DNA ladders from NEB were used as molecular weight markers.

2.5.6 Purification of DNA from agarose

DNA was electrophoresed through an appropriate percentage agarose gel containing ethidium bromide, then visualized with a UV transilluminator for excision of the desired fragment. The PerfectPrep Gel Clean-up (Eppendorf) kit was utilized for purification of DNA from the agarose. This method was used as per the manufacturer’s instructions without modification.

2.5.7 Restriction endonuclease digestion of DNA

Restriction endonuclease digests were carried out as per the manufacturer’s recommendation. Restriction endonuclease digestion was performed at a maximum volume of 50 µL. The enzyme concentration was 20 U/µg DNA with the appropriate buffer used, supplied by the manufacturer. The digestion reaction was supplemented with BSA, when necessary. The incubation temperature was 37 ˚C, for the time advised by the supplier of the endonuclease, unless specified otherwise. Where appropriate, the enzyme was heat inactivated.

2.5.8 Polymerase Chain Reaction Procedures

2.5.8.1 Primers used in this study

Oligonucleotide primers were synthesized by Invitrogen and received in a lyophilized form. They were re-suspended in sterile distilled water (100 pmol/µL) and stored at -20˚C. Working stocks were created by making a one in four dilution of the original stock. Primers used in this study are listed in Table 2.4.

Table 2.4 List of primers used in the study

Primer name	Primer sequence	Restriction site incorporated	Reference	Purpose
T7	5’ TAATACGACTCACTATAGGG 3’	None	Promega	PCR screening to identify insert DNA in cloning, sequencing
SP6	5’ TATTTAGGTGACACTATAG 3’	None	Promega	PCR screening to identify insert DNA in cloning, sequencing
sgcp-*Nde*I-F	5’ CATATGAAAAATCTTATCCTAGCTA 3’	*Nde*I	This study	Amplification of *Cj1344c* for cloning into pGEM-T Easy
sgcp-XhoI-R	5’ CTCGAGCTTTTTTCATCTATATCCTTG 3’	*Xho*I	This study	Amplification of *Cj1344c* for cloning into pGEM-T Easy
kana-*Bgl*II-F	5’ GAAGATCTGCTCGGAATTAACCCT 3’	*Bgl*II	This study	Amplification of *km*^{R} gene including the promoter sequence
kana-*Bgl*II-R	5’ GAAGATCTGCTCGGAATTAACCCT 3’	*Bgl*II	This study	Amplification of *km*^{R} gene including the promoter sequence
NP800-km-*Bgl*II-F	5’ GAAGATCTCATGGCTAAAATGAGAATATC 3’	*Bgl*II	This study	Amplification of *km*^{R} gene excluding the promoter sequence
ampprom-*Xba*I-F	5’ GCTCTAGACTTGATATCGAATTCCTGCAGCCC 3’	*Xba*I	This study	Amplification of *oxa-61* promoter sequence
ampprom-*Xba*I-R2	5’ GCTCTAGACACAAAATATCTTTCTATTTAAAT 3’	*Xba*I	This study	Amplification of *oxa-61* promoter sequence

Table 2.4 List of primers used in the study (continued)

Primer name	Primer sequence	Restriction site incorporated	Reference	Purpose
Cat-*Bgl*II-F	5’ AGATCTCATGATGCAATTCACAAAGATT 3’	*Bgl*II	This study	Amplification of *cm*^{R} gene excluding the promoter sequence
Cat-*Bgl*II-R	5’ AGATCTTTATTTATTCAGCAAGTCTTG 3’	*Bgl*II	This study	Amplification of *cm*^{R} gene excluding the promoter sequence
inv-sgcp-*Bgl*II-F	5’ AAGATCTGAGCTTTTAGCAAGTACAAATGATGATAGC 3’	*Bgl*II	This study	Creation of unique restriction enzyme site within *Cj1344c* for insertion of antibiotic resistance gene
inv-sgcp-*Bgl*II-R	5’ AAGATCTCCACCACTAACAAGCAAAATTCCCATATCT 3’	*Bgl*II	This study	Creation of unique restriction enzyme site within *Cj1344c* for insertion of antibiotic resistance gene
Cj1343c-R	5’ GGATCCTTACATTCCCCCTATTAAAAG 3’	*Bam*HI	This study	Screening for cross over event during *C. jejuni* transformations
Kana-intr-F	5’ TCCAAAGGTCCTGCACTTTGAACG 3’	-	This study	Screening for cross over event during *C. jejuni* transformations

2.5.8.2 DNA template preparation

DNA template for PCR reactions were performed by creating a crude lysis preparation (section 2.5.1). Generally 2-5 µL of crude lysate was used per 20 µL reaction. A water control was always included for the PCR reactions.

2.5.8.3 PCR cycling conditions

PCR conditions are described in Sambrook, *et al.*, 2001. Reaction temperatures and times were as follows, 94 ˚C for 1 min, 45-55 ˚C for 1 min and 68 ˚C for 2-5 min, for 35 cycles. All PCRs were carried out using an Eppendorf Mastercycler Personal.

2.5.9 DNA sequencing

Plasmid DNA products were sequenced using the ABI Big Dye Terminator Version 3.1, performed by the Australian Genome Research Facility, Brisbane. Purified DNA samples to be sequenced were sent with 6.4 pmol of the appropriate sequencing primer. DNA sequence chromatograms were visualized and sequences analysed using MacVector software (MacVector, Inc.).

2.6 General cloning techniques

2.6.1 Cloning DNA into the pGEM®-T Easy Vector

A general procedure was followed for cloning into the pGEM®-T Easy vector as per the manufacturer's instructions. The PCR product was ligated into pGEM®-T Easy and transformed into competent XL1-Blue or DH5α *E. coli* cells. Transformations were plated onto media containing X-gal/IPTG for blue/white screening and ampicillin. Transformants were screened for recombinant plasmid by PCR using the *T7* and *SP6* primers.

2.6.2 Dephosphorylation of plasmid DNA

To prevent plasmid re-ligation, the 5' phosphates were removed using either Shrimp Alkaline Phosphatase (Promega) or Antarctic Phosphatase (NEB). 1 U/µg DNA of enzyme was used in the reaction. Samples were incubated for 10-15 minutes at 37 ˚C. Enzyme was heat-killed at 65 ˚C for 20 minutes. Dephosphorylated vector was then used for ligation.

2.6.3 Ligation of DNA into a vector

DNA ligation was performed at a maximum volume of 10 µL. The ligation reactions contained 10 x Ligation Buffer (Roche), 50 ng linearised plasmid DNA, 20-75 ng insert DNA (depending on size of fragment), T4 DNA ligase (3 Weiss U/µL)(Roche). For maximum transformants, the reaction was incubated overnight at 16 ˚C.

2.6.4 Preparation of *E. coli* competent cells

The method used was a modification of the Inoue method (Inoue *et al.*, 1990), where a single, fresh *E. coli* colony was inoculated into 5 mL LB and incubated for 18 hours at 37 ˚C, 200 rpm. This primary culture was added to 95 mL warm LB and shaken at 37 ˚C for another 2.5-3 hours until the OD_{600} was 0.4-0.6. The flask was then placed on ice for 10 minutes to cool. Culture was aliquoted into chilled 50 mL centrifuge tubes and spun at 3,000 g for 10 minutes at 4 ˚C. Supernatant was discarded and the pellets re-suspended in a total of 40 mL ice cold **Tbf1 buffer** (30 mM potassium acetate, 100 mM potassium chloride, 60 mM calcium chloride and 15% glycerol, pH was adjusted to 5.8 with 0.2 M acetic acid and 0.2 µm filter sterilized). The cells were incubated on ice for 10 minutes and then centrifuged at 3,000 g for 10 minutes at 4 ˚C. The supernatant was discarded and pellets re-suspended and pooled in 4 mL ice cold **Tbf2 buffer** (10 mM MOPS, 75 mM Calcium chloride, 10 mM Potassium chloride, 15% glycerol. pH was adjusted to 6.5 with 1 M KOH and 0.2 µm filter sterilized.) The cells were left on ice for further 15 minutes and then stored in small aliquots -80 ˚C.

2.6.5 Transformation of competent cells

Transformations were carried out as per the pGEM®-T Easy vector manual (Promega). Plasmid DNA and 50 µL competent cells were aliquoted into a tube on ice and gently mixed. The reactions were incubated for 20 minutes on ice and then heat shocked for 45 seconds at 42 ˚C and immediately returned to ice for 2 min. 950 µL of SOC medium at room temperature was added and incubated for 1.5 hours at 37 ˚C on a shaker. 100µL of the transformation culture was plated onto LB plates containing ampicillin, IPTG and X-gal or appropriate antibiotic. The plates were incubated overnight at 37 ˚C and stored at 4 ˚C to facilitate blue/white screening.

2.6.6 Natural transformation of *C. jejuni*

C. jejuni was grown on 1% Columbia agar supplemented with 5% defibrinated horse blood for 20 hours at 37 ˚C under microaerobic conditions. The cells were subcultured onto 4-6 fresh plates for 15-18 hours at 37 ˚C under microaerobic conditions. The bacteria were harvested using 1 mL sterile Heart Infusion broth (Oxoid) and the OD_{600} was adjusted to 0.6. 1 mL of the bacterial suspension was added to 2 mL eppendorf tubes half filled with 2.0% Heart Infusion agar, and incubated at 37 ˚C for 3 hours under microaerobic conditions. 1-2 μg of plasmid DNA was added to the bacterial suspension, and mixed by gently pipetting. The bacterial suspension was then incubated for an additional 5 hours at 37 ˚C under microaerobic conditions. The bacteria were then gently removed from the tube, ensuring that all attached bacteria was washed from the agar by gentle pipetting. 100 μL of the bacterial suspension was plated out onto selective media, containing appropriate antibiotic, and incubated for 48-72 hours at 37 ˚C under microaerobic conditions.

2.7 Protein techniques

2.7.1 Protein quantitation using the Bradford assay

Protein standards of BSA in PBS (pH 7.4) at concentrations of 0 mg/mL, 0.25 mg/mL, 0.5 mg/mL, 0.75 mg/mL, 1.0 mg/mL, 1.25 mg/mL and 1.4 mg/mL were prepared. 5 µL of each standard in duplicate was added to a 96 well plate. 5 µL of each sample to be quantitated was added in duplicate also to this 96 well plate. 250 µL of room temperature Bradford reagent was the added to each well and incubated for 15-30 minutes before being read by a plate reader at 595nm (VICTOR WALLAC 2). A linear regression was used to generate a standard curve and determine the protein concentration.

2.7.2 Trichloroacetic acid (TCA) precipitation

The TCA precipitation protocol was adopted entirely from the pET System Manual (Novagen, 2010). In the case of culture supernatant precipitation, 100 mL of culture was precipitated and resuspended in total volume of 1 mL of PBS.

2.7.3 Polyethylene Glycol (PEG) concentration

Concentration of a protein sample was performed using a dialysis membrane with a 10 kDa molecular mass cutoff. The membrane was prepared according to manufacturer’s instructions.

One end was sealed with clips or string and the protein solution was poured into the tube. The tube was placed in 30% PEG, and dyalisis was performed until the sample was concentrated to required volume. Once the concentration was finished, the outside of the tubing was gently washed with distilled water, and the concentrated solution of protein was removed by aspiration with a micropipette.

2.7.4 Preparation of samples for SDS-PAGE

SDS-PAGE sample buffer containing fresh 2-mercaptoethanol was added to the samples to a final 1X concentration as per Sambrook *et al*., 2001. Samples were boiled for 5 minutes to denature proteins and were briefly spun down at 14,000 rpm before loading onto the appropriate percentage SDS-PAGE gel using a Hamilton microlitre syringe.

2.7.5 SDS-PAGE resolution of proteins

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed as per Sambrook *et al.*, 2001 using the Laemmli (1970) method of discontinuous buffer system and addition of 0.1% SDS to all components of the system. 12% resolving gels and 5% stacking gels were used in all instances unless otherwise stated. Electrophoresis of samples was performed in a MINI-Protean 3 (BioRad) electrophoresis tank in a tris-glycine running buffer (0.025 M Tris, 0.250 M glycine, 0.1% SDS). A voltage of 8 V/cm was applied as the samples moved through the stacking gel and 15 V/cm as samples moved through the resolving gel. Electrophoresis was halted when the bromophenol blue dye front reached approximately 0.5 cm from the bottom of the gel. The Precision Plus All Blue Standard Prestained molecular weight marker was used for protein size determination (Bio Rad).

2.7.6 Staining of SDS-polyacrylamide gels after electrophoresis

Proteins in the SDS-polyacrylamide gels were stained by immersing the gel in at least 5 volumes of Coomassie brilliant blue stain (0.1% brilliant blue in 50% methanol, 40% deionised water, 10% acetic acid) and placing it on a slowly rotating platform for a minimum of 4 hours at room temperature. Gels were destained by soaking in the methanol:acetic acid solution (50% methanol, 40% water, 10% acetic acid) on a slowly rotating platform for 2-3 hours, changing the destaining solution three or four times.

2.7.7 Western Blot Analysis of Proteins

Protein samples were resolved on a 12% SDS-PAGE and western blot was performed by the transfer of the protein resolved in the acrylamide gel to a 0.45 μM PVDF membrane by Transblot semi-dry transfer cell (BioRad). The membrane was initially washed in methanol, rinsed with deionised water and soaked in Towbin buffer. Filter paper and gels were also soaked in Towbin buffer. Transfer was performed using Towbin buffer at 20 V for 60 minutes.

After transfer, the membrane was incubated in 25 mL blocking buffer (1XTBS, 0.1% Tween20, 1% skim milk powder) for 1 hour. Membrane was subsequently washed 3 times for 5 minutes each in 15 mL Tris-buffered saline containing 0.1% Tween20. Overnight incubation with the anti-His (NEB) antibody in 10 mL blocking buffer at a dilution of 1:10,000 was undertaken at 4 ˚C with agitation, as per the manufacturer’s instructions. 3 washes of 5 minutes, each with 15 mL Tris-buffered saline containing 0.1% Tween20 were performed. The membrane was incubated with a secondary goat-anti-mouse HRP conjugate (Biorad) at a ratio of 1:5,000 for 1 hour at room temperature, as per the manufacturer's instructions. The membrane was washed 3 times 5 minutes each with 15 mL Tris-buffered saline containing 0.1% Tween20. Chemiluminescence detection was subsequently undertaken using the SuperSignal West Pico detection solutions (Thermo Scientific) as per manufacturer's instructions. Membranes were then exposed to X-ray film (CL-Xposure, Kodak) using an X-ray cassette. This X-ray film was developed (KodakGBX developer and replenisher), washed in deionised water and fixed (KodakGBX fixative and replenisher) before a final wash in deionised water.

2.7.8 Dot Blot Analysis of Proteins

The procedure for dot blots was adopted from abcam® technical manual. Briefly, 2 μL of samples were spotted onto the nitrocellulose membrane. The membrane was left to dry. Non-specific sites were blocked by soaking in 5% BSA in TBS-T (0.5-1 hr, RT). The membrane was incubated with primary antibody (1:1,000 to 1:100,000 dilution for antiserum), dissolved in BSA/TBS-T for 30 min at RT, followed by three times washes with TBS-T (5 minutes each). The membrane was then incubated with secondary antibody conjugated with HRP (optimum dilution was performed as per the manufacturer's recommendation) for 30 min at RT. The membrane was washed three times with TBS-T (15 min x 1, 5 min x 2). Chemiluminescence detection was subsequently undertaken as described above in the Western blot protocol.

2.7.9 Small scale protein expression

The recombinant plasmid pGU0513 was transformed into the *E. coli* expression strain BL21(DE3), to form BL21(DE3)pGU0513. 50 µL of an overnight culture of *E. coli* BL21(DE3)pGU0513 was added to 950 µL of LB containing ampicillin, and incubated at 37 ˚C with shaking at 200 rpm for 2 hours. IPTG was added to a final concentration of 1 mM, for an additional 2 hours to induce expression of the His-fusion protein. 100 µL of the cell suspension was spun down and resuspended in an equal volume of PBS. SDS-PAGE sample buffer containing fresh 2-mercaptoethanol was added to samples to a final 1X concentration and samples were boiled for 5 minutes to denature the proteins. 20 µL of the suspension was loaded onto a 12% SDS-PAGE gel and electrophoresed for 90 min at 100 V. The protein bands were visualised by Coomassie Brilliant Blue stain.

2.7.10 Cellular localization of recombinant protein

Analysis of soluble cytoplasmic fraction and the insoluble cytoplasmic fraction, including inclusion bodies is needed to determine if the recombinant protein is present in the soluble cytoplasmic fraction needed for the further purification of the recombinant protein. 1 mL of an overnight culture of *E. coli* BL21(DE3)pGU0513 was added to 19 mL of LB broth supplemented with ampicillin. The culture was incubated at 37 ˚C for two hours with shaking. 500 µL of the culture was set aside as an un-induced sample while the rest of the culture was induced with IPTG (1 mM final concentration). The culture was incubated for an additional 2 hours and 500 µL was set aside as an induced sample. The cell pellet was removed by centrifugation at 4,000 *g* for 5 min. The supernatant was removed and the pellet was resuspended in 5 mL sterile Phosphate Buffered Saline (PBS). The cell pellet was lysed by the addition of lysozyme (200 µg/mL) and sonication. The unlysed cells were removed by centrifugation at 4,000 *g* for 5 min. The lysed cells were then centrifuged at 100,000 *g* for 90 min at 4 ˚C. The cell pellet and supernatant were analysed by SDS-PAGE gel electrophoresis.

2.7.11 Optimisation of purification protocol.

Purification of the His-Cj1344c protein was attempted using the following purification procedures. Firstly, purification was attempted using AKTA FPLC in conjunction with His-trap FF 5 mL column (GE Health Care). This procedure resulted in elution of the protein off the column, but there was a significant amount of co-eluting proteins present in elution samples. A pre-elution wash step incorporated in the method did not significantly reduce the amount of contaminating proteins. The yield of the recovery of the His-Cj1344c was also low. His-select Nickel affinity resin (Sigma) and Talon Cobalt affinity resin (Clontech) were assessed as means of purification of the His-Cj1344c protein. Protocols supplied by the manufacturer were followed. The purification protocols showed significantly less contaminating proteins present, when compared to AKTA FPLC. The yield of recovery of the His-Cj1344c was significantly higher using His-select Nickel affinity resin.

Modification of the manufacturers' existing protocols with the addition of 20 mM β-mercaptoethanol, 0.5% Tween20 and 50 mM imidazole addition in various buffers, reduced non-specific interactions, yielding eluted recombinant proteins with 80-90% purity.

2.7.12 Large scale protein expression.

In order to produce a large amount of the protein a 100 mL overnight culture of BL21(DE3)pGU0513 was used to inoculate one litre of LB containing ampicillin at 100 µg/mL and was incubated at 37 ˚C with aeration. Protein expression was induced using 1 mM IPTG when the OD_{600} of the culture reached 0.5 and was incubated for an additional 6 hours at 37 ˚C with aeration. The culture was centrifuged at 8,000 *g* for 20 min. The supernatant was removed and the cell pellet was resuspended in 30 mL of **binding buffer** (50 mM $NaHPO_4$, 0.3 M NaCl, 0.5% Triton X-100, 10 mM imidazole, pH 7.0) and lysed by the addition of lysozyme (0.2 mg/mL) overnight at 4 ˚C with subsequent sonication. An additional freeze/thaw step was performed to aid in cell lysis. DNaseI and protease inhibitors were added and the insoluble cell debris removed by centrifugation at 100,000 *g* for 80 min. The 30 mL of clarified supernatant was added to 1 mL of His-select nickel affinity resin (Sigma) and rotated overnight at 4 ˚C using a rotational mixer. The slurry mix was then packed by gravity into a 10 mL Bio-Rad chromatography column. The column was washed once with 25 mL binding buffer, then washed with 25 mL of **wash buffer** (50 mM $NaHPO_4$, 1.0 M NaCl, 20 mM β-mercaptoethanol, 50 mM imidazole, pH 7.0) and the bound His-tagged protein was eluted with **elution buffer** (50 mM $NaHPO_4$, 0.3 M NaCl, 500 mM imidazole, pH 7.0) in 0.5 mL volumes. All buffers were sterilised using a 0.22 µM filter and were kept at 4 ˚C. The procedure was performed at 4 ˚C to minimise protein degradation. Further purification of the His-Cj1344c fusion protein was achieved by Cobalt affinity chromatography. Imidazole was removed from the sample by dialysis in PBS, overnight at 4 ˚C and then mixed with 1 mL of the Talon® His-Tag Purification Resin (Clontech) overnight to ensure maximum binding. Purification was performed according to the manufacturer's protocols. The purity of the sample was confirmed by SDS-PAGE analysis and Western Blot using anti-His antibodies (Bio-Rad).

2.8 STD NMR spectroscopy

2.8.1 Preparation of protein samples for NMR spectroscopy.

Purified His-Cj1344c protein was concentrated using an Amicon Ultra 5K centrifugal concentrator (Millipore), previously washed with deuterium oxide (D_2O), 0.1 M NaOH/D_2O, then D_2O to remove NaOH. A series of five 5,000 *g* centrifugation steps at 4 ˚C were aimed to replace H_2O with D_2O. Protein concentration was estimated by Bradford assay.

2.8.2 STD NMR spectroscopy

NMR samples were prepared by the addition of ligand (10 mM in D_2O) to 600 µL of His-Cj1344c, to give a His-Cj1344c:ligand ratio of 1:100. The ligands investigated included L-methionine, L-lysine and L-arginine (Sigma). All ^{1}H NMR experiments were acquired on a Bruker Avance 600 MHz spectrometer equipped with a 5 mm TCI cryoprobe with Z-axis gradients at 288 K. Spectra were acquired with 1k scans, and a relaxation delay at least >1 X T_1 value of the longest T_1 of protons in the ligand being studied. ^{1}H T_1 values were determined using the inversion recovery method. For the STD experiments the protein was saturated on-resonance at a frequency of –600 Hz in the aliphatic region of the spectrum and off-resonance at 20,000 Hz with a cascade of 40 selective Gaussian shaped pulses of 50 ms with a 100 µs delay between each pulse resulting in a total saturation time of 2 seconds. As a control experiment, identical ligand-only spectra were acquired; no signal was observed in the ligand-only spectra. On-resonance and off-resonance spectra were subtracted to obtain a difference spectrum containing the STD signals of binding ligands.

2.9 Arrays

2.9.1 Amino Acid Arrays.

Amino acid arrays were performed as described by Day *et al.*, 2009 (Day *et al.*, 2009). Amino acids were solubilised in water and spotted on to epoxy functionalised glass slides (SuperEpoxy, ArrayIt) using a Piezorray (Perkin-Elmer) non-contact array printer in spots of 0.3-0.6 nL at a concentration of 10 mg/mL. Appendix A lists the amino acids tested. The slides were neutralised as per manufacturer's instructions and stored at 4 ˚C under desiccating conditions. The purified His-Cj1344c fusion protein in PBS was pre-complexed with primary, secondary and tertiary antibodies labelled with Alexa-488 as described previously (Blixt *et al.*, 2004). The antibodies used were Penta-His Alexa-488 mouse IgG (Qiagen) for the primary and the Signal amplification kit for mouse antibodies (Molecular Probes) for the secondary and tertiary. The labelled His-Cj1344c fusion protein was diluted to 25 μg/mL in PBS and Tween-20 was added to a final concentration of 0.025% (v/v). Each subarray was contained within a 65 μL adhesive frame (Abgene) to allow 3 simultaneous hybridisations. Hybridisation of the labelled His-Cj1344c fusion protein was performed at 37 ˚C for 15 minutes in the dark. The wash buffers, PBS with Tween-20 0.05% (v/v) (buffer 1) and PBS (buffer 2), were filtered (22 µm) and pre-warmed before use. The array was placed in a 50 mL tube and washed in 45 mL of buffer 1 for 15 minutes and then washed in buffer 2 for 1 minute prior to being dipped into filtered (22 µm) and pre-warmed water 3 times. The array was dried by centrifugation at 200 *g* for 2 minutes in a 50 mL conical centrifuge tube. In the event of high background signals, an additional wash of 15 minutes in buffer 1 was performed. The array was scanned by Proscan array scanner and the results analysed using Proscan software (Perkin-Elmer).

2.9.2 Glycan array.

Glycan arrays were performed as previously described by Day *et al.*, 2008, and Appendix B lists the compounds tested. Test compounds printed on the array comprised of various carbohydrates and other glycoproteins.

2.10 Purification of native protein using Dynabeads® M-280 Sheep anti-Rabbit IgG

Attempts to isolate the native Cj1344c protein using the combination of Dynabeads® M-280 Sheep anti-Rabbit IgG system and polyclonal antibody raised in a rabbit were performed using the culture supernatant of *C. jejuni* and the cytoplasmic fraction of the cell. The procedure was performed as per manufacturer's instructions without modifications.

2.11 Glycoprotease assays

An aliquot of MUC2 and bovine lactoferrin were incubated with various amounts of His-Cj1344c in 50 mM HEPES buffer pH 7.4, total volume 50 µL, for 16 hours at 37 ˚C. The substrate and products were separated by 1% agarose-SDS gel and 12% SDS-PAGE gel followed by Coomassie Blue staining.

2.12 *In vivo* methods

2.12.1 Animals

Male 129X1/SvJ mice aged between 6-8 weeks were purchased from Animal Resource Centre, Western Australia and were housed under clean conventional conditions in groups of 8 or 4. Food and water were provided *ad libitum*. Once the mice were inoculated with *C. jejuni*, they were kept in isolation. During this time faecal samples were routinely cultured as described below to be certain the animals were free of campylobacter. The experiments were approved and conducted according to the principles set forth by the Griffith University Animal Ethics Committee (Approval number:BDD/03/08/AEC).

2.12.2 Mice vaccination

The animals were inoculated intra-peritoneally, subcutaneously and intra-nasally (3 groups of 8 animals), with 5 μg of purified His-Cj1344c previously mixed with an appropriate adjuvant. In each experiment a control group, kept in isolation, was included and was vaccinated (intra-peritoneally, subcutaneously and intra-nasally) with adjuvant alone. A total of 3 vaccinations for each administration (intra-peritoneally, subcutaneously and intra-nasally) were delivered at 2-weeks interval. In case of intra-peritoneal and subcutaneous vaccination Freund's complete adjuvant (Sigma-Aldrich) was used for the first vaccination. Freund's incomplete adjuvant (Sigma-Aldrich) was used for the subsequent vaccinations. Cholera Toxin Subunit B (5 μg/injection) (Sigma-Aldrich) was used as an adjuvant for intranasal vaccination.

2.12.3 Blood sample collection and processing

100 µL of blood was collected from mice by the submandibular bleed method using the Goldenrod animal lancet (MEDIpoint International) 2 weeks after the second and third vaccination and the serum was separated and stored at -20 ˚C until assayed for immunoglobulin A (IgA), IgG and IgM.

2.12.4 Faecal sample collection and processing.

Faecal excretion of *C. jejuni* was routinely monitored during vaccination period on a weekly basis and daily for 7 days after bacterial challenge by culturing 100 µL of faecal homogenates (2 faecal pellets dissolved in 1 mL of Brucella broth) onto a campylobacter-selective agar.

2.12.5 Determination of immune responses by ELISA.

Serum IgA, IgG and IgM antibody responses to His-Cj1344c were quantitated using an enzyme immunosorbent assay (ELISA). The wells of the ELISA plates were coated with purified His-Cj1344c (1 µg/mL, 200 µL/well) in PBS at 4 ˚C overnight. Plates were washed three times with wash solution consisting of PBS with 0.05% Tween 20, and blocked with 200 µL/well of blocking buffer (3% BSA in Tris-buffered saline, 0.05% Tween 20 (TBS-T)) for two hours at room temperature. Plates were washed three times with wash solution, and 200 µL of serum samples diluted in PBS (with 0.05% Tween 20) was added to the wells. Dilution fractions ranged from 1:20 to 1:65,376. Sera collected prior to the vaccination served as a negative control. The plates were incubated at room temperature for two hours, followed by three washes. For the mouse studies peroxidase-conjugated goat anti-mouse IgM (µ chain;

0.2 μg/mL), IgG (γ chain; 0.125 μg/mL), or IgA (α chain; 0.25 μg/mL) were used as detecting antibodies (Sigma-Aldrich) during a 2 hour incubation at 37˚C. Bound conjugate was detected using the TMB Peroxidase EIA Substrate Kit (BioRad). The IgM, IgG, and IgA endpoint titres for individual mice were determined (reciprocal of the highest dilution showing a net optical density at 405 nm of 0.10).

2.12.6 Preparation of *C. jejuni* inoculums and mice inoculation.

C. jejuni cells were harvested off Columbia agar plates in 1 mL of Brucella broth and the concentration was adjusted to 3.3 x 10^9 CFU/mL using spectrophotometry and viable count. 129X1/SvJ mice were orally inoculated with 30μL Brucella broth containing 1 x 10^8 cfu bacterial cells and were monitored for signs of sickness for 7 days. Bacterial faecal load was monitored on a daily basis as explained above. Putative campylobacter colonies were confirmed by morphology. Mice were sacrificed 7 days post inoculation by cervical dislocation and bacterial load was determined in systemic organs and small and large intestines as outlined below.

2.12.7 Homogenisation of organs

After each organ was aseptically removed from each animal, it was weighed and placed in a 5 mL tube containing 2 mL of sterile Brucella broth. Each sample was homogenised until a homogenous sample was achieved. A sample of the homogenate was serially diluted and plated in triplicate on campylobacter-selective agar. The bacterial load per gram of organ was enumerated by viable count.

2.13 Statistical analysis

The mean of the groups for bacteria load in each organ ($n \geq 5$) were individually compared to that of control groups at the same time point. Significance was determined by un-paired t-tests with an alpha of 0.05.

CHAPTER 3

Expression, purification and characterisation of the putative *O*-sialoglycoprotease from *Campylobacter jejuni*

3.1 Introduction

The proteolytic enzyme glycoprotease (gcp) was first discovered in the culture supernatant of *Mannheimia (Pasteurella) haemolytica* A1 associated with bovine pneumonic pasteurellosis (Otulakowski *et al.*, 1983). The glycoprotease of *M. haemolytica* is an enzyme highly specific for *O*-linked sialoterminal oligosaccharides of glycoproteins such as the sialylated membrane glycoprotein, glycophorin A of human red blood cells (Abdullah *et al.*, 1992, Mellors & Jiang, 1998). The enzyme has a marked specificity for human CD34 (an antigen expressed on panhematopoietic stem cells in the bone marrow), human CD43 (a sialomucin that has been implicated in immune and human cell function and cell-signalling phenomena) and human CD44 (Sutherland *et al.*, 1992). In addition, Hu *et al*. have demonstrated that the enzyme degrades epitectin and other mucin-type sialoglycoproteins (Hu *et al.*, 1994).

A number of *M. haemolytica* glycoprotease-susceptible substrates become resistant to cleavage by the glycoprotease after treatment with sialidase from *Clostridium perfirogens* or *Vibrio cholerae* (Sutherland *et al.*, 1992). Thus, the proteolysis seems to be dependent on the presence of terminal sialic acid residues. Abolition or reduction of glycoprotease activity by desialylation has been shown for glycophorin A (Abdullah *et al.*, 1992), CD34 and CD44 (Sutherland *et al.*, 1992). The enzyme was also inhibited by zinc ions (Cladman *et al.*, 1996) even though the predicted amino acid sequence does not contain metal binding sites similar to any of the 12 known classes of metalloproteases (Hooper, 1994).

The Blastp search of *C. jejuni* genome strain NCTC11168 has revealed a gene with similarity to the *M. haemolytica O*-sialoglycoprotease gene (Altschul *et al.*, 1997). This chapter describes the isolation of the putative glycoprotease of *C. jejuni*

NCTC11168-O (Cj1344c). Amino acid and carbohydrate arrays were utilised to investigate the ligand-binding specificities of Cj1344c. Characterisation of the *Cj1344c* isogenic mutant is described in Chapter 4.

3.2 Results

3.2.1 *In silico* analysis of Cj1344c

Comparative bioinformatics analyses of the of *C. jejuni* NCTC11168-GS glycoprotease predicted amino acid sequence (Cj1344c) were performed using the Blastp program and conserved domain search service as described by Marchler and Bryant (Marchler-Bauer & Bryant, 2004). Comparative amino acid analysis of the Cj1344c indicated that it was similar (36% identity / 55% similarity) to Gcp of *M. haemolytica* A1, an enzyme highly specific for *O*-linked sialoterminal oligosaccharides of glycoproteins (Otulakowski *et al.*, 1983). A *Cj1344c* homologue was present in all to date sequenced strains of *C. jejuni* with higher than 97% amino acid identity and an orthologue was present in other *Campylobacter* species with greater than 70% amino acid similarity (Table 3.1).

Table 3.1 Comparison of *C. jejuni* NCTC11168 Cj1344c to the putative Gcp proteins from *Campylobacter* strains.

Bacterial species	Strain	Gene annotation	Amino acid % Identity / % Similarity
C. jejuni* subspecies *jejuni	RM1221	CJE1533	100/100
	CF93-6	CJJCF936_1435	100/100
	CG8421	Cj8421_1389	99/100
	84-25	CJJ8425_1424	97/99
C. jejuni* subspecies *doylei	269.97	JJD26997_0366	97/99
C. coli	RM2228	CCO1450	84/92
C. lari	RM2100	CLA0148	72/85
C. upsaliensis	RM3195	CUP0251	71/83
C. concisus	13826	CCC13826_0417	60/77
C. rectus	RM3267	CAMRE0001_0426	73/58
C. curvus	525.92	CCV52592_0591	59/72
C. hominis	ATCC	CHAB381_0120	59/77
C. fetus* subspecies *fetus	82-40	CFF8240_0215	58/76

Source: CampyDB

The glycoprotease was also present in the genomes of *Campylobacter* related species, such as *Helicobacter* and *Wollinella*. In *Helicobacter* spp, putative *gcp* homologues showed approximately 70% amino acid similarity to the putative Gcp protein in *C. jejuni*. *Wollinella* spp homologues have approximately 65% similarity to *C. jejuni* Gcp. Comparative analysis also identified presence of orthologues in different bacterial species such as *Bacillus anthracis*, *Staphylococcus aureus* and *Haemophilus influenzae*, with more than 49% amino acid similarity to the predicted Cj1344c amino acid sequence. Figure 3.1 shows grouping of different bacterial species according to the similarities observed in the predicted amino acid sequence of gcp homologues. The Cj1344c amino acid sequence analysis did not identify specific conserved domains within the amino acid sequence.

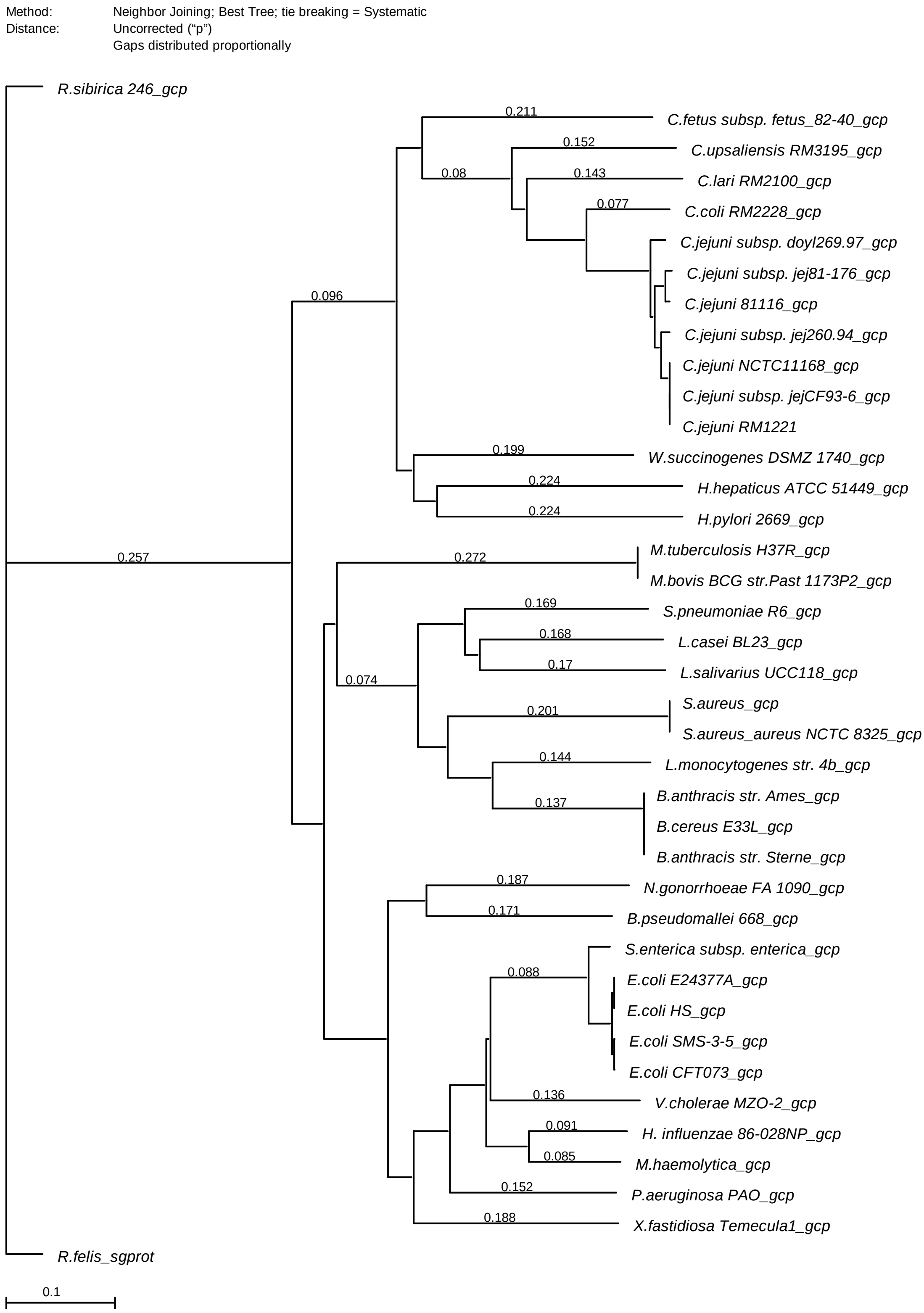


Figure 3.1 Bacterial species grouping according to the gcp amino acid sequence similarity

3.2.2 Amplification of the *Cj1344c* gene

In order to express and purify the putative *C. jejuni* glycoprotease, the gene encoding the protein was amplified by PCR for incorporation into a cloning plasmid intermediate, and subsequently into an expression vector. For PCR amplification of the *Cj1344c* coding region, the forward and reverse primers were designed based on the sequence of *C. jejuni* NCTC11168-GS (Parkhill *et al.*, 2000a) to include the translation initiation codon and the stop codon of the gene (Materials and Methods; Table 2.1). In addition, restriction endonuclease sites at the 5’ (*Nde*I) and 3’ (*Xho*I) termini were included to aid in subsequent DNA manipulation. The generated PCR product was visualised by gel electrophoresis as a DNA fragment of 1027 bp (Figure 3.2). The amplified DNA product was excised from the gel and DNA purified for subsequent cloning.

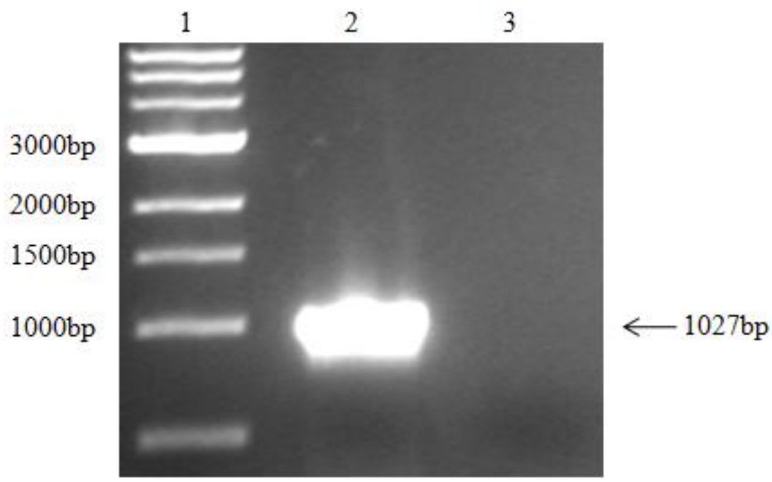


Figure 3.2 Gel electrophoresis of PCR amplification of *Cj1344c*

Legend: **Lane 1** – 1Kb DNA ladder, **Lane 2** – PCR product of *Cj1344c* at 1027bp, **Lane 3** – negative control

3.2.3 Cloning of *Cj1344c* into pGEM-T Easy

In order to generate a cloning intermediate for subsequent manipulation, a PCR generated DNA fragment encoding *Cj1344c* was ligated into the cloning vector pGEM-T Easy within the *lac*Z gene, and transformed into *E. coli* DH5α using standard cloning protocols (Materials and Methods; Section 2.6). Colonies containing recombinant plasmids were selected by growing on LB agar supplemented with ampicillin and IPTG/X-gal utilising disruption of the *lac*Z gene and allowing blue/white colony selection. The integrity of recombinant plasmid was verified by PCR amplification using *Cj1344c* specific primers (Materials and Methods; Table 2.4), which confirmed the presence of the expected 1027 bp fragment in the recombinant plasmids isolated from the 5 out of 8 colonies screened (Figure 3.3).

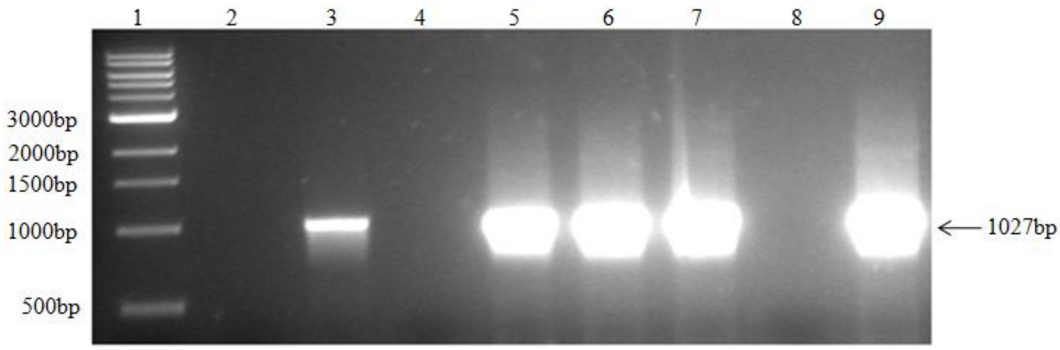


Figure 3.3 PCR amplifications of *Cj1344c* gene from putative recombinant plasmids

Legend: **Lane 1** – 1Kbp ladder; **Lane 2→9** – amplification of *Cj1344c* from plasmid templates isolated from eight colonies screened for the presence of the recombinant plasmid

To further confirm the integrity of the recombinant plasmids from three of these colonies (Lane 5, 6 and 7; Figure 3.3), plasmid DNA was isolated from bacterial cells as described in Materials and Methods; Section 2.5.2, and restricted with the enzymes *Nde*I and *Xho*I, which released the insert from the pGEM-T Easy backbone (Figure 3.4). Sequencing and sequence analysis of the recombinant plasmid DNA using the set of primers positioned in the multiple cloning site of pGEM-T Easy vector (T7 and SP6 (Table 2.4)) confirmed that the *Cj1344c* insert had the correct nucleotide sequence without nucleotide substitution or deletion. This recombinant plasmid was named pGU0501 (nucleotide sequence and map are shown in Appendix C).

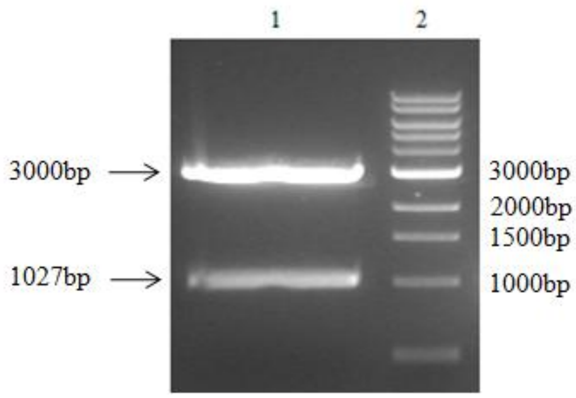


Figure 3.4 Gel electrophoresis of recombinant plasmid pGU0501 *Nde*I/*Xho*I restriction enzyme cleavage

Legend: **Lane 1** – pGU0513 restriction enzyme cleavage, showing 3.0Kb pGEM-T Easy backbone and 1.0Kb gcp insert; **Lane 2** – 1Kbp ladder

3.2.4 Cloning of *Cj1344c* into the expression vector pET-19b

The pET-19b expression vector was used to enable expression of a fusion protein consisting of a recombinant Cj1344c and an N-terminal histidine tag in an *E. coli* bacterial system.

In order to generate such a recombinant plasmid, the pGU0501 cloning intermediate and pET-19b expression vector were cleaved with *Nde*I and *Xho*I restriction enzymes, which released the *Cj1344c* from pGU0501 and linearised pET-19b vector. The excised *Cj1344c* DNA fragment and linearised pET-19b were ligated and recombinant plasmid was transformed into *E. coli* host cells.

The resultant *E. coli* colonies were initially screened by PCR, using *Cj1344c* specific primers (Materials and Methods; Table 2.4) and four colonies carrying recombinant plasmids were identified (Data not shown). Plasmids were isolated from the four colonies containing the 1027 bp insert and were cleaved by *Swa*I restriction enzyme (restriction site present within *Cj1344c* gene). The product of the restriction enzyme digest was the linearised recombinant plasmid DNA of 6.7 Kb as shown in Figure 3.5. Subsequent sequencing of selected plasmids was performed to verify the integrity of the new construct; and DNA sequence analysis indicated that the His-tag sequence was cloned in frame with the open reading frame sequence of *Cj1344c*. The plasmid was named pGU0513 and was used to express the His-Cj1344c fusion polypeptide (nucleotide sequence map in Appendix D).

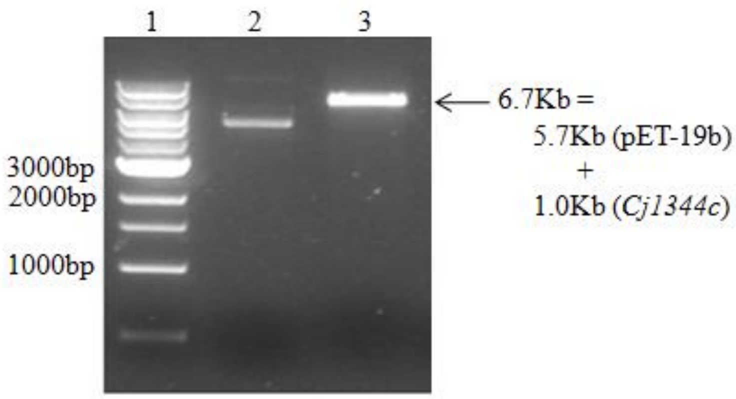


Figure 3.5 Gel electrophoresis of recombinant plasmid pGU0513 restriction enzyme cleavage with *Swa*I

Legend: **Lane 1** – 1Kbp ladder; **Lane 2** – undigested recombinant plasmid; Lane 3 - pGU0513 restriction enzyme cleavage, linearised plasmid DNA at 6.7Kb

3.2.5 Small scale protein expression and confirmation

To enable expression of the fusion His-Cj1344c protein, the recombinant plasmid pGU0513 was transformed into the *E. coli* strain BL21(DE3) as described in the Materials and Methods; Section 2.6. Culture medium containing *E. coli* BL21(DE3)pGU0513 was supplemented with 1 mM IPTG, which enabled over-expression of Cj1344c as a fusion protein with an N-terminal polyhistidine tag (His-Cj1344c) which was detected as a 40 KDa protein on a Coomassie stained SDS-PAGE gel (Figure 3.6).

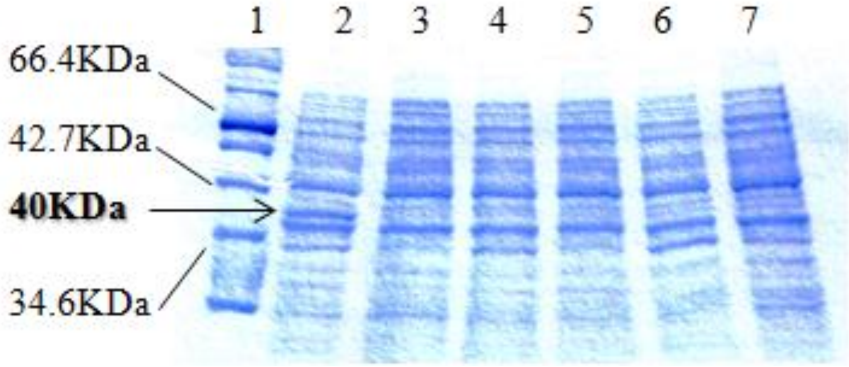


Figure 3.6 Small scale protein expression of His-Cj1344c fusion protein.

Legend: **Lane 1** – protein marker; **Lane 2** – BL21(DE3) pGU0513 induced showing over expressed protein at 38 KDa; **Lane 3** – BL21(DE3) pGU0513 uninduced; **Lane 4** – BL21(DE3) pET-19b induced; **Lane 5** – BL21(DE3) pET-19b uninduced; **Lane 6** – BL21(DE3) induced; **Lane 7** – BL21(DE3) uninduced

Western blot analysis using an anti-His antibody was used to confirm expression of the His-Cj1344c. This test confirmed that the 40 KDa protein in Coomassie stained SDS-PAGE contained a polyhistidine tag (Data not shown). To verify the localisation of the His-Cj1344c protein in the cell, different cell fractions were prepared as described in Materials and Methods; section 2.7. Western blot analysis using anti-His tag antibodies showed that most of the His-Cj1344c protein was present in the soluble cytoplasmic fraction. In addition, the His-Cj1344c protein was also present in the insoluble cytoplasmic fraction and the culture media (Figure 3.7).

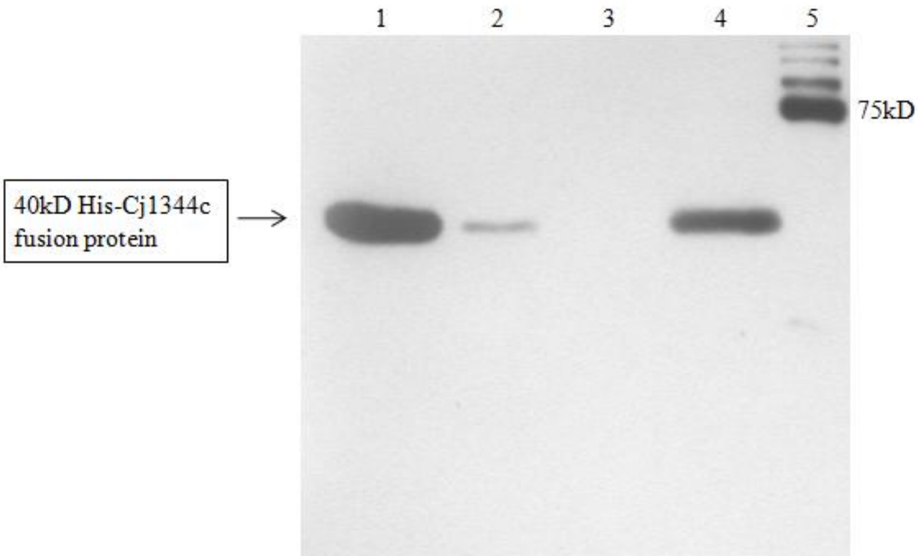


Figure 3.7 Western blot analysis using an anti-His antibody of the different cell fractions

<u>**Legend:**</u> **Lane 1** – soluble cytoplasmic fraction; **Lane 2** – pelleted insoluble cytoplasmic fraction; **Lane 3** – culture supernatant; **Lane 4** – TCA precipitated sample of supernatant; **Lane 5** – Precision Plus Protein Marker

3.2.6 Optimisation of the recombinant protein purification protocols

AKTA FPLC in conjunction with His-trap column, His-select Nickel affinity resin and Talon Cobalt affinity resin were the three methods trialled for His-Cj1344c purification (Materials and Methods; section 2.7). The His-select Nickel affinity resin produced the best results, as the yield and purity of the recombinant protein was the highest of the three methods tested (data not shown). Modification of the existing manufacturer's protocols with the addition of 20 mM β-mercaptoethanol, 0.5 % Tween20 and 50 mM imidazole in buffers, reduced non-specific interactions, yielding eluted recombinant proteins with 80-90% purity (Figure 3.8). The concentration of the protein was estimated using the automated spectrophotometry software Victor[3], and comparing the protein sample against known protein standards, determined the purified His-Cj1344c protein had a concentration of 0.5 mg/mL.

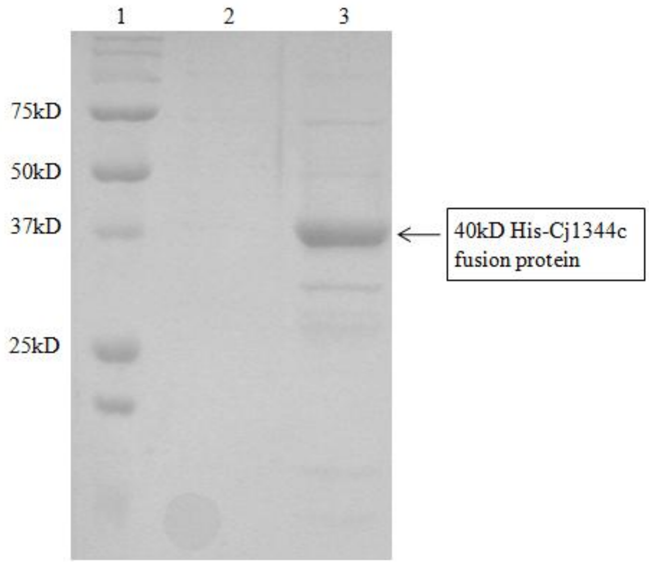


Figure 3.8 Coomassie brilliant blue stained 12 % SDS-PAGE gel showing purification of His-Cj1344c.

Legend: **Lane 1** – Precision plus protein marker (Bio-Rad), **Lane 2** – empty lane; **Lane 3** – purified His-Cj1344c

3.2.7 Protein precipitation and solubility.

Over-expression and purification of the His-Cj1344c protein using the pET-19b system and His-select Nickel affinity resin produced a high concentration of the target protein (0.5 mg/mL). However, the protein precipitated shortly after purification. In order to maintain the protein in a soluble form that was required for further analysis and testing, different parameters were tested to maintain protein in solution. Induction time (3h-8h), induction temperature (37 ˚C versus 25 ˚C), IPTG concentrations (gradient concentrations from 0.2 mM to 5 mM) did not change the solubility of the expressing protein and all resulted in protein precipitation shortly after elution off the resin. Changing the pH of the elution buffer (from pH8.0 to pH7.0) resulted in a soluble protein that stayed in solution for more than 2 weeks at 4 ˚C.

3.2.8 Identification of the substrate specificity by amino acid, glycan and glycoprotein array.

The substrate specificity of His-Cj1344c was investigated by testing substrate-protein interactions using amino acid, glycan and selected glycoprotein arrays (Materials and Methods; section 2.9). In addition, a sample of the whole cell *C. jejuni* lysate was also added on an array slide to test possible interaction of Cj1344c with *C. jejuni* cell proteins. The purified His-Cj1344c fusion protein in PBS was pre-complexed with primary, secondary and tertiary antibodies as described by Blixt *et al.*, 2004. The labelled protein complex was hybridised against a printed amino acid, glycan and glycoprotein array (Appendix A and B list of amino acids, glycans and glycoproteins used in array).

Amino acid array hybridisation with His-Cj1344c identified an interaction between methionine, arginine, lysine and the His-tagged Cj1344c protein with binding of greater than 2000 fluorescent units observed (Figure 3.9). The His-Cj1344c antibody complex showed no significant binding to any of the other amino acids present on the array.

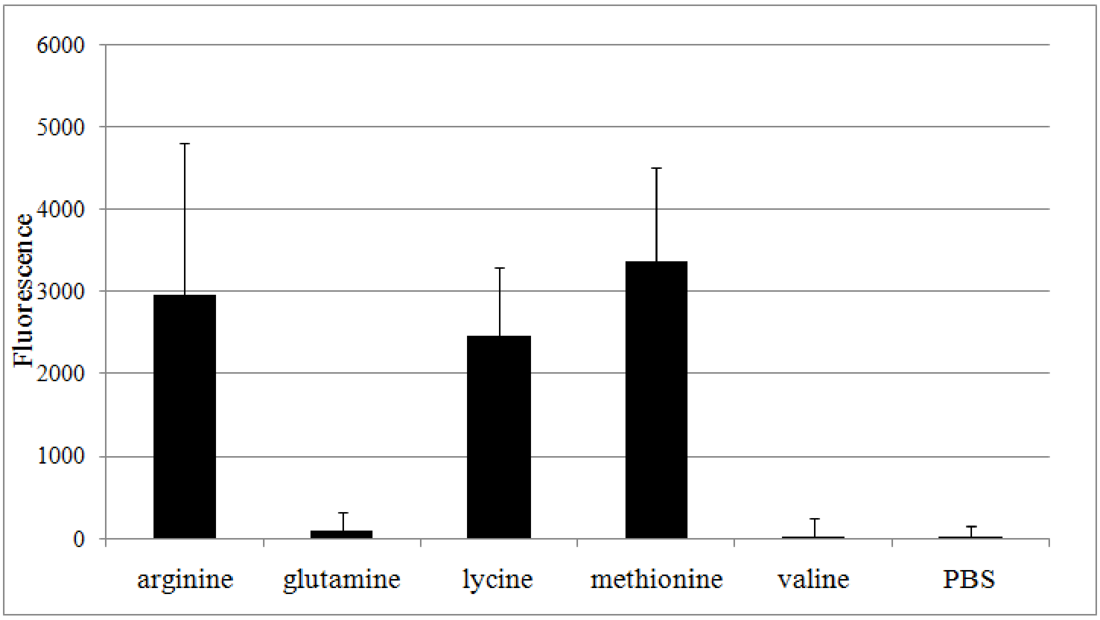


Figure 3.9 Comparative Fluorescence Results of Amino Acid Array Hybridisation.

No significant binding of His-Cj1344c to any of the tested glycan structures was detected. However, the glycoprotein array identified a significant interaction between His-Cj1344c and bovine lactoferrin (Figure 3.10). Binding of His-Cj1344c to recombinant lactoferrin was also recorded, but it was lower than binding to native bovine lactoferrin. In addition, significant binding was identified to *C. jejuni* whole cell lysate and MUC2 (Figure 3.10).

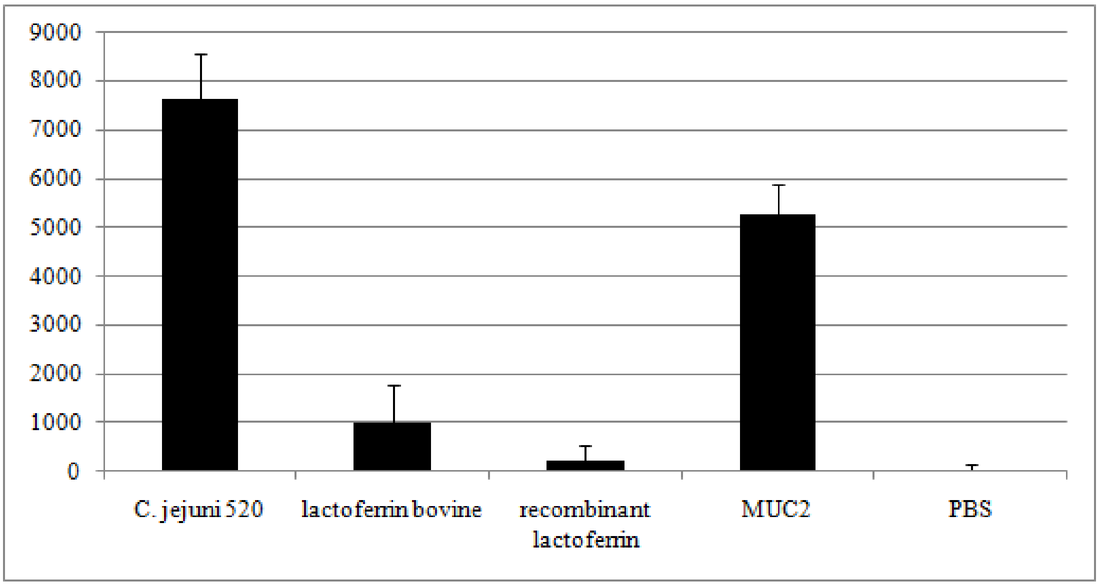


Figure 3.10 Comparative Fluorescence Results of Glycoprotein Array hybridisation

3.2.9 Attempt to confirm the ligand-binding specificity by Saturation transfer difference NMR.

STD-NMR spectroscopy was used to further verify the protein-ligand interactions observed on amino acid arrays. ^{1}H NMR spectra for each of the amino acids under investigation were acquired and used as reference spectra; a ^{1}H NMR spectrum of the His-Cj1344c fusion protein was also acquired. One-dimensional STD spectra were acquired with ligand-only. These spectra served as controls to ensure that any signals observed in the ligand:His-Cj1344c STD spectrum resulted solely due to ligand binding to His-Cj1344c. STD spectra of L-methionine, L-lysine, and L-arginine in the presence of His-Cj1344c gave spectra with no signals attributable to the amino acids being examined, which suggest that these amino acids either did not bind His-Cj1344c or that the binding to these substrates is too high affinity and therefore not detectable by this method (Data not shown).

STD-NMR spectroscopy was not performed for bovine lactoferrin, MUC2 and whole cell *C. jejuni* lysates because the size of these proteins was a limiting factor in the procedure. Enzymatic protein cleavage studies of bovine lactoferrin and human MUC2 were used instead to determine the activity of the enzyme.

3.2.10 Assessment of the enzymatic activity of His-Cj1344c fusion protein

Glycoprotein array studies identified bovine lactoferrin and MUC2 as potential substrates for His-Cj1344c. To determine the activity of the enzyme, these substrates were used in enzymatic digests by His-Cj1344c. Enzyme digests were performed as detailed in the Materials and Methods section 2.11. After enzymatic digest, MUC2 and lactoferrin were analysed by Coomassie Blue staining of agarose SDS gel and SDS-PAGE gels respectively. The results of these experiments showed no significant enzymatic activity, however a change in the MUC2 protein banding pattern before and after digest could be noted on the agarose gel (Figure 3.11). There was no notable effect of His-Cj1344c on bovine lactoferrin as assayed by SDS-PAGE analysis (data not shown).

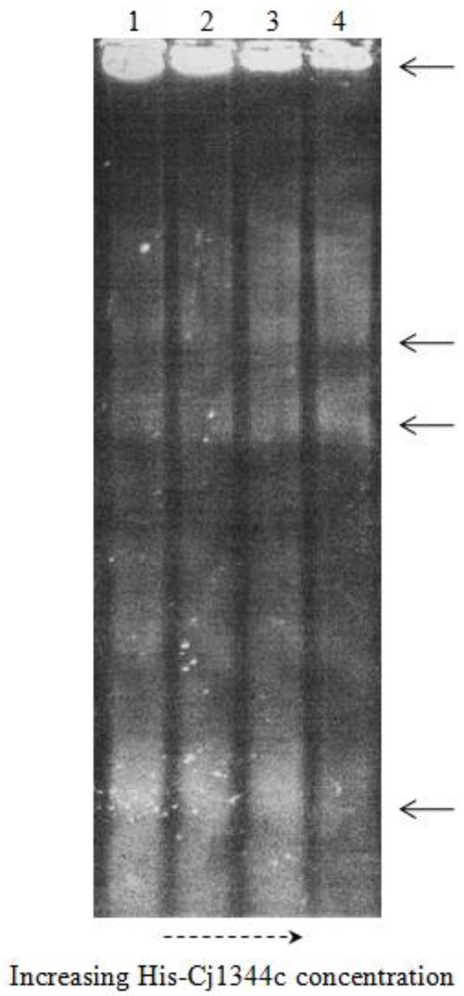


Figure 3.11 His-Cj1344c digests of human MUC2

Legend: Lane 1 – MUC2; **Lane 2 – Lane 4** enzymatic digest of MUC2 with an increase of MUC2:His-Cj1344c ratio. Solid arrows point to the changes in the banding pattern

3.3 Discussion

The predicted amino acid sequence of the putative *C. jejuni* glycoprotease (Cj1344c) shows 55% similarity to *M. haemolytica* gcp which shows marked specificity for *O*-glycosylated sialoglycoproteins (Abdullah *et al.*, 1992, Mellors & Jiang, 1998). In a study by Abdullah *et al*. digestion of glycophorin A identified the amino acid recognition sequence of *M. haemolytica* gcp. The major cleavage site of glycophorin A occurs at the Arg-31–Asp-32 peptide bond. Other cleavage site include Glu-60–Arg-61, Arg-31–Asp-32, Ala-65–His-66 and Tyr-34–Ala-35 (Abdullah *et al.*, 1992). To determine the specificity of the recombinant fusion protein, His-Cj1344c was used in amino acid array studies to identify amino acids recognised by this enzyme. The results of these experiments identified an interaction of recombinant Cj1344c with methionine, lysine and arginine. Recognition of these amino acids by His-Cj1344c may indicate amino acids within a polypeptide sequence recognised by *C. jejuni* glycoprotease and potential cleavage sites of this enzyme, and its difference to *M. haemolytica* gcp.

STD-NMR method was employed in an attempt to confirm the interactions between the amino acids and His-Cj1344c observed as binding partners in the amino acid array. The ligand binding studies performed with methionine, lysine and arginine failed to confirm the interactions observed with the amino acid array technology. The absence of signal in these studies could be attributed to the lack of interaction between these ligands and His-Cj1344c. However, the strong signal observed in amino acid arrays (greater than 2,000 fluorescent units and significantly higher than binding to PBS which served as a negative control) suggests that the reason for the absence of the STD-NMR signal was probably a strong binding interaction between these ligands and His-Cj1344c, which cannot be detected by the STD-NMR and is one of the limiting factors of this method (Haselhorst *et al.*, 2009).

The glycoprotein array analysis identified bovine lactoferrin as the potential ligand for Cj1344c. Binding of Cj1344c to lactoferrin, an iron binding glycoprotein, may suggest a potential role in inactivation of this molecule; and thus provides a protective mechanism for bacteria against this component of the host innate system. Interaction of His-Cj1344c was also observed with recombinant lactoferrin, but the levels of interaction were significantly lower than interaction of His-Cj1344c to the native bovine lactoferrin. The difference in the binding ability of His-Cj1344c to bovine lactoferrin compared to recombinant lactoferrin may be due to the lack of posttranslational modification of the recombinant lactoferrin or absence of sialic acid residues (Lönnerdal & Iyer, 1995). The increased binding to native bovine lactoferrin compared to recombinant lactoferrin, expressed in *E. coli*, indicates a possible requirement of the Cj1344c for *O*-linked glycosylation and presence of sialic acid, both of which are absent in the recombinant lactoferrin structure. Studies conducted with *M. haemolytica* gcp show that the removal of sialic acid from glycophorin A significantly reduces the cleavage of the glycoprotein by this enzyme (Abdullah *et al.*, 1992).

In addition to bovine lactoferrin, glycoprotein array experiments also identified an interaction between His-Cj1344c and human MUC2. MUC2 is the main gel-forming mucin of the small and large intestines (Toribara *et al.*, 1991) and is heavily glycosylated with *O*-linked oligosaccharides. In addition, mucin oligosaccharides are extensively decorated by sialic acid residues (Holmen Larsson *et al.*, 2009). The binding of His-Cj1344c to purified MUC2 may indicate a role for this enzyme in the degradation of intestinal tract MUC2. *C. jejuni* gene expression studies conducted by Tu and associates suggest an involvement of Cj1344c in MUC2 degradation as the levels of expression of the *Cj1344c* gene were upregulated when the bacteria were grown in the presence of MUC2 (Tu *et al.,* 2008), further strengthening the hypothesis that Cj1344c may be important in bacterial pathogenesis.

Confirmation of the His-Cj1344c interaction with lactoferrin and MUC2 was attempted by digesting these substrates with His-Cj1344c. Enzymatic digestion of these substrates could not confirm the biological activity of the enzyme as the SDS-PAGE of the substrates after digestion did not show distinct digestion patterns, suggesting that Cj1344c expressed in *E. coli* may be biologically inactive. It may be speculated that the enzyme required metal ions for activation, as sequence analysis identified a putative Zn^{2+} binding motif within the predicted amino acid sequence, similar to a predicted Zn^{2+} binding motif identified within *M. haemolytica* gcp amino acid sequence (Abdullah *et al.,* 1991). The *M. haemolytica* A1 *O*-sialoglycoprotein endopeptidase was also shown to be inactive when expressed in *E. coli* (Watt *et al.*, 1997b). Refolding of the recombinant rGcp by mammalian protein disulfide isomerise or by *E. coli* chaperones can restore the biological activity of the enzyme (Watt *et al.*, 1997a). It has been postulated by Watt *et al*. that the presence of glycoprotease inhibitors in the *E. coli* cytoplasm may contribute to lack the of biological activity of the gcp enzyme, as these inhibitors may inactivate the enzyme during the purification process (Watt *et al.*, 1997a).

Binding of His-Cj1344c to *C. jejuni* whole cell lysate samples suggests multiple roles of this enzyme. Extracellularly, the enzyme is speculated to be involved in pathogenesis processes such as adherence and mucin degradation, as well as bacterial protection against the host immune system through its binding to lactoferrin. Intracellularly, the enzyme may be involved in housekeeping cell processes. Conserved amino acid sequences of 98-100% identity encoding the putative glycoprotease were identified in all fully and partially sequenced *C. jejuni* strains, showing that the glycoprotease is likely to be conserved throughout *C. jejuni*. Disruption of the glycoprotease gene results in the inability to recover isogenic mutants in *M. haemolytica* (Mellors, personal communication) and *C. jejuni* (Chapter 4, this study). Downregulation of the gene in *S. aureus* causes growth defects (Zheng *et al.*, 2005) and eliminates autolysis, (Zheng *et al.*, 2007) while in *E. coli* the enzyme is involved in modulation of the macromolecular operon (Nesin *et al.*, 1987).

3.3.1 Conclusion

This chapter describes the expression, purification and enzymatic assessment of *C. jejuni* putative glycoprotease (Cj1344c). Utilising a pET-19b system enabled the overexpression of the protein in *E. coli* host and its subsequent purification as a fusion protein with an N-terminal polyhistidine tag. Purified protein was used for initial screening of protein binding to amino acids and selected glycans and glycoproteins, utilising modified amino acid and glycan array analysis (Day *et al.*, 2009). Analysis of the purified His-Cj1344c binding capability determined that *C. jejuni* Cj1344c could bind to methionine, lysine and arginine, suggesting that these amino acids are present in the sequences within glycoproteins that are recognised by Cj1344c. In addition, the specificity of Cj1344c to MUC2 and lactoferrin may suggest a putative role for Cj1344c in the pathogenesis of *C. jejuni* during the initial stages of adherence and invasion of epithelial cells, as well as protection of the bacterial cells against the host innate immune system.

CHAPTER 4

Construction and characterisation of the *Cj1344c* isogenic mutant of *Campylobacter jejuni*

4.1 Introduction

Genes encoding glycoprotease enzymes have been identified in many Gram-positive and Gram-negative pathogens, including *Bacillus anthracis*, *Streptococcus pyogens*, *Manheimia haemolytica* A1 (Otulakowski *et al.*, 1983) and *Escherichia coli* (Nesin *et al.*, 1987). Glycoproteases have a variety of functions. The first characterised glycoprotease was that of *M. haemolytica* A1, an enzyme highly specific for *O*-glycosylated glycoproteins (Abdullah *et al.*, 1992). The glycoprotease homologue in *E. coli* may be involved in the modulation of the rpsU-dnaG-rpoD macromolecular-synthesis operon (Nesin *et al.*, 1987). In the Cyanobacterium *Synechocystis* sp., mutation of the glycoprotease gene reduces salt tolerance, alters pigmentation and changes cyanophycin accumulation (Zuther *et al.*, 1998, Karandashova *et al.*, 2002). In *Staphylococcus aureus*, it is essential for bacterial survival, however, the function and the reasons it is required for growth are still unclear (Zheng *et al.*, 2005). In addition, the enzyme was confirmed to be essential for the survival of *Salmonella typhimurium* (Nichols *et al.*, 2006) and *M. haemolytica* (Mellors, personal communication) since any mutation of the gene was lethal to the cells.

In order to characterise the *C. jejuni* putative glycoprotease it is necessary to generate an isogenic mutant strain for comparative analysis. Creation of isogenic mutants is an extensively used method to determine potential gene function through analysis of the effect of gene mutation on bacteria. Many *C. jejuni* virulence factors such as iron acquisition (Palyada *et al.*, 2004), lipooligosaccharides (Fry *et al.*, 1998) and motility and chemotaxis (Nachamkin *et al.*, 1993a) as well as chicken intestinal colonisation (Hendrixson & DiRita, 2004) have been investigated using isogenic mutants, deleting the genes of interest. Kanamycin (*aph(3')-III*) and Chloramphenicol (*cat*) resistance genes are often used for the creation of isogenic constructs in *C. jejuni*.

Many *C. jejuni* genes are organised into long tandem clusters, but genes within the cluster are usually not functionally related and often have individual control sequences (Parkhill *et al.*, 2000a). The lack of well defined promoter sequence in *C. jejuni* genome is an acknowledged problem in the genetic analysis of *Campylobacter* spp (Wosten *et al.*, 1998). The insertion of an antibiotic resistance gene within the open reading frame (ORF) of the *Cj1344c* gene may cause a polar mutation effect by inactivating genes downstream from the insertion point. To negate this effect, different antibiotic resistance cassette constructs needed to be generated. Kanamycin resistance cassettes expressed in both *E. coli* and *Campylobacter* spp (Wosten, M., personal communication) and chloramphenicol resistance cassettes (Taylor & Wang, 1990), have previously been used in site specific mutations in campylobacter genes and in the creation of suicide constructs in *E. coli* (Yao *et al.*, 1993). This chapter focuses on generation of an isogenic mutant of *Cj1344c* aimed at the initial characterisation of this gene and its function.

4.2 Results

4.2.1 *In silico* analysis of the *Cj1344c* gene locus

Analysis of *Cj1344c* within the published sequence *C. jejuni* NCTC11168 (Parkhill *et al.*, 2000a) revealed that the gene is flanked by genes of an unknown function (Figure 4.1).

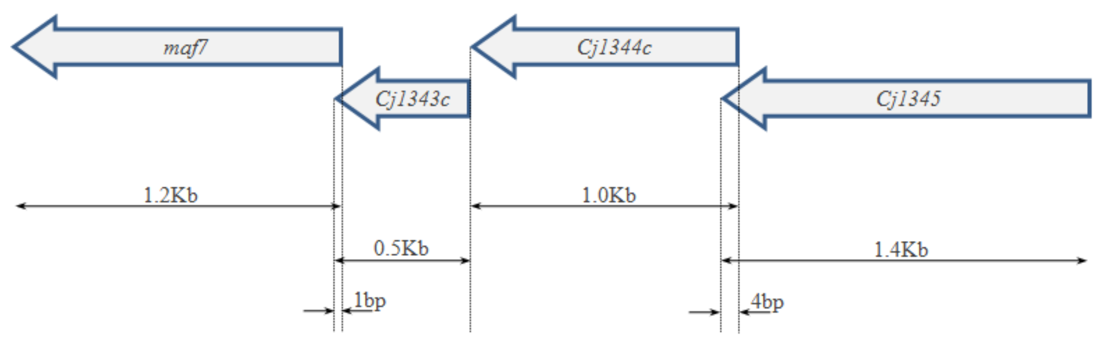


Figure 4.1 Schematic diagram of the *C. jejuni* NCTC11168 *Cj1344c* gene cluster

Further analysis of the genetic organisation of the *Cj1344c* gene conducted using consensus sequences for control elements according to Wosten and colleagues (Wosten *et al.*, 1998) identified a putative -35 box (117bp upstream from the start codon of *Cj1344c*), a *C. jejuni* consensus -16 box (ATTTGGAT), -10 box (TAAATAC) and ribosomal binding site (GTGGA); beginning 73bp, 64bp and 11bp upstream from the start codon of *Cj1344c*, and within the ORF of *Cj1345c*. In addition, the putative promoter sequence elements of *Cj1343c* were identified within the ORF of *Cj1344c* (a putative -35 box (60bp upstream from the start codon of *Cj1343c*), a *C. jejuni* consensus -16 box (TTTAAGCC), -10 box (TAAAAAT) and ribosomal binding site (AAGGA); beginning 37bp, 28bp and 10bp upstream from the start codon of *Cj1343c* (The map of the *Cj1344c/Cj1343c* promoter sequence elements is provided in Appendix E). Immediately following the transcriptional termination signal of the *Cj1344c* gene, there is the start codon of the *Cj1343c* gene. However, the overlapping start and stop codons of the *Cj1344c* and *Cj1343c* genes and a weak ribosomal binding sequence in front of the *Cj1344c* are strong indications that genes are located in an operon. This could present problems in the mutagenesis of *Cj1344c,* as any insertional inactivation of *Cj1344c* gene could potentially affect the expression of downstream genes.

4.2.2 Strategy for positioning mutation sites to circumvent potential polar effects resulting from insertion of an antibiotic resistance cassette within ORF of *Cj1344c*

Plasmid constructs intended for *Cj1344c* mutagenesis were made by insertion of an antibiotic resistance gene within the open reading frame of *Cj1344c* leaving 500bp of *Cj1344c* DNA on each side of antibiotic resistance gene insertion site to maximise the chances for a successful cross-over event and incorporation of foreign DNA into the *C. jejuni* genome (Wassenaar *et al.*, 1993). Special consideration was given to the fact that the gene is positioned in an operon-like organisation with other genes of unknown function. The insertion of the *Cj1344c* gene may potentially create polar mutations by inactivating downstream genes. To circumvent the possibility of polar mutation, four different antibiotic resistance gene cassettes, two of which had to be constructed, were used to generate of suicide plasmids for *Cj1344c* mutagenesis (Figure 4.2). The transcriptional termination signal was removed from all cassettes to prevent disassociation of ribosomes during the transcription process.

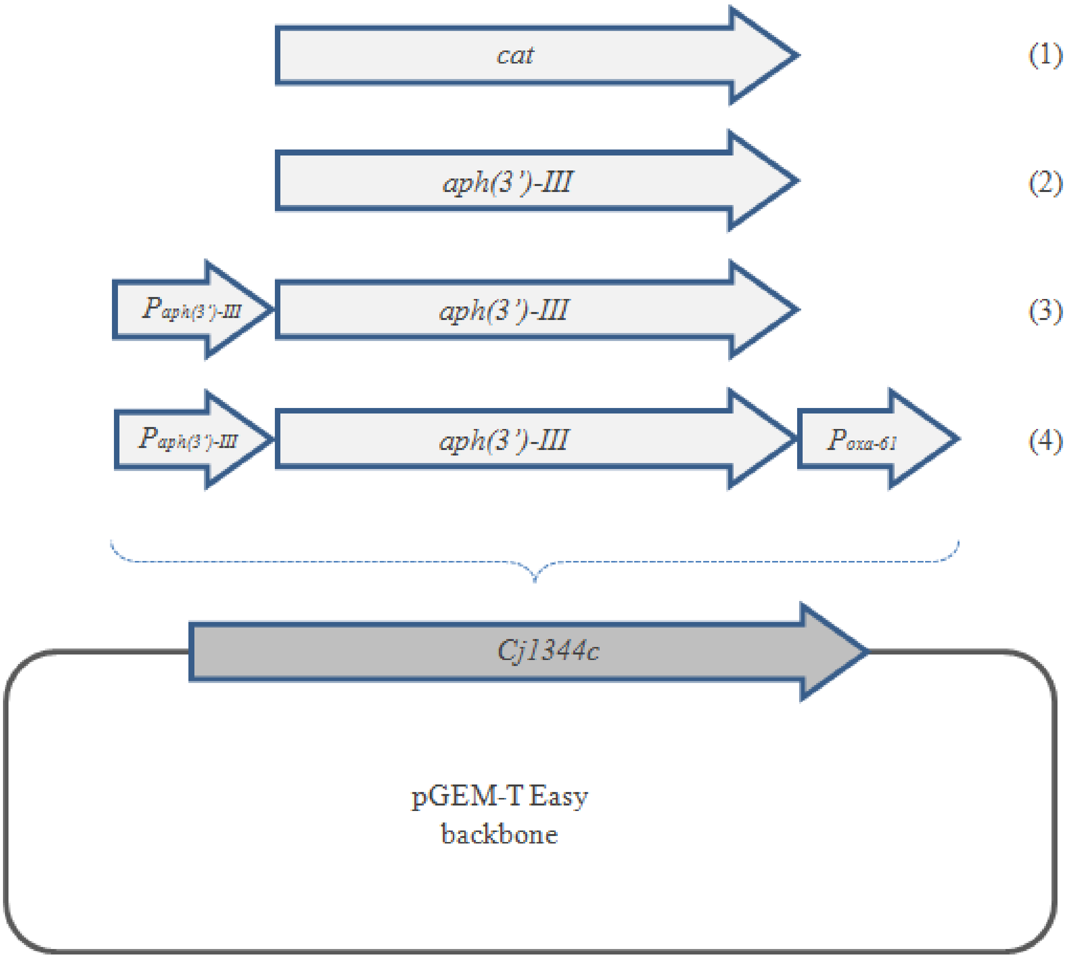


Figure 4.2 Schematic presentation of the different antibiotic resistance gene constructs used for creation of the suicide plasmids for *Cj1344c* mutagenesis

<u>Legend:</u> (1) promoter-less *cat* gene (Karin *et al.*, 2003)

(2) promoter-less *aph(3')-III* gene (This study)

(3) *aph(3')-III* gene including *aph(3')-III* gene promoter (Alfredson & Korolik, 2003)

(4) *aph(3')-III* gene including *aph(3')-III* promoter and *oxa-61* promoter downstream (This study)

4.2.3 Construction of the suicide plasmid backbone for insertion of an antibiotic resistance cassette

In order to create a suicide plasmid for mutagenesis; containing the interrupted *Cj1344c* gene by one of the antibiotic resistance cassettes, a previously created cloning intermediate pGU0501 (Chapter 3; Section 3.3.3) was used as a template to create plasmid backbone for insertion of the antibiotic cassette. A unique *Bgl*II restriction enzyme site within the *Cj1344c* gene was created by the inverse PCR method by incorporating restriction endonuclease sites at the 5’ and 3’ termini of the primer sequences. The primers used in this study (Materials and Methods; Table 2.4) were designed based on the genome sequence of *C. jejuni* NCTC 11168 and were positioned within the open reading frame of the *Cj1344c* gene creating a small deletion (40bp) within the gene (Figure 4.3). *Bgl*II restriction endonuclease sites sequence and 40bp deletion were carefully positioned so that the subsequent insertion of an antibiotic resistance cassette at the newly created *Bgl*II site within *Cj1344c* would be in frame with the start and stop codons of *Cj1344c*.

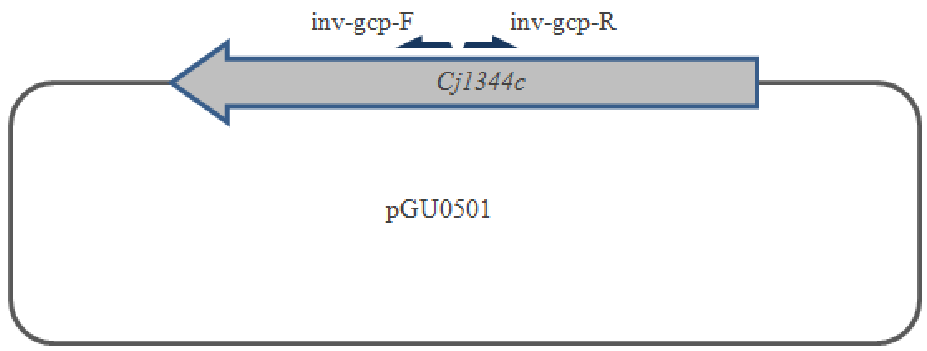


Figure 4.3 Schematic diagram of inverse PCR primers position within pGU0501 plasmid

The inverse PCR product was amplified using standard PCR conditions stated in Chapter 2 (Figure 4.4). Self ligation of the PCR product that had been cleaved with *Bgl*II resulted in the formation of plasmid pGU0501*Bgl*II. pGU0501*Bgl*II was transformed into *E. coli* DH5α using standard cloning protocols. *E. coli* colonies were screened to confirm the presence of the insert by PCR using the *Cj1344c* gene specific primers followed by restriction enzyme digest to ensure the presence of a newly created Bg*l*II site.

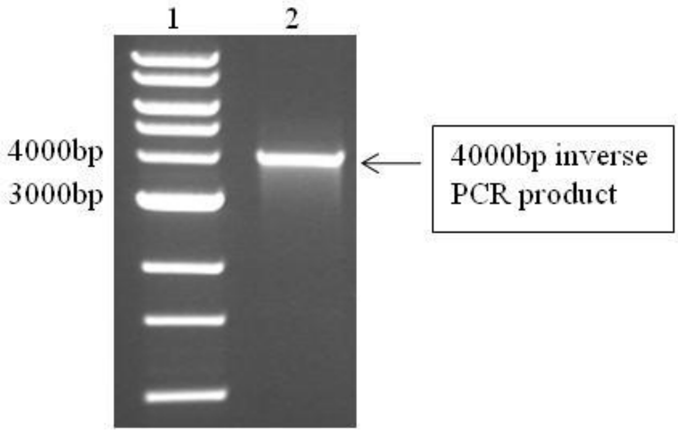


Figure 4.4 Amplification of the inverse PCR product using pGU0501 as a template

4.2.4 Construction of suicide plasmids for mutagenesis of *Cj1344c*

4.2.4.1 Using the promoter-less *cat* gene cassette

The advantage of insertion of the coding region of the chloramphenicol resistance gene without its promoter and in frame with the start codon of *Cj1344c* is expected co-expression of the *cat* gene and *Cj1344c* without interruption of the expression of downstream genes.

The disadvantage of this approach is the possibility that the newly formed fusion protein of *Cj1344c* and chloramphenicol acetyltransferase (encoded by *cat*) may fold incorrectly resulting in the absence of chloramphenicol resistant phenotype. In addition, it needs to be considered that the strength of *Cj1344c* promoter is unknown, which raises a possibility of low levels of expression of *cat* gene product.

Construction of the cassette. The *C. jejuni* chloramphenicol resistance gene was amplified from the pAV110 plasmid (Ketley, J., personal communication) by PCR. The primer set (Table 2.4; Materials and Methods) was designed based on the sequence of pAV110 plasmid to amplify the chloramphenicol gene excluding the transcriptional termination signal. In addition, *Bgl*II restriction endonuclease sites at the 5'and 3' termini of the primers were included to aid in cloning of the PCR product into pGU0501*Bgl*II plasmid. The generated PCR product was cloned into the pGEM-T Easy and the recombinant plasmid was transformed into *E. coli* DH5α using standard cloning protocols (Materials and Methods; Section 2.6). Colonies containing recombinant plasmids were selected by growing on LB agar supplemented with ampicillin and IPTG/X-gal which allowed for white/blue colony selection. *E. coli* colonies were screened for the presence of the insert by PCR using the *cat* specific primers; and subsequently by *Bgl*II restriction enzyme digest which released the cloned fragment from the pGEM-T Easy backbone. The new recombinant plasmid was named pGU0804.

Construction of the suicide plasmid. The *cat* cassette released from the pGU0804 plasmid by *Bgl*II digest was ligated into the linearised pGU0501*Bgl*II plasmid, and then transformed into *E. coli* DH5α using standard cloning protocols (Materials and Methods; Section 2.6). *E. coli* colonies transformed with recombinant plasmid carrying promoterless *cat* cassette could not be recovered when grown on media containing chloramphenicol. Consequently *E. coli* clones were grown on media supplemented with ampicillin relying on the ampicillin resistance gene of the cloning plasmid for selection. The poor expression of the campylobacter specific antibiotic resistance genes in *E. coli* was previously reported (Alfredson & Korolik, 2005). However, the levels of expression of these genes recovered to wild type level once the plasmid construct was transformed into *C. jejuni* (Alfredson & Korolik, 2005). Recombinant plasmid DNA was isolated from *E. coli* cells (Materials and Method; Section 2.5). *Not*I restriction enzyme digest of newly created suicide plasmid was used to confirm the insertion of *cat* cassette within *Cj1344c* gene (Figure 4.5). The plasmid construct was named pGU0805 (nucleotide sequence and map are shown in Appendix F).

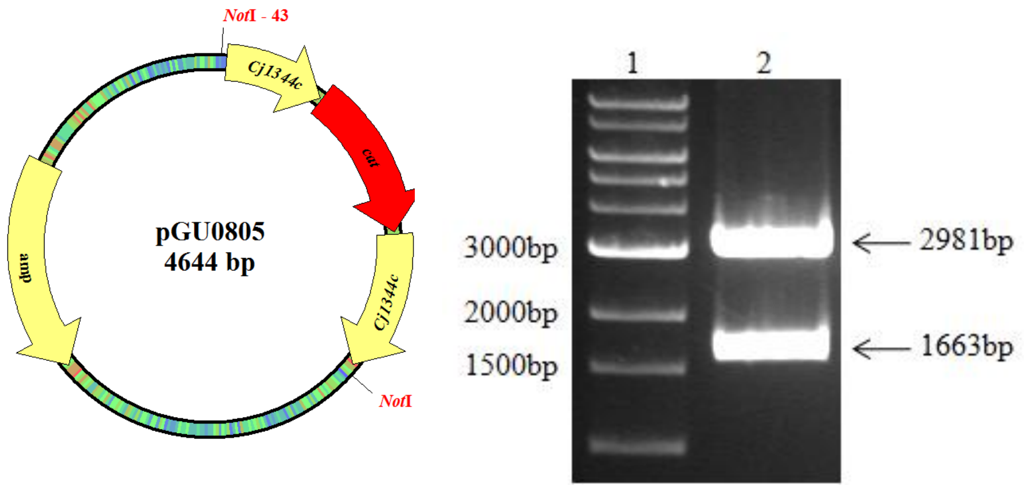


Figure 4.5 *Not*I restriction enzyme digest showing excision of a mutated copy of *Cj1344c* gene by insertion of *cat* cassette

Mutagenesis of *C. jejuni* strain NCTC11168. In order to mutate the *Cj1344c* gene pGU0805 plasmid was transformed into *C. jejuni* strain NCTC11168 using natural transformation and electro-transformation (Materials and Methods). Five attempts to create isogenic mutants of *Cj1344c* using natural and electro-transformation failed to produce viable colonies. The transformation efficiency was tested in each case by transforming *C. jejuni* cells with the pBF6 vector (Bleumink-Pluym *et al.*, 1999). pBF6 is a suicide vector constructed by inserting the *flaA* and *flaB* genes of *C. jejuni* 81116 separated by a kanamycin resistance gene cassette in pBluescript vector. In every instance, pBF6 transformations produced kanamycin resistant *C. jejuni* colonies.

4.2.4.2 Using the promoter-less *aph(3')-III* gene cassette

A Campylobacter-derived kanamycin resistance gene coding region (*aph(3')-III*) is commonly used for mutagenesis of *C. jejuni* genes (Bleumink-Pluym *et al.*, 1999, Myers & Kelly, 2005, Hartley *et al.*, 2009).

The advantage of the insertion of the coding region of the kanamycin resistance gene without its promoter and in frame with the start codon of *Cj1344c* is the expected co-expression of the *aph(3')-III* gene and *Cj1344c* without interruption of the expression of downstream genes.

The disadvantage of this approach, as in case with *cat* gene, is the possibility of low expression due to the strength of *Cj1344c* promoter or incorrect folding of the fusion protein resulting in a non-kanamycin resistant phenotype.

Construction of the cassette. The *C. jejuni* promoter-less kanamycin resistance gene cassette was amplified from the pBF6 plasmid (Bleumink-Pluym *et al.*, 1999) by the PCR method. Primer sets were designed to amplify the coding region of the *aph(3')-III* (Figure 4.2). The transcriptional termination signal of the resistance gene was not amplified. *Bgl*II restriction endonuclease sites were incorporated into the primer sequence at the 5' and 3' termini to aid in cloning of the product into the pGU0501*Bgl*II plasmid. The generated PCR product was visualised by gel electrophoresis (Figure 4.6). The product was excised from the gel and purified for subsequent cloning into pGEM-T Easy vector, followed by transformation into *E. coli* DH5α using standard cloning protocols (Materials and Methods; Section 2.6).

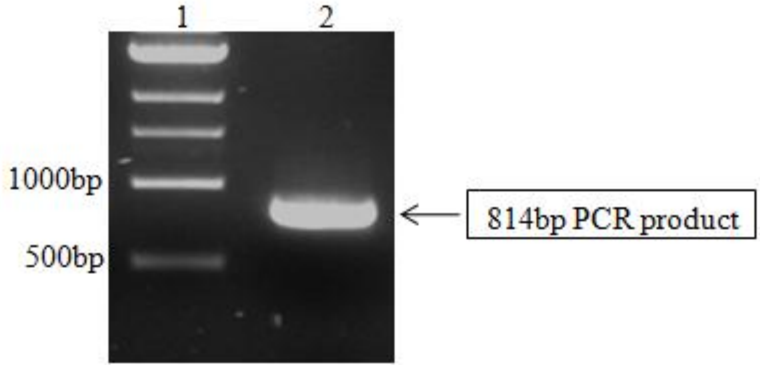


Figure 4.6 Amplification of promoter-less *aph(3')-III* gene cassette

Legend: **Lane 1** – 1Kbp DNA marker; **Lane 2** – PCR amplification of coding part of *aph(3')-III* gene excluding the promoter sequence

The *E. coli* colonies were grown on LB agar supplemented with ampicillin and IPTG/X-gal utilising disruption of *lac*Z gene by an insert and allowing blue/white colony selection. Screening of the colonies was done by PCR using *aph(3')-III* gene specific primers; and subsequently by *Bgl*II restriction enzyme digest which released the cloned fragment from the pGEM-T Easy backbone. The new recombinant plasmid was called pGU0706.

Construction of the suicide plasmids. The kanamycin cassette released from the pGU0706 plasmid by *Bgl*II digest was ligated into the linearised pGU0501*Bgl*II plasmid, and transformed into *E. coli* DH5α using standard cloning protocols (Materials and Methods; Section 2.6). As was the case with the promoter-less *cat* cassette, *E. coli* colonies transformed with recombinant plasmid carrying promoter-less kanamycin cassette could not be grown on media supplemented with kanamycin; and were therefore grown on ampicillin supplemented media, utilising ampicillin resistance gene of the cloning vector for selection. Recombinant plasmid DNA was isolated from *E. coli* cells using plasmid mini prep kit (Materials and Method; Section 2.5) and screened by *Not*I restriction enzyme digest (not shown), which confirmed the insertion of the promoter-less kanamycin cassette within *Cj1344c*. The plasmids construct was named pGU0707 (nucleotide sequence and map are shown in Appendix G).

Mutagenesis of *C. jejuni* strain NCTC11168. Mutagenesis of *Cj1344c* gene was carried out by transformation of pGU0707 plasmid into *C. jejuni* NCTC 11168 using natural transformation and electro-transformation methods (Materials and Methods). Five attempts to create isogenic mutants of *Cj1344c* using these procedures failed to produce viable *C. jejuni* colonies. Transformation efficiency was tested in each case by transforming *C. jejuni* cells with pBF6 vector (Bleumink-Pluym *et al.*, 1999). In every instance, pBF6 transformations produced kanamycin resistant *C. jejuni* colonies.

4.2.4.3 Using the promoted *aph(3')-III* gene cassette

The advantage of using a promoted cassette would be an independent expression of *aph(3')-III* gene ensured by the presence of the gene promoter sequence, and thus ensuring the kanamycin resistant genotype in case of successful cross-over event.

The disadvantage of this approach is the potential creation of a polar mutation as the expression of genes downstream from the cassette insertion may be reduced or eliminated due to an increased distance between their start codon and putative common promoter upstream from *aph(3')-III* insertion point. On the other hand, the expression of these genes may be upregulated due to a presence of a strong *aph(3')-III* promoter.

Construction of the cassettes. *C. jejuni* kanamycin resistance cassette gene cassette was amplified from pBF6 plasmid (Bleumink-Pluym *et al.*, 1999) by PCR method. Primer sets were designed to amplify the coding region and the promoter sequence of the *aph(3')-III* gene (Figure 4.2). As in the case with all other cassettes, the transcriptional termination signal of the resistance gene was excluded. *Bgl*II restriction endonuclease sites were incorporated into the primer sequence at the 5' and 3' termini to aid in cloning of the product into pGU0501*Bgl*II plasmid. The generated PCR product was visualised by gel electrophoresis (Figure 4.7). The PCR product was excised from the gel, purified and cloned into the pGEM-T Easy vector, followed by transformation into *E. coli* DH5α using standard cloning protocols (Materials and Methods; Section 2.6).

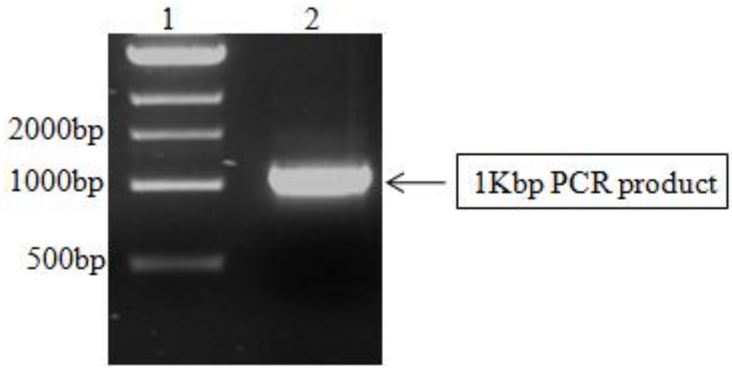


Figure 4.7 Amplification of promoted *aph(3')-III* gene cassette

Legend: Lane 1 – 1Kbp DNA marker; **Lane 2** – PCR amplification of fully functional *aph(3')-III* gene including its promoter

The *E. coli* colonies were grown on LB agar supplemented with ampicillin and IPTG/X-gal utilising disruption of *lac*Z gene by an insert and allowing blue/white colony selection, as well as kanamycin antibiotic. Screening of the colonies was done by PCR using *aph(3')-III* gene specific primers; and subsequently by *Bgl*II restriction enzyme digest which released the cloned fragment from the pGEM-T Easy backbone. The new recombinant plasmid was called pGU0509.

Construction of the suicide plasmids. The kanamycin cassette released from the pGU0509 plasmid by *Bgl*II digest were ligated into the linearised pGU0501*Bgl*II plasmid, and transformed into *E. coli* DH5α using standard cloning protocols (Materials and Methods; Section 2.6). Recombinant plasmid DNA was isolated from *E. coli* cells (Materials and Method; Section 2.5) and screened by *Not*I restriction enzyme digest (not shown), which confirmed the insertion of the promoted kanamycin cassette within *Cj1344c*. The plasmids construct was named pGU0613 (nucleotide sequence and map are shown in Appendix H).

Mutagenesis of *C. jejuni* strain NCTC11168. Mutagenesis of *Cj1344c* gene was performed by transformation of pGU0613 plasmid into *C. jejuni* NCTC 11168 using natural transformation and electro-transformation methods (Materials and Methods). Five attempts to create isogenic mutants of *Cj1344c* using these procedures failed to produce viable *C. jejuni* colonies. Transformation efficiency was tested in each case by transforming *C. jejuni* cells with the pBF6 vector (Bleumink-Pluym *et al.*, 1999). In every instance, the pBF6 transformations produced kanamycin resistant *C. jejuni* colonies.

4.2.4.4 The promoted *aph(3')-III* gene cassette including additional *oxa-61* promoter downstream

The advantage of using this cassette construct was two-fold; the presence of a fully functional promoter sequence of *aph(3')-III* gene ensures expression of the gene while addition of another campylobacter promoter sequence immediately downstream from the *aph(3')-III* gene maximises the chances of expression of downstream genes, which may otherwise be effected by the insertion of *aph(3')-III* gene.

The disadvantages of this method are the potential polar mutation effects due to dis-regulation of genes downstream from the insertion point as additional *oxa-61* promoter can lead to an upregulated gene expression of the downstream genes that usually may be expressed at low levels.

Construction of the cassettes. In order to construct the kanamycin resistance cassette containing the additional promoter sequence, the ampicilin (*oxa-61*) gene promoter region ($P_{oxa\text{-}61}$) was amplified by PCR using a set of primers (Table 2.4) designed based on the sequence of pGU0401 plasmid (Alfredson & Korolik, 2005). The *Xba*I restriction endonuclease sites at the 5'and 3' termini were included in the primer design to allow cloning of the DNA fragment downstream of the kanamycin resistance gene in the pGU0509 plasmid. The 174bp PCR product was cleaved with *Xba*I and cloned into *Xba*I linearised pGU0509 to allow insertion of $P_{oxa\text{-}61}$ DNA fragment downstream from the kanamycin resistance gene (Materials and Methods). The resulting recombinant plasmid DNA was transformed into *E. coli* competent cells using standard cloning procedures (Materials and Methods). Recombinant plasmid DNA was isolated from *E. coli* cells and screened by PCR using a combination of kanamycin specific primers and P_{oxa-61} specific primers to confirm the correct orientation of the P_{oxa-61} insert in relation to the kanamycin gene (Figure 4.8). The new plasmid construct was named pGU0522.

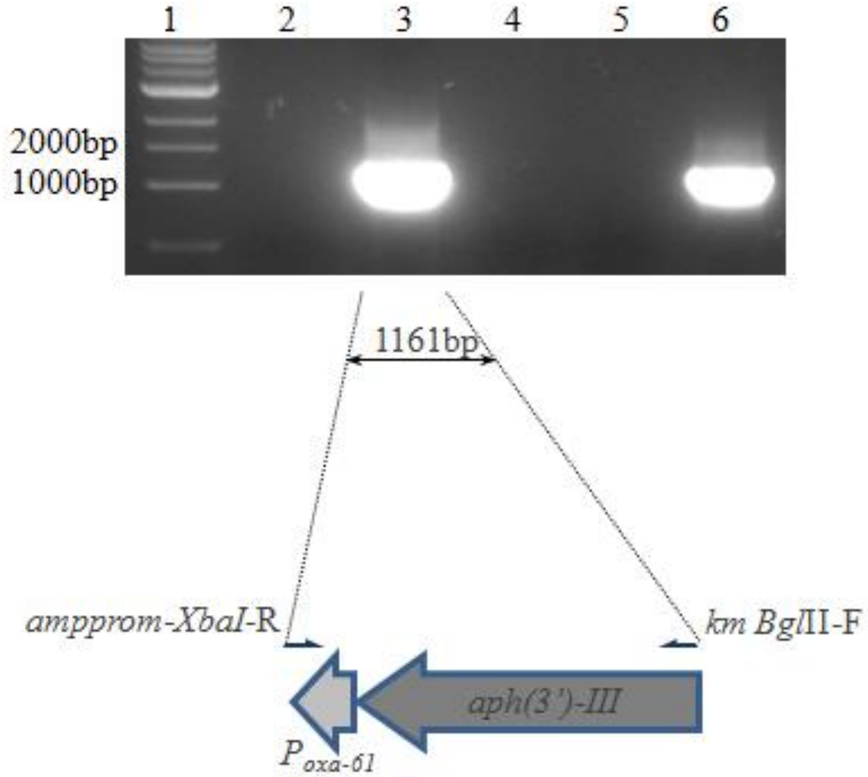


Figure 4.8 Screening for the correct orientation of P_{oxa-61} in relation to *aph(3')-III* gene

Construction of the suicide plasmid. The kanamycin cassette containing the *oxa-61* promoter sequence was released from the pGU0522 plasmid by *Bgl*II digest; and the DNA fragment was ligated into the linearised pGU0501*Bgl*II plasmid, and transformed into *E. coli* DH5α using standard cloning protocols (Materials and Methods; Section 2.6). Recombinant plasmid DNA was isolated from *E. coli* (Materials and Method; Section 2.5) and screened by *Not*I restriction enzyme digest (not shown). The newly created suicide plasmids showed insertion of the kanamycin cassette containing the *oxa-61* promoter within *Cj1344c*. The plasmid construct was named pGU0523 (nucleotide sequence and map are shown in Appendix I).

Mutagenesis of *C. jejuni* strain NCTC11168. Mutagenesis of *Cj1344c* gene was carried out by transformation of pGU0523 plasmid into *C. jejuni* NCTC11168 using natural transformation and electro transformation methods (Materials and Methods). Five attempts to create isogenic mutants of *Cj1344c* using these two methods failed to produce viable colonies. The transformation efficiency was tested in each case by transforming *C. jejuni* cells with pBF6 vector (Bleumink-Pluym *et al.*, 1999). In every instance, pBF6 transformations produced kanamycin resistant *C. jejuni* colonies.

In summary, the mutagenesis of *Cj1344c* gene was attempted using four different suicide plasmids: pGU0523, pGU0613, pGU0707 and pGU0804; all of which contain a mutated copy of the *Cj1344c* gene by different antibiotic resistance cassette (Table 4.1). The plasmids were transformed into *C. jejuni* NCTC 11168 using natural transformation and electro-transformation. Five attempts of isogenic mutant creation by each method failed to produce viable *C. jejuni* colonies. Every transformation attempt was accompanied by transformation of *C. jejuni* cells with a control plasmid pBF6 which produced kanamycin resistant *C. jejuni* colonies. It was speculated that the reasons for the unsuccessful recovery of the *Cj1344c* isogenic mutant could be a cross-over inefficiency during transformation which was further examined.

Table 4.1 Suicide vectors used for mutagenesis of *Cj1344c*

Recombinant suicide plasmid name	*Cj1344c* mutated by	Promoter region included	Mutagenesis of *C. jejuni*
pGU0523	*aph(3')-III*	$P_{aph(3')\text{-}III}$ / $P_{oxa\text{-}61}$	unsuccessful
pGU0613	*aph(3')-III*	$P_{aph(3')\text{-}III}$	unsuccessful
pGU0707	*aph(3')-III*	none	unsuccessful
pGU0805	*cat*	none	unsuccessful

4.2.5 Strategy for analysis of non-recoverable cross-over events during *C. jejuni* transformations

Isogenic mutant creation by inactivating a gene may result in an inability to recover isogenic mutants when the gene is essential for bacterial survival. To confirm that the potential lack of recovery of the isogenic mutant was a result of a lethal mutation, and not an error during the transformation procedure, the possibility of a cross-over event needed to be analysed. *C. jejuni* requires 202bp of homologous DNA for a successful cross-over event and incorporation of foreign DNA into the genome (Wassenaar *et al.*, 1993). By analysing a transformation mix by PCR using a set of primers positioned in the genome, outside the integration event (to avoid amplification of a PCR product from a suicide vector used in transformation) and antibiotic resistance gene cassette it would be possible to confirm a successful cross-over event in the transformation mixture (Figure 4.9).

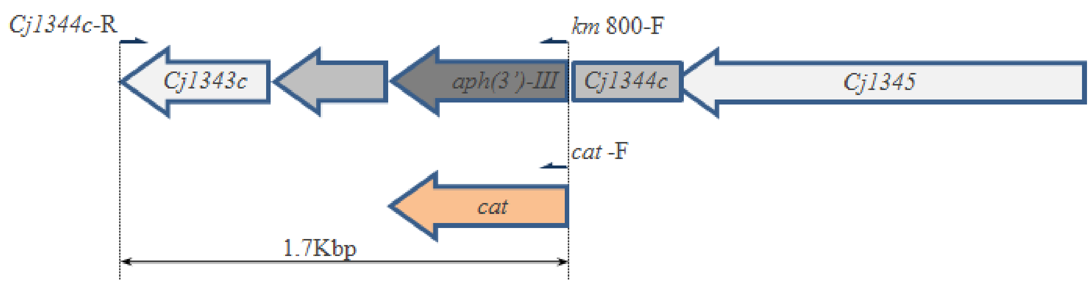


Figure 4.9 Schematic representation of positions of a primer pair used for confirmation of double cross-over event in cases of the *aph(3')-III* gene cassette (top) and the *cat* cassette (bottom)

4.2.6 Assessment of the cross-over event during *C. jejuni* transformation.

To confirm the cross-over event and incorporation of a mutated version of the gene into the *C. jejuni* genome, the transformation mix was analysed by PCR as described in the strategy section 4.2.5. PCR primers used in the experiment were designed based on the sequence of genome strain of *C. jejuni* NCTC 11168.

DNA from the transformation mix was isolated as stated in Materials and Methods. PCR analysis was performed using this DNA and the set of primers positioned in the *Cj1344c* downstream gene and antibiotic resistance cassette. The results of this experiment confirmed that the cross-over event did occur and the mutated gene was incorporated into *C. jejuni* genome (Figure 4.10). The inability to produce isogenic mutants suggests that *Cj1344c* gene product is essential for bacterial survival. Supplementation of the media with this gene product could potentially enable recovery of the isogenic mutants.

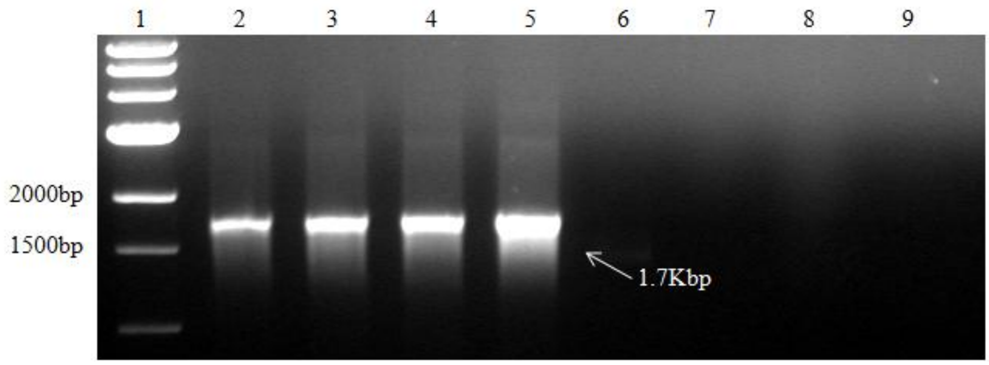


Figure 4.10 DNA amplification from transformation mixture, confirming double cross-over event during *C. jejuni* transformation

Legend: **Lane 1** – 1Kbp DNA marker; **Lane 2** – transformation mixture pGU0523; **Lane 3** – transformation mixture pGU0613; **Lane 4** – transformation mixture pGU0707; **Lane 5** – transformation mixture pGU0805; **Lane 6** – control pGU0523; **Lane 7** – control pGU0613; **Lane 8** – control pGU0707; **Lane 9** – control pGU0805

4.2.7 Attempts to culture isogenic mutants of *Cj1344c* in conditioned media

Preliminary studies have identified that the glycoprotease is secreted into the culture media (Chapter 3). It was therefore, postulated that growing putative isogenic mutants in media where wild type *C. jejuni* was cultured and then removed (conditioned media) may enhance the chances of recovery of isogenic mutants. Conditioned media was prepared as described in the Materials and Methods; Section 1.4.2 and was supplemented with kanamycin or chloramphenicol for selection of isogenic mutants. *C. jejuni* cells were transformed with pGU0523, pGU0613, pGU0707, pGU0804 and pBF6 as a control and the transformation mixes were incubated in conditioned media under standard incubation conditions (Materials and Methods; Section 2.6). OD_{600} readings were taken immediately after transformation and 48h post transformation to monitor *C. jejuni* growth. After 48h incubation no growth was detected in any of the cultures transformed with suicide plasmids. *C. jejuni* cells transformed with control plasmid pBF6 (Bleumink-Pluym *et al.*, 1999) had significant growth at 48hr time point compared to zero point time (Figure 4.11). This experiment further confirmed that the mutation of *Cj1344c* is lethal and that presence of the Cj1344c in culture media cannot aid recovery of *Cj1344c* isogenic mutants, suggesting the intra-cellular function of the enzyme or inability of Cj1344c uptake into *C. jejuni* from culture media.

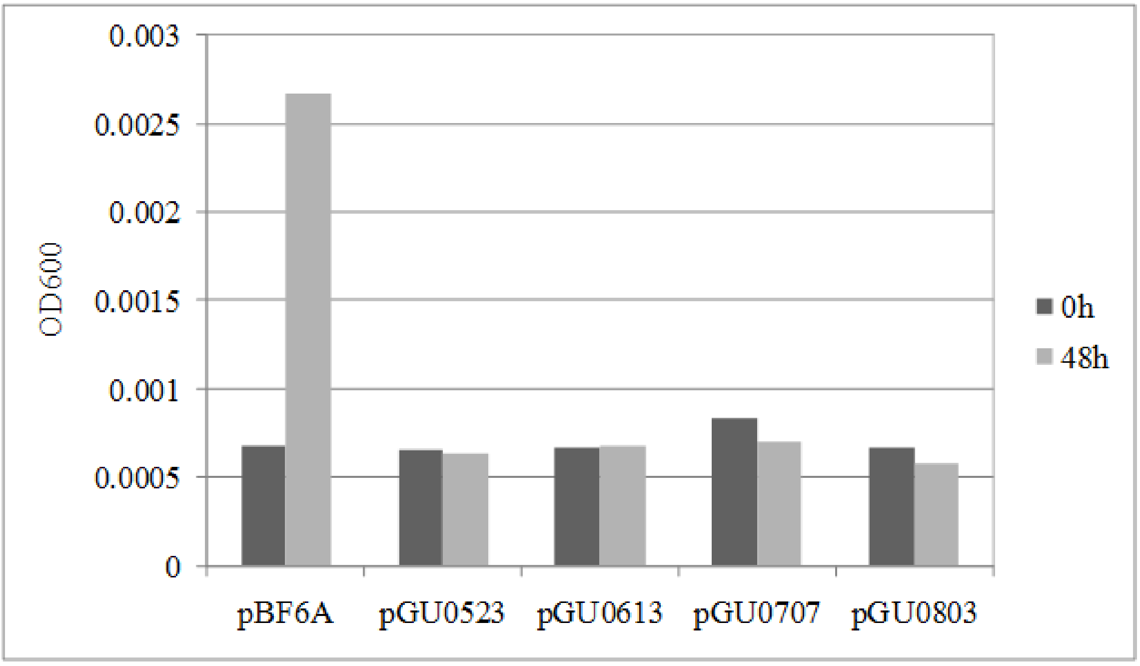


Figure 4.11 Putative *C. jejuni* isogenic mutants growth in conditioned media

4.3 Discussion

An insertional inactivation mutagenesis approach was taken in order to perform the functional analysis of *C. jejuni* Cj1344c, which shows 36% identity and 55% similarity to *O*-sialoglycoprotease of *Mannheimia (Pasteurella) haemolytica* A1. Five attempts to mutate the gene by insertional inactivation failed to produce a *Cj1344c* isogenic mutant suggesting that the gene may be essential for *C. jejuni* growth. These results are similar to those found by Zheng and colleagues who found glycoprotease to be essential for growth of *S. aureus* using regulated gene expression (Zheng *et al.*, 2005). Down-regulation of *S. aureus gcp* resulted in increased sensitivity of the bacteria to Zn^{2+}. Similarly, a *gcp* knockout mutant in *M. haemolytica* was found to be not viable (Mellors, personal communication), while downregulation of *gcp* expression in *S. aureus* inhibited bacterial growth (Zheng *et al.*, 2005) suggesting that the *gcp* gene may be one of the housekeeping genes necessary for bacterial survival.

Attempts to inactivate *Cj1344c* by four different antibiotic resistance cassettes with different characteristics did not result in the creation of the *Cj1344c* isogenic mutant, indicating that inactivation of *Cj1344c* is lethal for *C. jejuni* survival *in vitro*. It needs to be considered that the absence of fully characterised promoter sequences and genes organised in an operon-like structures pose a significant issue when constructing gene knock-out mutants in *C. jejuni* (Parkhill *et al.*, 2000a). By employing the four antibiotic resistance gene cassettes this study attempted addressing different polarity issues that may arise during the creation of the *Cj1344c* isogenic mutant, such as inactivation of the genes downstream from the insertion of an antibiotic resistance cassette.

The mutagenesis of the gene was performed using natural transformation and electro-transformation of *C. jejuni* cells with the suicide plasmids to account for the reported variability in transformation efficiency observed among *C. jejuni* strains (Wassenaar *et al.*, 1993). In addition, transformation efficiency of *C. jejuni* was tested by transforming the cells with a control plasmid pBF6, carrying *aph(3')-III* gene positioned between *C. jejuni flaA* and *flaB* genes which produced kanamycin resistant *C. jejuni* colonies in every experiment. This finding confirmed that *C. jejuni* cells used in the transformation were naturally transformable or electro-transformable as reported in literature (Miller *et al.*, 1988) and that non-recovery of isogenic mutant was not due to transformation inefficiency, but due to an essential gene inactivation.

The five unsuccessful attempts to mutate the gene posed the question of cross-over efficiency during transformation., Analysis of the transformation mix by PCR was performed to answer this question. The minimum of homologous base pairs required for successful cross over event and incorporation of antibiotic resistance-encoding genes into genome was reported to be 202bp (Wassenaar *et al.*, 1993), and this study followed that recommendation by allowing 500bp of homologous DNA. Utilisation of kanamycin resistance gene specific primers and genome specific primers showed that the double cross-over event took place and that the mutated copy of *Cj1344c* gene was incorporated into the *C. jejuni* genome by a homologous recombination event. This experiment confirmed that the non-recovery of isogenic mutants was not due to the cross-over inefficiency but an essential role of this gene in *C. jejuni*.

Growing newly transformed *C. jejuni* cells in media where wild type *C. jejuni* was previously grown was hypothesised to aid in the isogenic mutant recovery as the presence of Cj1344c in the culture media may complement the lack of the enzyme in isogenic mutants, as reported in case of phosphonate degradation genes (Hartley *et al.*, 2009). Growing transformed *C. jejuni* cells in conditioned media, however, did not result in the recovery of isogenic mutants suggesting that the Cj1344c transfer mechanism across the cell wall may be in one direction only. In addition, the medium was supplemented with the *E. coli* expressed and purified His-Cj1344c protein without success in mutant recovery. This experiment also suggests that Cj1344c may have different functions, both intra and extra-cellularly.

4.3.1 Conclusion

This chapter describes construction and utilisation of different antibiotic resistance gene cassettes to mutate *C. jejuni* putative glycoprotease gene. The position of the gene within the genome in close proximity with other genes of unknown function posed a problem in creating an isogenic mutant of the gene, due to possibility of polar mutation. Inactivation of the gene by four different antibiotic resistance cassettes confirmed the essential function of the gene, as isogenic mutants could not be recovered, confirming data published by other research groups in different bacterial species.

CHAPTER 5

Raising antibodies against the recombinant Cj1344c protein in a rabbit; native protein purification and enzymatic analysis

5.1 Introduction

Expression of the recombinant *C. jejuni gcp* gene (*Cj1344c*) in an *E. coli* host did not yield a protein with a detectable level of activity, probably due to lack of posttranslational modification, incorrect protein folding or lack of a co-factor (This study, Chapter 3). This is similar to the *M. haemolytica* recombinant protein expressed in an *E. coli* system, which was reported to show a marked decrease in activity as assayed by glycophorin A digest (Watt *et al.*, 1997a).

To purify and determine the enzymatic function of native *C. jejuni* Cj1344c, polyclonal antibodies raised against recombinant Cj1344c protein in a rabbit could be utilised to isolate the native *C. jejuni* protein. Polyclonal antibodies raised against recombinant protein expressed in *E. coli* have been previously shown to recognise the native form of similar proteins (Ermolova *et al.*, 2003) by recognising similar epitomes on both the recombinant and native forms of the protein.

Polyclonal antibodies may also assist in the identification of Cj1344c localisation within the *C. jejuni* cell as the *C. jejuni* gcp homologue cell localisation is not known. *M. haemolytica* gcp was shown to be present in the culture supernatant (Otulakowski *et al.*, 1983) despite the absence of the conventional peptide secretion signal sequence. In addition, the antibodies may be used to identify *C. jejuni* Cj1344c homologues in related bacterial species.

5.2 Results

5.2.1 Purification and preparation of His-tagged Cj1344c recombinant protein

In order to produce polyclonal antibodies against *C. jejuni* Cj1344c, the protein was expressed in *E. coli* BL21 host using the pET19-b expression system and purified using Nickel affinity resin, as described in Chapter 3. The apparent molecular weight of the purified recombinant protein produced by BL21 cells was approximately 40KDa, as determined by SDS-PAGE (Chapter 3; Figure 3.8). This is consistent with the predicted molecular weight of 39,850Da for a fusion protein which includes molecular weight 37,066Da of Cj1344c and 2,784 from the His tag.

The purification procedure using the Nickel affinity resin method yielded 90% pure His tagged protein. In order to minimise the rabbit antibody response to *E. coli* proteins co-eluted with His-Cj1344c, further purification of His-Cj1344c was performed by size separation of proteins on an SDS-PAGE as detailed in Chapter 2; Section 2.7. The protein band of 40KDa was confirmed to contain His-Cj1344c protein by Western blot analysis using anti-His antibody. This band was excised from the gel and sent to the IMVS facility to raise antibodies against His-Cj1344c (data not shown).

5.2.2 Production and specificity of the anti-His-Cj1344c antibodies

Polyclonal antibodies against His-Cj1344c were raised in a rabbit as described in the Materials and Methods Section 2.11. Serum samples were collected on the day of the primary immunization, in week 7 after the 3rd immunisation, and in week 10 after the 4th and final immunisation.

In order to determine the specificity of the rabbit serum to His-Cj1344c purified protein and native Cj1344c in *C. jejuni* whole cell lysate, Western blot analysis was performed using the rabbit serum as a primary antibody as described in the Materials and Methods section. The Western blot analysis identified a 38KDa protein in the *C. jejuni* lysate which was hypothesised to be the native Cj1344c based on the expected protein molecular weight) in addition to the His-Cj1344c protein (Figure 5.1). The cross reactivity between the His-Cj1344c protein and native Cj1344c protein was subsequently used as a basis for native protein purification of *C. jejuni* Cj1344c from a whole cell lysate using the rabbit serum.

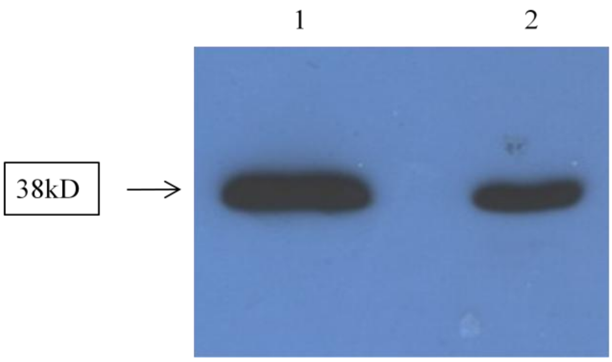


Figure 5.1 Specificity of rabbit sera to Cj1344c protein present in the *C. jejuni* whole bacterial lysate

Legend: Lane 1 – sample of purified His-Cj1344c *E. coli* expressed protein, **Lane 2** – sample of *C. jejuni* lysate showing Cj1344c protein band

Once the specificity of the antibodies in the rabbit serum to Cj1344c was confirmed, serum IgG titres were determined by direct ELISA. His-Cj1344c protein was used to coat ELISA plate wells to evaluate the protein-specific immune response of the primary antibody in the rabbit serum. The His-Cj1344c specific serum IgG titres increased with the number of vaccinations, reaching 1:65,000 after the 4th vaccination (Figure 5.2).

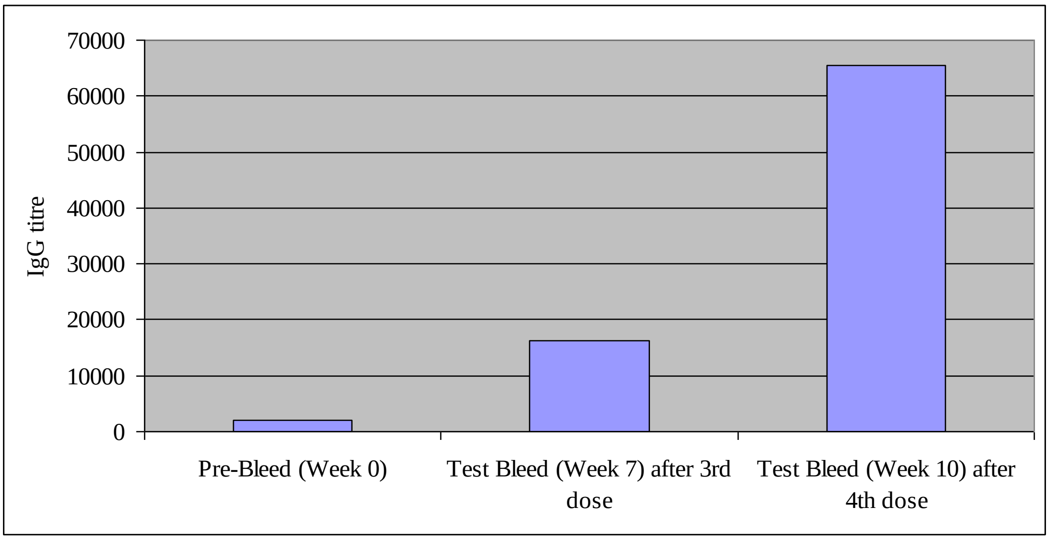


Figure 5.2 Rabbit sera IgG response to His-Cj1344c protein vaccination at different time intervals

5.2.3 Cell localisation of Cj1344c in *C. jejuni* and homologue detection in other bacterial species

In order to investigate the possibility that the *C. jejuni* Cj1344c is secreted into the culture supernatant, the predicted amino acid sequence was analysed for the presence of a signal sequence that would direct the polypeptide to the secretion pathway. The SignalP 3.0 program (Emanuelsson *et al.*, 2007) did not show the presence of the conventional secretion signal sequence nor the cleavage sites within the polypeptide, suggesting that the protein is not secreted via Sec pathway, but this did not preclude secretion of the protein by some alternative mechanism.

To investigate the localisation of Cj1344c protein within the *C. jejuni* cell; and its possible secretion into the culture supernatant, different cell fractions were prepared and analysed for the presence of the protein as described in Chapter 2. Samples of the cytoplasmic fraction, the insoluble cell wall fraction, the periplasmic fraction and the culture media were resolved by SDS-PAGE. Western blot analysis using rabbit anti-His-Cj1344c was used to determine the presence of the protein in the different fractions prepared. The results of these experiments showed the presence of Cj1344c protein in the soluble cytoplasmic fraction and the culture media (Figure 5.3).

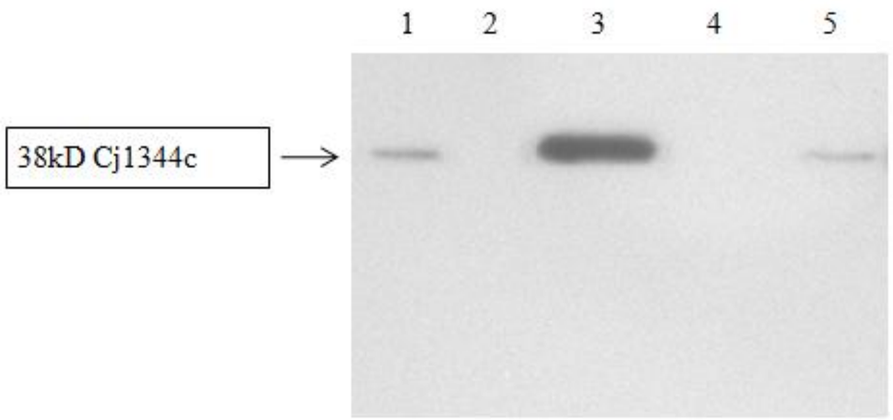


Figure 5.3 Western blot analysis of the different cell fractions for the presence of Cj1344c protein

Legend: **Lane 1** – control His-Cj1344c, **Lane 2** – sample of insoluble cell wall fraction, **Lane 3** – sample of soluble cytoplasmic fraction, **Lane 4** – sample of periplasmic fraction, **Lane 5** – sample of concentrated culture medium (100x)

5.2.4 Cj1344c homologue identification in selected bacterial species

In silico comparative analysis of the Cj1344c predicted amino acid sequence identified homologues of Cj1344c in all the other bacterial species that were analysed (Chapter 3). To determine the cross reactivity of the rabbit antibodies between different *Campylobacter* spp. species, selected bacterial whole cell lysates were analysed using Western Blot method utilising the rabbit anti-His-Cj1344c antibody as described in the Materials and Methods section. The results of the analysis identified a protein of the expected size (37kDa) present in all *C. jejuni* strains assessed (*C. jejuni* 11168 shown here). In addition, the homologue of Cj1344c was also present in all *C. coli* and *C. fetus* strains as well as in *H. pylori* (Figure 5.6).

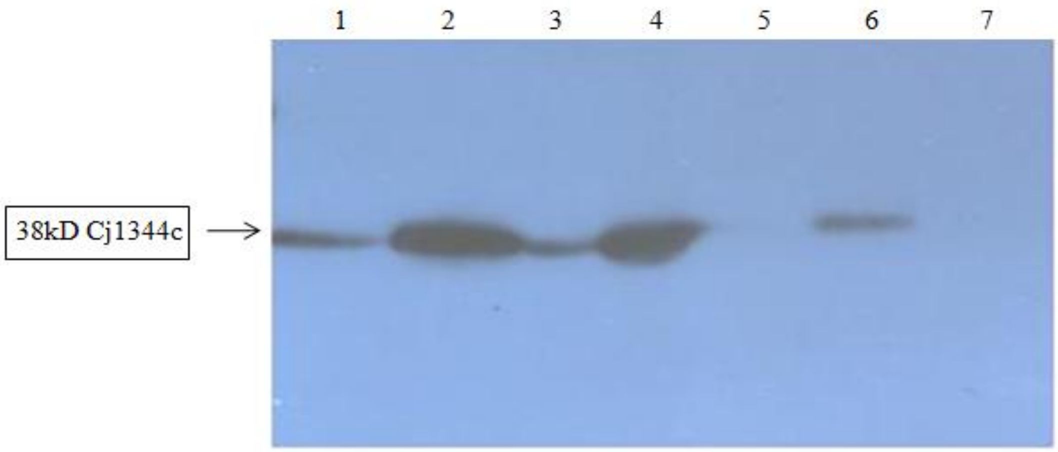


Figure 5.4 Western blot analysis of different bacterial lysates using an anti-His-Cj1344c rabbit antibody

<u>**Legend:**</u> **Lane 1** – purified His-Cj1344c; **Lane 2** – *C. jejuni* NCTC11168; **Lane 3** – *C. coli*; **Lane 4 -** *H. pylori*; **Lane 5** – *S. aureus*; **Lane 6** – *C. jejuni* 81116; **Lane 7** – *E. coli*

5.2.5 Native protein purification

In order to assess the enzymatic activity of the native *C. jejuni* Cj1344c protein, rabbit antibodies raised against recombinant His-Cj1344c protein were used to isolate the native protein from the bacterial lysate. Before the attempt to purify the native protein using anti-His-Cj1344c antibodies, different cell lysates (non-denatured) were tested on a dot blot to determine the specificity of the rabbit antibodies against the native *C. jejuni* Cj1344c (Figure 5.5). The results of the blot show that the antibody raised agains the His-Cj1344c in rabbit, recognises the native Cj1344c protein, or its homologue, in *C. jejuni, C. coli* and *H. pylori*, while at the same time does not exhibit specificity to *E. coli* and *S. aureus* lysates.

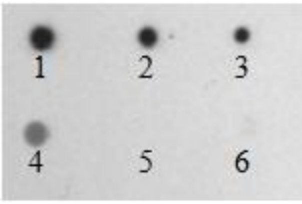


Figure 5.5 Dot blot analysis of different bacterial lysates using an anti-His-Cj1344c rabbit antibody

<u>Legend:</u> **Spot 1** – purified His-Cj1344c; **Spot 2** – *C. jejuni* NCTC11168; **Spot 3** – *C. coli*; **Spot 4 -** *H. pylori*; **Spot 5** – *S. aureus*; **Spot 6** – *E. coli*

The specificity of the raised antibodies enabled their use in native protein isolation using the Dynabead M-280 purification system, as described in the Materials and Methods. The anti-His-Cj1344c antibodies, bound covalently to sheep anti-rabbit IgG antibodies on Dynabeads were used to isolate native Cj1344c from the *C. jejuni* cell lysate. The native protein recognised by the anti-His-Cj1344c was subsequently eluted off the beads by disrupting the covalent antigen-antibody bonds.

This elution step did not result in the elution of the native protein from the beads. Despite multiple attempts to purify the native protein using this system, native Cj1344c protein could not be isolated from the whole cell *C. jejuni* lysate. Alternative methods will have to be employed to accomplish this experiment in the future.

5.3 Discussion

To determine the activity and a potential role of the *C. jejuni* Cj1344c protein in bacterial pathogenesis, an attempt to isolate the native protein was undertaken by utilising antibodies raised against the recombinant protein.

Over-expression of the *C. jejuni* protein in *E. coli* resulted in a lack of enzymatic activity (Chapter 3), possibly due to the lack of posttranslational modification of the protein in *E. coli*, incorrect folding of the polypeptide during biosynthesis or lack of an unidentified enzyme co-factor. Similarly, over-expression studies of *M. haemolytica* gcp in *E. coli* found that the enzyme lacked or showed significantly reduced activity when compared with the native protein (Watt *et al.*, 1997b). Attempts to refold the protein did not result in a significant increase in enzyme activity (Watt *et al.*, 1997a).

To isolate the native protein, His-Cj1344c produced in *E. coli* was used to immunise a rabbit and produce antibodies against this polypeptide. Rabbit sera collected after the third injection of recombinant protein showed high titres of anti-His-Cj1344c antibodies (1:65,000), which was significantly higher than the results obtained from vaccination studies by other *C. jejuni* proteins (Lee *et al.*, 1999, Monteiro *et al.*, 2009). It was hypothesised that the polyclonal antibodies raised against the recombinant protein would recognise different epitomes on the protein surface of both the recombinant protein and the native *C. jejuni* protein, as similar studies in different bacteria have demonstrated cross-reactivity (Ermolova *et al.*, 2003). This cross-reactivity was to be used in isolating native *C. jejuni* Cj1344c protein from the whole cell lysate.

The Western blot analysis of different cellular fractions showed that Cj1344c was present in the cytoplasmic fraction of the cell and the culture media. The predicted amino acid analysis of Cj1344c did not reveal a conventional signal secretion sequence. This is also the case for *M. haemolytica*, gcp where secretion occurs by an alternative secretion mechanism not involving an amino terminus signal (Otulakowski *et al.*, 1983). Detection of the enzyme in the culture media could be attributed to the release of the cytoplasmic enzyme due to cell death and this finding needs to be further investigated. The detection of the enzyme in cytoplasmic fraction of the *C. jejuni* cell as well as in the culture medium, may suggest a dual role of the enzyme in *C. jejuni*. The detection of the enzyme in cytoplasmic fraction, in addition to the inability to create the Cj1344c mutant (Chapter 4) and the interaction of His-Cj1344c with the whole cell fraction of *C. jejuni* (Chapter 3) may suggest an intracellular role of the enzyme in cell homeostasis. On the other hand the presence of the enzyme in the culture supernatant may suggest an extracellular role of the enzyme in bacterial pathogenesis.

Multiple attempts to isolate the native protein using Dynabead M-280 sheep anti-rabbit IgG failed to produce sufficient quantities of the protein to perform enzymatic studies. The standard Western Blot analyses showed that rabbit IgG antibodies cross-reacted with native Cj1344c protein, suggesting that the possible reason for the inefficient isolation of the native protein could have been unsuccessful elution of the protein from IgG antibody coated Dynabeads.

Western blot analysis of whole cell protein from different *C. jejuni* strains using anti-His-Cj1344c rabbit antibody identified cross-reactivity with all *C. jejuni* strains examined, a result that confirms similarity matches identified (100-90% predicted amino acid similarity) in the bioinformatics studies (Chapter 3). In addition, the cross-reactivity was also identified in other *Campylobacter* species tested as well as *H. pylori*. The presence of the Cj1344c homologue in all tested bacterial species may suggest an important role that the enzyme plays in bacterial homeostasis. In addition, its role may extend further in bacterial pathogenesis considering that the enzyme is secreted into the culture supernatant; and is hypothesised to have a role in glycoprotein degradation (Abdullah *et al.*, 1992, Tu *et al.*, 2008).

5.3.1 Conclusions

Recombinant Cj1344c showed a high IgG response in rabbit with titres higher than 1:60,000 which allowed the use of the enzyme in the mice vaccination and protection trials. The specificity of the raised antibody to native *C. jejuni* enzyme was used to detect the homologues of the enzyme in other *Campylobacter* spp as well as in the related microorganism *H. pylori.* Analysis of different cell fractions identified the enzyme in the cytoplasm and culture supernatant of *C. jejuni*, suggesting a possible dual function of the enzyme. The absence of the conventional secretion signal sequence, however, suggests a non-conventional secretion mechanism, probably via chaperone molecules. The enzymatic activity of Cj1344c could not be determined as the purification of the native protein using the Dynabead M-280 system did not produce sufficient quantities of the enzyme. Further investigation of the activity of the enzyme needs to be carried out to determine its specificity and elucidate its role in *C. jejuni*.

CHAPTER 6

Evaluation of a His-tag purified glycoprotease homologue vaccine against *Campylobacter jejuni* infection in a mouse model

6.1 Introduction.

Campylobacter jejuni and *Campylobacter coli* are among the most frequently isolated causes of bacterial diarrhoea worldwide (Tauxe, 2002). Several reports show that prior infection with *C. jejuni* can result in acquisition of immunity (Black *et al.*, 1988, Martin *et al.*, 1989) suggesting protective immunogenic epitopes on the surface of the cell. Vaccine development against *C. jejuni* has been hindered due to a number of factors that include a lack of understanding of the basic virulence mechanisms, the antigenic complexity of *C. jejuni*; and the lack of small-animal models suitable for vaccine evaluation.

Different antigens have been evaluated as possible vaccines using different animal models for protection against *C. jejuni* (reviewed by (Jagusztyn-Krynicka *et al.*, 2009). The vaccine candidates evaluated include killed whole cells (Baqar *et al.*, 1995b), maltose-binding protein (MBP) of *Escherichia coli* fused to flagellin protein (Lee *et al.*, 1999), periplasmic binding protein (Prokhorova *et al.*, 2006), flagellum-secreted proteins, FlaC, FspA1 and FspA2 (Baqar *et al.*, 2008), conjugated capsular polysaccharides (Monteiro *et al.*, 2009), an adherence protein expressed on and delivered by an attenuated *Salmonella enterica* serovar Typhimurium strain (Sizemore *et al.*, 2006), the amino acid binding protein CjaA, the aspartate/glutamate-binding ABC transporter Peb1A also delivered by an attenuated *S. enterica* serovar Typhimurium strain (Buckley *et al.*, 2010).

Moreover, a limited number of virulence factors that might be useful for subunit vaccine candidates have been identified in *C. jejuni.* The putative glycoprotease has been identified in all strains of *C. jejuni* and shows a high degree of homology within the species of *Campylobacter*. Inactivation of the gene results in a non-viable phenotype suggesting an essential role for the enzyme in bacterial survival. Vaccination of calves with recombinant *M. haemolytica* glycoprotease showed some protection against the bacterium and reduced severity of the disease (Shewen *et al.*, 2003). The demonstrated relationship between serum antibodies (Ab) to Gcp and resistance to pneumonia (Lee *et al.*, 1994) suggested that the Gcp may be involved in the disease process and as such was a good vaccine candidate.

High antibody titres obtained in rabbit immunisation with His-Cj1344c provided encouraging preliminary results for the investigation of possible protective role of the Cj1344c against *C. jejuni* infection. Vaccination with Cj1344c, the homologues of which are present in many bacterial species, some of which are significant human pathogens, may hypothetically provide some protection against these pathogenic agents and therefore increase benefit of this vaccine.

An experimental vaccination trial was therefore conducted in mice using His-Cj1344c as a vaccine antigen, in combination with different routes of administration (subcutaneous, intraperitoneal and intra-nasal) and different adjuvants used in the trial; followed by *C. jejuni* challenge to ascertain capacity of His-Cj1344c to protect mice from *C. jejuni* infect.

6.2 Results:

6.2.1 Protein expression and purification

To test the immunogenicity of the *C. jejuni* putative glycoprotease in a murine model, the enzyme had to be expressed and purified. The *Cj1344c* gene from *C. jejuni* 11168 was expressed and purified as described in chapter 3.

6.2.2 anti-His-Cj1344c antibodies in mice prior to immunisation regimen

The immunogenicity of His-Cj1344c was tested in mice via delivery of the protein by three different routes of administration. The His-Cj1344c was delivered to three groups of animals by subcutaneous, intraperitoneal and intranasal vaccination to determine the best route of antigen delivery. Freunds adjuvant was used with His-Cj1344c in cases of subcutaneous and intraperitoneal injection (complete Freunds adjuvant for the first injection, incomplete Freunds adjuvant for subsequent booster injections). Cholera toxin subunit B was used as an adjuvant in cases of intranasal delivery. PBS was substituted for His-Cj1344c in the control groups and administered by same delivery as the antigen. The fourth group served as a negative control and did not receive any treatment.

Prior to antigen administration, mice sera were collected and assayed for the presence of His-Cj1344c specific IgG by ELISA. 5/40 mice showed the presence of Cj1344c specific antibodies with titres less than 1:32 (Figure 6.1). The experiment was continued and animals that showed antibody response to Cj1344c were noted.

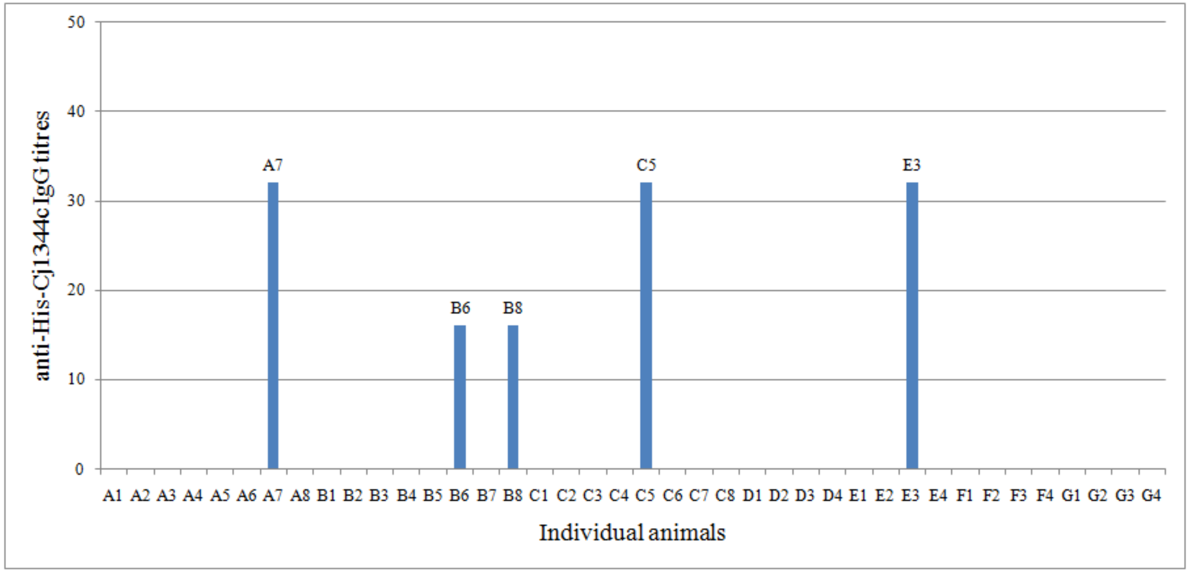


Figure 6.1 Mouse sera IgG response to His-Cj1344c protein prior to vaccination

Legend:

A1-A7 – subcutaneous His-Cj1344c immunisation

B1-B8 – intraperitoneal His-Cj1344c immunisation

C1-C8 – intranasal His-Cj1344c immunisation

D1-D4 – subcutaneous PBS control immunisation

E1-E4 – intraperitoneal PBS control immunisation

F1-F4 – intranasal PBS control immunisation

G1-G4 – non-immunised controls

6.2.3 Immunisation with His-Cj1344c using subcutaneous delivery method

Immunisation: A dose of 5µg of purified His-Cj1344c mixed with an appropriate adjuvant was delivered subcutaneously at 2 weeks intervals and resulted in high IgG titres (compared to PBS controls and negative controls) 2 weeks after the fourth dose, as determined by direct ELISA using His-Cj1344c for coating the wells of plates. For all immunisation regimens, the His-Cj1344c specific serum IgG titres increased with number of vaccinations, however, the final titres varied between different animals (Table 6.1). The highest end point dilution titre of anti-His-Cj1344c IgG serum was 1:65,000 (3 out of 8 animals). The IgM and IgA titres, on the other hand, showed no significant difference when compared to controls (Table 6.1).

Table 6.1 His-Cj1344c specific immunoglobulin response in mice after 4th booster injection delivered subcutaneously

	Animal	IgA	IgG	IgM
His-Cj1344c immunised	A1	0	1:1,024	1:128
	A2	1:64	1:8,172	1:512
	A3	0	1:16,344	1:64
	A4	1:128	1:65,376	1:32
	A5	1:32	1:16,344	1:128
	A6	0	1:65,376	0
	A7	1:64	1:2,048	1:128
	A8	1:1,024	1:65,376	1:1,024
PBS immunised	D1	0	1:128	1:64
	D2	0	1:32	0
	D3	0	1:4,086	0
	D4	0	1:512	1:512

Protection studies: To determine if the anti-His-Cj1344c antibodies in mice were protective against *C. jejuni* colonisation, immunised mice were challenged with 10^8 *C. jejuni* cells fourteen days following the last immunisation dose. All animals were challenged orally with wild type *C. jejuni* NCTC 11168 as described in Materials and Methods chapter. Animals were monitored for signs and symptoms of disease such as, ruffling of fur, the presence of diarrhoea and general well being on a daily basis by two animal attendants and their well being was scored based on the severity of the illness as per Monteiro *et al*, 2009 for the duration of the experiment. Animals did not show any signs of illness for the entire duration of the experiment.

Faecal samples were collected daily and *C. jejuni* load was determined which was indication of the intestinal colonisation with the bacteria. Oral inoculation with *C. jejuni* resulted in colonization of the gastrointestinal tract of all mice as assessed by positive faecal cultures in all animals. Bacterial faecal load observed during the course of experiment showed a difference between His-Cj1344c vaccinated and control mice. Though there are variations between individual animals, generally His-Cj1344c vaccinated mice show lower bacterial counts when compared with PBS vaccinated or non-vaccinated mice, though the differences were not significant as determined by un-paired t-test (Figure 6.2).

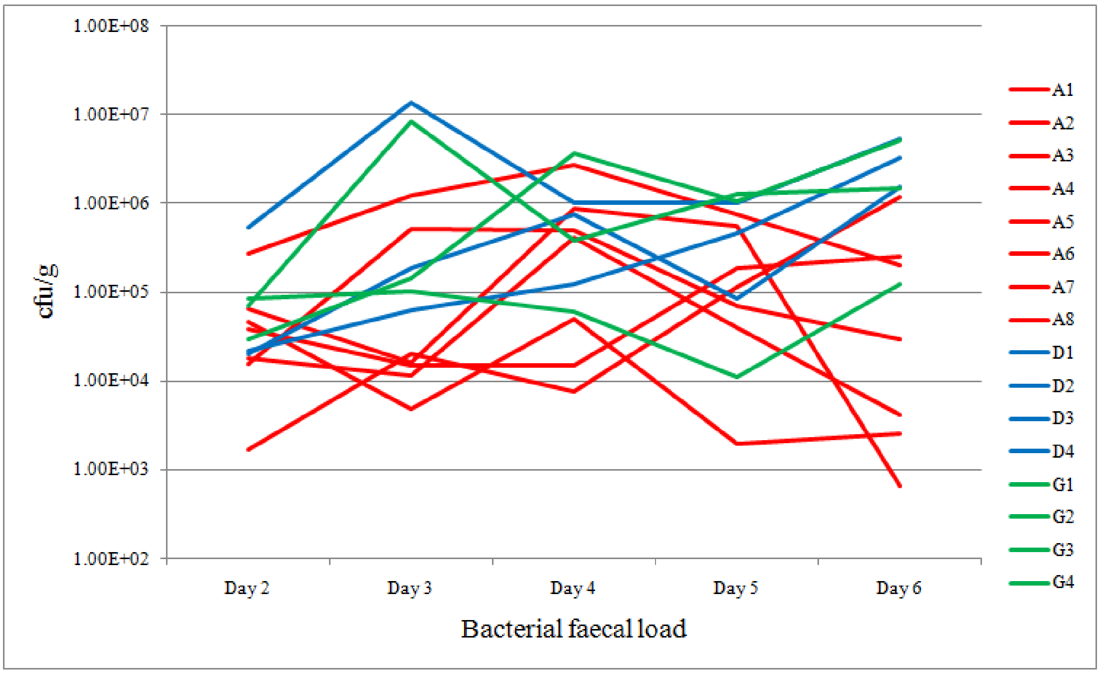


Figure 6.2 *C. jejuni* faecal load in subcutaneously vaccinated mice and their respective controls over the period of 6 days

<u>**Legend:**</u>

A1-A7 – subcutaneous His-Cj1344c immunisation

D1-D4 – subcutaneous PBS control immunization

G1-G4 – non-immunised controls

At day 7 post infection mice were sacrificed, at which point different systemic organs and samples of small and large intestines were analysed for the presence of *C. jejuni*. *C. jejuni* was not cultured from systemic organs (liver, lungs, or spleen) from the animals in any group. However, samples of small and large intestines in the test and control animals showed a difference in bacterial load. Bacterial counts isolated from small and large intestines of His-Cj1344c vaccinated mice were lower than those observed in control animals thought the significance was not statistically different as determined by un-paired t-tests (Figure 6.3).

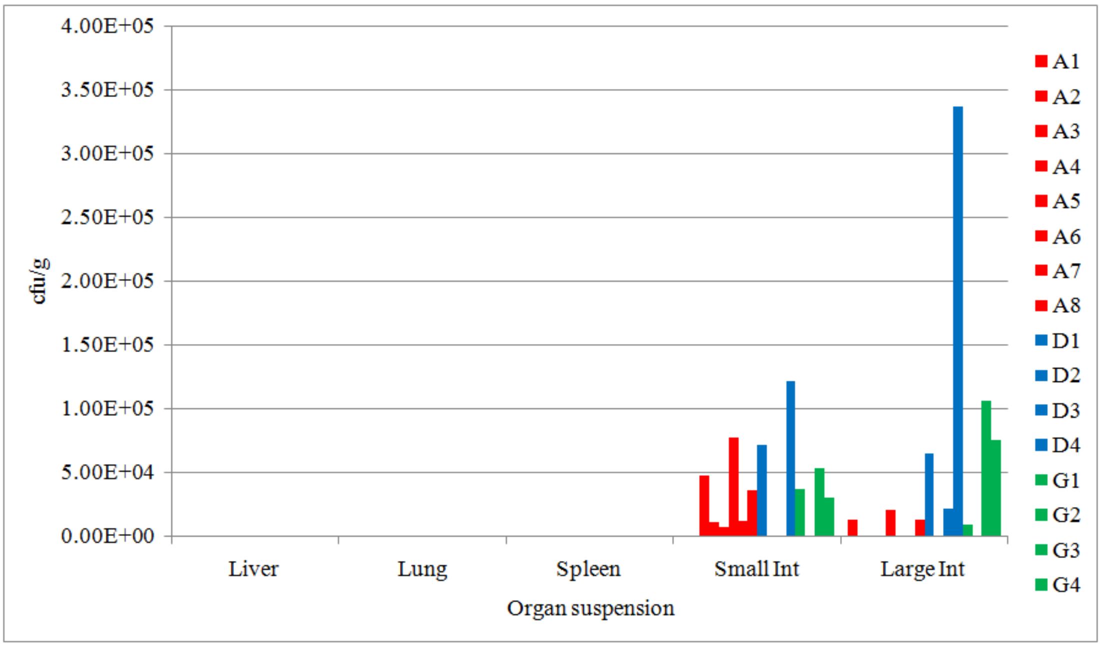


Figure 6.3 *C. jejuni* bacterial load in different organs in subcutaneously vaccinated mice

Legend:

A1-A7 – subcutaneous His-Cj1344c immunisation

D1-D4 – subcutaneous PBS control immunization

G1-G4 – non-immunised controls

6.2.4 Immunisation with His-Cj1344c using intraperitoneal delivery method

Immunisation: Intraperitoneal delivery of 5µg of purified His-Cj1344c mixed with Freund's adjuvant at 2 weeks intervals resulted in high IgG titres (compared to PBS controls and negative controls) 2 weeks after the fourth dose (Table 6.2). In addition, mean IgA titres for 8 mice showed higher levels compared to controlled mice. The IgM titres, however, showed no significant difference when compared to controls (Table 6.2). The His-Cj1344c specific serum immunoglobulin titres showed a steady increase during the course of the vaccination regimen. The highest end point dilution titre of anti-His-Cj1344c IgG serum was 1:65,000 (1 out of 8 animals), while the lowest point dilution was 1:256 (2 out of 8 animals) showing the substantial antibody response variation between individual animals.

Table 6.2 His-Cj1344c specific immunoglobulin response in mice after 4th booster injection delivered intraperitoneally

	Animal	IgA	IgG	IgM
His-Cj1344c immunised	B1	0	1:256	1:64
	B2	1:4,086	1:65,376	1:64
	B3	1:64	1:1,024	1:128
	B4	1:32	1:256	1:64
	B5	1:64	1;2,048	1:256
	B6	1:64	1:2,048	1:256
	B7	1:64	1;1,024	1:128
	B8	1:64	1:1,024	1:64
PBS immunised	E1	0	1:2,048	1:256
	E2	1:64	1:4,086	1:2,048
	E3	1:64	1:2,048	1:256
	E4	0	1:512	1:128

Protection studies: All animals were challenged orally with wild type *C. jejuni* NCTC 11168 same as the subcutaneously vaccinated mice. Daily observations of animals did not record any signs of illness for the entire duration of the experiment.

Bacterial load was assessed in faeces of mice receiving His-Cj1344c intraperitoneally. The bacterial load remained stable during the course of the experiment and did not show a marked difference between control animals and His-Cj1344c vaccinated animals (Figure 6.4)

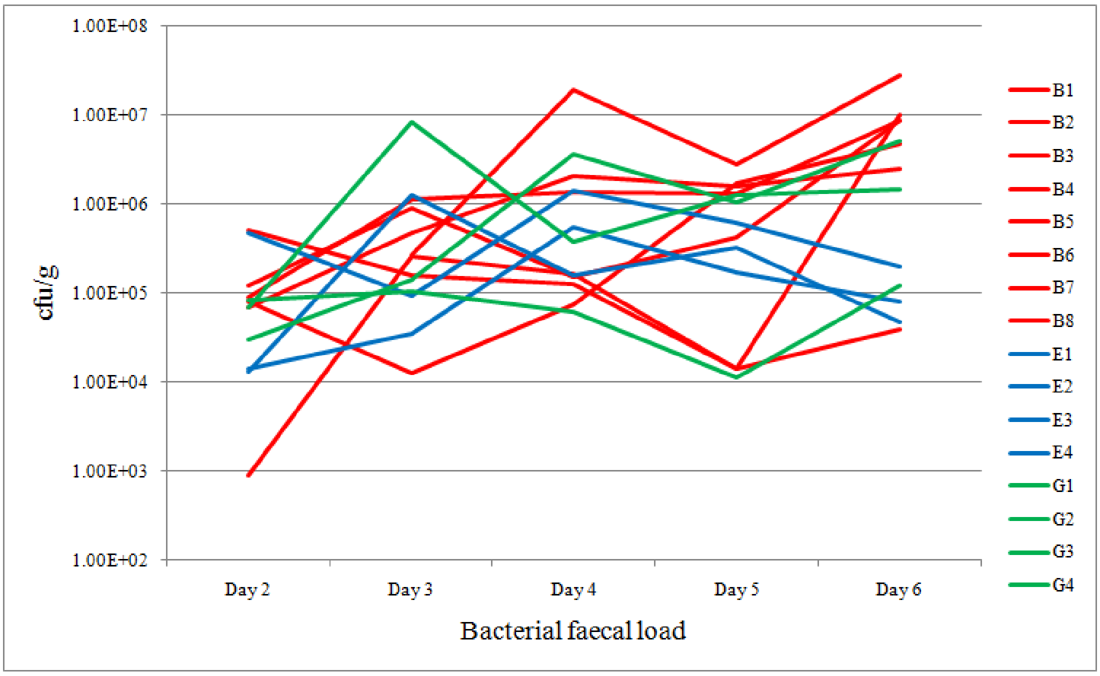


Figure 6.4 *C. jejuni* faecal load in intraperitoneally vaccinated mice over the period of 6 days and their respective controls

<u>**Legend:**</u>

B1-B8 – intraperitoneal His-Cj1344c immunisation
E1-E4 – intraperitoneal PBS control immunisation
G1-G4 – non-immunised controls

Similarly to the group of animals that were immunised with His-Cj1344c subcutaneously, 7 days post infection all animals in this group showed no presence of *C. jejuni* in systemic organs. However, bacterial numbers in the small and large intestines in this group of animals was notably higher than those in the subcutaneous group. In addition, there was no statistical difference in bacterial counts between His-Cj1344c vaccinated animals and the control animals (Figure 6.5).

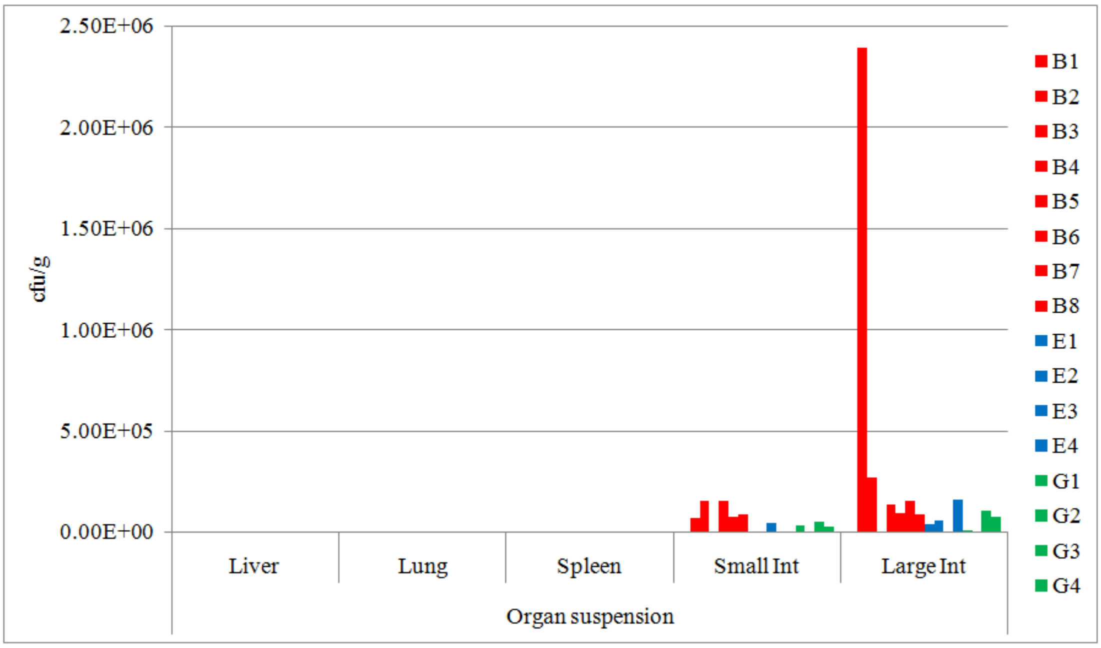


Figure 6.5 *C. jejuni* bacterial load in different organs in intraperitoneally vaccinated mice

Legend:

B1-B8 – intraperitoneal His-Cj1344c immunisation
E1-E4 – intraperitoneal PBS control immunisation
G1-G4 – non-immunised controls

6.2.5 Immunisation with His-Cj1344c using intranasal delivery method

Immunisation: Administration of His-Cj1344c intranasally mixed with Cholera Toxin Subunit B at 2 weeks intervals resulted in low IgG titres (compared to subcutaneous and intraperitoneal vaccination regimen). The IgG titres were not significantly higher than those observed in control groups (Table 6.3). In addition, mean IgA and IgM titres showed no significant difference when compared to controls. The His-Cj1344c specific serum IgG titres showed steady increase during the course of vaccination regimen reaching highest end point dilution titre of anti-His-Cj1344c IgG serum of 1:4,000 (1 out of 8 animals). The lowest point dilution was 1:64 (3 out of 8 animals) suggesting that this method of antigen delivery was the least successful in raising His-Cj1344c specific antibodies in mice.

Table 6.3 His-Cj1344c specific immunoglobulin response in mice after 4th booster injection delivered intranasally

	Animal	IgA	IgG	IgM
His-Cj1344c immunised	C1	0	1:64	1:128
	C2	16	1:2,048	1:256
	C3	0	1:128	1:64
	C4	0	1:64	1:64
	C5	128	1:4,086	1:4,086
	C6	0	1:64	0
	C7	16	1:1,024	1:64
	C8	0	1:512	1:128
PBS immunised	F1	1:32	1:256	1:128
	F2	1:128	1:1,024	1:512
	F3	0	1:256	1:32
	F4	1:32	1:256	1:64

Protection studies: All animals were challenged orally with wild type *C. jejuni* NCTC 11168 same as the subcutaneously vaccinated mice. Daily observations of animals did not record any signs of illness for the duration of the experiment.

Administration of His-Cj1344c intranasally seems to confer some protection against *C. jejuni* infection as vaccinated mice in this group show lower faecal bacterial load during the course of the infection when compared to control animals, though great variation between animals was observed (Figure 6.6).

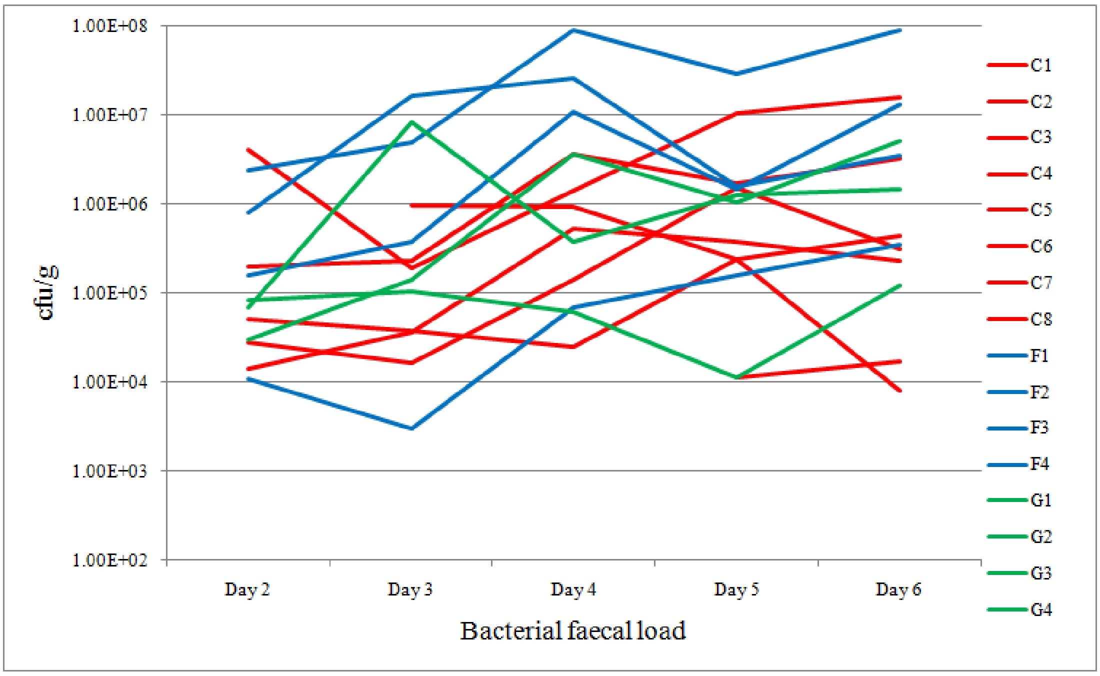


Figure 6.6 *C. jejuni* faecal load in intranasally vaccinated mice over the period of 6 days and their respective controls

__Legend:__ C1-C8 – intranasal His-Cj1344c immunisation
F1-F4 – intranasal PBS control immunisation
G1-G4 – non-immunised controls

Examination of the systemic organs of animals in this group for the presence of *C. jejuni* identified one PBS vaccinated animal that had *C. jejuni* present in its liver. The systemic organs of the rest of the animals in this group were free of *C. jejuni*. The bacterial loads in the small and large intestines showed no significant difference between His-Cj1344c vaccinated and the control groups (Figure 6.7)

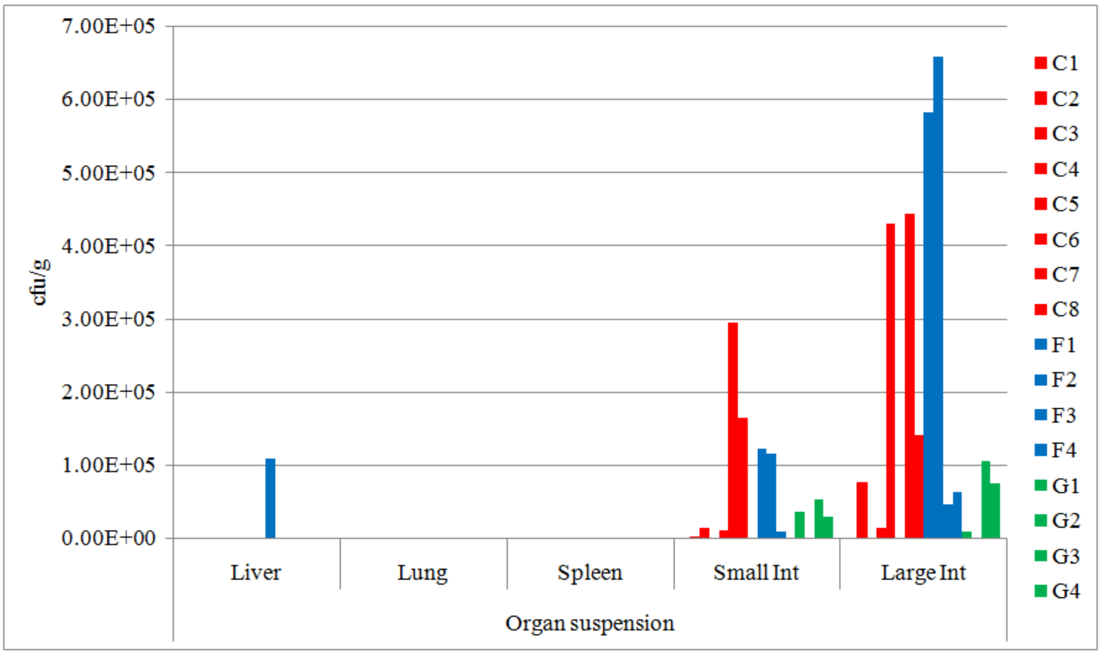


Figure 6.7 *C. jejuni* bacterial load in different organs in intranasally vaccinated mice

Legend: C1-C8 – intranasal His-Cj1344c immunisation
F1-F4 – intranasal PBS control immunisation
G1-G4 – non-immunised controls

6.2.6 Mice exposure to transient infection

The presence of relatively high titres of anti-His-Cj1344c antibodies in the sera of control animals suggested a transient infection of animals with *Campylobacter* spp prior or during the course of vaccination. In addition, it provided hypothetical explanation for the non-significant difference in bacterial counts between the vaccinated and control groups observed in the *C. jejuni* challenge experiments.

To determine if pre-bleed sera taken before the initial vaccination and after the last dose of vaccination in PBS vaccinated control animal mice contained *C. jejuni* specific antibodies, Western blot analysis was performed using different bacterial cell lysates. This experiment confirmed transient infection of mice with *Campylobacter* spp. or a campylobacter related species (Figure 6.8) as results indicate cross reactivity of mice sera with different bacterial species and different bacterial proteins within the lysates. In addition, recognition of purified His-Cj1344c by sera of sham vaccinated mice suggests that antibody titres observed in sera of control mice in this experiment were raised against native Cj1344c protein or its homologue in other bacterial species during transient infection with *C. jejuni* or related organism.

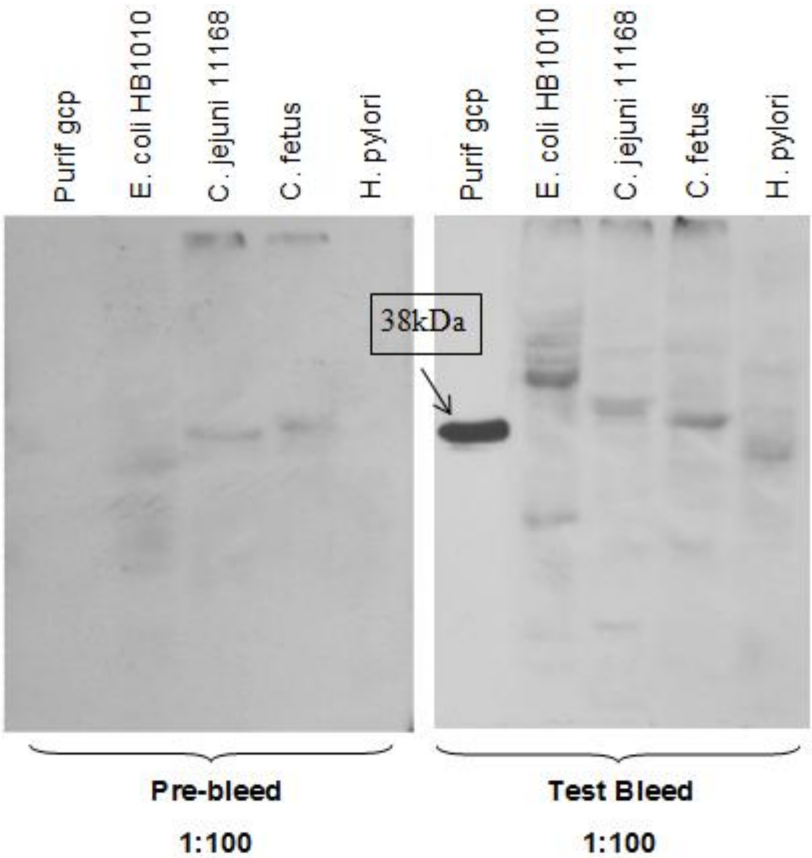


Figure 6.8 Western blot analyses of different bacterial lysates and His-Cj1344c protein using mice sera pre and post vaccination

6.3 Discussion

Despite its importance as a human diarrheal pathogen and its strong association with GBS (Jacobs *et al.*, 2008) a commercial vaccine against *C. jejuni* is not available. Studies aimed at developing a human anti-*Campylobacter* vaccine based on killed cells (*Campylobacter* whole-cell [CWC] vaccine) have been undertaken in mice, ferrets and non-human primates (Baqar *et al.*, 1995a, Baqar *et al.*, 1995b, Burr *et al.*, 2005). Introduction of anti-CWC vaccines, however, seems rather unlikely for several reasons. One of the main limitations of this vaccine approach is the high level of genetic diversity of the *Campylobacter* strains resulting in strain-specific vaccines. Secondly, introduction of a human anti-*Campylobacter* vaccine containing whole attenuated bacterial cells to the pharmaceutical market before the elucidation of the detailed mechanisms of the autoimmune disease caused by the mimicry between bacterial LOS and human gangliosides, is considered too risky (Ang *et al.*, 2003). In addition, the association of *C. jejuni* infection and the development of reactive arthritis further complicates the use of the whole cell killed vaccine. The pathophysiology of the *Campylobacter* disease and the *C. jejuni* surface structures involved in the process are largely unknown (Pope *et al.*, 2007). On the other hand, subunit vaccine approaches generally utilize antigens that play a role in virulence, however, the *C. jejuni* pathogenesis remains poorly understood (Jagusztyn-Krynicka *et al.*, 2009) making the choice of antigen candidates problematic.

In this study the issues described above were considered and an enzyme with an intra- and extra-cellular function was chosen as a potential candidate for the *C. jejuni* vaccine. High antibody titres obtained in rabbit immunisation with His-Cj1344c provided encouraging preliminary results for the investigation of the possible protective role of the putative glycoprotease (Cj1344c) against *C. jejuni* infection. The use of *E. coli* His-tag expressed protein was benefitial in terms of vaccination safety as the protein shows the absence of enzymatic activity expected from *C. jejuni* native enzyme (Chapter 3), but at the same time antibodies raised against the His-Cj1344c show cross-reactivity with native *C. jejuni* Cj1344c protein (Chapter 5). Mouse immunisation trials with His-Cj1344c were undertaken to determine the best route for antigen delivery and to provide animal safety data of the vaccination regimen using three different routes of delivery: subcutaneous, intraperitoneal and intranasal. The data presented in this chapter demonstrate the subcutaneous delivery of the His-Cj1344c, when administered with Freund's adjuvant, is the most effective in eliciting the immune response, when compared with other two routes of delivery. Vaccination of mice with His-Cj1344c resulted in varied antibody titre levels in the animals, confirming that antibody levels depend on the delivery method as well as the host immune response.

The IgG His-Cj1344c specific antibody titres reaching up to 1:65,000 in animals in **the subcutaneous** delivery group are encouraging for future vaccine development, as these levels of antibody titres are higher or similar to those reported in the literature for other *C. jejuni* antigens. The magnitude of the serum immune responses detected in this study can be compared to those reported for flagellum-secreted proteins in mice by Baqar *et al.* The IgG levels obtained in His-Cj1344c study (1:65,000) are several magnitudes higher than those obtained in vaccination with flagellum-secreted protein FlaC (1:164), which reached the titre levels of this study only after vaccinating mice with 25 µg of the FlaC antigen. On the other hand, the His-Cj1344c specific IgG titre levels reported here are significantly lower than the ones reported in case of flagellum-secreted proteins, FspA1 and FspA2 tested in mice, which reached levels of 1:2,400,000 and 1:300,000 respectively (Baqar *et al.*, 2008). However, the protection provided in the case of these two flagellum-secreted proteins is likely to be strain specific as these proteins show considerable diversity among *C. jejuni* strains (Poly *et al.*, 2007). Cj1344c protein, on the other hand, shows significant conservation among strains (90-100% amino acid similarity) and as such is likely to be a better vaccine candidate. The high His-Cj1344c specific IgG antibody titres observed here were also similar to the high antibody titres observed in a His-Cj1344c rabbit immunisation trial (Chapter 5) showing that the high antibody response is not species specific. The animals in subcutaneous delivery group also had a reasonably high serum IgA and IgM titre values compared to the other two delivery method groups, averaging 1:1,000. These levels of IgA and IgM are within the levels reported in the literature which range from 1:30 in case of MBP-FlaA (Lee *et al.*, 1999) to 1:32,000 in case of FspA1 protein immunisation (Baqar *et al.*, 2008).

The data resulting from a subsequent protection trial shows that His-Cj1344c is capable of eliciting a limited protective immune response against *C. jejuni* as measured in an oral challenge mouse model. The subcutaneous administration of 5 µg of His-Cj1344c resulted in the immunised animals showing a lowered faecal bacterial count during intestinal colonization compared to control animals (Figure 6.2), though the results were not statistically significant. In addition, the bacterial numbers in the faeces have a general trend of reduction at day 6 of the experiment in the His-Cj1344c immunised group of animals. The results of bacterial counts in the small and large intestines after bacterial challenge (Figure 6.3), showed reduced numbers of *C. jejuni* in the His-Cj1344c immunised mice, compared to control mice.

The animals in **the intraperitoneally** administered His-Cj1344c group had generally lower titre counts, suggesting that this antigen delivery method was less successful. The faecal bacterial counts data collected during the challenge study showed that this immunisation regimen provided no protection against *C. jejuni* colonisation in mice.

Similar to the animals in the intraperitoneally administered His-Cj1344c group, animals immunised **intranasally** had low titre counts compared to the subcutaneously immunised mice. The IgG titres in this group of animals averaged 1:1,000, while IgA titres averaged 1:20. The low IgA titres were somewhat surprising as the mucosal antigen delivery was thought to stimulate production of IgA antibody subclass. Though, the direct mucosal antibody response was not determined, the serum IgA was significantly lower than reported for other *C. jejuni* antigens administered by this method. In the study of flagellum-secreted protein, FspA1, IgG titre levels of 1:32,000 have been reported (Baqar *et al.*, 2008). The use of LTR adjuvant in the FspA1 study as opposed to CT adjuvant in this study may have attributed to the reduced immune response in this group. In addition, the methodology of administering the adjuvant to the external nares of the mice, where some of the vaccine can be swallowed rather than inhaled by an animal, could contribute to the differences in titres observed in this study. The IgM titres in this group did not show significant difference when compared to titres from the control animals, further confirming that the administration of the antigen by this route was not optimal for inducing an immune response.

The bacterial challenge in this group of animals shows some, not statistically significant, reduction of the bacteria numbers in the faeces of immunised animals, while the bacterial counts in the small and large intestines showed no significant difference between the immunised and control animals. This confirms that the immune response raised in this group was not sufficient to reduce the gastrointestinal bacterial colonisation in mice.

This pilot study of His-Cj1344c immunisation of mice to determine the protection efficacy against *C. jejuni* intestinal colonisation has identified a number of problems that need to be addressed in future studies. Primarily, the low antibody response to the antigen in some animals could be explained by the minimal amount of antigen used in the immunisation trials (5 µg); compared to the amounts used in other studies (up to 50 µg of antigen). In the immunisation study of MBP-FlaA fusion protein, the use of 50 µg of the antigen produced significantly increased serum IgG response of 1:2,000,000 compared to 1:450,000 obtained after immuniation with 6 µg of the antigen (Lee *et al.*, 1999). The increased use of the antigen in His-Cj1344c immunisation study could significantly increase antibody response in animals and potentially provide protection against *C. jejuni* intestinal colonisation.

A minimal dose of the antigen was used in order to avoid possible toxicity of as its function is not fully elucidated. The native protein was speculated to degrade *O*-sialoglycoproteins which are the main constituent of the mammalian mucosal layer, and as such was not recommended for use in this study. The His-Cj1344c, on the other hand, was used as it was shown in this study that it was not fully functional due to misfolding or a lack of enzyme co-factor (Chapter 2). However, the noted minimal activity of the recombinant His-Cj1344c protein during glycan array studies and MUC2 digest studies had to be taken into consideration, resulting in the minimal amount of the protein used in the pilot study. The 5 µg dose of His-Cj1344c during immunisation did not produce any adverse effects in the mice, so future studies could potentially use a higher dose of antigen and thereby potentially increase antibody titre levels.

This immunisation study has also identified that the non-vaccinated mice had His-Cj1344c specific antibodies, similar to findings of Shewen *et al*. in the *M. haemolytica* gcp vaccination trials (Shewen *et al.*, 2003). This finding suggests that the His-Cj1344c specific antibodies have been produced in the non-vaccinated group of animals through a transient infection or an infection previous to immunisation with *C. jejuni* or a related species against the native protein. Analysis of the control mice sera by Western blot analysis after the immunisation regimen using different bacterial whole cell lysates showed specificity of the sera to the whole cell proteins which was not observed in the sera of the mice before vaccination. Though the faecal samples were collected and analysed by plating on a *Campylobacter* selective media on a weekly basis for the duration of the study, transient infection was not detected. The presence of the *H. hepaticus*, however, a bacterium related to *C. jejuni,* could have attributed to the rise of His-Cj1344c specific antibodies as this bacterium also possesses a Cj1344c homologue. Due to its localisation in mouse liver, this bacterium would not be detected by the faecal sampling performed in the study. Considering that the control group of animals has His-Cj1344c specific antibodies in their sera the comparison between the immunised and control groups is difficult and conclusions drawn from it need to be carefully considered.

6.3.1 Conclusion

Rabbit immunisation with His-Cj1344c provided a good immune response to the antigen with IgG titres reaching 1:65,000. In addition, the finding by Shewen *et al.* with *M. haemolytica* gcp calf vaccination which resulted in decreased pneumonic tissue necropsy. These findings initiated immunisation trial with His-Cj1344c in mice to determine the protective ability of the antigen against the *C. jejuni* infection. This preliminary immunisation study has identified a high IgG immune response to His-Cj1344c in mice, as well as a relatively high IgA response in the subcutaneous delivery group. The trial of the three different antigen administration routes has determined the best immune response to the His-Cj1344c to be with subcutaneous immunisation. The protection studies against *C. jejuni* infection have determined that the mice immunised with His-Cj1344c showed a lower number of *C. jejuni* cells in their faeces and small and large intestines, which was indicative of lower colonisation. Though the results of this study were not statistically different, the trend observed in the results is encouraging to warrant further examination of this antigen as a vaccine candidate against *C. jejuni* infection.

CHAPTER 7

General discussion

Campylobacter jejuni infection is one of the most commonly identified bacterial causes of acute gastroenteritis worldwide. Data obtained from the United States, Europe, and Australia reveals that 50-70% of all *Campylobacter* infections have been attributed to the consumption and handling of contaminated poultry (Hall *et al.*, 2005). The increasing number of human infections with *C. jejuni* and *C. coli*, strains resistant to the antibiotics commonly used in human therapy (i.e., macrolides, quinolones and tetracycline) constitutes a serious medical problem (Moore *et al.*, 2006). Emerging strains of *Campylobacter* resistant to the commonly used antibiotics, significantly impair the process of combating campylobacteriosis by prolonging therapy and adding cost to an already burdened medical system. The development of an effective vaccine against *C. jejuni* infections is desirable. Vaccine safety remains a major consideration in the development of the *C. jejuni* vaccine. The molecular mimicry between bacterial LOS structures and human gangliosides presents a serious problem as it is associated with the development of a reactive arthritis/arthropathy and neuroparalytic syndromes such as Guillain-Barré Syndrome (Dingle *et al.*, 2001). This makes the use of the whole cell killed *C. jejuni* vaccine risky in the view of paucity of knowledge concerning *C. jejuni* physiology and pathogenesis.

The aim of this study was to determine the potential of the *C. jejuni* putative glycoprotease encoded by the *Cj1344c* gene as a vaccine candidate, and to characterise the role of the protein in bacterial pathogenesis. A *C. jejuni* glycoprotease enzyme orthologue was identified based on amino acid similarity to *M. haemolytica O*-sialoglycoprotease (Gcp) (Abdullah *et al.*, 1991). Though the enzymatic specificity for *O*-sialoglycoproteins has been determined for the *M. haemoltica* Gcp, its role in bacterial pathogenesis is currently unknown (Abdullah *et* *al.*, 1992). To determine the role of the Cj1344c in *C. jejuni* pathogenesis, creation of the isogenic mutant was attempted in order to assess the effect of the mutation on the bacterial cell and *in vivo* colonisation models. The activity and the specificity of Cj1344c were examined by expression and purification of the protein using *E. coli* and performing enzymatic studies with recombinant protein. To determine its potential use as a vaccine candidate against *C. jejuni,* studies to investigate the immunogenicity of the His-Cj1344c protein and the possibility of its use in animal immunisation trials were performed.

7.1 Mutagenesis

The methodology for the *Cj1344c* gene mutagenesis needed to be carefully considered as the *Cj1344c* gene is positioned within the *C. jejuni* chromosome in an operon-like locus with other genes of unknown function. To determine the effect of the *Cj1344c* mutagenesis on *C. jejuni* and potentially determine the role of the gene product in bacterial pathogenesis, insertional inactivation of the gene was attempted. The insertional inactivation of the *Cj1344c* gene posed a potential problem as it had a capacity to cause a polar mutation by inactivating genes downstream from the insertion point of the antibiotic resistance gene cassette. This is not unusual for the dense genome of *C. jejuni*, where 90% of the genome encodes proteins (Parkhill *et al.*, 2000a). Many of the *C. jejuni* genes have an unknown function, which further complicates the mutagenesis studies where there is a possibility of a polar mutation. In this instance, the *Cj1343c* gene, downstream from the putative glycoprotease *Cj1344c* gene, encodes for a putative periplasmic protein and is identified as a gene potentially involved in type II secretion system (Wiesner *et al.*, 2003). In addition, the *Cj1344c* is positioned in close proximity to the *flaA* and *flaB* genes prompting speculation that the Cj1344c protein uses the flagellar secretion system. The polar mutation caused by insertional activation of *Cj1344c* could, therefore, significantly change the bacterial phenotype of the resulting isogenic mutant in which case the true effect of *Cj1344c* mutation would not be determined. To circumvent this, different plasmid suicide constructs (pGU0523, pGU0613, pGU0707 and pGU0804) were prepared and utilised in an attempt to generate the *Cj1344c* isogenic mutant. The suicide plasmid constructs contained the *Cj1344c* gene interrupted by one of the following different antibiotic resistance cassettes (promoter-less kanamycin resistance gene cassette, promoted kanamycin resistance gene cassette, promoted kanamycin resistance gene cassette with *oxa-61* gene promoter downstream and promoter-less chloramphenicol gene resistance cassette) (Table 4.1).

Multiple attempts to mutate the *Cj1344c* gene did not produce a viable *C. jejuni* isogenic mutant, although the construct DNA was proven to be integrated into the *C. jejuni* genome during transformation, suggesting that the inactivation of this gene is lethal for the bacterium. This finding is similar to the results of mutagenesis studies of the *gcp* homologues in *S. aureus* (Zheng *et al.*, 2005), *M. haemolytica* (Mellors, 2008) and *H. influenzae* (Apicella, 2010) which identified that the enzyme is essential for bacterial growth, but were unable to determine the function of the enzyme. The only studies that provide insight into the function of the *gcp* gene homologue product are conditional mutagenesis studies during which the expression of the gene is reduced. These studies utilised a controllable promoter element to down-regulate the expression of the gene, and subsequently examine the effect of the gene down-regulation on the bacterial phenotype.

The gene down-regulation studies in different bacteria have identified the importance of the glycoprotease gene homologue for bacterial survival, but have not always identified a function of the protein and its role in the bacterial pathogenesis. However, a few studies have determined or suggested a role for the enzyme in different bacteria. The down-regulation of *gcp* expression in *S. aureus* had a lethal effect on bacterial growth and Gcp was demonstrated to be a critical mediator involved in the modification of cell wall biosynthesis through modification of cell wall peptidoglycans (Zheng *et al.*, 2007). A recent publication by Katz *et al*., using this approach, identified the role of the glycoprotease encoded by the *E. coli ygjY* gene in the metabolism of toxic products of glycation (Amadori-modified proteins (AMPs) and advanced glycated end products (AGEs)). Though the enzymatic activity of the *E. coli* enzyme in this case has not been examined, these findings suggest that the enzyme in *E. coli* does not have a specificity for the *O*-sialoglycoproteins observed in *M. haemolytica* gcp (Abdullah *et al.*, 1992). The role of the gcp enzyme homologue in removal of the toxic components of glycation provides an explanation why it was not possible to create isogenic mutants of *Cj1344c* in *C. jejuni*, but does not explain the specificity of the enzyme for *O*-sialoglycoproteins observed in *M. haemolytica* gcp studies and this study.

Unfortunately, a controllable promoter element has not been identified in *C. jejuni*, preventing the examination of the role of Cj1344c by the method of gene down-regulation. Due to time constraints, other methods, such as the chromosomal *Cj1344c* gene deletion and complementation by the presence of a low-copy-number plasmid carrying the *Cj1344c* coding sequence were not be explored. This approach, as well as antisense-mRNA approach to down-regulate expression of the gene on RNA level, could be explored in future studies to determine the role of Cj1344c in *C. jejuni*.

From the mutagenesis studies of *C. jejuni Cj1344c* reported here and mutagenesis studies in other bacterial species it can be concluded that the *Cj1344c* gene inactivation is lethal for the bacteria. The functions of the gcp within a bacterial cell seem to be complex and vary between bacterial species. The role of gcp in metabolism of toxic products of glycation in *E. coli* suggests an intracellular function for the enzyme (Katz *et al.*, 2010) and gives a plausible explanation for the essential role of the enzyme. On the other hand, the role of the enzyme homologue in *S. aureus,* in the cell wall peptidoglycan synthesis through the regulation of murein hydrolases, involved in this process, suggests an extracellular function of the enzyme. The finding of the gcp in the culture supernatant of *M. haemolytica* reported by Otulakowski *et al*. and this study, in *C. jejuni*, may suggest the role of the enzyme in the degradation/modification of the host cell glycoproteins.

7.2 Enzymatic specificity and function

Based on the study of *M. haemolytica* gcp specificity to *O*-linked sialoglycoproteins and its involvement in the pathology of calf pulmonary infections (Abdullah *et al.*, 1992, Shewen *et al.*, 2003), it was proposed in this study that the *C. jejuni* putative glycoprotease homologue, Cj1344c, was involved in bacteria-host interactions by exerting an effect on host glycopropteins, specifically the mucin glycoproteins in the host gastrointestinal tract. The role of *C. jejuni* gcp homologue was hypothesised to be involved in degradation of the mucous layer in the gastrointestinal tract allowing bacterial penetration and potentially enhancing adherence to the epithelial cells. The possibility of the modification of this layer by the enzyme was also hypothesised. New glycoprotein structures created by the action of enzyme cleavage may provide more intimate adherence of the bacterial cells to the mucous layer. The interaction between mucins and bacterial cells was demonstrated for *C. upsaliensis*, a closely related species, and was speculated to influence access of the bacteria to cell membrane receptors and thereby influence host resistance to infection (Sylvester *et al.*, 1996). The modification of glycoproteins was hypothesised to enhance this initial interaction and aid in subsequent invasion of bacteria.

The initial characterisation of the enzyme specificity was performed by the ligand-binding interactions between *E. coli* expressed His-Cj1344c and a library of amino acids and glycoproteins. The results of this study identified methionine, lysine and arginine to be amino acids recognised by the enzyme. In addition, His-Cj1344c also demonstrated specificity to human MUC2 and bovine lactoferrin, an iron-binding *O*-sialoglycoprotein present in mammalian milk.

To confirm the interactions identified between the His-Cj1344c and the amino acids, STD-NMR method was employed. The ligand binding studies performed, failed to confirm the interactions initially observed with the amino acid array technology, which was hypothesised to be due to a strong binding interaction between these amino acids and His-Cj1344c, as this is one of the limiting factors of the method (Haselhorst *et al.*, 2009). These three amino acids were, therefore, hypothesised to indicate amino acid residues within a polypeptide recognised by the enzyme as a cleavage site. The *M. haemolytica* gcp enzyme was shown to have a specific amino acid sequence cleavage site and *C. jejuni* Cj1344c was thought to be no exception to this. The experiments involving degradation of the glycophorin A by *M. haemolytica* Gcp have identified the amino acid recognition sequence of the enzyme to be Arg-31–Asp-32 (Abdullah *et al.*, 1992). The *C. jejuni* Cj1344c amino acid recognition specificity, identified here to be methionine, lysine and arginine, however, may be different to *M. haemolytica* gcp considering that the predicted amino acid sequence of Cj1344c shows only 37% similarity to *M. haemolytica* Gcp (Chapter 3). The substrate specificity difference between *C. jejuni* Cj1344c and *M. haemolytica* Gcp may be due to the different environmental niches of these two organisms, considering that *M. haemolytica* is a typical bovine lung pathogen, while *C. jejuni* is a typical gastrointestinal bacterium. The substrate specificity difference may also reflect the different roles these enzymes play in the pathogenesis of these two organisms.

A MUC2 digest with the recombinant *C. jejuni* glycoprotease, expressed in *E. coli* was attempted to confirm the interactions of His-Cj1344c and mucin observed with the glycoprotein array method. The findings reported here, however, did not conclusively prove the interaction of the enzyme with MUC2, as the results of the enzymatic digest did not show a distinct cleavage pattern of MUC2 as assayed by SDS-PAGE analysis. The problem was hypothesised to be due to the incomplete modification of the His-Cj1344c enzyme in *E. coli*, lack of enzyme co-factors or to instability of the enzyme in *E. coli*. These problems have also been observed in enzymatic studies with *M. haemolytica* gcp expressed in *E. coli* (Abdullah *et al.*, 1991). Abdullah *et al*. have hypothesised that gcp may need posttranslational modification or additional chaperones, absent in *E. coli*, to be activated. They have compared it to leukotoxin, another secreted protein of *M. haemolytica* Al where the leukotoxin determinant is composed of four contiguous genes, *lktCABD*. The expression of *lktC* functions in the activation of leukotoxin (LktA), while proteins encoded by *lktB* and *lktD* are involved in the secretion of leukotoxin (Abdullah *et al.*, 1991). It is not known whether the glycoprotease requires a similar activation mechanism, which might explain the lower activity of the enzyme expressed in

E. coli. Examination of the *C. jejuni* DNA sequence immediately downstream from the *Cj1344c* gene showed an open reading frame encoding a putative periplasmic chaperone-like protein of about 18.6 kDa which might be involved in the secretion of the glycoprotease. This data suggest that the glycoprotease may require chaperon molecules for secretion, or posttranslational modification for its activation. Due to time constrains, studies involving the interactions between Cj1344c and potential chaperones were not further examined.

The strong interactions between His-Cj1344c and MUC2 observed using the glycan array technology, in addition to the limited activity of the recombinant enzyme during the enzymatic studies involving MUC2 digest, suggest that Cj1344c is most likely interacting with mucins and potentially assists during the initial stages of *C. jejuni* pathogenesis. In addition to the findings of this study, the possible interactions between Cj1344c and MUC2 have also been observed in other published studies. Tu *et al*. showed that the MUC2 presence in the media growing *C. jejuni* causes upregulation of the *Cj1344c* gene suggesting that the gene product may be involved in the degradation of this molecule (Tu *et al.*, 2008). On the other hand, the high concentration of the MUC2 in culture media is inhibiting *C. jejuni* growth, so it can also be speculated that the expression of the Cj1344c is one of the protective mechanisms of the bacterium against MUC2.

7.3 Immunisation trial

Development of an effective vaccine against *C. jeju*ni is both, necessary and desirable but is complicated due to a number of factors that include the tremendous antigenic diversity of the organism, a lack of understanding of the nature of acquired immunity, a lack of small animal models suitable for vaccine evaluation, as well as the fact that the protective epitopes are not clearly defined (Scot, 1997). The subunit vaccine approaches generally utilize antigens that play a role in virulence, but *C. jejuni* pathogenesis remains poorly understood, in spite of an intensive study (Jagusztyn-Krynicka *et al.*, 2009), which makes identification of potential vaccine antigens extremely difficult.

The His-Cj1344c immunisation trial was performed after the rabbit immunisation with His-Cj1344c provided a reasonably high antibody titre when compared to published immunisation studies with various *C. jejuni* antigens, such as polysaccharides (Monteiro *et al.*, 2009), whole *C. jejuni* cells (Burr *et al.*, 2005) or flagellum-secreted proteins (Baqar *et al.*, 2008). In addition, the potential role of the enzyme in the pathogenesis of *C. jejuni* and its essential role in bacterial survival identified through mutagenesis studies, as well as the high degree of homology observed between the *Campylobacter* species made it a potentially good candidate for the immunisation trial. In *M. haemolytica*, the immunisation trial with a recombinant fusion protein expressed by *E. coli* (Gcp-F) enhanced the protection of calves against infection. Animals vaccinated with Gcp-F had a significantly lower percent pneumonic tissue than unvaccinated controls and a lower percent pneumonic tissue necropsy (Shewen *et al.*, 2003).

The preliminary data presented in this study indicate the subcutaneous delivery of the His-Cj1344c, when administered with Freund's adjuvant, was most effective in eliciting an immune response, when compared with the intraperitoneal and intranasal routes of delivery. Though similar antibody titres have been reported against different *C. jejuni* antigens such flagellum-secreted FspA1 and FspA2 (Baqar *et al.*, 2008), these proteins show great diversity among *C. jejuni* strains, which would limit their use in the vaccination trial, despite the protective role they may exert against *C. jejuni* infection. This is not the case with Cj1344c, which shows the high degree of homology between strains and is therefore a good vaccine candidate.

In addition, this study has identified that mice vaccinated with His-Cj1344c showed lower bacterial numbers in faeces observed for the duration of the experiment, as well as lower bacterial counts in the small and large intestines compared to non-vaccinated mice. Though the data were not statistically different, this study shows a more objective, quantitative measurement of the immunisation efficiency by enumerating bacteria present in different organs and tissues, compared to other immunisation studies which determine the efficacy of the vaccine by observation of animal well being. In the study by Baqar *et al*. using flagellum-secreted *C. jejuni* proteins, immunisation provided no, or limited strain-specific protection against disease caused by intranasal bacterial challenge, but not against intestinal colonisation (Baqar *et al.*, 2008). The intranasal challenge model used in this study produced some signs of illness in mice, but this model is still debatable (as reviewed (Jagusztyn-Krynicka *et al.*, 2009) as signs and symptoms of the *C. jejuni* infection in mice are usually mild and difficult to observe.

This immunisation study has also identified that the non-vaccinated mice had His-Cj1344c specific antibodies, similar to findings of Shewen *et al*. in *M. haemolytica* gcp vaccination trials (Shewen *et al.*, 2003). This finding suggests that the His-Cj1344c specific antibodies have been produced in the non-vaccinated group of animals through a transient infection with *C. jejuni* or related species against the native protein. More importantly, these studies confirm that the enzyme is expressed by *C. jejuni in vivo* and is likely involved in the bacterial pathogensis. These findings were also recorded by Lee *at al*. who documented the presence of anti-Gcp Ab in bovine serum, which arose spontaneously, as a result of natural exposure to *M.*

haemolytica. They also showed that animals with an anti-Gcp response had less pneumonia at necropsy, suggesting that the presence of anti-Gcp activity may induce protective immunity, enhance resistance to pneumonia and neutralise the effect of gcp on the lung tissue during the infection, thus reducing the severity of the disease (Lee *et al.*, 1994).

7.4 Conclusion and further studies

From the mutagenesis studies of *C. jejuni Cj1344c* reported here it can be concluded that the *Cj1344c* gene inactivation is lethal for *C. jejuni* survival. These findings suggest an intracellular function of the protein, potentially in degradation of the toxic by products of the cell metabolism. On the other hand, the presence of the protein in culture supernatant of *C. jejuni*, as well as the interaction of the His-Cj1344c with MUC2 observed during glycan array studies suggest a possible function of the enzyme in MUC2 degradation during bacterial colonisation/infection. It can be also speculated that the enzyme is involved in protection of the bacterial cells by means of degrading mucin which has been reported as being inhibitory for the *C. jejuni* growth. However, these hypotheses suggesting the functions of the gcp within a bacterial cell could not be confirmed due to time constraints and technical difficulties encountered during the project. The His-Cj1344c specific antibodies could be used in the future to block the activity of the native protein in *C. jejuni* during adherence/invasion studies which would provide an insight into the function of the protein *in vitro* and provide an alternative for creating the isogenic mutant.

Characterisation of *C. jejuni* Cj1344c, glycoprotease homologue has revealed the binding capacity of the protein to methionine, lysine and arginine, suggesting that these amino acids are present in the sequences within glycoproteins that are recognised by Cj1344c. The specificity of His-Cj1344c to MUC2 observed during glycan array studies could not be conclusively confirmed by the enzymatic digest of the MUC2 with His-Cj1344c. This was hypothesised to be caused by the inactivity of the recombinant protein due to misfolding, instability in *E. coli* or lack of enzyme co-factors. Utilisation of the native protein which was detected in the *C. jejuni* culture supernatant may circumvent the problems associated with the recombinant protein expressed in *E. coli*. The His-Cj1344c specific antibodies can potentially be utilised to block the activity of the enzyme in the negative controls. This approach would address the problem of the presence of other proteases exerting an effect on the MUC2 and masking the true effect of Cj1344c which was experienced in other studies using culture supernatant.

The preliminary mouse immunisation study has identified that subcutaneous immunisation provided the best immune response to His-Cj1344c. The protection studies against *C. jejuni* infection have determined that mice immunised with His-Cj1344c show a lower number of *C. jejuni* cells in their faeces and small and large intestines, which was indicative of lower colonisation. Though the results of this study were not statistically different, the trend observed in the results is suggesting it is worthwhile continuing to examine this antigen as the vaccine candidate against *C. jejuni* infection. The minimal dose of 5 µg His-Cj1344c used during immunisation did not produce any adverse effects in mice, so the future studies could potentially use a higher dose of the antigen and potentially increase the antibody titre levels. Examination of the pathological changes in the gastrointestinal tract of immunised vs. non-immunised animals could provide an insight into the changes in the gastrointestinal pathology caused by immunisation with His-Cj1344c and subsequent *C. jejuni* challenge. These findings could potentially provide an insight into the role and action of the glycoprotease on the mucous layer *in vivo*.

Appendix A

Amino Acid Array

Amino Acid	1-letter	Description
Alanine	A	Nonpolar, hydrophobic
Arginine	R	Polar, basic, hydrophilic
Asparagine	N	Polar, hydrophilic
Aspartic acid potassium salt	D	Polar, acidic, hydrophilic
Cysteine	C	Polar, hydrophilic, acidic
Glutamic acid sodium salt	E	Polar, acidic, hydrophilic
Glutamine	Q	Polar, hydrophilic
Glycine	G	Nonpolar, hydrophobic
Histidine	H	Polar, basic, hydrophilic
Isoleucine	I	Nonpolar, hydrophobic
Leucine	L	Nonpolar, hydrophobic
Lysine	K	Polar, basic, hydrophilic
Methionine	M	Nonpolar, hydrophobic
Phenylalanine	F	Nonpolar, hydrophobic
Proline	P	Nonpolar, hydrophobic
Serine	S	Polar, hydrophilic
Threonine	T	Polar, hydrophilic
Tryptophane	W	Nonpolar, hydrophobic
Tyrosine	Y	Polar, hydrophilic
Valine	V	Nonpolar, hydrophobic

Appendix B

Table of Glycans

Code	Name	Structure
	Terminal Galactose	
1A.	Lacto-*N*-Biose I	Galβ1-3Glc*N*Ac
1B.	N-Acetyllactosamine	Galβ1-4Glc*N*Ac
1C.	β1-4galactosyl-galactose	Galβ1-4Gal
1D.	β1-6galactosyl-*N*-acetylglucosamine	Galβ1-6Glc*N*Ac
1E.	β1-3galactosyl-*N*-acetylgalactosamine	Galβ1-3Gal*N*Ac
1F.	Galb1-3Gal*N*Acβ1-4Galβ1-4Glc	
1G	Lacto-*N*-tetrose	Galβ1-3Glc*N*Acβ1-3Galβ1-4Glc
1H	Lacto-*N*-neotetrose	Galβ1-4Glc*N*Acβ1-3Galβ1-4Glc
1I.	Lacto-*N*-neohexose	Galβ1-4Glc*N*Acβ1-6(Galβ1-4Glc*N*Acβ1-3)Galβ1-4Glc
1J.	Lacto-*N*-hexose	Galβ1-4Glc*N*Acβ1-6(Galβ1-3Glc*N*Acβ1-3)Galβ1-4Glc
1K.	Globotriose	Gal□1-4Galβ1-4Glc
1L.	Tn Antigen	Gal*N*Ac□1-*O*-Ser
1M.	Galactosyl-Tn Antigen	Gal□1-3Gal*N*Acα1-*O*-Ser
1N.	α1-3 Galactobiose	Gal□1-3Gal
1O.	Linear B-2 Trisaccharide	Gal□1-3Galβ1-4Glc*N*Ac
1P.	Linear B-6 Trisaccharide	Gal□1-3Galβ1-4Glc
2A.	α1-3, β1-4, α1-3 Galactotetrose	Gal□1-3Galβ1-4Galα1-3Gal
2B.	β1-6Galactobiose	Galβ1-6Gal
2C.	Terminal disaccharide of globotriose	Gal*N*Acβ1-3Gal
2D.	Receptor for pili of *P. aeruginosa*	Gal*N*Acβ1-4Gal
2E.	P1 Antigen	Galα1-4Galβ1-4Glc*N*Ac
2F.	α-D-*N*-acetylgalactosaminyl-1-3Gal-β1-4Glc	Gal*N*Acα1-3Galβ1-4Glc
2G.	iso-Lacto-*N*-octose	Galβ1-3Glc*N*Acβ1-3Galβ1-4Glc*N*Acβ1-6(Galβ1-3Glc*N*Acβ1-3)Galβ1-4Glc
2H.	*para*-Lacto-*N*-hexose	Galβ1-3Glc*N*Acβ1-3Galβ1-4Glc*N*Acβ1-3Galβ1-4Glc

Appendix B (continued)

Code	Name	Structure
	Terminal N`Acetyl glucosamine	
4A.	*N*,*N'*-Diacetyl chitobiose	Glc*N*Acβ1-4Glc*N*Ac
4B.	*N*,*N'*,*N''*-Triacetyl chitotriose	Glc*N*Acβ1-4Glc*N*Acβ1-4Glc*N*Ac
4C.	*N*,*N'*,*N''*,*N'''*-Tetraacetyl chitotetrose	Glc*N*Acβ1-4Glc*N*Acβ1-4Glc*N*Acβ1-4Glc*N*Ac
4D.	*N*,*N'*,*N''*,*N'''*,*N''''*,*N'''''*-Hexaacetyl chitohexose	Glc*N*Acβ1-4Glc*N*Acβ1-4Glc*N*Acβ1-4Glc*N*Acβ1-4Glc*N*Acβ1-4Glc*N*Ac
4E.	Bacterial cell wall muramyl discaccharide	Glc*N*Acβ1-4Mur*N*Ac
	Mannose containing structures	
5A.	β1-2-*N*-Acetylglucosamine-mannose	Glc*N*Acβ1-2Man
5B.	Bianntennary *N*-linked core pentasaccharide	Glc*N*Acβ1-2Manα1-6(Glc*N*Acβ1-2Manα1-3)Man
5C.	α1-2-Mannobiose	Manα1-2Man
5D.	α1-3-Mannobiose	Manα1-3Man
5E.	α1-4-Mannobiose	Manα1-4Man
5F.	α1-6-Mannobiose	Manα1-6Man
5G.	α1-3, α1-6-Mannobiose	Manα1-6(Manα1-3)Man
5H.	α1-3, α1-3, α1-6-Mannopentaose	Manα1-6(Manα1-3)Manα1-6(Manα1-3)Man

Appendix B (continued)

Code	Name	Structure
	Fucosylated structures	
7A.	Lacto-*N*-fucopentose I	Fucα1-2Galβ1-3Glc*N*Acβ1-3Galβ1-4Glc
7B.	Lacto-*N*-fucopentose II	Galβ1-3(Fucα1-4)Glc*N*Acβ1-3Galβ1-4Glc
7C.	Lacto-*N*-fucopentose III	Galβ1-4(Fucα1-3)Glc*N*Acβ1-3Galβ1-4Glc
7D.	Lacto-*N*-difucohexose I	Fucα1-2Galβ1-3(Fucα1-4)Glc*N*Acβ1-3Galβ1-4Glc
7E.	Lacto-*N*-difucohexose II	Galβ1-3(Fucα1-4)Glc*N*Acβ1-3Galβ1-4(Fucα1-3)Glc
7F.	H-disaccharide	Fucα1-2Gal
7G.	2'-Fucosyllactose	Fucα1-2Galβ1-4Glc
7H.	3'-Fucosyllactose	Galβ1-4(Fucα1-3)Glc
7I.	Lewis^{x}	Galβ1-4(Fucα1-3)Glc*N*Ac
7J.	Lewis^{a}	Galβ1-3(Fucα1-4)Glc*N*Ac
7K.	Blood Group A-trisaccharide	Gal*N*Acα1-3(Fucα1-2)Gal
7L.	Lactodifucotetrose	Fucα1-2Galβ1-4(Fucα1-3)Glc
7M.	Blood Group B-Trisaccharide	Galβ1-3(Fucα1-2)Gal
7N.	Lewis^{y}	Fucα1-2Galβ1-4(Fucα1-3)Glc*N*Ac
7O.	Blood Group H Type II Trisaccharide	Fucα1-2Galβ1-3Glc*N*Ac
7P.	Lewisb tetrasaccharide	Fucα1-2Galβ1-3(Fucα1-4)Glc*N*Ac
8A.	Sulpho Lewis^{a}	SO_3-3Galβ1-3(Fucα1-4)Glc*N*Ac
8B.	Sulpho Lewis^{x}	SO_3-3Galβ1-4(Fucα1-3)Glc*N*Ac
8C.	Monofucosyl-para-Lacto-*N*-hexose IV	Galβ1-3Glc*N*Acβ1-3Galβ1-4(Fucα1-3)Glc*N*Acβ1-3Galβ1-4Glc
8D.	Monofucosyllacto-*N*-hexose III	Galβ1-4(Fucα1-3)Glc*N*Acβ1-6(Galβ1-3Glc*N*Acβ1-3)Galβ1-4Glc
8E.	Difucosyllacto-*N*-hexose	Galβ1-4(Fucα1-3)Glc*N*Acβ1-6(Fucα1-2Galβ1-3Glc*N*Acβ1-3)Galβ1-4Glc
8F.	Trifucosyllacto-*N*-hexose	Galβ1-4(Fucα1-3)Glc*N*Acβ1-6(Fucα1-2Galβ1-3(Fucα1-4)Glc*N*Acβ1-3)Galβ1-4Glc

Appendix B (continued)

Code	Name	Structure
	Neu5Ac containing structures	
10A.	Sialyl Lewis[a]	Neu5Acα2-3Galβ1-3(Fucα1-4)Glc*N*Ac
10B.	Sialyl Lewis[x]	Neu5Acα2-3Galβ1-4(Fucα1-3)Glc*N*Ac
10C.	Sialyllacto-*N*-tetrose a	Neu5Acα2-3Galβ1-3Glc*N*Acβ1-3Galβ1-4Glc
10D.	Monosialyl, monofucosyllacto-*N*-neohexose	Galβ1-4(Fucα1-3)Glc*N*Acβ1-6(Neu5Acα2-6Galβ1-4Glc*N*Acβ1-3)Galβ1-4Glc
10K.	2,3'-Sialyllactosamine	Neu5Acα2-3Galβ1-4Glc*N*Ac
10L.	2,6'-Sialyllactosamine	Neu5Acα2-6Galβ1-4Glc*N*Ac
10M.	LS-Tetrasaccharide a	
10N.	LS-Tetrasaccharide b	Galβ1-3(Neu5Acα2-6)Glc*N*Acβ1-3Galβ1-4Glc
10O.	LS-Tetrasaccharide c	Neu5Acα2-6Galβ1-4Glc*N*Acβ1-3Galβ1-4Glc
10P.	Disialyllacto-*N*-tetrose	Neu5Acα2-3Galβ1-3(Neu5Acα2-6)Glc*N*Acβ1-3Galβ1-4Glc
11A.	2,3'-Sialyllactose	Neu5Acα2-3Galβ1-4Glc
11B.	2,6'-Sialyllactose	Neu5Acα2-6Galβ1-4Glc
11C.	Colominic acid	(Neu5Acα2-8Neu5Ac)n (n<50)
11D.	Biantennary 2,6-sialylated-*N*-glycan-Asn	Neu5Acα2-6Galβ1-4Glc*N*Acβ1-2Manα1-6(Neu5Acα2-6Galβ1-4Glc*N*Acβ1-2Manα1-6)Manβ1-4Glc*N*Acβ1-4Glc*N*Ac-Asn

Appendix B (continued)

Code	Name	Structure
	Carageenan and Glycoaminoglycans (GAGs)	
12A.	Neocarratetrose-41, 3-di-*O*-sulphate (Na^+)	$C_{24}H_{36}O_{25}S_2Na_2$ (Mixed anomers. Tetrasaccharide of regular κ - carrageenan)
12B.	Neocarratetrose-41-*O*-sulphate (Na^+)	$C_{24}H_{37}O_{22}SNa$ (Mixed anomers. Derived from C1003 by removal of the non-reducing terminal 4-sulphate)
12C.	Neocarrahexose-24,41, 3, 5-tetra-*O*-sulphate (Na^+)	$C_{36}H_{52}O_{40}S_4Na_4$ (Mixed anomers. A hybrid sequence comprising carrageenan disaccharides in the order κ–ι–κ, derived from the carrageenan from *Chondrus crispus*)
12D.	Neocarrahexose-41, 3, 5-tri-*O*-sulphate (Na^+)	$C_{36}H_{53}O_{37}S_3Na_3$ (Mixed anomers. Hexasaccharide of regular κ-carrageenan)
12E.	Neocarraoctose-41, 3, 5, 7-tetra-*O*-sulphate (Na^+)	$C_{48}H_{70}O_{49}S_4Na_4$ (Mixed anomers. Octasaccharide of regular κ-carrageenan)
12F.	Neocarradecose-41, 3, 5, 7, 9-penta-*O*-sulphate (Na^+)	$C_{60}H_{87}O_{61}S_5Na_5$ (Mixed anomers. Decasaccharide of regular κ-carrageenan)
12G.	ΔUA-2S → Glc*N*S-6S Na_4 (I-S)	$C_{12}H_{15}NO_{19}S_3Na_4$ (Predominant disaccharide produced from heparin by heparinase I and II)
12H.	ΔUA → Gluc*N*S-6S Na_3 (II-S)	$C_{12}H_{16}NO_{16}S_2Na_3$ (Produced from heparinase II digestion of heparin and heparin sulphate)
12I.	ΔUA → 2S-Glc*N*S Na_3 (III-S)	$C_{12}H_{16}NO_{16}S_2Na_3$ (Produced from heparin by digestion with heparinase I and II)
12J.	ΔUA → 2S-Glc*N*Ac-6S Na_3 (I-A)	$C_{14}H_{18}NO_{17}S_2Na_3$ (Minor component produced from heparin by heparinase II)

Appendix B (continued)

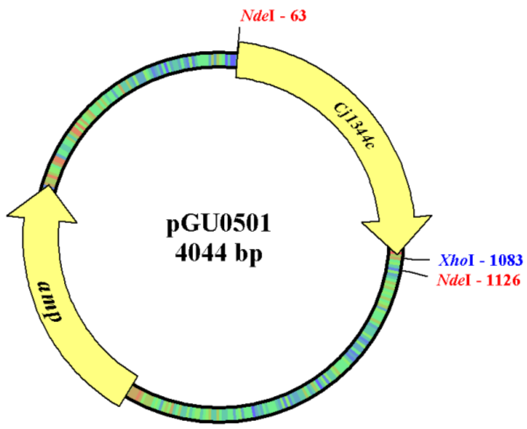
Code	Name	Structure
	Carageenan and Glycoaminoglycans (GAGs)	
12K.	ΔUA → Glc*N*Ac-6S Na_2 (II-A)	$C_{14}H_{19}NO_{14}SNa_2$ (Product of the action of heparinases II and III on heparin and heparan sulphate)
12L.	ΔUA → 2S-Glc*N*Ac Na_2 (III-A)	$C_{14}H_{19}NO_{14}SNa_2$ (Minor product of the action of heparinase II on heparin)
12M.	ΔUA → Glc*N*Ac Na (IV-A)	$C_{14}H_{20}NO_{11}Na$ (Produced from heparin sulphate by digestion With heparinase III)
12N.	ΔUA → Gal*N*Ac-4S Na_2 (ΔDi-4S)	$C_{14}H_{19}NO_{14}SNa_2$ (Produced from various chondroitin sulphates By the action of chondroitinases ABC, B and AC-1)
12O.	ΔUA → Gal*N*Ac-6S Na2 (ΔDi-6S)	$C_{14}H_{19}NO_{14}SNa_2$ (Produced from various chondroitin sulphates By the action of chondroitinases ABC, AC-1 and C)
12P.	ΔUA → Gal*N*Ac-4S,6S Na_3 (ΔDi-disE)	$C_{14}H_{18}NO_{17}S_2Na_3$ (Produced from various chondroitin sulphates By the action of chondroitinases ABC, B and AC-1)
13A.	ΔUA → 2S-Gal*N*Ac-4S Na_2 (ΔDi-disB)	$C_{14}H_{18}NO_{17}S_2Na_3$ (Produced from various chondroitin sulphates by action of chondroitinase ABC and/or B. Most typically from chondroitin sulphate B (dermatan sulphate))
13B.	ΔUA → 2S-Gal*N*Ac-6S Na_3 (ΔDi-disD)	$C_{14}H_{18}NO_{17}S_2Na_3$ (Produced from various chondroitin sulphates by the action of chondroitinase ABC)
13C.	ΔUA → 2S-Gal*N*Ac-4S-6S Na_4 (ΔDi-tisS)	$C_{14}H_{17}NO_{20}S_3Na_4$ (Produced as a minor component by the action of chondroitinase ABC on various chondroitin sulphates, particularly B)

Appendix B (continued)

Code	Name	Structure
	Carageenan and Glycoaminoglycans (GAGs)	
13E.	ΔUA → Glc*N*Ac Na (ΔDi-HA)	$C_{14}H_{20}NO_{11}Na$ (The only unsaturated disaccharide produced from hyaluronic acid by the action of chondroitinase ABC or AC-1)
13F.	Hyaluronan fragments (4mer)	(GlcAβ1-3Glc*N*Acβ1-4)n (n=4)
13G.	Hyaluronan fragments (8mer)	(GlcAβ1-3Glc*N*Acβ1-4)n (n=8)
13H.	Hyaluronan fragments (10mer)	(GlcAβ1-3Glc*N*Acβ1-4)n (n=10)
13I.	Hyaluronan fragments (12mer)	(GlcAβ1-3Glc*N*Acβ1-4)n (n=12)
13J.	Heparin	(GlcA/IdoAα/β1-4Glc*N*Acα1-4)n (n=200)
13K.	Chondroitin sulfate	(GlcA/IdoAβ1-3(±4/6S)Gal*N*Acβ1-4)n (n<250)
13L.	Dermatan sulfate	((±2S)GlcA/IdoAα/b1-3(±4S)Gal*N*Acβ1-4)n (n<250)
13M.	Chondroitin 6-Sulfate	(GlcA/IdoAβ1-3(±6S)Gal*N*Acβ1-4)n (n<250)

Appendix C

pGU0501 (pGEM T-Easy::*Cj1344c*)




```
  1 GGGCGAATTG GGCCCGACGT CGCATGCTCC CGGCCGCCAT GGCGGCCGCG
    CCCGCTTAAC CCGGGCTGCA GCGTACGAGG GCCGGCGGTA CCGCCGGCGC

                NdeI
 51 GGAATTCGAT TCATATGAAA AATCTTATCC TAGCTATAGA AAGTTCTTGT
    CCTTAAGCTA AGTATACTTT TTAGAATAGG ATCGATATCT TTCAAGAACA
                   M  K   N  L  I  L   A  I  E   S  S  C   Cj1344c

101 GATGATAGTT CTATAGCTAT CATTGATAAA AACACCTTAG AATGTAAATT
    CTACTATCAA GATATCGATA GTAACTATTT TTGTGGAATC TTACATTTAA
    D  D  S  S   I  A  I   I  D  K   N  T  L  E   C  K  F  Cj1344c

151 TCATAAAAAA ATTTCCCAAG AATTAGATCA TAGTATCTAT GGGGGAGTGG
    AGTATTTTTT TAAAGGGTTC TTAATCTAGT ATCATAGATA CCCCCTCACC
     H  K  K   I  S  Q  E   L  D  H   S  I  Y   G  G  V  V Cj1344c

201 TACCTGAACT TGCTGCAAGA CTTCATAGCG AGGCTTTACC AAAGATGCTT
    ATGGACTTGA ACGACGTTCT GAAGTATCGC TCCGAAATGG TTTCTACGAA
      P  E  L   A  A  R   L  H  S  E   A  L  P   K  M  L   Cj1344c

251 AAGCAATGCA AAGAGCATTT TAAAAATCTT TGTGCCATAG CTGTGACAAA
    TTCGTTACGT TTCTCGTAAA ATTTTTAGAA ACACGGTATC GACACTGTTT
    K  Q  C  K   E  H  F   K  N  L   C  A  I  A   V  T  N  Cj1344c

301 TGAACCTGGA CTTAGTGTTT CTTTGCTCAG TGGAATTTCT ATGGCAAAAA
    ACTTGGACCT GAATCACAAA GAAACGAGTC ACCTTAAAGA TACCGTTTTT
     E  P  G   L  S  V  S   L  L  S   G  I  S   M  A  K  T Cj1344c

351 CCTTAGCAAG TGCGCTAAAT TTACCCTTAA TCCCTATAAA TCATCTTAAA
    GGAATCGTTC ACGCGATTTA AATGGGAATT AGGGATATTT AGTAGAATTT
      L  A  S   A  L  N   L  P  L  I   P  I  N   H  L  K   Cj1344c
```

```
 401 GGTCATATTT ATAGTCTTTT TTTGGAAGAA AAAATTTCTT TAGATATGGG
     CCAGTATAAA TATCAGAAAA AAACCTTCTT TTTTAAAGAA ATCTATACCC
     G  H  I  Y   S  L  F   L  E  E   K  I  S  L   D  M  G  Cj1344

 451 AATTTTGCTT GTTAGTGGTG GGCATACCAT GGTGCTTTAT CTTAAAGATG
     TTAAAACGAA CAATCACCAC CCGTATGGTA CCACGAAATA GAATTTCTAC
      I  L  L   V  S  G  G   H  T  M   V  L  Y   L  K  D  D Cj1344c

 501 ATGCAAGCTT AGAGCTTTTA GCAAGTACAA ATGATGATAG CTTTGGAGAA
     TACGTTCGAA TCTCGAAAAT CGTTCATGTT TACTACTATC GAAACCTCTT
       A  S  L   E  L  L   A  S  T  N   D  D  S   F  G  E   Cj1344c

 551 AGTTTTGATA AAGTGGCTAA AATGATGAAT TTAGGTTACC CTGGTGGGGT
     TCAAAACTAT TTCACCGATT TTACTACTTA AATCCAATGG GACCACCCCA
     S  F  D  K   V  A  K   M  M  N   L  G  Y  P   G  G  V  Cj1344c

 601 CATCATAGAA AATTTAGCAA AAAATGCCAA ACTTAAAAAT ATCTCTTTTA
     GTAGTATCTT TTAAATCGTT TTTTACGGTT TGAATTTTTA TAGAGAAAAT
      I  I  E   N  L  A  K   N  A  K   L  K  N   I  S  F  N Cj1344c

 651 ACACACCTTT AAAGCATTCT AAAGAACTCG CTTTCAGTTT TTCAGGGCTT
     TGTGTGGAAA TTTCGTAAGA TTTCTTGAGC GAAAGTCAAA AAGTCCCGAA
       T  P  L   K  H  S   K  E  L  A   F  S  F   S  G  L   Cj1344c

 701 AAAAATGCAG TGCGTTTGGA AATTTTAAAA CATGAAAATT TAAATGAAGA
     TTTTTACGTC ACGCAAACCT TTAAAATTTT GTACTTTTAA ATTTACTTCT
     K  N  A  V   R  L  E   I  L  K   H  E  N  L   N  E  D  Cj1344c

 751 CACAAAAGCA GAAATAGCCT ATGCCTTTGA AAATACAGCT TGTGATCATA
     GTGTTTTCGT CTTTATCGGA TACGGAAACT TTTATGTCGA ACACTAGTAT
      T  K  A   E  I  A  Y   A  F  E   N  T  A   C  D  H  I Cj1344c

 801 TCATGGATAA ATTAGAAAAA ATTTTTAATC TTTATAAATT TAAAAATTTT
     AGTACCTATT TAATCTTTTT TAAAAATTAG AAATATTTAA ATTTTTAAAA
       M  D  K   L  E  K   I  F  N  L   Y  K  F   K  N  F   Cj1344c

 851 GGCGTTGTAG GTGGAGCTAG TGCAAATCTT AACTTGCGTT CGCGTTTGCA
     CCGCAACATC CACCTCGATC ACGTTTAGAA TTGAACGCAA GCGCAAACGT
     G  V  V  G   G  A  S   A  N  L   N  L  R  S   R  L  Q  Cj1344c

 901 AAATTTATGT CAAAAATATA ATGCAAATTT AAAACTAGCT CCTTTAAAAT
     TTTAAATACA GTTTTTATAT TACGTTTAAA TTTTGATCGA GGAAATTTTA
      N  L  C   Q  K  Y  N   A  N  L   K  L  A   P  L  K  F Cj1344c

 951 TCTGCTCTGA TAATGCTTTG ATGATAGCAA GAGCCGCAGT TGATGCTTAT
     AGACGAGACT ATTACGAAAC TACTATCGTT CTCGGCGTCA ACTACGAATA
       C  S  D   N  A  L   M  I  A  R   A  A  V   D  A  Y   Cj1344c

1001 GAAAAAAAGG AATTTGTAAG TGTAGAAGAA GATATTTTAA GCCCTAAAAA
     CTTTTTTTCC TTAAACATTC ACATCTTCTT CTATAAAATT CGGGATTTTT
     E  K  K  E   F  V  S   V  E  E   D  I  L  S   P  K  N  Cj1344c

                                        XhoI
1051 TAAAAATTTT TCAAGGATAT AGATGAAAAA AGCTCGAGAA TCACTAGTGA
     ATTTTTAAAA AGTTCCTATA TCTACTTTTT TCGAGCTCTT AGTGATCACT
      K  N  F   S  R  I  *   Cj1344c
```

```
                                 NdeI
1101 ATTCGCGGCC GCCTGCAGGT CGACCATATG GGAGAGCTCC CAACGCGTTG
     TAAGCGCCGG CGGACGTCCA GCTGGTATAC CCTCTCGAGG GTTGCGCAAC

1151 GATGCATAGC TTGAGTATTC TATAGTGTCA CCTAAATAGC TTGGCGTAAT
     CTACGTATCG AACTCATAAG ATATCACAGT GGATTTATCG AACCGCATTA

1201 CATGGTCATA GCTGTTTCCT GTGTGAAATT GTTATCCGCT CACAATTCCA
     GTACCAGTAT CGACAAAGGA CACACTTTAA CAATAGGCGA GTGTTAAGGT

1251 CACAACATAC GAGCCGGAAG CATAAAGTGT AAAGCCTGGG GTGCCTAATG
     GTGTTGTATG CTCGGCCTTC GTATTTCACA TTTCGGACCC CACGGATTAC

1301 AGTGAGCTAA CTCACATTAA TTGCGTTGCG CTCACTGCCC GCTTTCCAGT
     TCACTCGATT GAGTGTAATT AACGCAACGC GAGTGACGGG CGAAAGGTCA

1351 CGGGAAACCT GTCGTGCCAG CTGCATTAAT GAATCGGCCA ACGCGCGGGG
     GCCCTTTGGA CAGCACGGTC GACGTAATTA CTTAGCCGGT TGCGCGCCCC

1401 AGAGGCGGTT TGCGTATTGG GCGCTCTTCC GCTTCCTCGC TCACTGACTC
     TCTCCGCCAA ACGCATAACC CGCGAGAAGG CGAAGGAGCG AGTGACTGAG

1451 GCTGCGCTCG GTCGTTCGGC TGCGGCGAGC GGTATCAGCT CACTCAAAGG
     CGACGCGAGC CAGCAAGCCG ACGCCGCTCG CCATAGTCGA GTGAGTTTCC

1501 CGGTAATACG GTTATCCACA GAATCAGGGG ATAACGCAGG AAAGAACATG
     GCCATTATGC CAATAGGTGT CTTAGTCCCC TATTGCGTCC TTTCTTGTAC

1551 TGAGCAAAAG GCCAGCAAAA GGCCAGGAAC CGTAAAAAGG CCGCGTTGCT
     ACTCGTTTTC CGGTCGTTTT CCGGTCCTTG GCATTTTTCC GGCGCAACGA

1601 GGCGTTTTTC CATAGGCTCC GCCCCCCTGA CGAGCATCAC AAAAATCGAC
     CCGCAAAAAG GTATCCGAGG CGGGGGGACT GCTCGTAGTG TTTTTAGCTG

1651 GCTCAAGTCA GAGGTGGCGA AACCCGACAG GACTATAAAG ATACCAGGCG
     CGAGTTCAGT CTCCACCGCT TTGGGCTGTC CTGATATTTC TATGGTCCGC

1701 TTTCCCCCTG GAAGCTCCCT CGTGCGCTCT CCTGTTCCGA CCCTGCCGCT
     AAAGGGGGAC CTTCGAGGGA GCACGCGAGA GGACAAGGCT GGGACGGCGA

1751 TACCGGATAC CTGTCCGCCT TTCTCCCTTC GGGAAGCGTG GCGCTTTCTC
     ATGGCCTATG GACAGGCGGA AAGAGGGAAG CCCTTCGCAC CGCGAAAGAG

1801 ATAGCTCACG CTGTAGGTAT CTCAGTTCGG TGTAGGTCGT TCGCTCCAAG
     TATCGAGTGC GACATCCATA GAGTCAAGCC ACATCCAGCA AGCGAGGTTC

1851 CTGGGCTGTG TGCACGAACC CCCCGTTCAG CCCGACCGCT GCGCCTTATC
     GACCCGACAC ACGTGCTTGG GGGGCAAGTC GGGCTGGCGA CGCGGAATAG

1901 CGGTAACTAT CGTCTTGAGT CCAACCCGGT AAGACACGAC TTATCGCCAC
     GCCATTGATA GCAGAACTCA GGTTGGGCCA TTCTGTGCTG AATAGCGGTG

1951 TGGCAGCAGC CACTGGTAAC AGGATTAGCA GAGCGAGGTA TGTAGGCGGT
     ACCGTCGTCG GTGACCATTG TCCTAATCGT CTCGCTCCAT ACATCCGCCA

2001 GCTACAGAGT TCTTGAAGTG GTGGCCTAAC TACGGCTACA CTAGAAGAAC
     CGATGTCTCA AGAACTTCAC CACCGGATTG ATGCCGATGT GATCTTCTTG

2051 AGTATTTGGT ATCTGCGCTC TGCTGAAGCC AGTTACCTTC GGAAAAAGAG
     TCATAAACCA TAGACGCGAG ACGACTTCGG TCAATGGAAG CCTTTTTCTC
```

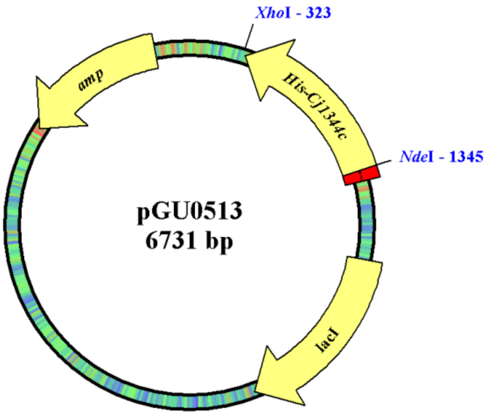
```
2101 TTGGTAGCTC TTGATCCGGC AAACAAACCA CCGCTGGTAG CGGTGGTTTT
     AACCATCGAG AACTAGGCCG TTTGTTTGGT GGCGACCATC GCCACCAAAA

2151 TTTGTTTGCA AGCAGCAGAT TACGCGCAGA AAAAAAGGAT CTCAAGAAGA
     AAACAAACGT TCGTCGTCTA ATGCGCGTCT TTTTTTCCTA GAGTTCTTCT

2201 TCCTTTGATC TTTTCTACGG GGTCTGACGC TCAGTGGAAC GAAAACTCAC
     AGGAAACTAG AAAAGATGCC CCAGACTGCG AGTCACCTTG CTTTTGAGTG

2251 GTTAAGGGAT TTTGGTCATG AGATTATCAA AAAGGATCTT CACCTAGATC
     CAATTCCCTA AAACCAGTAC TCTAATAGTT TTTCCTAGAA GTGGATCTAG

2301 CTTTTAAATT AAAAATGAAG TTTTAAATCA ATCTAAAGTA TATATGAGTA
     GAAAATTTAA TTTTTACTTC AAAATTTAGT TAGATTTCAT ATATACTCAT

2351 AACTTGGTCT GACAGTTACC AATGCTTAAT CAGTGAGGCA CCTATCTCAG
     TTGAACCAGA CTGTCAATGG TTACGAATTA GTCACTCCGT GGATAGAGTC
                          *  W  H  K  I  L  S  A  G  I  E   ampR

2401 CGATCTGTCT ATTTCGTTCA TCCATAGTTG CCTGACTCCC CGTCGTGTAG
     GCTAGACAGA TAAAGCAAGT AGGTATCAAC GGACTGAGGG GCAGCACATC
     A  I  Q  R  N  R  E  D  M  T  A  Q  S  G  T  T  Y   ampR

2451 ATAACTACGA TACGGGAGGG CTTACCATCT GGCCCCAGTG CTGCAATGAT
     TATTGATGCT ATGCCCTCCC GAATGGTAGA CCGGGGTCAC GACGTTACTA
      I  V  V  I  R  S  P  K  G  D  P  G  L  A  A  I  I  ampR

2501 ACCGCGAGAC CCACGCTCAC CGGCTCCAGA TTTATCAGCA ATAAACCAGC
     TGGCGCTCTG GGTGCGAGTG GCCGAGGTCT AAATAGTCGT TATTTGGTCG
       G  R  S  G  R  E  G  A  G  S  K  D  A  I  F  W    ampR

2551 CAGCCGGAAG GGCCGAGCGC AGAAGTGGTC CTGCAACTTT ATCCGCCTCC
     GTCGGCCTTC CCGGCTCGCG TCTTCACCAG GACGTTGAAA TAGGCGGAGG
     G  A  P  L  A  S  R  L  L  P  G  A  V  K  D  A  E   ampR

2601 ATCCAGTCTA TTAATTGTTG CCGGGAAGCT AGAGTAAGTA GTTCGCCAGT
     TAGGTCAGAT AATTAACAAC GGCCCTTCGA TCTCATTCAT CAAGCGGTCA
      M  W  D  I  L  Q  Q  R  S  A  L  T  L  L  E  G  T  ampR

2651 TAATAGTTTG CGCAACGTTG TTGCCATTGC TACAGGCATC GTGGTGTCAC
     ATTATCAAAC GCGTTGCAAC AACGGTAACG ATGTCCGTAG CACCACAGTG
       L  L  K  R  L  T  T  A  M  A  V  P  M  T  T  D    ampR

2701 GCTCGTCGTT TGGTATGGCT TCATTCAGCT CCGGTTCCCA ACGATCAAGG
     CGAGCAGCAA ACCATACCGA AGTAAGTCGA GGCCAAGGGT TGCTAGTTCC
     R  E  D  N  P  I  A  E  N  L  E  P  E  W  R  D  L   ampR

2751 CGAGTTACAT GATCCCCCAT GTTGTGCAAA AAAGCGGTTA GCTCCTTCGG
     GCTCAATGTA CTAGGGGGTA CAACACGTTT TTTCGCCAAT CGAGGAAGCC
      R  T  V  H  D  G  M  N  H  L  F  A  T  L  E  K  P  ampR

2801 TCCTCCGATC GTTGTCAGAA GTAAGTTGGC CGCAGTGTTA TCACTCATGG
     AGGAGGCTAG CAACAGTCTT CATTCAACCG GCGTCACAAT AGTGAGTACC
       G  G  I  T  T  L  L  L  N  A  A  T  N  D  S  M    ampR

2851 TTATGGCAGC ACTGCATAAT TCTCTTACTG TCATGCCATC CGTAAGATGC
     AATACCGTCG TGACGTATTA AGAGAATGAC AGTACGGTAG GCATTCTACG
     T  I  A  A  S  C  L  E  R  V  T  M  G  D  T  L  H   ampR
```

```
2901 TTTTCTGTGA CTGGTGAGTA CTCAACCAAG TCATTCTGAG AATAGTGTAT
     AAAAGACACT GACCACTCAT GAGTTGGTTC AGTAAGACTC TTATCACATA
      K  E  T   V  P  S  Y   E  V  L   D  N  Q   S  Y  H  I ampR

2951 GCGGCGACCG AGTTGCTCTT GCCCGGCGTC AATACGGGAT AATACCGCGC
     CGCCGCTGGC TCAACGAGAA CGGGCCGCAG TTATGCCCTA TTATGGCGCG
       R  R  G   L  Q  E   Q  G  A  D   I  R  S   L  V  A   ampR

3001 CACATAGCAG AACTTTAAAA GTGCTCATCA TTGGAAAACG TTCTTCGGGG
     GTGTATCGTC TTGAAATTTT CACGAGTAGT AACCTTTTGC AAGAAGCCCC
     G  C  L  L   V  K  F   T  S  M   M  P  F  R   E  E  P  ampR

3051 CGAAAACTCT CAAGGATCTT ACCGCTGTTG AGATCCAGTT CGATGTAACC
     GCTTTTGAGA GTTCCTAGAA TGGCGACAAC TCTAGGTCAA GCTACATTGG
      R  F  S   E  L  I  K   G  S  N   L  D  L   E  I  Y  G ampR

3101 CACTCGTGCA CCCAACTGAT CTTCAGCATC TTTTACTTTC ACCAGCGTTT
     GTGAGCACGT GGGTTGACTA GAAGTCGTAG AAAATGAAAG TGGTCGCAAA
       V  R  A   G  L  Q   D  E  A  D   K  V  K   V  L  T   ampR

3151 CTGGGTGAGC AAAAACAGGA AGGCAAAATG CCGCAAAAAA GGGAATAAGG
     GACCCACTCG TTTTTGTCCT TCCGTTTTAC GGCGTTTTTT CCCTTATTCC
     E  P  H  A   F  V  P   L  C  F   A  A  F  F   P  I  L  ampR

3201 GCGACACGGA AATGTTGAAT ACTCATACTC TTCCTTTTTC AATATTATTG
     CGCTGTGCCT TTACAACTTA TGAGTATGAG AAGGAAAAAG TTATAATAAC
      A  V  R   F  H  Q  I   S  M  ampR

3251 AAGCATTTAT CAGGGTTATT GTCTCATGAG CGGATACATA TTTGAATGTA
     TTCGTAAATA GTCCCAATAA CAGAGTACTC GCCTATGTAT AAACTTACAT

3301 TTTAGAAAAA TAAACAAATA GGGGTTCCGC GCACATTTCC CCGAAAAGTG
     AAATCTTTTT ATTTGTTTAT CCCCAAGGCG CGTGTAAAGG GGCTTTTCAC

3351 CCACCTGATG CGGTGTGAAA TACCGCACAG ATGCGTAAGG AGAAAATACC
     GGTGGACTAC GCCACACTTT ATGGCGTGTC TACGCATTCC TCTTTTATGG

3401 GCATCAGGAA ATTGTAAGCG TTAATATTTT GTTAAAATTC GCGTTAAATT
     CGTAGTCCTT TAACATTCGC AATTATAAAA CAATTTTAAG CGCAATTTAA

3451 TTTGTTAAAT CAGCTCATTT TTTAACCAAT AGGCCGAAAT CGGCAAAATC
     AAACAATTTA GTCGAGTAAA AAATTGGTTA TCCGGCTTTA GCCGTTTTAG

3501 CCTTATAAAT CAAAAGAATA GACCGAGATA GGGTTGAGTG TTGTTCCAGT
     GGAATATTTA GTTTTCTTAT CTGGCTCTAT CCCAACTCAC AACAAGGTCA

3551 TTGGAACAAG AGTCCACTAT TAAAGAACGT GGACTCCAAC GTCAAAGGGC
     AACCTTGTTC TCAGGTGATA ATTTCTTGCA CCTGAGGTTG CAGTTTCCCG

3601 GAAAAACCGT CTATCAGGGC GATGGCCCAC TACGTGAACC ATCACCCTAA
     CTTTTTGGCA GATAGTCCCG CTACCGGGTG ATGCACTTGG TAGTGGGATT

3651 TCAAGTTTTT TGGGGTCGAG GTGCCGTAAA GCACTAAATC GGAACCCTAA
     AGTTCAAAAA ACCCCAGCTC CACGGCATTT CGTGATTTAG CCTTGGGATT

3701 AGGGAGCCCC CGATTTAGAG CTTGACGGGG AAAGCCGGCG AACGTGGCGA
     TCCCTCGGGG GCTAAATCTC GAACTGCCCC TTTCGGCCGC TTGCACCGCT

3751 GAAAGGAAGG GAAGAAAGCG AAAGGAGCGG GCGCTAGGGC GCTGGCAAGT
     CTTTCCTTCC CTTCTTTCGC TTTCCTCGCC CGCGATCCCG CGACCGTTCA
```

```
3801 GTAGCGGTCA CGCTGCGCGT AACCACCACA CCCGCCGCGC TTAATGCGCC
     CATCGCCAGT GCGACGCGCA TTGGTGGTGT GGGCGGCGCG AATTACGCGG

3851 GCTACAGGGC GCGTCCATTC GCCATTCAGG CTGCGCAACT GTTGGGAAGG
     CGATGTCCCG CGCAGGTAAG CGGTAAGTCC GACGCGTTGA CAACCCTTCC

3901 GCGATCGGTG CGGGCCTCTT CGCTATTACG CCAGCTGGCG AAAGGGGGAT
     CGCTAGCCAC GCCCGGAGAA GCGATAATGC GGTCGACCGC TTTCCCCCTA

3951 GTGCTGCAAG GCGATTAAGT TGGGTAACGC CAGGGTTTTC CCAGTCACGA
     CACGACGTTC CGCTAATTCA ACCCATTGCG GTCCCAAAAG GGTCAGTGCT

4001 CGTTGTAAAA CGACGGCCAG TGAATTGTAA TACGACTCAC TATA
     GCAACATTTT GCTGCCGGTC ACTTAACATT ATGCTGAGTG ATAT
```

Appendix D

pGU0513 (pET-19b::*Cj1344c*)




```
  1 TTCTCATGTT TGACAGCTTA TCATCGATAA GCTTTAATGC GGTAGTTTAT
    AAGAGTACAA ACTGTCGAAT AGTAGCTATT CGAAATTACG CCATCAAATA

 51 CACAGTTAAA TTGCTAACGC AGTCAGGCAC CGTGTATGAA ATCTAACAAT
    GTGTCAATTT AACGATTGCG TCAGTCCGTG GCACATACTT TAGATTGTTA

101 GCGCTCATCG TCATCCTCGG CACCGTCACC CTGGATGCTG TAGGCATAGG
    CGCGAGTAGC AGTAGGAGCC GTGGCAGTGG GACCTACGAC ATCCGTATCC

151 CTTGGTTATG CCGGTACTGC CGGGCCTCTT GCGGGATATC CGGATATAGT
    GAACCAATAC GGCCATGACG GCCCGGAGAA CGCCCTATAG GCCTATATCA

201 TCCTCCTTTC AGCAAAAAAC CCCTCAAGAC CCGTTTAGAG GCCCCAAGGG
    AGGAGGAAAG TCGTTTTTTG GGGAGTTCTG GGCAAATCTC CGGGGTTCCC

251 GTTATGCTAG TTATTGCTCA GCGGTGGCAG CAGCCAACTC AGCTTCCTTT
    CAATACGATC AATAACGAGT CGCCACCGTC GTCGGTTGAG TCGAAGGAAA

                                 XhoI
301 CGGGCTTTGT TAGCAGCCGG ATCTCGAGCT TTTTTCATCT ATATCCTTGA
    GCCCGAAACA ATCGTCGGCC TAGAGCTCGA AAAAAGTAGA TATAGGAACT
                                                *  I  R  S Cj1344c

351 AAAATTTTTA TTTTTAGGGC TTAAAATATC TTCTTCTACA CTTACAAATT
    TTTTAAAAAT AAAAATCCCG AATTTTATAG AAGAAGATGT GAATGTTTAA
      F  N  K   N  K  P   S  L  I  D   E  E  V   S  V  F   Cj1344c

401 CCTTTTTTTC ATAAGCATCA ACTGCGGCTC TTGCTATCAT CAAAGCATTA
    GGAAAAAAAG TATTCGTAGT TGACGCCGAG AACGATAGTA GTTTCGTAAT
    E  K  K  E   Y  A  D   V  A  A   R  A  I  M   L  A  N  Cj1344c

451 TCAGAGCAGA ATTTTAAAGG AGCTAGTTTT AAATTTGCAT TATATTTTTG
    AGTCTCGTCT TAAAATTTCC TCGATCAAAA TTTAAACGTA ATATAAAAAC
     D  S  C   F  K  L  P   A  L  K   L  N  A   N  Y  K  Q Cj1344c
```

```
 501 ACATAAATTT TGCAAACGCG AACGCAAGTT AAGATTTGCA CTAGCTCCAC
     TGTATTTAAA ACGTTTGCGC TTGCGTTCAA TTCTAAACGT GATCGAGGTG
       C  L  N   Q  L  R   S  R  L  N   L  N  A   S  A  G   Cj1344c

 551 CTACAACGCC AAAATTTTTA AATTTATAAA GATTAAAAAT TTTTTCTAAT
     GATGTTGCGG TTTTAAAAAT TTAAATATTT CTAATTTTTA AAAAAGATTA
     G  V  V  G   F  N  K   F  K  Y   L  N  F  I   K  E  L Cj1344c

 601 TTATCCATGA TATGATCACA AGCTGTATTT TCAAAGGCAT AGGCTATTTC
     AATAGGTACT ATACTAGTGT TCGACATAAA AGTTTCCGTA TCCGATAAAG
      K  D  M   I  H  D  C   A  T  N   E  F  A   Y  A  I  E Cj1344c

 651 TGCTTTTGTG TCTTCATTTA AATTTTCATG TTTTAAAATT TCCAAACGCA
     ACGAAAACAC AGAAGTAAAT TTAAAAGTAC AAAATTTTAA AGGTTTGCGT
       A  K  T   D  E  N   L  N  E  H   K  L  I   E  L  R   Cj1344c

 701 CTGCATTTTT AAGCCCTGAA AAACTGAAAG CGAGTTCTTT AGAATGCTTT
     GACGTAAAAA TTCGGGACTT TTTGACTTTC GCTCAAGAAA TCTTACGAAA
     V  A  N  K   L  G  S   F  S  F   A  L  E  K   S  H  K Cj1344c

 751 AAAGGTGTGT TAAAAGAGAT ATTTTTAAGT TTGGCATTTT TTGCTAAATT
     TTTCCACACA ATTTTCTCTA TAAAAATTCA AACCGTAAAA AACGATTTAA
      L  P  T   N  F  S  I   N  K  L   K  A  N   K  A  L  N Cj1344c

 801 TTCTATGATG ACCCCACCAG GGTAACCTAA ATTCATCATT TTAGCCACTT
     AAGATACTAC TGGGGTGGTC CCATTGGATT TAAGTAGTAA AATCGGTGAA
       E  I  I   V  G  G   P  Y  G  L   N  M  M   K  A  V   Cj1344c

 851 TATCAAAACT TTCTCCAAAG CTATCATCAT TTGTACTTGC TAAAAGCTCT
     ATAGTTTTGA AAGAGGTTTC GATAGTAGTA AACATGAACG ATTTTCGAGA
     K  D  F  S   E  G  F   S  D  D   N  T  S  A   L  L  E Cj1344c

 901 AAGCTTGCAT CATCTTTAAG ATAAAGCACC ATGGTATGCC CACCACTAAC
     TTCGAACGTA GTAGAAATTC TATTTCGTGG TACCATACGG GTGGTGATTG
      L  S  A   D  D  K  L   Y  L  V   M  T  H   G  G  S  V Cj1344c

 951 AAGCAAAATT CCCATATCTA AAGAAATTTT TTCTTCCAAA AAAAGACTAT
     TTCGTTTTAA GGGTATAGAT TTCTTTAAAA AAGAAGGTTT TTTTCTGATA
       L  L  I   G  M  D   L  S  I  K   E  E  L   F  L  S   Cj1344c

1001 AAATATGACC TTTAAGATGA TTTATAGGGA TTAAGGGTAA ATTTAGCGCA
     TTTATACTGG AAATTCTACT AAATATCCCT AATTCCCATT TAAATCGCGT
     Y  I  H  G   K  L  H   N  I  P   I  L  P  L   N  L  A Cj1344c

1051 CTTGCTAAGG TTTTTGCCAT AGAAATTCCA CTGAGCAAAG AAACACTAAG
     GAACGATTCC AAAAACGGTA TCTTTAAGGT GACTCGTTTC TTTGTGATTC
      S  A  L   T  K  A  M   S  I  G   S  L  L   S  V  S  L Cj1344c

1101 TCCAGGTTCA TTTGTCACAG CTATGGCACA AAGATTTTTA AAATGCTCTT
     AGGTCCAAGT AAACAGTGTC GATACCGTGT TTCTAAAAAT TTTACGAGAA
       G  P  E   N  T  V   A  I  A  C   L  N  K   F  H  E   Cj1344c

1151 TGCATTGCTT AAGCATCTTT GGTAAAGCCT CGCTATGAAG TCTTGCAGCA
     ACGTAACGAA TTCGTAGAAA CCATTTCGGA GCGATACTTC AGAACGTCGT
     K  C  Q  K   L  M  K   P  L  A   E  S  H  L   R  A  A Cj1344c

1201 AGTTCAGGTA CCACTCCCCC ATAGATACTA TGATCTAATT CTTGGGAAAT
     TCAAGTCCAT GGTGAGGGGG TATCTATGAT ACTAGATTAA GAACCCTTTA
      L  E  P   V  V  G  G   Y  I  S   H  D  L   E  Q  S  I Cj1344c
```

```
1251 TTTTTTATGA AATTTACATT CTAAGGTGTT TTTATCAATG ATAGCTATAG
     AAAAAATACT TTAAATGTAA GATTCCACAA AAATAGTTAC TATCGATATC
       K  K  H   F  K  C   E  L  T  N   K  D  I   I  A  I   Cj1344c

                                                   NdeI
1301 AACTATCATC ACAAGAACTT TCTATAGCTA GGATAAGATT TTTCATATGC
     TTGATAGTAG TGTTCTTGAA AGATATCGAT CCTATTCTAA AAAGTATACG
     S  S  D  D   C  S  S   E  I  A   L  I  L  N   K  M  H  Cj1344c

1351 TTGTCGTCGT CGTCGATATG GCCGCTGCTG TGATGATGAT GATGATGATG
     AACAGCAGCA GCAGCTATAC CGGCGACGAC ACTACTACTA CTACTACTAC
      K  D  D   D  D  I  H   G  S  S   H  H  H   H  H  H  H

1401 ATGATGATGG CCCATGGTAT ATCTCCTTCT TAAAGTTAAA CAAAATTATT
     TACTACTACC GGGTACCATA TAGAGGAAGA ATTTCAATTT GTTTTAATAA
       H  H  H   G  M
            His-tag

1451 TCTAGAGGGG AATTGTTATC CGCTCACAAT TCCCCTATAG TGAGTCGTAT
     AGATCTCCCC TTAACAATAG GCGAGTGTTA AGGGGATATC ACTCAGCATA

1501 TAATTTCGCG GGATCGAGAT CTCGATCCTC TACGCCGGAC GCATCGTGGC
     ATTAAAGCGC CCTAGCTCTA GAGCTAGGAG ATGCGGCCTG CGTAGCACCG
      ←T7 promoter

1551 CGGCATCACC GGCGCCACAG GTGCGGTTGC TGGCGCCTAT ATCGCCGACA
     GCCGTAGTGG CCGCGGTGTC CACGCCAACG ACCGCGGATA TAGCGGCTGT

1601 TCACCGATGG GGAAGATCGG GCTCGCCACT TCGGGCTCAT GAGCGCTTGT
     AGTGGCTACC CCTTCTAGCC CGAGCGGTGA AGCCCGAGTA CTCGCGAACA
                                              M   S  A  C   lacI

1651 TTCGGCGTGG GTATGGTGGC AGGCCCCGTG GCCGGGGGAC TGTTGGGCGC
     AAGCCGCACC CATACCACCG TCCGGGGCAC CGGCCCCCTG ACAACCCGCG
     F  G  V  G   M  V  A   G  P  V   A  G  G  L   L  G  A  lacI

1701 CATCTCCTTG CATGCACCAT TCCTTGCGGC GGCGGTGCTC AACGGCCTCA
     GTAGAGGAAC GTACGTGGTA AGGAACGCCG CCGCCACGAG TTGCCGGAGT
      I  S  L   H  A  P  F   L  A  A   A  V  L   N  G  L  N lacI

1751 ACCTACTACT GGGCTGCTTC CTAATGCAGG AGTCGCATAA GGGAGAGCGT
     TGGATGATGA CCCGACGAAG GATTACGTCC TCAGCGTATT CCCTCTCGCA
       L  L  L   G  C  F   L  M  Q  E   S  H  K   G  E  R   lacI

1801 CGAGATCCCG GACACCATCG AATGGCGCAA AACCTTTCGC GGTATGGCAT
     GCTCTAGGGC CTGTGGTAGC TTACCGCGTT TTGGAAAGCG CCATACCGTA
     R  D  P  G   H  H  R   M  A  Q   N  L  S  R   Y  G  M  lacI

1851 GATAGCGCCC GGAAGAGAGT CAATTCAGGG TGGTGAATGT GAAACCAGTA
     CTATCGCGGG CCTTCTCTCA GTTAAGTCCC ACCACTTACA CTTTGGTCAT
      I  A  P   G  R  E  S   I  Q  G   G  E  C   E  T  S  N lacI

1901 ACGTTATACG ATGTCGCAGA GTATGCCGGT GTCTCTTATC AGACCGTTTC
     TGCAATATGC TACAGCGTCT CATACGGCCA CAGAGAATAG TCTGGCAAAG
       V  I  R   C  R  R   V  C  R  C   L  L  S   D  R  F   lacI

1951 CCGCGTGGTG AACCAGGCCA GCCACGTTTC TGCGAAAACG CGGGAAAAAG
     GGCGCACCAC TTGGTCCGGT CGGTGCAAAG ACGCTTTTGC GCCCTTTTTC
     P  R  G  E   P  G  Q   P  R  F   C  E  N  A   G  K  S  lacI
```

```
2001 TGGAAGCGGC GATGGCGGAG CTGAATTACA TTCCCAACCG CGTGGCACAA
     ACCTTCGCCG CTACCGCCTC GACTTAATGT AAGGGTTGGC GCACCGTGTT
      G  S  G   D  G  G  A   E  L  H   S  Q  P   R  G  T  T lacI

2051 CAACTGGCGG GCAAACAGTC GTTGCTGATT GGCGTTGCCA CCTCCAGTCT
     GTTGACCGCC CGTTTGTCAG CAACGACTAA CCGCAACGGT GGAGGTCAGA
       T  G  G   Q  T  V   V  A  D  W   R  C  H   L  Q  S   lacI

2101 GGCCCTGCAC GCGCCGTCGC AAATTGTCGC GGCGATTAAA TCTCGCGCCG
     CCGGGACGTG CGCGGCAGCG TTTAACAGCG CCGCTAATTT AGAGCGCGGC
     G  P  A  R   A  V  A   N  C  R   G  D  * lacI

2151 ATCAACTGGG TGCCAGCGTG GTGGTGTCGA TGGTAGAACG AAGCGGCGTC
     TAGTTGACCC ACGGTCGCAC CACCACAGCT ACCATCTTGC TTCGCCGCAG

2201 GAAGCCTGTA AAGCGGCGGT GCACAATCTT CTCGCGCAAC GCGTCAGTGG
     CTTCGGACAT TTCGCCGCCA CGTGTTAGAA GAGCGCGTTG CGCAGTCACC

2251 GCTGATCATT AACTATCCGC TGGATGACCA GGATGCCATT GCTGTGGAAG
     CGACTAGTAA TTGATAGGCG ACCTACTGGT CCTACGGTAA CGACACCTTC

2301 CTGCCTGCAC TAATGTTCCG GCGTTATTTC TTGATGTCTC TGACCAGACA
     GACGGACGTG ATTACAAGGC CGCAATAAAG AACTACAGAG ACTGGTCTGT

2351 CCCATCAACA GTATTATTTT CTCCCATGAA GACGGTACGC GACTGGGCGT
     GGGTAGTTGT CATAATAAAA GAGGGTACTT CTGCCATGCG CTGACCCGCA

2401 GGAGCATCTG GTCGCATTGG GTCACCAGCA AATCGCGCTG TTAGCGGGCC
     CCTCGTAGAC CAGCGTAACC CAGTGGTCGT TTAGCGCGAC AATCGCCCGG

2451 CATTAAGTTC TGTCTCGGCG CGTCTGCGTC TGGCTGGCTG GCATAAATAT
     GTAATTCAAG ACAGAGCCGC GCAGACGCAG ACCGACCGAC CGTATTTATA

2501 CTCACTCGCA ATCAAATTCA GCCGATAGCG GAACGGGAAG GCGACTGGAG
     GAGTGAGCGT TAGTTTAAGT CGGCTATCGC CTTGCCCTTC CGCTGACCTC

2551 TGCCATGTCC GGTTTTCAAC AAACCATGCA AATGCTGAAT GAGGGCATCG
     ACGGTACAGG CCAAAAGTTG TTTGGTACGT TTACGACTTA CTCCCGTAGC

2601 TTCCCACTGC GATGCTGGTT GCCAACGATC AGATGGCGCT GGGCGCAATG
     AAGGGTGACG CTACGACCAA CGGTTGCTAG TCTACCGCGA CCCGCGTTAC

2651 CGCGCCATTA CCGAGTCCGG GCTGCGCGTT GGTGCGGATA TCTCGGTAGT
     GCGCGGTAAT GGCTCAGGCC CGACGCGCAA CCACGCCTAT AGAGCCATCA

2701 GGGATACGAC GATACCGAAG ACAGCTCATG TTATATCCCG CCGTTAACCA
     CCCTATGCTG CTATGGCTTC TGTCGAGTAC AATATAGGGC GGCAATTGGT

2751 CCATCAAACA GGATTTTCGC CTGCTGGGGC AAACCAGCGT GGACCGCTTG
     GGTAGTTTGT CCTAAAAGCG GACGACCCCG TTTGGTCGCA CCTGGCGAAC

2801 CTGCAACTCT CTCAGGGCCA GGCGGTGAAG GGCAATCAGC TGTTGCCCGT
     GACGTTGAGA GAGTCCCGGT CCGCCACTTC CCGTTAGTCG ACAACGGGCA

2851 CTCACTGGTG AAAAGAAAAA CCACCCTGGC GCCCAATACG CAAACCGCCT
     GAGTGACCAC TTTTCTTTTT GGTGGGACCG CGGGTTATGC GTTTGGCGGA

2901 CTCCCCGCGC GTTGGCCGAT TCATTAATGC AGCTGGCACG ACAGGTTTCC
     GAGGGGCGCG CAACCGGCTA AGTAATTACG TCGACCGTGC TGTCCAAAGG
```

```
2951 CGACTGGAAA GCGGGCAGTG AGCGCAACGC AATTAATGTA AGTTAGCTCA
     GCTGACCTTT CGCCCGTCAC TCGCGTTGCG TTAATTACAT TCAATCGAGT

3001 CTCATTAGGC ACCGGGATCT CGACCGATGC CCTTGAGAGC CTTCAACCCA
     GAGTAATCCG TGGCCCTAGA GCTGGCTACG GGAACTCTCG GAAGTTGGGT

3051 GTCAGCTCCT TCCGGTGGGC GCGGGGCATG ACTATCGTCG CCGCACTTAT
     CAGTCGAGGA AGGCCACCCG CGCCCCGTAC TGATAGCAGC GGCGTGAATA

3101 GACTGTCTTC TTTATCATGC AACTCGTAGG ACAGGTGCCG GCAGCGCTCT
     CTGACAGAAG AAATAGTACG TTGAGCATCC TGTCCACGGC CGTCGCGAGA

3151 GGGTCATTTT CGGCGAGGAC CGCTTTCGCT GGAGCGCGAC GATGATCGGC
     CCCAGTAAAA GCCGCTCCTG GCGAAAGCGA CCTCGCGCTG CTACTAGCCG

3201 CTGTCGCTTG CGGTATTCGG AATCTTGCAC GCCCTCGCTC AAGCCTTCGT
     GACAGCGAAC GCCATAAGCC TTAGAACGTG CGGGAGCGAG TTCGGAAGCA

3251 CACTGGTCCC GCCACCAAAC GTTTCGGCGA GAAGCAGGCC ATTATCGCCG
     GTGACCAGGG CGGTGGTTTG CAAAGCCGCT CTTCGTCCGG TAATAGCGGC

3301 GCATGGCGGC CGACGCGCTG GGCTACGTCT TGCTGGCGTT CGCGACGCGA
     CGTACCGCCG GCTGCGCGAC CCGATGCAGA ACGACCGCAA GCGCTGCGCT

3351 GGCTGGATGG CCTTCCCCAT TATGATTCTT CTCGCTTCCG GCGGCATCGG
     CCGACCTACC GGAAGGGGTA ATACTAAGAA GAGCGAAGGC CGCCGTAGCC

3401 GATGCCCGCG TTGCAGGCCA TGCTGTCCAG GCAGGTAGAT GACGACCATC
     CTACGGGCGC AACGTCCGGT ACGACAGGTC CGTCCATCTA CTGCTGGTAG

3451 AGGGACAGCT TCAAGGATCG CTCGCGGCTC TTACCAGCCT AACTTCGATC
     TCCCTGTCGA AGTTCCTAGC GAGCGCCGAG AATGGTCGGA TTGAAGCTAG

3501 ACTGGACCGC TGATCGTCAC GGCGATTTAT GCCGCCTCGG CGAGCACATG
     TGACCTGGCG ACTAGCAGTG CCGCTAAATA CGGCGGAGCC GCTCGTGTAC

3551 GAACGGGTTG GCATGGATTG TAGGCGCCGC CCTATACCTT GTCTGCCTCC
     CTTGCCCAAC CGTACCTAAC ATCCGCGGCG GGATATGGAA CAGACGGAGG

3601 CCGCGTTGCG TCGCGGTGCA TGGAGCCGGG CCACCTCGAC CTGAATGGAA
     GGCGCAACGC AGCGCCACGT ACCTCGGCCC GGTGGAGCTG GACTTACCTT

3651 GCCGGCGGCA CCTCGCTAAC GGATTCACCA CTCCAAGAAT TGGAGCCAAT
     CGGCCGCCGT GGAGCGATTG CCTAAGTGGT GAGGTTCTTA ACCTCGGTTA

3701 CAATTCTTGC GGAGAACTGT GAATGCGCAA ACCAACCCTT GGCAGAACAT
     GTTAAGAACG CCTCTTGACA CTTACGCGTT TGGTTGGGAA CCGTCTTGTA

3751 ATCCATCGCG TCCGCCATCT CCAGCAGCCG CACGCGGCGC ATCTCGGGCA
     TAGGTAGCGC AGGCGGTAGA GGTCGTCGGC GTGCGCCGCG TAGAGCCCGT

3801 GCGTTGGGTC CTGGCCACGG GTGCGCATGA TCGTGCTCCT GTCGTTGAGG
     CGCAACCCAG GACCGGTGCC CACGCGTACT AGCACGAGGA CAGCAACTCC

3851 ACCCGGCTAG GCTGGCGGGG TTGCCTTACT GGTTAGCAGA ATGAATCACC
     TGGGCCGATC CGACCGCCCC AACGGAATGA CCAATCGTCT TACTTAGTGG

3901 GATACGCGAG CGAACGTGAA GCGACTGCTG CTGCAAAACG TCTGCGACCT
     CTATGCGCTC GCTTGCACTT CGCTGACGAC GACGTTTTGC AGACGCTGGA
```

```
3951 GAGCAACAAC ATGAATGGTC TTCGGTTTCC GTGTTTCGTA AAGTCTGGAA
     CTCGTTGTTG TACTTACCAG AAGCCAAAGG CACAAAGCAT TTCAGACCTT

4001 ACGCGGAAGT CAGCGCCCTG CACCATTATG TTCCGGATCT GCATCGCAGG
     TGCGCCTTCA GTCGCGGGAC GTGGTAATAC AAGGCCTAGA CGTAGCGTCC

4051 ATGCTGCTGG CTACCCTGTG GAACACCTAC ATCTGTATTA ACGAAGCGCT
     TACGACGACC GATGGGACAC CTTGTGGATG TAGACATAAT TGCTTCGCGA

4101 GGCATTGACC CTGAGTGATT TTTCTCTGGT CCCGCCGCAT CCATACCGCC
     CCGTAACTGG GACTCACTAA AAAGAGACCA GGGCGGCGTA GGTATGGCGG

4151 AGTTGTTTAC CCTCACAACG TTCCAGTAAC CGGGCATGTT CATCATCAGT
     TCAACAAATG GGAGTGTTGC AAGGTCATTG GCCCGTACAA GTAGTAGTCA

4201 AACCCGTATC GTGAGCATCC TCTCTCGTTT CATCGGTATC ATTACCCCCA
     TTGGGCATAG CACTCGTAGG AGAGAGCAAA GTAGCCATAG TAATGGGGGT

4251 TGAACAGAAA TCCCCCTTAC ACGGAGGCAT CAGTGACCAA ACAGGAAAAA
     ACTTGTCTTT AGGGGGAATG TGCCTCCGTA GTCACTGGTT TGTCCTTTTT

4301 ACCGCCCTTA ACATGGCCCG CTTTATCAGA AGCCAGACAT TAACGCTTCT
     TGGCGGGAAT TGTACCGGGC GAAATAGTCT TCGGTCTGTA ATTGCGAAGA

4351 GGAGAAACTC AACGAGCTGG ACGCGGATGA ACAGGCAGAC ATCTGTGAAT
     CCTCTTTGAG TTGCTCGACC TGCGCCTACT TGTCCGTCTG TAGACACTTA

4401 CGCTTCACGA CCACGCTGAT GAGCTTTACC GCAGCTGCCT CGCGCGTTTC
     GCGAAGTGCT GGTGCGACTA CTCGAAATGG CGTCGACGGA GCGCGCAAAG

4451 GGTGATGACG GTGAAAACCT CTGACACATG CAGCTCCCGG AGACGGTCAC
     CCACTACTGC CACTTTTGGA GACTGTGTAC GTCGAGGGCC TCTGCCAGTG

4501 AGCTTGTCTG TAAGCGGATG CCGGGAGCAG ACAAGCCCGT CAGGGCGCGT
     TCGAACAGAC ATTCGCCTAC GGCCCTCGTC TGTTCGGGCA GTCCCGCGCA

4551 CAGCGGGTGT TGGCGGGTGT CGGGGCGCAG CCATGACCCA GTCACGTAGC
     GTCGCCCACA ACCGCCCACA GCCCCGCGTC GGTACTGGGT CAGTGCATCG

4601 GATAGCGGAG TGTATACTGG CTTAACTATG CGGCATCAGA GCAGATTGTA
     CTATCGCCTC ACATATGACC GAATTGATAC GCCGTAGTCT CGTCTAACAT

4651 CTGAGAGTGC ACCATATATG CGGTGTGAAA TACCGCACAG ATGCGTAAGG
     GACTCTCACG TGGTATATAC GCCACACTTT ATGGCGTGTC TACGCATTCC

4701 AGAAAATACC GCATCAGGCG CTCTTCCGCT TCCTCGCTCA CTGACTCGCT
     TCTTTTATGG CGTAGTCCGC GAGAAGGCGA AGGAGCGAGT GACTGAGCGA

4751 GCGCTCGGTC GTTCGGCTGC GGCGAGCGGT ATCAGCTCAC TCAAAGGCGG
     CGCGAGCCAG CAAGCCGACG CCGCTCGCCA TAGTCGAGTG AGTTTCCGCC

4801 TAATACGGTT ATCCACAGAA TCAGGGGATA ACGCAGGAAA GAACATGTGA
     ATTATGCCAA TAGGTGTCTT AGTCCCCTAT TGCGTCCTTT CTTGTACACT

4851 GCAAAAGGCC AGCAAAAGGC CAGGAACCGT AAAAAGGCCG CGTTGCTGGC
     CGTTTTCCGG TCGTTTTCCG GTCCTTGGCA TTTTTCCGGC GCAACGACCG

4901 GTTTTTCCAT AGGCTCCGCC CCCCTGACGA GCATCACAAA AATCGACGCT
     CAAAAAGGTA TCCGAGGCGG GGGGACTGCT CGTAGTGTTT TTAGCTGCGA
```

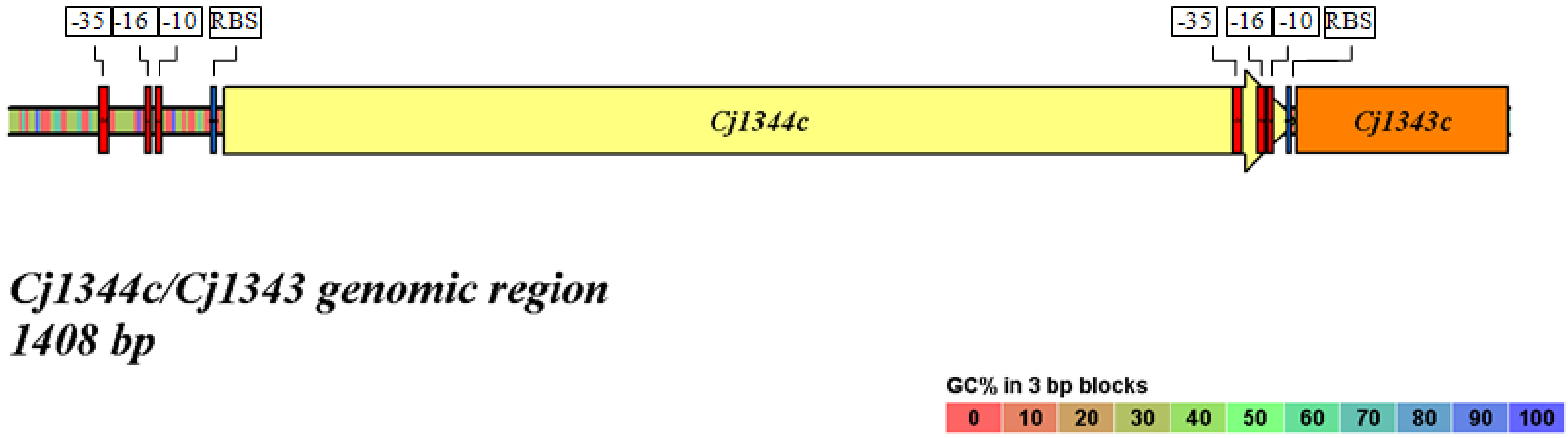
```
4951 CAAGTCAGAG GTGGCGAAAC CCGACAGGAC TATAAAGATA CCAGGCGTTT
     GTTCAGTCTC CACCGCTTTG GGCTGTCCTG ATATTTCTAT GGTCCGCAAA

5001 CCCCCTGGAA GCTCCCTCGT GCGCTCTCCT GTTCCGACCC TGCCGCTTAC
     GGGGGACCTT CGAGGGAGCA CGCGAGAGGA CAAGGCTGGG ACGGCGAATG

5051 CGGATACCTG TCCGCCTTTC TCCCTTCGGG AAGCGTGGCG CTTTCTCATA
     GCCTATGGAC AGGCGGAAAG AGGGAAGCCC TTCGCACCGC GAAAGAGTAT

5101 GCTCACGCTG TAGGTATCTC AGTTCGGTGT AGGTCGTTCG CTCCAAGCTG
     CGAGTGCGAC ATCCATAGAG TCAAGCCACA TCCAGCAAGC GAGGTTCGAC

5151 GGCTGTGTGC ACGAACCCCC CGTTCAGCCC GACCGCTGCG CCTTATCCGG
     CCGACACACG TGCTTGGGGG GCAAGTCGGG CTGGCGACGC GGAATAGGCC

5201 TAACTATCGT CTTGAGTCCA ACCCGGTAAG ACACGACTTA TCGCCACTGG
     ATTGATAGCA GAACTCAGGT TGGGCCATTC TGTGCTGAAT AGCGGTGACC

5251 CAGCAGCCAC TGGTAACAGG ATTAGCAGAG CGAGGTATGT AGGCGGTGCT
     GTCGTCGGTG ACCATTGTCC TAATCGTCTC GCTCCATACA TCCGCCACGA

5301 ACAGAGTTCT TGAAGTGGTG GCCTAACTAC GGCTACACTA GAAGGACAGT
     TGTCTCAAGA ACTTCACCAC CGGATTGATG CCGATGTGAT CTTCCTGTCA

5351 ATTTGGTATC TGCGCTCTGC TGAAGCCAGT TACCTTCGGA AAAAGAGTTG
     TAAACCATAG ACGCGAGACG ACTTCGGTCA ATGGAAGCCT TTTTCTCAAC

5401 GTAGCTCTTG ATCCGGCAAA CAAACCACCG CTGGTAGCGG TGGTTTTTTT
     CATCGAGAAC TAGGCCGTTT GTTTGGTGGC GACCATCGCC ACCAAAAAAA

5451 GTTTGCAAGC AGCAGATTAC GCGCAGAAAA AAAGGATCTC AAGAAGATCC
     CAAACGTTCG TCGTCTAATG CGCGTCTTTT TTTCCTAGAG TTCTTCTAGG

5501 TTTGATCTTT TCTACGGGGT CTGACGCTCA GTGGAACGAA AACTCACGTT
     AAACTAGAAA AGATGCCCCA GACTGCGAGT CACCTTGCTT TTGAGTGCAA

5551 AAGGGATTTT GGTCATGAGA TTATCAAAAA GGATCTTCAC CTAGATCCTT
     TTCCCTAAAA CCAGTACTCT AATAGTTTTT CCTAGAAGTG GATCTAGGAA

5601 TTAAATTAAA AATGAAGTTT TAAATCAATC TAAAGTATAT ATGAGTAAAC
     AATTTAATTT TTACTTCAAA ATTTAGTTAG ATTTCATATA TACTCATTTG

5651 TTGGTCTGAC AGTTACCAAT GCTTAATCAG TGAGGCACCT ATCTCAGCGA
     AACCAGACTG TCAATGGTTA CGAATTAGTC ACTCCGTGGA TAGAGTCGCT
                    *  W   H  K  I  L   S  A  G   I  E  A   ampR

5701 TCTGTCTATT TCGTTCATCC ATAGTTGCCT GACTCCCCGT CGTGTAGATA
     AGACAGATAA AGCAAGTAGG TATCAACGGA CTGAGGGGCA GCACATCTAT
     I  Q  R  N   R  E  D   M  T  A   Q  S  G  T   T  Y  I   ampR

5751 ACTACGATAC GGGAGGGCTT ACCATCTGGC CCCAGTGCTG CAATGATACC
     TGATGCTATG CCCTCCCGAA TGGTAGACCG GGGTCACGAC GTTACTATGG
      V  V  I   R  S  P  K   G  D  P   G  L  A   A  I  I  G ampR

5801 GCGAGACCCA CGCTCACCGG CTCCAGATTT ATCAGCAATA AACCAGCCAG
     CGCTCTGGGT GCGAGTGGCC GAGGTCTAAA TAGTCGTTAT TTGGTCGGTC
       R  S  G   R  E  G   A  G  S  K   D  A  I   F  W  G   ampR

5851 CCGGAAGGGC CGAGCGCAGA AGTGGTCCTG CAACTTTATC CGCCTCCATC
     GGCCTTCCCG GCTCGCGTCT TCACCAGGAC GTTGAAATAG GCGGAGGTAG
     A  P  L  A   S  R  L   L  P  G   A  V  K  D   A  E  M   ampR
```

```
5901 CAGTCTATTA ATTGTTGCCG GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA
     GTCAGATAAT TAACAACGGC CCTTCGATCT CATTCATCAA GCGGTCAATT
      W  D  I   L  Q  Q  R   S  A  L   T  L  L   E  G  T  L ampR

5951 TAGTTTGCGC AACGTTGTTG CCATTGCTGC AGGCATCGTG GTGTCACGCT
     ATCAAACGCG TTGCAACAAC GGTAACGACG TCCGTAGCAC CACAGTGCGA
       L  K  R   L  T  T   A  M  A  A   P  M  T   T  D  R   ampR

6001 CGTCGTTTGG TATGGCTTCA TTCAGCTCCG GTTCCCAACG ATCAAGGCGA
     GCAGCAAACC ATACCGAAGT AAGTCGAGGC CAAGGGTTGC TAGTTCCGCT
     E  D  N  P   I  A  E   N  L  E   P  E  W  R   D  L  R  ampR

6051 GTTACATGAT CCCCCATGTT GTGCAAAAAA GCGGTTAGCT CCTTCGGTCC
     CAATGTACTA GGGGGTACAA CACGTTTTTT CGCCAATCGA GGAAGCCAGG
      T  V  H   D  G  M  N   H  L  F   A  T  L   E  K  P  G ampR

6101 TCCGATCGTT GTCAGAAGTA AGTTGGCCGC AGTGTTATCA CTCATGGTTA
     AGGCTAGCAA CAGTCTTCAT TCAACCGGCG TCACAATAGT GAGTACCAAT
       G  I  T   T  L  L   L  N  A  A   T  N  D   S  M  T   ampR

6151 TGGCAGCACT GCATAATTCT CTTACTGTCA TGCCATCCGT AAGATGCTTT
     ACCGTCGTGA CGTATTAAGA GAATGACAGT ACGGTAGGCA TTCTACGAAA
     I  A  A  S   C  L  E   R  V  T   M  G  D  T   L  H  K  ampR

6201 TCTGTGACTG GTGAGTACTC AACCAAGTCA TTCTGAGAAT AGTGTATGCG
     AGACACTGAC CACTCATGAG TTGGTTCAGT AAGACTCTTA TCACATACGC
      E  T  V   P  S  Y  E   V  L  D   N  Q  S   Y  H  I  R ampR

6251 GCGACCGAGT TGCTCTTGCC CGGCGTCAAC ACGGGATAAT ACCGCGCCAC
     CGCTGGCTCA ACGAGAACGG GCCGCAGTTG TGCCCTATTA TGGCGCGGTG
       R  G  L   Q  E  Q   G  A  D  V   R  S  L   V  A  G   ampR

6301 ATAGCAGAAC TTTAAAAGTG CTCATCATTG GAAAACGTTC TTCGGGGCGA
     TATCGTCTTG AAATTTTCAC GAGTAGTAAC CTTTTGCAAG AAGCCCCGCT
     C  L  L  V   K  F  T   S  M  M   P  F  R  E   E  P  R  ampR

6351 AAACTCTCAA GGATCTTACC GCTGTTGAGA TCCAGTTCGA TGTAACCCAC
     TTTGAGAGTT CCTAGAATGG CGACAACTCT AGGTCAAGCT ACATTGGGTG
      F  S  E   L  I  K  G   S  N  L   D  L  E   I  Y  G  V ampR

6401 TCGTGCACCC AACTGATCTT CAGCATCTTT TACTTTCACC AGCGTTTCTG
     AGCACGTGGG TTGACTAGAA GTCGTAGAAA ATGAAAGTGG TCGCAAAGAC
       R  A  G   L  Q  D   E  A  D  K   V  K  V   L  T  E   ampR

6451 GGTGAGCAAA AACAGGAAGG CAAAATGCCG CAAAAAAGGG AATAAGGGCG
     CCACTCGTTT TTGTCCTTCC GTTTTACGGC GTTTTTTCCC TTATTCCCGC
     P  H  A  F   V  P  L   C  F  A   A  F  F  P   I  L  A  ampR

6501 ACACGGAAAT GTTGAATACT CATACTCTTC CTTTTTCAAT ATTATTGAAG
     TGTGCCTTTA CAACTTATGA GTATGAGAAG GAAAAAGTTA TAATAACTTC
      V  R  F   H  Q  I  S   M   ampR

6551 CATTTATCAG GGTTATTGTC TCATGAGCGG ATACATATTT GAATGTATTT
     GTAAATAGTC CCAATAACAG AGTACTCGCC TATGTATAAA CTTACATAAA

6601 AGAAAAATAA ACAAATAGGG GTTCCGCGCA CATTTCCCCG AAAAGTGCCA
     TCTTTTTATT TGTTTATCCC CAAGGCGCGT GTAAAGGGGC TTTTCACGGT

6651 CCTGACGTCT AAGAAACCAT TATTATCATG ACATTAACCT ATAAAAATAG
     GGACTGCAGA TTCTTTGGTA ATAATAGTAC TGTAATTGGA TATTTTTATC
```

```
6701 GCGTATCACG AGGCCCTTTC GTCTTCAAGA A
     CGCATAGTGC TCCGGGAAAG CAGAAGTTCT T
```

Appendix E

Cj1344c/Cj1343c genomic region and promoter elements

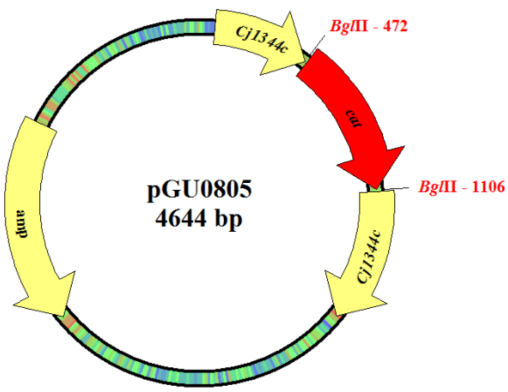


```
  1 CAAACACAAT ACTCTTTAGA CATGGCAGGT AAAATTTATA GAGCGGAATT
    GTTTGTGTTA TGAGAAATCT GTACCGTCCA TTTTAAATAT CTCGCCTTAA

                                            -35
 51 TTATGAGTTA AGAGGTGCAA ATTTGCAACA ACTTTTAGAG GCAAATATCA
    AATACTCAAT TCTCCACGTT TAAACGTTGT TGAAAATCTC CGTTTATAGT

                                   -16        -10
101 ATAGCAAGTT GATCAAAAAC GCTAAAAATT TGGATTTAAA TACTCTTAAA
    TATCGTTCAA CTAGTTTTTG CGATTTTTAA ACCTAAATTT ATGAGAATTT

                                                RBS
151 ATGGCAAGAT CAAAAGATAA ATTCTTAGGT TCTATTTTAG TGGAATTTGA
    TACCGTTCTA GTTTTCTATT TAAGAATCCA AGATAAAATC ACCTTAAACT

    START
201 ATGAAAAATC TTATCCTAGC TATAGAAAGT TCTTGTGATG ATAGTTCTAT
    TACTTTTTAG AATAGGATCG ATATCTTTCA AGAACACTAC TATCAAGATA
    M  K  N  L   I  L  A   I  E  S   S  C  D  D   S  S  I  Cj1344c

251 AGCTATCATT GATAAAAACA CCTTAGAATG TAAATTTCAT AAAAAAATTT
    TCGATAGTAA CTATTTTTGT GGAATCTTAC ATTTAAAGTA TTTTTTTAAA
     A  I  I   D  K  N  T   L  E  C   K  F  H   K  K  I  S Cj1344c

301 CCCAAGAATT AGATCATAGT ATCTATGGGG GAGTGGTACC TGAACTTGCT
    GGGTTCTTAA TCTAGTATCA TAGATACCCC CTCACCATGG ACTTGAACGA
      Q  E  L   D  H  S   I  Y  G  G   V  V  P   E  L  A   Cj1344c

351 GCAAGACTTC ATAGCGAGGC TTTACCAAAG ATGCTTAAGC AATGCAAAGA
    CGTTCTGAAG TATCGCTCCG AAATGGTTTC TACGAATTCG TTACGTTTCT
    A  R  L  H   S  E  A   L  P  K   M  L  K  Q   C  K  E  Cj1344c

401 GCATTTTAAA AATCTTTGTG CCATAGCTGT GACAAATGAA CCTGGACTTA
    CGTAAAATTT TTAGAAACAC GGTATCGACA CTGTTTACTT GGACCTGAAT
     H  F  K   N  L  C  A   I  A  V   T  N  E   P  G  L  S Cj1344c

451 GTGTTTCTTT GCTCAGTGGA ATTTCTATGG CAAAAACCTT AGCAAGTGCG
    CACAAAGAAA CGAGTCACCT TAAAGATACC GTTTTTGGAA TCGTTCACGC
      V  S  L   L  S  G   I  S  M  A   K  T  L   A  S  A   Cj1344c

501 CTAAATTTAC CCTTAATCCC TATAAATCAT CTTAAAGGTC ATATTTATAG
    GATTTAAATG GGAATTAGGG ATATTTAGTA GAATTTCCAG TATAAATATC
    L  N  L  P   L  I  P   I  N  H   L  K  G  H   I  Y  S  Cj1344c

551 TCTTTTTTTG GAAGAAAAAA TTTCTTTAGA TATGGGAATT TTGCTTGTTA
    AGAAAAAAAC CTTCTTTTTT AAAGAAATCT ATACCCTTAA AACGAACAAT
     L  F  L   E  E  K  I   S  L  D   M  G  I   L  L  V  S Cj1344c

601 GTGGTGGGCA TACCATGGTG CTTTATCTTA AAGATGATGC AAGCTTAGAG
    CACCACCCGT ATGGTACCAC GAAATAGAAT TTCTACTACG TTCGAATCTC
      G  G  H   T  M  V   L  Y  L  K   D  D  A   S  L  E   Cj1344c

651 CTTTTAGCAA GTACAAATGA TGATAGCTTT GGAGAAAGTT TTGATAAAGT
    GAAAATCGTT CATGTTTACT ACTATCGAAA CCTCTTTCAA AACTATTTCA
    L  L  A  S   T  N  D   D  S  F   G  E  S  F   D  K  V  Cj1344c
```

```
 701 GGCTAAAATG ATGAATTTAG GTTACCCTGG TGGGGTCATC ATAGAAAATT
     CCGATTTTAC TACTTAAATC CAATGGGACC ACCCCAGTAG TATCTTTTAA
      A  K  M   M  N  L  G   Y  P  G   G  V  I   I  E  N  L Cj1344c

 751 TAGCAAAAAA TGCCAAACTT AAAAATATCT CTTTTAACAC ACCTTTAAAG
     ATCGTTTTTT ACGGTTTGAA TTTTTATAGA GAAAATTGTG TGGAAATTTC
       A  K  N   A  K  L   K  N  I  S   F  N  T   P  L  K   Cj1344c

 801 CATTCTAAAG AACTCGCTTT CAGTTTTTCA GGGCTTAAAA ATGCAGTGCG
     GTAAGATTTC TTGAGCGAAA GTCAAAAAGT CCCGAATTTT TACGTCACGC
     H  S  K  E   L  A  F   S  F  S   G  L  K  N   A  V  R  Cj1344c

 851 TTTGGAAATT TTAAAACATG AAAATTTAAA TGAAGACACA AAAGCAGAAA
     AAACCTTTAA AATTTTGTAC TTTTAAATTT ACTTCTGTGT TTTCGTCTTT
      L  E  I   L  K  H  E   N  L  N   E  D  T   K  A  E  I Cj1344c

 901 TAGCCTATGC CTTTGAAAAT ACAGCTTGTG ATCATATCAT GGATAAATTA
     ATCGGATACG GAAACTTTTA TGTCGAACAC TAGTATAGTA CCTATTTAAT
       A  Y  A   F  E  N   T  A  C  D   H  I  M   D  K  L   Cj1344c

 951 GAAAAAATTT TTAATCTTTA TAAATTTAAA AATTTTGGCG TTGTAGGTGG
     CTTTTTTAAA AATTAGAAAT ATTTAAATTT TTAAAACCGC AACATCCACC
     E  K  I  F   N  L  Y   K  F  K   N  F  G  V   V  G  G  Cj1344c

1001 AGCTAGTGCA AATCTTAACT TGCGTTCGCG TTTGCAAAAT TTATGTCAAA
     TCGATCACGT TTAGAATTGA ACGCAAGCGC AAACGTTTTA AATACAGTTT
      A  S  A   N  L  N  L   R  S  R   L  Q  N   L  C  Q  K Cj1344c

1051 AATATAATGC AAATTTAAAA CTAGCTCCTT TAAAATTCTG CTCTGATAAT
     TTATATTACG TTTAAATTTT GATCGAGGAA ATTTTAAGAC GAGACTATTA
       Y  N  A   N  L  K   L  A  P  L   K  F  C   S  D  N   Cj1344c

1101 GCTTTGATGA TAGCAAGAGC CGCAGTTGAT GCTTATGAAA AAAAGGAATT
     CGAAACTACT ATCGTTCTCG GCGTCAACTA CGAATACTTT TTTTCCTTAA
     A  L  M  I   A  R  A   A  V  D   A  Y  E  K   K  E  F  Cj1344c

      -35                   -16       -10                RBS
1151 TGTAAGTGTA GAAGAAGATA TTTTAAGCCC TAAAAATAAA AATTTTTCAA
     ACATTCACAT CTTCTTCTAT AAAATTCGGG ATTTTTATTT TTAAAAAGTT
      V  S  V   E  E  D  I   L  S  P   K  N  K   N  F  S  R Cj1344c

   Cj1344c STOP
1201 GGATATAGAT GAAAAAAGCT TTTACTATAT TAGAACTTGT TTTTGTGATC
     CCTATATCTA CTTTTTTCGA AAATGATATA ATCTTGAACA AAAACACTAG
       I  *  START Cj1343c

1251 GTTATTTTGG CTTAAAATCG GCGATAACGA AACGGTTTTT ACTCAAGTTC
     CAATAAAACC GAATTTTAGC CGCTATTGCT TTGCCAAAAA TGAGTTCAAG

1301 ATTCCTACTT CGACTTCAAT GCAAGTCTTT AAACAACCTT AAAACTTTGA
     TAAGGATGAA GCTGAAGTTA CGTTCAGAAA TTTGTTGGAA TTTTGAAACT

1351 TCAATGATAT AAGCATTTAT ACTTTAAAAA TACTAGTAAA TTCGAGATAT
     AGTTACTATA TTCGTAAATA TGAAATTTTT ATGATCATTT AAGCTCTATA

1401 TTTTGCTA
     AAAACGAT
```

Appendix F

pGU0805 (pGU0501Δ*Cj1344c*::*cat*)




```
  1 GGGCGAATTG GGCCCGACGT CGCATGCTCC CGGCCGCCAT GGCGGCCGCG
    CCCGCTTAAC CCGGGCTGCA GCGTACGAGG GCCGGCGGTA CCGCCGGCGC

                   START Cj1344c
 51 GGAATTCGAT TCATATGAAA AATCTTATCC TAGCTATAGA AAGTTCTTGT
    CCTTAAGCTA AGTATACTTT TTAGAATAGG ATCGATATCT TTCAAGAACA
                   M  K   N  L  I  L   A  I  E   S  S  C

101 GATGATAGTT CTATAGCTAT CATTGATAAA AACACCTTAG AATGTAAATT
    CTACTATCAA GATATCGATA GTAACTATTT TTGTGGAATC TTACATTTAA
    D  D  S  S   I  A  I   I  D  K   N  T  L  E   C  K  F

151 TCATAAAAAA ATTTCCCAAG AATTAGATCA TAGTATCTAT GGGGGAGTGG
    AGTATTTTTT TAAAGGGTTC TTAATCTAGT ATCATAGATA CCCCCTCACC
     H  K  K   I  S  Q  E   L  D  H   S  I  Y   G  G  V  V

201 TACCTGAACT TGCTGCAAGA CTTCATAGCG AGGCTTTACC AAAGATGCTT
    ATGGACTTGA ACGACGTTCT GAAGTATCGC TCCGAAATGG TTTCTACGAA
      P  E  L   A  A  R   L  H  S  E   A  L  P   K  M  L

251 AAGCAATGCA AAGAGCATTT TAAAAATCTT TGTGCCATAG CTGTGACAAA
    TTCGTTACGT TTCTCGTAAA ATTTTTAGAA ACACGGTATC GACACTGTTT
    K  Q  C  K   E  H  F   K  N  L   C  A  I  A   V  T  N

301 TGAACCTGGA CTTAGTGTTT CTTTGCTCAG TGGAATTTCT ATGGCAAAAA
    ACTTGGACCT GAATCACAAA GAAACGAGTC ACCTTAAAGA TACCGTTTTT
     E  P  G   L  S  V  S   L  L  S   G  I  S   M  A  K  T

351 CCTTAGCAAG TGCGCTAAAT TTACCCTTAA TCCCTATAAA TCATCTTAAA
    GGAATCGTTC ACGCGATTTA AATGGGAATT AGGGATATTT AGTAGAATTT
      L  A  S   A  L  N   L  P  L  I   P  I  N   H  L  K
```

```
 401 GGTCATATTT ATAGTCTTTT TTTGGAAGAA AAAATTTCTT TAGATATGGG
     CCAGTATAAA TATCAGAAAA AAACCTTCTT TTTTAAAGAA ATCTATACCC
     G  H  I  Y   S  L  F   L  E  E   K  I  S  L   D  M  G

                                         START cat (in frame)
 451 AATTTTGCTT GTTAGTGGTG GAGATCTCAT GATGCAATTC ACAAAGATTG
     TTAAAACGAA CAATCACCAC CTCTAGAGTA CTACGTTAAG TGTTTCTAAC
      I  L  L   V  S  G  G   D  L  M   M  Q  F   T  K  I  D

 501 ATATAAATAA TTGGACACGA AAAGAGTATT TCGACCACTA TTTTGGCAAT
     TATATTTATT AACCTGTGCT TTTCTCATAA AGCTGGTGAT AAAACCGTTA
       I  N  N   W  T  R   K  E  Y  F   D  H  Y   F  G  N

 551 ACGCCCTGCA CATATAGTAT GACGGTAAAA CTCGATATTT CTAAGTTGAA
     TGCGGGACGT GTATATCATA CTGCCATTTT GAGCTATAAA GATTCAACTT
     T  P  C  T   Y  S  M   T  V  K   L  D  I  S   K  L  K

 601 AAAGGATGGA AAAAAGTTAT ACCCAACTCT TTTATATGGA GTTACAACGA
     TTTCCTACCT TTTTTCAATA TGGGTTGAGA AAATATACCT CAATGTTGCT
      K  D  G   K  K  L  Y   P  T  L   L  Y  G   V  T  T  I

 651 TCATCAATCG ACATGAAGAG TTCAGGACCG CATTAGATGA AAACGGACAG
     AGTAGTTAGC TGTACTTCTC AAGTCCTGGC GTAATCTACT TTTGCCTGTC
       I  N  R   H  E  E   F  R  T  A   L  D  E   N  G  Q

 701 GTAGGCGTTT TTTCAGAAAT GCTGCCTTGC TACACAGTTT TTCATAAGGA
     CATCCGCAAA AAAGTCTTTA CGACGGAACG ATGTGTCAAA AAGTATTCCT
     V  G  V  F   S  E  M   L  P  C   Y  T  V  F   H  K  E

 751 AACTGAAACC TTTTCGAGTA TTTGGACTGA GTTTACAGCA GACTATACTG
     TTGACTTTGG AAAAGCTCAT AAACCTGACT CAAATGTCGT CTGATATGAC
      T  E  T   F  S  S  I   W  T  E   F  T  A   D  Y  T  E

 801 AGTTTCTTCA GAACTATCAA AAGGATATAG ACGCTTTTGG TGAACGAATG
     TCAAAGAAGT CTTGATAGTT TTCCTATATC TGCGAAAACC ACTTGCTTAC
       F  L  Q   N  Y  Q   K  D  I  D   A  F  G   E  R  M

 851 GGAATGTCCG CAAAGCCTAA TCCTCCGGAA AACACTTTCC CTGTTTCTAT
     CCTTACAGGC GTTTCGGATT AGGAGGCCTT TTGTGAAAGG GACAAAGATA
     G  M  S  A   K  P  N   P  P  E   N  T  F  P   V  S  M

 901 GATACCGTGG ACAAGCTTTG AAGGCTTTAA CTTAAATCTA AAAAAAGGAT
     CTATGGCACC TGTTCGAAAC TTCCGAAATT GAATTTAGAT TTTTTTCCTA
      I  P  W   T  S  F  E   G  F  N   L  N  L   K  K  G  Y

 951 ATGACTATCT ACTGCCGATA TTTACGTTTG GGAAGTATTA TGAGGAGGGC
     TACTGATAGA TGACGGCTAT AAATGCAAAC CCTTCATAAT ACTCCTCCCG
       D  Y  L   L  P  I   F  T  F  G   K  Y  Y   E  E  G

1001 GGAAAATACT ATATTCCCTT ATCGATTCAA GTGCATCATG CCGTTTGTGA
     CCTTTTATGA TATAAGGGAA TAGCTAAGTT CACGTAGTAC GGCAAACACT
     G  K  Y  Y   I  P  L   S  I  Q   V  H  H  A   V  C  D

1051 CGGCTTTCAT GTTTGCCGTT TTTTGGATGA ATTACAAGAC TTGCTGAATA
     GCCGAAAGTA CAAACGGCAA AAAACCTACT TAATGTTCTG AACGACTTAT
      G  F  H   V  C  R  F   L  D  E   L  Q  D   L  L  N  K

       STOP cat
1101 AATAAAGATC TGAGCTTTTA GCAAGTACAA ATGATGATAG CTTTGGAGAA
     TTATTTCTAG ACTCGAAAAT CGTTCATGTT TACTACTATC GAAACCTCTT
       *
```

```
1151 AGTTTTGATA AAGTGGCTAA AATGATGAAT TTAGGTTACC CTGGTGGGGT
     TCAAAACTAT TTCACCGATT TTACTACTTA AATCCAATGG GACCACCCCA
                            M  M  N   L  G  Y  P   G  G  V

1201 CATCATAGAA AATTTAGCAA AAAATGCCAA ACTTAAAAAT ATCTCTTTTA
     GTAGTATCTT TTAAATCGTT TTTTACGGTT TGAATTTTTA TAGAGAAAAT
      I  I  E   N  L  A  K   N  A  K   L  K  N   I  S  F  N

1251 ACACACCTTT AAAGCATTCT AAAGAACTCG CTTTCAGTTT TTCAGGGCTT
     TGTGTGGAAA TTTCGTAAGA TTTCTTGAGC GAAAGTCAAA AAGTCCCGAA
       T  P  L   K  H  S   K  E  L  A   F  S  F   S  G  L

1301 AAAAATGCAG TGCGTTTGGA AATTTTAAAA CATGAAAATT TAAATGAAGA
     TTTTTACGTC ACGCAAACCT TTAAAATTTT GTACTTTTAA ATTTACTTCT
     K  N  A  V   R  L  E   I  L  K   H  E  N  L   N  E  D

1351 CACAAAAGCA GAAATAGCCT ATGCCTTTGA AAATACAGCT TGTGATCATA
     GTGTTTTCGT CTTTATCGGA TACGGAAACT TTTATGTCGA ACACTAGTAT
      T  K  A   E  I  A  Y   A  F  E   N  T  A   C  D  H  I

1401 TCATGGATAA ATTAGAAAAA ATTTTTAATC TTTATAAATT TAAAAATTTT
     AGTACCTATT TAATCTTTTT TAAAAATTAG AAATATTTAA ATTTTTAAAA
       M  D  K   L  E  K   I  F  N  L   Y  K  F   K  N  F

1451 GGCGTTGTAG GTGGAGCTAG TGCAAATCTT AACTTGCGTT CGCGTTTGCA
     CCGCAACATC CACCTCGATC ACGTTTAGAA TTGAACGCAA GCGCAAACGT
     G  V  V  G   G  A  S   A  N  L   N  L  R  S   R  L  Q

1501 AAATTTATGT CAAAAATATA ATGCAAATTT AAAACTAGCT CCTTTAAAAT
     TTTAAATACA GTTTTTATAT TACGTTTAAA TTTTGATCGA GGAAATTTTA
      N  L  C   Q  K  Y  N   A  N  L   K  L  A   P  L  K  F

1551 TCTGCTCTGA TAATGCTTTG ATGATAGCAA GAGCCGCAGT TGATGCTTAT
     AGACGAGACT ATTACGAAAC TACTATCGTT CTCGGCGTCA ACTACGAATA
       C  S  D   N  A  L   M  I  A  R   A  A  V   D  A  Y

1601 GAAAAAAAGG AATTTGTAAG TGTAGAAGAA GATATTTTAA GCCCTAAAAA
     CTTTTTTTCC TTAAACATTC ACATCTTCTT CTATAAAATT CGGGATTTTT
     E  K  K  E   F  V  S   V  E  E   D  I  L  S   P  K  N

                          STOP Cj1344c
1651 TAAAAATTTT TCAAGGATAT AGATGAAAAA AGCTCGAGAA TCACTAGTGA
     ATTTTTAAAA AGTTCCTATA TCTACTTTTT TCGAGCTCTT AGTGATCACT
      K  N  F   S  R  I  *

1701 ATTCGCGGCC GCCTGCAGGT CGACCATATG GGAGAGCTCC CAACGCGTTG
     TAAGCGCCGG CGGACGTCCA GCTGGTATAC CCTCTCGAGG GTTGCGCAAC

1751 GATGCATAGC TTGAGTATTC TATAGTGTCA CCTAAATAGC TTGGCGTAAT
     CTACGTATCG AACTCATAAG ATATCACAGT GGATTTATCG AACCGCATTA

1801 CATGGTCATA GCTGTTTCCT GTGTGAAATT GTTATCCGCT CACAATTCCA
     GTACCAGTAT CGACAAAGGA CACACTTTAA CAATAGGCGA GTGTTAAGGT

1851 CACAACATAC GAGCCGGAAG CATAAAGTGT AAAGCCTGGG GTGCCTAATG
     GTGTTGTATG CTCGGCCTTC GTATTTCACA TTTCGGACCC CACGGATTAC

1901 AGTGAGCTAA CTCACATTAA TTGCGTTGCG CTCACTGCCC GCTTTCCAGT
     TCACTCGATT GAGTGTAATT AACGCAACGC GAGTGACGGG CGAAAGGTCA
```

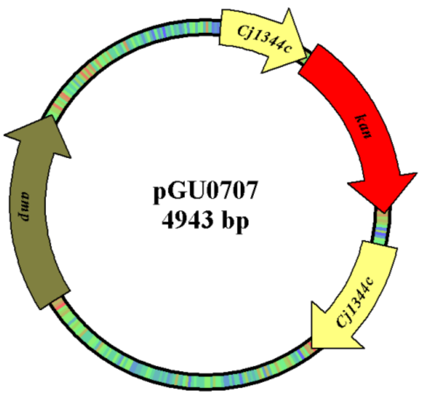
```
1951 CGGGAAACCT GTCGTGCCAG CTGCATTAAT GAATCGGCCA ACGCGCGGGG
     GCCCTTTGGA CAGCACGGTC GACGTAATTA CTTAGCCGGT TGCGCGCCCC

2001 AGAGGCGGTT TGCGTATTGG GCGCTCTTCC GCTTCCTCGC TCACTGACTC
     TCTCCGCCAA ACGCATAACC CGCGAGAAGG CGAAGGAGCG AGTGACTGAG

2051 GCTGCGCTCG GTCGTTCGGC TGCGGCGAGC GGTATCAGCT CACTCAAAGG
     CGACGCGAGC CAGCAAGCCG ACGCCGCTCG CCATAGTCGA GTGAGTTTCC

2101 CGGTAATACG GTTATCCACA GAATCAGGGG ATAACGCAGG AAAGAACATG
     GCCATTATGC CAATAGGTGT CTTAGTCCCC TATTGCGTCC TTTCTTGTAC

2151 TGAGCAAAAG GCCAGCAAAA GGCCAGGAAC CGTAAAAAGG CCGCGTTGCT
     ACTCGTTTTC CGGTCGTTTT CCGGTCCTTG GCATTTTTCC GGCGCAACGA

2201 GGCGTTTTTC CATAGGCTCC GCCCCCCTGA CGAGCATCAC AAAAATCGAC
     CCGCAAAAAG GTATCCGAGG CGGGGGGACT GCTCGTAGTG TTTTTAGCTG

2251 GCTCAAGTCA GAGGTGGCGA AACCCGACAG GACTATAAAG ATACCAGGCG
     CGAGTTCAGT CTCCACCGCT TTGGGCTGTC CTGATATTTC TATGGTCCGC

2301 TTTCCCCCTG GAAGCTCCCT CGTGCGCTCT CCTGTTCCGA CCCTGCCGCT
     AAAGGGGGAC CTTCGAGGGA GCACGCGAGA GGACAAGGCT GGGACGGCGA

2351 TACCGGATAC CTGTCCGCCT TTCTCCCTTC GGGAAGCGTG GCGCTTTCTC
     ATGGCCTATG GACAGGCGGA AAGAGGGAAG CCCTTCGCAC CGCGAAAGAG

2401 ATAGCTCACG CTGTAGGTAT CTCAGTTCGG TGTAGGTCGT TCGCTCCAAG
     TATCGAGTGC GACATCCATA GAGTCAAGCC ACATCCAGCA AGCGAGGTTC

2451 CTGGGCTGTG TGCACGAACC CCCCGTTCAG CCCGACCGCT GCGCCTTATC
     GACCCGACAC ACGTGCTTGG GGGGCAAGTC GGGCTGGCGA CGCGGAATAG

2501 CGGTAACTAT CGTCTTGAGT CCAACCCGGT AAGACACGAC TTATCGCCAC
     GCCATTGATA GCAGAACTCA GGTTGGGCCA TTCTGTGCTG AATAGCGGTG

2551 TGGCAGCAGC CACTGGTAAC AGGATTAGCA GAGCGAGGTA TGTAGGCGGT
     ACCGTCGTCG GTGACCATTG TCCTAATCGT CTCGCTCCAT ACATCCGCCA

2601 GCTACAGAGT TCTTGAAGTG GTGGCCTAAC TACGGCTACA CTAGAAGAAC
     CGATGTCTCA AGAACTTCAC CACCGGATTG ATGCCGATGT GATCTTCTTG

2651 AGTATTTGGT ATCTGCGCTC TGCTGAAGCC AGTTACCTTC GGAAAAAGAG
     TCATAAACCA TAGACGCGAG ACGACTTCGG TCAATGGAAG CCTTTTTCTC

2701 TTGGTAGCTC TTGATCCGGC AAACAAACCA CCGCTGGTAG CGGTGGTTTT
     AACCATCGAG AACTAGGCCG TTTGTTTGGT GGCGACCATC GCCACCAAAA

2751 TTTGTTTGCA AGCAGCAGAT TACGCGCAGA AAAAAAGGAT CTCAAGAAGA
     AAACAAACGT TCGTCGTCTA ATGCGCGTCT TTTTTTCCTA GAGTTCTTCT

2801 TCCTTTGATC TTTTCTACGG GGTCTGACGC TCAGTGGAAC GAAAACTCAC
     AGGAAACTAG AAAAGATGCC CCAGACTGCG AGTCACCTTG CTTTTGAGTG

2851 GTTAAGGGAT TTTGGTCATG AGATTATCAA AAAGGATCTT CACCTAGATC
     CAATTCCCTA AAACCAGTAC TCTAATAGTT TTTCCTAGAA GTGGATCTAG

2901 CTTTTAAATT AAAAATGAAG TTTTAAATCA ATCTAAAGTA TATATGAGTA
     GAAAATTTAA TTTTTACTTC AAAATTTAGT TAGATTTCAT ATATACTCAT
```

```
2951 AACTTGGTCT GACAGTTACC AATGCTTAAT CAGTGAGGCA CCTATCTCAG
     TTGAACCAGA CTGTCAATGG TTACGAATTA GTCACTCCGT GGATAGAGTC
                       *   W  H  K  I   L  S  A   G  I  E   ampR

3001 CGATCTGTCT ATTTCGTTCA TCCATAGTTG CCTGACTCCC CGTCGTGTAG
     GCTAGACAGA TAAAGCAAGT AGGTATCAAC GGACTGAGGG GCAGCACATC
     A  I  Q  R   N  R  E   D  M  T   A  Q  S  G   T  T  Y  ampR

3051 ATAACTACGA TACGGGAGGG CTTACCATCT GGCCCCAGTG CTGCAATGAT
     TATTGATGCT ATGCCCTCCC GAATGGTAGA CCGGGGTCAC GACGTTACTA
      I  V  V   I  R  S  P   K  G  D   P  G  L   A  A  I  I ampR

3101 ACCGCGAGAC CCACGCTCAC CGGCTCCAGA TTTATCAGCA ATAAACCAGC
     TGGCGCTCTG GGTGCGAGTG GCCGAGGTCT AAATAGTCGT TATTTGGTCG
       G  R  S   G  R  E   G  A  G  S   K  D  A   I  F  W   ampR

3151 CAGCCGGAAG GGCCGAGCGC AGAAGTGGTC CTGCAACTTT ATCCGCCTCC
     GTCGGCCTTC CCGGCTCGCG TCTTCACCAG GACGTTGAAA TAGGCGGAGG
     G  A  P  L   A  S  R   L  L  P   G  A  V  K   D  A  E  ampR

3201 ATCCAGTCTA TTAATTGTTG CCGGGAAGCT AGAGTAAGTA GTTCGCCAGT
     TAGGTCAGAT AATTAACAAC GGCCCTTCGA TCTCATTCAT CAAGCGGTCA
      M  W  D   I  L  Q  Q   R  S  A   L  T  L   L  E  G  T ampR

3251 TAATAGTTTG CGCAACGTTG TTGCCATTGC TACAGGCATC GTGGTGTCAC
     ATTATCAAAC GCGTTGCAAC AACGGTAACG ATGTCCGTAG CACCACAGTG
       L  L  K   R  L  T   T  A  M  A   V  P  M   T  T  D   ampR

3301 GCTCGTCGTT TGGTATGGCT TCATTCAGCT CCGGTTCCCA ACGATCAAGG
     CGAGCAGCAA ACCATACCGA AGTAAGTCGA GGCCAAGGGT TGCTAGTTCC
     R  E  D  N   P  I  A   E  N  L   E  P  E  W   R  D  L  ampR

3351 CGAGTTACAT GATCCCCCAT GTTGTGCAAA AAAGCGGTTA GCTCCTTCGG
     GCTCAATGTA CTAGGGGGTA CAACACGTTT TTTCGCCAAT CGAGGAAGCC
      R  T  V   H  D  G  M   N  H  L   F  A  T   L  E  K  P ampR

3401 TCCTCCGATC GTTGTCAGAA GTAAGTTGGC CGCAGTGTTA TCACTCATGG
     AGGAGGCTAG CAACAGTCTT CATTCAACCG GCGTCACAAT AGTGAGTACC
       G  G  I   T  T  L   L  L  N  A   A  T  N   D  S  M   ampR

3451 TTATGGCAGC ACTGCATAAT TCTCTTACTG TCATGCCATC CGTAAGATGC
     AATACCGTCG TGACGTATTA AGAGAATGAC AGTACGGTAG GCATTCTACG
     T  I  A  A   S  C  L   E  R  V   T  M  G  D   T  L  H  ampR

3501 TTTTCTGTGA CTGGTGAGTA CTCAACCAAG TCATTCTGAG AATAGTGTAT
     AAAAGACACT GACCACTCAT GAGTTGGTTC AGTAAGACTC TTATCACATA
      K  E  T   V  P  S  Y   E  V  L   D  N  Q   S  Y  H  I ampR

3551 GCGGCGACCG AGTTGCTCTT GCCCGGCGTC AATACGGGAT AATACCGCGC
     CGCCGCTGGC TCAACGAGAA CGGGCCGCAG TTATGCCCTA TTATGGCGCG
       R  R  G   L  Q  E   Q  G  A  D   I  R  S   L  V  A   ampR

3601 CACATAGCAG AACTTTAAAA GTGCTCATCA TTGGAAAACG TTCTTCGGGG
     GTGTATCGTC TTGAAATTTT CACGAGTAGT AACCTTTTGC AAGAAGCCCC
     G  C  L  L   V  K  F   T  S  M   M  P  F  R   E  E  P  ampR

3651 CGAAAACTCT CAAGGATCTT ACCGCTGTTG AGATCCAGTT CGATGTAACC
     GCTTTTGAGA GTTCCTAGAA TGGCGACAAC TCTAGGTCAA GCTACATTGG
      R  F  S   E  L  I  K   G  S  N   L  D  L   E  I  Y  G ampR

3701 CACTCGTGCA CCCAACTGAT CTTCAGCATC TTTTACTTTC ACCAGCGTTT
```

```
     GTGAGCACGT GGGTTGACTA GAAGTCGTAG AAAATGAAAG TGGTCGCAAA
       V  R  A   G  L  Q   D  E  A  D   K  V  K   V  L  T   ampR

3751 CTGGGTGAGC AAAAACAGGA AGGCAAAATG CCGCAAAAAA GGGAATAAGG
     GACCCACTCG TTTTTGTCCT TCCGTTTTAC GGCGTTTTTT CCCTTATTCC
     E  P  H  A   F  V  P   L  C  F   A  A  F  F   P  I  L  ampR

3801 GCGACACGGA AATGTTGAAT ACTCATACTC TTCCTTTTTC AATATTATTG
     CGCTGTGCCT TTACAACTTA TGAGTATGAG AAGGAAAAAG TTATAATAAC
      A  V  R   F  H  Q  I   S  M  ampR

3851 AAGCATTTAT CAGGGTTATT GTCTCATGAG CGGATACATA TTTGAATGTA
     TTCGTAAATA GTCCCAATAA CAGAGTACTC GCCTATGTAT AAACTTACAT

3901 TTTAGAAAAA TAAACAAATA GGGGTTCCGC GCACATTTCC CCGAAAAGTG
     AAATCTTTTT ATTTGTTTAT CCCCAAGGCG CGTGTAAAGG GGCTTTTCAC

3951 CCACCTGATG CGGTGTGAAA TACCGCACAG ATGCGTAAGG AGAAAATACC
     GGTGGACTAC GCCACACTTT ATGGCGTGTC TACGCATTCC TCTTTTATGG

4001 GCATCAGGAA ATTGTAAGCG TTAATATTTT GTTAAAATTC GCGTTAAATT
     CGTAGTCCTT TAACATTCGC AATTATAAAA CAATTTTAAG CGCAATTTAA

4051 TTTGTTAAAT CAGCTCATTT TTTAACCAAT AGGCCGAAAT CGGCAAAATC
     AAACAATTTA GTCGAGTAAA AAATTGGTTA TCCGGCTTTA GCCGTTTTAG

4101 CCTTATAAAT CAAAAGAATA GACCGAGATA GGGTTGAGTG TTGTTCCAGT
     GGAATATTTA GTTTTCTTAT CTGGCTCTAT CCCAACTCAC AACAAGGTCA

4151 TTGGAACAAG AGTCCACTAT TAAAGAACGT GGACTCCAAC GTCAAAGGGC
     AACCTTGTTC TCAGGTGATA ATTTCTTGCA CCTGAGGTTG CAGTTTCCCG

4201 GAAAAACCGT CTATCAGGGC GATGGCCCAC TACGTGAACC ATCACCCTAA
     CTTTTTGGCA GATAGTCCCG CTACCGGGTG ATGCACTTGG TAGTGGGATT

4251 TCAAGTTTTT TGGGGTCGAG GTGCCGTAAA GCACTAAATC GGAACCCTAA
     AGTTCAAAAA ACCCCAGCTC CACGGCATTT CGTGATTTAG CCTTGGGATT

4301 AGGGAGCCCC CGATTTAGAG CTTGACGGGG AAAGCCGGCG AACGTGGCGA
     TCCCTCGGGG GCTAAATCTC GAACTGCCCC TTTCGGCCGC TTGCACCGCT

4351 GAAAGGAAGG GAAGAAAGCG AAAGGAGCGG GCGCTAGGGC GCTGGCAAGT
     CTTTCCTTCC CTTCTTTCGC TTTCCTCGCC CGCGATCCCG CGACCGTTCA

4401 GTAGCGGTCA CGCTGCGCGT AACCACCACA CCCGCCGCGC TTAATGCGCC
     CATCGCCAGT GCGACGCGCA TTGGTGGTGT GGGCGGCGCG AATTACGCGG

4451 GCTACAGGGC GCGTCCATTC GCCATTCAGG CTGCGCAACT GTTGGGAAGG
     CGATGTCCCG CGCAGGTAAG CGGTAAGTCC GACGCGTTGA CAACCCTTCC

4501 GCGATCGGTG CGGGCCTCTT CGCTATTACG CCAGCTGGCG AAAGGGGGAT
     CGCTAGCCAC GCCCGGAGAA GCGATAATGC GGTCGACCGC TTTCCCCCTA

4551 GTGCTGCAAG GCGATTAAGT TGGGTAACGC CAGGGTTTTC CCAGTCACGA
     CACGACGTTC CGCTAATTCA ACCCATTGCG GTCCCAAAAG GGTCAGTGCT

4601 CGTTGTAAAA CGACGGCCAG TGAATTGTAA TACGACTCAC TATA
     GCAACATTTT GCTGCCGGTC ACTTAACATT ATGCTGAGTG ATAT
```

Appendix G

pGU0707 (pGU0501Δ*Cj1344c*::*kan*)




```
  1 GGGCGAATTG GGCCCGACGT CGCATGCTCC CGGCCGCCAT GGCGGCCGCG
    CCCGCTTAAC CCGGGCTGCA GCGTACGAGG GCCGGCGGTA CCGCCGGCGC

                   START Cj1344c
 51 GGAATTCGAT TCATATGAAA AATCTTATCC TAGCTATAGA AAGTTCTTGT
    CCTTAAGCTA AGTATACTTT TTAGAATAGG ATCGATATCT TTCAAGAACA
                   M  K   N  L  I  L   A  I  E   S  S  C   Cj1344c

101 GATGATAGTT CTATAGCTAT CATTGATAAA AACACCTTAG AATGTAAATT
    CTACTATCAA GATATCGATA GTAACTATTT TTGTGGAATC TTACATTTAA
    D  D  S  S   I  A  I   I  D  K   N  T  L  E   C  K  F  Cj1344c

151 TCATAAAAAA ATTTCCCAAG AATTAGATCA TAGTATCTAT GGGGGAGTGG
    AGTATTTTTT TAAAGGGTTC TTAATCTAGT ATCATAGATA CCCCCTCACC
     H  K  K   I  S  Q  E   L  D  H   S  I  Y   G  G  V  V Cj1344c

201 TACCTGAACT TGCTGCAAGA CTTCATAGCG AGGCTTTACC AAAGATGCTT
    ATGGACTTGA ACGACGTTCT GAAGTATCGC TCCGAAATGG TTTCTACGAA
      P  E  L   A  A  R   L  H  S  E   A  L  P   K  M  L   Cj1344c

251 AAGCAATGCA AAGAGCATTT TAAAAATCTT TGTGCCATAG CTGTGACAAA
    TTCGTTACGT TTCTCGTAAA ATTTTTAGAA ACACGGTATC GACACTGTTT
    K  Q  C  K   E  H  F   K  N  L   C  A  I  A   V  T  N  Cj1344c

301 TGAACCTGGA CTTAGTGTTT CTTTGCTCAG TGGAATTTCT ATGGCAAAAA
    ACTTGGACCT GAATCACAAA GAAACGAGTC ACCTTAAAGA TACCGTTTTT
     E  P  G   L  S  V  S   L  L  S   G  I  S   M  A  K  T Cj1344c

351 CCTTAGCAAG TGCGCTAAAT TTACCCTTAA TCCCTATAAA TCATCTTAAA
    GGAATCGTTC ACGCGATTTA AATGGGAATT AGGGATATTT AGTAGAATTT
      L  A  S   A  L  N   L  P  L  I   P  I  N   H  L  K   Cj1344c
```

```
 401 GGTCATATTT ATAGTCTTTT TTTGGAAGAA AAAATTTCTT TAGATATGGG
     CCAGTATAAA TATCAGAAAA AAACCTTCTT TTTTAAAGAA ATCTATACCC
     G  H  I  Y   S  L  F   L  E  E   K  I  S  L   D  M  G  Cj1344c

                                START kan (in frame)
 451 AATTTTGCTT GTTAGTGGTG GAGATCTCAT GGCTAAAATG AGAATATCAC
     TTAAAACGAA CAATCACCAC CTCTAGAGTA CCGATTTTAC TCTTATAGTG
      I  L  L   V  S  G  G   D  L  M   A  K  M   R  I  S  P kan

 501 CGGAATTGAA AAAACTGATC GAAAAATACC GCTGCGTAAA AGATACGGAA
     GCCTTAACTT TTTTGACTAG CTTTTTATGG CGACGCATTT TCTATGCCTT
       E  L  K   K  L  I   E  K  Y  R   C  V  K   D  T  E   kan

 551 GGAATGTCTC CTGCTAAGGT ATATAAGCTG GTGGGAGAAA ATGAAAACCT
     CCTTACAGAG GACGATTCCA TATATTCGAC CACCCTCTTT TACTTTTGGA
     G  M  S  P   A  K  V   Y  K  L   V  G  E  N   E  N  L  kan

 601 ATATTTAAAA ATGACGGACA GCCGGTATAA AGGGACCACC TATGATGTGG
     TATAAATTTT TACTGCCTGT CGGCCATATT TCCCTGGTGG ATACTACACC
      Y  L  K   M  T  D  S   R  Y  K   G  T  T   Y  D  V  E kan

 651 AACGGGAAAA GGACATGATG CTATGGCTGG AAGGAAAGCT GCCTGTTCCA
     TTGCCCTTTT CCTGTACTAC GATACCGACC TTCCTTTCGA CGGACAAGGT
       R  E  K   D  M  M   L  W  L  E   G  K  L   P  V  P   kan

 701 AAGGTCCTGC ACTTTGAACG GCATGATGGC TGGAGCAATC TGCTCATGAG
     TTCCAGGACG TGAAACTTGC CGTACTACCG ACCTCGTTAG ACGAGTACTC
     K  V  L  H   F  E  R   H  D  G   W  S  N  L   L  M  S  kan

 751 TGAGGCCGAT GGCGTCCTTT GCTCGGAAGA GTATGAAGAT GAACAAAGCC
     ACTCCGGCTA CCGCAGGAAA CGAGCCTTCT CATACTTCTA CTTGTTTCGG
      E  A  D   G  V  L  C   S  E  E   Y  E  D   E  Q  S  P kan

 801 CTGAAAAGAT TATCGAGCTG TATGCGGAGT GCATCAGGCT CTTTCACTCC
     GACTTTTCTA ATAGCTCGAC ATACGCCTCA CGTAGTCCGA GAAAGTGAGG
       E  K  I   I  E  L   Y  A  E  C   I  R  L   F  H  S   kan

 851 ATCGACATAT CGGATTGTCC CTATACGAAT AGCTTAGACA GCCGCTTAGC
     TAGCTGTATA GCCTAACAGG GATATGCTTA TCGAATCTGT CGGCGAATCG
     I  D  I  S   D  C  P   Y  T  N   S  L  D  S   R  L  A  kan

 901 CGAATTGGAT TACTTACTGA ATAACGATCT GGCCGATGTG GATTGCGAAA
     GCTTAACCTA ATGAATGACT TATTGCTAGA CCGGCTACAC CTAACGCTTT
      E  L  D   Y  L  L  N   N  D  L   A  D  V   D  C  E  N kan

 951 ACTGGGAAGA AGACACTCCA TTTAAAGATC CGCGCGAGCT GTATGATTTT
     TGACCCTTCT TCTGTGAGGT AAATTTCTAG GCGCGCTCGA CATACTAAAA
       W  E  E   D  T  P   F  K  D  P   R  E  L   Y  D  F   kan

1001 TTAAAGACGG AAAAGCCCGA AGAGGAACTT GTCTTTTCCC ACGGCGACCT
     AATTTCTGCC TTTTCGGGCT TCTCCTTGAA CAGAAAAGGG TGCCGCTGGA
     L  K  T  E   K  P  E   E  E  L   V  F  S  H   G  D  L  kan

1051 GGGAGACAGC AACATCTTTG TGAAAGATGG CAAAGTAAGT GGCTTTATTG
     CCCTCTGTCG TTGTAGAAAC ACTTTCTACC GTTTCATTCA CCGAAATAAC
      G  D  S   N  I  F  V   K  D  G   K  V  S   G  F  I  D kan

1101 ATCTTGGGAG AAGCGGCAGG GCGGACAAGT GGTATGACAT TGCCTTCTGC
     TAGAACCCTC TTCGCCGTCC CGCCTGTTCA CCATACTGTA ACGGAAGACG
       L  G  R   S  G  R   A  D  K  W   Y  D  I   A  F  C   kan
```

```
1151 GTCCGGTCGA TCAGGGAGGA TATCGGGGAA GAACAGTATG TCGAGCTATT
     CAGGCCAGCT AGTCCCTCCT ATAGCCCCTT CTTGTCATAC AGCTCGATAA
     V  R  S  I   R  E  D   I  G  E   E  Q  Y  V   E  L  F  kan

1201 TTTTGACTTA CTGGGGATCA AGCCTGATTG GGAGAAAATA AAATATTATA
     AAAACTGAAT GACCCCTAGT TCGGACTAAC CCTCTTTTAT TTTATAATAT
      F  D  L   L  G  I  K   P  D  W   E  K  I   K  Y  Y  I kan

1251 TTTTACTGGA TGAATTGTTT TAGTACCTAG ATTTAGATGT CTAAAAAGCT
     AAAATGACCT ACTTAACAAA ATCATGGATC TAAATCTACA GATTTTTCGA
       L  L  D   E  L  F   *  kan

1301 TGATATCGAA TTCCTGCAGC CCGGGGGATC CACTAGTTCT AGAGCGGCCG
     ACTATAGCTT AAGGACGTCG GGCCCCCTAG GTGATCAAGA TCTCGCCGGC

1351 CCACCGCGGT GGAGCTCCAG CTTTTGTTCC CTTTAGTGAG GGTTAATTCC
     GGTGGCGCCA CCTCGAGGTC GAAAACAAGG GAAATCACTC CCAATTAAGG

1401 GAGCAGATCT GAGCTTTTAG CAAGTACAAA TGATGATAGC TTTGGAGAAA
     CTCGTCTAGA CTCGAAAATC GTTCATGTTT ACTACTATCG AAACCTCTTT

                                Cj1344 (split)
1451 GTTTTGATAA AGTGGCTAAA ATGATGAATT TAGGTTACCC TGGTGGGGTC
     CAAAACTATT TCACCGATTT TACTACTTAA ATCCAATGGG ACCACCCCAG
                           M  M  N  L   G  Y  P   G  G  V   Cj1344c

1501 ATCATAGAAA ATTTAGCAAA AAATGCCAAA CTTAAAAATA TCTCTTTTAA
     TAGTATCTTT TAAATCGTTT TTTACGGTTT GAATTTTTAT AGAGAAAATT
     I  I  E  N   L  A  K   N  A  K   L  K  N  I   S  F  N  Cj1344c

1551 CACACCTTTA AAGCATTCTA AAGAACTCGC TTTCAGTTTT TCAGGGCTTA
     GTGTGGAAAT TTCGTAAGAT TTCTTGAGCG AAAGTCAAAA AGTCCCGAAT
      T  P  L   K  H  S  K   E  L  A   F  S  F   S  G  L  K Cj1344c

1601 AAAATGCAGT GCGTTTGGAA ATTTTAAAAC ATGAAAATTT AAATGAAGAC
     TTTTACGTCA CGCAAACCTT TAAAATTTTG TACTTTTAAA TTTACTTCTG
       N  A  V   R  L  E   I  L  K  H   E  N  L   N  E  D   Cj1344c

1651 ACAAAAGCAG AAATAGCCTA TGCCTTTGAA AATACAGCTT GTGATCATAT
     TGTTTTCGTC TTTATCGGAT ACGGAAACTT TTATGTCGAA CACTAGTATA
     T  K  A  E   I  A  Y   A  F  E   N  T  A  C   D  H  I  Cj1344c

1701 CATGGATAAA TTAGAAAAAA TTTTTAATCT TTATAAATTT AAAAATTTTG
     GTACCTATTT AATCTTTTTT AAAAATTAGA AATATTTAAA TTTTTAAAAC
      M  D  K   L  E  K  I   F  N  L   Y  K  F   K  N  F  G Cj1344c

1751 GCGTTGTAGG TGGAGCTAGT GCAAATCTTA ACTTGCGTTC GCGTTTGCAA
     CGCAACATCC ACCTCGATCA CGTTTAGAAT TGAACGCAAG CGCAAACGTT
       V  V  G   G  A  S   A  N  L  N   L  R  S   R  L  Q   Cj1344c

1801 AATTTATGTC AAAAATATAA TGCAAATTTA AAACTAGCTC CTTTAAAATT
     TTAAATACAG TTTTTATATT ACGTTTAAAT TTTGATCGAG GAAATTTTAA
     N  L  C  Q   K  Y  N   A  N  L   K  L  A  P   L  K  F  Cj1344c

1851 CTGCTCTGAT AATGCTTTGA TGATAGCAAG AGCCGCAGTT GATGCTTATG
     GACGAGACTA TTACGAAACT ACTATCGTTC TCGGCGTCAA CTACGAATAC
      C  S  D   N  A  L  M   I  A  R   A  A  V   D  A  Y  E Cj1344c

1901 AAAAAAAGGA ATTTGTAAGT GTAGAAGAAG ATATTTTAAG CCCTAAAAAT
     TTTTTTTCCT TAAACATTCA CATCTTCTTC TATAAAATTC GGGATTTTTA
       K  K  E   F  V  S   V  E  E  D   I  L  S   P  K  N   Cj1344c
```

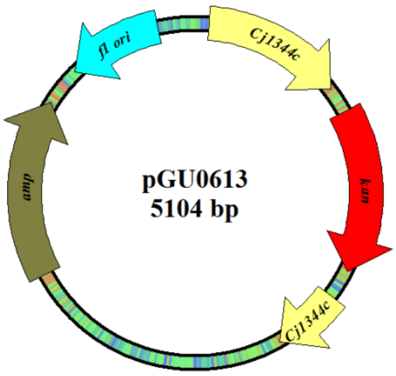
```
1951 AAAAATTTTT CAAGGATATA GATGAAAAAA GCTCGAGAAT CACTAGTGAA
     TTTTTAAAAA GTTCCTATAT CTACTTTTTT CGAGCTCTTA GTGATCACTT
     K  N  F  S   R  I  *  Cj1344c

2001 TTCGCGGCCG CCTGCAGGTC GACCATATGG GAGAGCTCCC AACGCGTTGG
     AAGCGCCGGC GGACGTCCAG CTGGTATACC CTCTCGAGGG TTGCGCAACC

2051 ATGCATAGCT TGAGTATTCT ATAGTGTCAC CTAAATAGCT TGGCGTAATC
     TACGTATCGA ACTCATAAGA TATCACAGTG GATTTATCGA ACCGCATTAG

2101 ATGGTCATAG CTGTTTCCTG TGTGAAATTG TTATCCGCTC ACAATTCCAC
     TACCAGTATC GACAAAGGAC ACACTTTAAC AATAGGCGAG TGTTAAGGTG

2151 ACAACATACG AGCCGGAAGC ATAAAGTGTA AAGCCTGGGG TGCCTAATGA
     TGTTGTATGC TCGGCCTTCG TATTTCACAT TTCGGACCCC ACGGATTACT

2201 GTGAGCTAAC TCACATTAAT TGCGTTGCGC TCACTGCCCG CTTTCCAGTC
     CACTCGATTG AGTGTAATTA ACGCAACGCG AGTGACGGGC GAAAGGTCAG

2251 GGGAAACCTG TCGTGCCAGC TGCATTAATG AATCGGCCAA CGCGCGGGGA
     CCCTTTGGAC AGCACGGTCG ACGTAATTAC TTAGCCGGTT GCGCGCCCCT

2301 GAGGCGGTTT GCGTATTGGG CGCTCTTCCG CTTCCTCGCT CACTGACTCG
     CTCCGCCAAA CGCATAACCC GCGAGAAGGC GAAGGAGCGA GTGACTGAGC

2351 CTGCGCTCGG TCGTTCGGCT GCGGCGAGCG GTATCAGCTC ACTCAAAGGC
     GACGCGAGCC AGCAAGCCGA CGCCGCTCGC CATAGTCGAG TGAGTTTCCG

2401 GGTAATACGG TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAACATGT
     CCATTATGCC AATAGGTGTC TTAGTCCCCT ATTGCGTCCT TTCTTGTACA

2451 GAGCAAAAGG CCAGCAAAAG GCCAGGAACC GTAAAAAGGC CGCGTTGCTG
     CTCGTTTTCC GGTCGTTTTC CGGTCCTTGG CATTTTTCCG GCGCAACGAC

2501 GCGTTTTTCC ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGACG
     CGCAAAAAGG TATCCGAGGC GGGGGGACTG CTCGTAGTGT TTTTAGCTGC

2551 CTCAAGTCAG AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT
     GAGTTCAGTC TCCACCGCTT TGGGCTGTCC TGATATTTCT ATGGTCCGCA

2601 TTCCCCCTGG AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT
     AAGGGGGACC TTCGAGGGAG CACGCGAGAG GACAAGGCTG GGACGGCGAA

2651 ACCGGATACC TGTCCGCCTT TCTCCCTTCG GGAAGCGTGG CGCTTTCTCA
     TGGCCTATGG ACAGGCGGAA AGAGGGAAGC CCTTCGCACC GCGAAAGAGT

2701 TAGCTCACGC TGTAGGTATC TCAGTTCGGT GTAGGTCGTT CGCTCCAAGC
     ATCGAGTGCG ACATCCATAG AGTCAAGCCA CATCCAGCAA GCGAGGTTCG

2751 TGGGCTGTGT GCACGAACCC CCCGTTCAGC CCGACCGCTG CGCCTTATCC
     ACCCGACACA CGTGCTTGGG GGGCAAGTCG GGCTGGCGAC GCGGAATAGG

2801 GGTAACTATC GTCTTGAGTC CAACCCGGTA AGACACGACT TATCGCCACT
     CCATTGATAG CAGAACTCAG GTTGGGCCAT TCTGTGCTGA ATAGCGGTGA

2851 GGCAGCAGCC ACTGGTAACA GGATTAGCAG AGCGAGGTAT GTAGGCGGTG
     CCGTCGTCGG TGACCATTGT CCTAATCGTC TCGCTCCATA CATCCGCCAC

2901 CTACAGAGTT CTTGAAGTGG TGGCCTAACT ACGGCTACAC TAGAAGAACA
     GATGTCTCAA GAACTTCACC ACCGGATTGA TGCCGATGTG ATCTTCTTGT
```

```
2951 GTATTTGGTA TCTGCGCTCT GCTGAAGCCA GTTACCTTCG GAAAAAGAGT
     CATAAACCAT AGACGCGAGA CGACTTCGGT CAATGGAAGC CTTTTTCTCA

3001 TGGTAGCTCT TGATCCGGCA AACAAACCAC CGCTGGTAGC GGTGGTTTTT
     ACCATCGAGA ACTAGGCCGT TTGTTTGGTG GCGACCATCG CCACCAAAAA

3051 TTGTTTGCAA GCAGCAGATT ACGCGCAGAA AAAAAGGATC TCAAGAAGAT
     AACAAACGTT CGTCGTCTAA TGCGCGTCTT TTTTTCCTAG AGTTCTTCTA

3101 CCTTTGATCT TTTCTACGGG GTCTGACGCT CAGTGGAACG AAAACTCACG
     GGAAACTAGA AAAGATGCCC CAGACTGCGA GTCACCTTGC TTTTGAGTGC

3151 TTAAGGGATT TTGGTCATGA GATTATCAAA AAGGATCTTC ACCTAGATCC
     AATTCCCTAA AACCAGTACT CTAATAGTTT TTCCTAGAAG TGGATCTAGG

3201 TTTTAAATTA AAAATGAAGT TTTAAATCAA TCTAAAGTAT ATATGAGTAA
     AAAATTTAAT TTTTACTTCA AAATTTAGTT AGATTTCATA TATACTCATT

3251 ACTTGGTCTG ACAGTTACCA ATGCTTAATC AGTGAGGCAC CTATCTCAGC
     TGAACCAGAC TGTCAATGGT TACGAATTAG TCACTCCGTG GATAGAGTCG

3301 GATCTGTCTA TTTCGTTCAT CCATAGTTGC CTGACTCCCC GTCGTGTAGA
     CTAGACAGAT AAAGCAAGTA GGTATCAACG GACTGAGGGG CAGCACATCT

3351 TAACTACGAT ACGGGAGGGC TTACCATCTG GCCCCAGTGC TGCAATGATA
     ATTGATGCTA TGCCCTCCCG AATGGTAGAC CGGGGTCACG ACGTTACTAT

3401 CCGCGAGACC CACGCTCACC GGCTCCAGAT TTATCAGCAA TAAACCAGCC
     GGCGCTCTGG GTGCGAGTGG CCGAGGTCTA AATAGTCGTT ATTTGGTCGG

3451 AGCCGGAAGG GCCGAGCGCA GAAGTGGTCC TGCAACTTTA TCCGCCTCCA
     TCGGCCTTCC CGGCTCGCGT CTTCACCAGG ACGTTGAAAT AGGCGGAGGT

3501 TCCAGTCTAT TAATTGTTGC CGGGAAGCTA GAGTAAGTAG TTCGCCAGTT
     AGGTCAGATA ATTAACAACG GCCCTTCGAT CTCATTCATC AAGCGGTCAA

3551 AATAGTTTGC GCAACGTTGT TGCCATTGCT ACAGGCATCG TGGTGTCACG
     TTATCAAACG CGTTGCAACA ACGGTAACGA TGTCCGTAGC ACCACAGTGC

3601 CTCGTCGTTT GGTATGGCTT CATTCAGCTC CGGTTCCCAA CGATCAAGGC
     GAGCAGCAAA CCATACCGAA GTAAGTCGAG GCCAAGGGTT GCTAGTTCCG

3651 GAGTTACATG ATCCCCCATG TTGTGCAAAA AAGCGGTTAG CTCCTTCGGT
     CTCAATGTAC TAGGGGGTAC AACACGTTTT TTCGCCAATC GAGGAAGCCA

3701 CCTCCGATCG TTGTCAGAAG TAAGTTGGCC GCAGTGTTAT CACTCATGGT
     GGAGGCTAGC AACAGTCTTC ATTCAACCGG CGTCACAATA GTGAGTACCA

3751 TATGGCAGCA CTGCATAATT CTCTTACTGT CATGCCATCC GTAAGATGCT
     ATACCGTCGT GACGTATTAA GAGAATGACA GTACGGTAGG CATTCTACGA

3801 TTTCTGTGAC TGGTGAGTAC TCAACCAAGT CATTCTGAGA ATAGTGTATG
     AAAGACACTG ACCACTCATG AGTTGGTTCA GTAAGACTCT TATCACATAC

3851 CGGCGACCGA GTTGCTCTTG CCCGGCGTCA ATACGGGATA ATACCGCGCC
     GCCGCTGGCT CAACGAGAAC GGGCCGCAGT TATGCCCTAT TATGGCGCGG

3901 ACATAGCAGA ACTTTAAAAG TGCTCATCAT TGGAAAACGT TCTTCGGGGC
     TGTATCGTCT TGAAATTTTC ACGAGTAGTA ACCTTTTGCA AGAAGCCCCG
```

```
3951 GAAAACTCTC AAGGATCTTA CCGCTGTTGA GATCCAGTTC GATGTAACCC
     CTTTTGAGAG TTCCTAGAAT GGCGACAACT CTAGGTCAAG CTACATTGGG

4001 ACTCGTGCAC CCAACTGATC TTCAGCATCT TTTACTTTCA CCAGCGTTTC
     TGAGCACGTG GGTTGACTAG AAGTCGTAGA AAATGAAAGT GGTCGCAAAG

4051 TGGGTGAGCA AAAACAGGAA GGCAAAATGC CGCAAAAAAG GGAATAAGGG
     ACCCACTCGT TTTTGTCCTT CCGTTTTACG GCGTTTTTTC CCTTATTCCC

4101 CGACACGGAA ATGTTGAATA CTCATACTCT TCCTTTTTCA ATATTATTGA
     GCTGTGCCTT TACAACTTAT GAGTATGAGA AGGAAAAAGT TATAATAACT

4151 AGCATTTATC AGGGTTATTG TCTCATGAGC GGATACATAT TTGAATGTAT
     TCGTAAATAG TCCCAATAAC AGAGTACTCG CCTATGTATA AACTTACATA

4201 TTAGAAAAAT AAACAAATAG GGGTTCCGCG CACATTTCCC CGAAAAGTGC
     AATCTTTTTA TTTGTTTATC CCCAAGGCGC GTGTAAAGGG GCTTTTCACG

4251 CACCTGATGC GGTGTGAAAT ACCGCACAGA TGCGTAAGGA GAAAATACCG
     GTGGACTACG CCACACTTTA TGGCGTGTCT ACGCATTCCT CTTTTATGGC

4301 CATCAGGAAA TTGTAAGCGT TAATATTTTG TTAAAATTCG CGTTAAATTT
     GTAGTCCTTT AACATTCGCA ATTATAAAAC AATTTTAAGC GCAATTTAAA

4351 TTGTTAAATC AGCTCATTTT TTAACCAATA GGCCGAAATC GGCAAAATCC
     AACAATTTAG TCGAGTAAAA AATTGGTTAT CCGGCTTTAG CCGTTTTAGG

4401 CTTATAAATC AAAAGAATAG ACCGAGATAG GGTTGAGTGT TGTTCCAGTT
     GAATATTTAG TTTTCTTATC TGGCTCTATC CCAACTCACA ACAAGGTCAA

4451 TGGAACAAGA GTCCACTATT AAAGAACGTG GACTCCAACG TCAAAGGGCG
     ACCTTGTTCT CAGGTGATAA TTTCTTGCAC CTGAGGTTGC AGTTTCCCGC

4501 AAAAACCGTC TATCAGGGCG ATGGCCCACT ACGTGAACCA TCACCCTAAT
     TTTTTGGCAG ATAGTCCCGC TACCGGGTGA TGCACTTGGT AGTGGGATTA

4551 CAAGTTTTTT GGGGTCGAGG TGCCGTAAAG CACTAAATCG GAACCCTAAA
     GTTCAAAAAA CCCCAGCTCC ACGGCATTTC GTGATTTAGC CTTGGGATTT

4601 GGGAGCCCCC GATTTAGAGC TTGACGGGGA AAGCCGGCGA ACGTGGCGAG
     CCCTCGGGGG CTAAATCTCG AACTGCCCCT TTCGGCCGCT TGCACCGCTC

4651 AAAGGAAGGG AAGAAAGCGA AAGGAGCGGG CGCTAGGGCG CTGGCAAGTG
     TTTCCTTCCC TTCTTTCGCT TTCCTCGCCC GCGATCCCGC GACCGTTCAC

4701 TAGCGGTCAC GCTGCGCGTA ACCACCACAC CCGCCGCGCT TAATGCGCCG
     ATCGCCAGTG CGACGCGCAT TGGTGGTGTG GGCGGCGCGA ATTACGCGGC

4751 CTACAGGGCG CGTCCATTCG CCATTCAGGC TGCGCAACTG TTGGGAAGGG
     GATGTCCCGC GCAGGTAAGC GGTAAGTCCG ACGCGTTGAC AACCCTTCCC

4801 CGATCGGTGC GGGCCTCTTC GCTATTACGC CAGCTGGCGA AAGGGGGATG
     GCTAGCCACG CCCGGAGAAG CGATAATGCG GTCGACCGCT TTCCCCCTAC

4851 TGCTGCAAGG CGATTAAGTT GGGTAACGCC AGGGTTTTCC CAGTCACGAC
     ACGACGTTCC GCTAATTCAA CCCATTGCGG TCCCAAAAGG GTCAGTGCTG

4901 GTTGTAAAAC GACGGCCAGT GAATTGTAAT ACGACTCACT ATA
     CAACATTTTG CTGCCGGTCA CTTAACATTA TGCTGAGTGA TAT
```

Appendix H

pGU0613 (pGU0501Δ*Cj1344c*::*kan*)




```
  1 GGGCGAATTG GGCCCGACGT CGCATGCTCC CGGCCGCCAT GGCGGCCGCG
    CCCGCTTAAC CCGGGCTGCA GCGTACGAGG GCCGGCGGTA CCGCCGGCGC

                   START Cj1344c
 51 GGAATTCGAT TCATATGAAA AATCTTATCC TAGCTATAGA AAGTTCTTGT
    CCTTAAGCTA AGTATACTTT TTAGAATAGG ATCGATATCT TTCAAGAACA
                   M  K   N  L  I  L   A  I  E   S  S  C   Cj1344c

101 GATGATAGTT CTATAGCTAT CATTGATAAA AACACCTTAG AATGTAAATT
    CTACTATCAA GATATCGATA GTAACTATTT TTGTGGAATC TTACATTTAA
    D  D  S  S   I  A  I   I  D  K   N  T  L  E   C  K  F  Cj1344c

151 TCATAAAAAA ATTTCCCAAG AATTAGATCA TAGTATCTAT GGGGGAGTGG
    AGTATTTTTT TAAAGGGTTC TTAATCTAGT ATCATAGATA CCCCCTCACC
     H  K  K   I  S  Q  E   L  D  H   S  I  Y   G  G  V  V Cj1344c

201 TACCTGAACT TGCTGCAAGA CTTCATAGCG AGGCTTTACC AAAGATGCTT
    ATGGACTTGA ACGACGTTCT GAAGTATCGC TCCGAAATGG TTTCTACGAA
      P  E  L   A  A  R   L  H  S  E   A  L  P   K  M  L   Cj1344c

251 AAGCAATGCA AAGAGCATTT TAAAAATCTT TGTGCCATAG CTGTGACAAA
    TTCGTTACGT TTCTCGTAAA ATTTTTAGAA ACACGGTATC GACACTGTTT
    K  Q  C  K   E  H  F   K  N  L   C  A  I  A   V  T  N  Cj1344c

301 TGAACCTGGA CTTAGTGTTT CTTTGCTCAG TGGAATTTCT ATGGCAAAAA
    ACTTGGACCT GAATCACAAA GAAACGAGTC ACCTTAAAGA TACCGTTTTT
     E  P  G   L  S  V  S   L  L  S   G  I  S   M  A  K  T Cj1344c

351 CCTTAGCAAG TGCGCTAAAT TTACCCTTAA TCCCTATAAA TCATCTTAAA
    GGAATCGTTC ACGCGATTTA AATGGGAATT AGGGATATTT AGTAGAATTT
      L  A  S   A  L  N   L  P  L  I   P  I  N   H  L  K   Cj1344c

401 GGTCATATTT ATAGTCTTTT TTTGGAAGAA AAAATTTCTT TAGATATGGG
    CCAGTATAAA TATCAGAAAA AAACCTTCTT TTTTAAAGAA ATCTATACCC
    G  H  I  Y   S  L  F   L  E  E   K  I  S  L   D  M  G  Cj1344c
```

```
 451 AATTTTGCTT GTTAGTGGTG GGCATACCAT GGTGCTTTAT CTTAAAGATG
     TTAAAACGAA CAATCACCAC CCGTATGGTA CCACGAAATA GAATTTCTAC
      I  L  L   V  S  G  G   H  T  M   V  L  Y   L  K  D  D Cj1344c

 501 ATGCAAGCTT AGAGCTTTTA GCAAGTACAA ATGATGATAG CTTTGGAGAA
     TACGTTCGAA TCTCGAAAAT CGTTCATGTT TACTACTATC GAAACCTCTT
       A  S  L   E  L  L   A  S  T  N   D  D  S   F  G  E   Cj1344c

 551 AGTTTTGATA AAGTGGCTAA AATGATGAAT TTAGGTTACC CTGGTGGGGT
     TCAAAACTAT TTCACCGATT TTACTACTTA AATCCAATGG GACCACCCCA
     S  F  D  K   V  A  K   M  M  N   L  G  Y  P   G  G  V  Cj1344c

 601 CATCATAGAA AATTTAGCAA AAAATGCCAA ACTTAAAAAT ATCTCTTTTA
     GTAGTATCTT TTAAATCGTT TTTTACGGTT TGAATTTTTA TAGAGAAAAT
      I  I  E   N  L  A  K   N  A  K   L  K  N   I  S  F  N Cj1344c

 651 ACACACCTTT AAAGCATTCT AAAGAACTCG CTTTCAGTTT TTCAGGGCTT
     TGTGTGGAAA TTTCGTAAGA TTTCTTGAGC GAAAGTCAAA AAGTCCCGAA
       T  P  L   K  H  S   K  E  L  A   F  S  F   S  G  L   Cj1344c

 701 AAAAATGCAG TGCGTTTGGA AATTTTAAAA CATGAAAATT TGAAGATCTC
     TTTTTACGTC ACGCAAACCT TTAAAATTTT GTACTTTTAA ACTTCTAGAG
     K  N  A  V   R  L  E   I  L  K   H  E  N  L   K  I  S  Cj1344c

 751 TGTTTTCTGG TATTTAAGGT TTTAGAATGC AAGGAACAGT GAATTGGAGT
     ACAAAAGACC ATAAATTCCA AAATCTTACG TTCCTTGTCA CTTAACCTCA
      V  F  W   Y  L  R  F   *  Cj1344c

 801 TCGTCTTGTT ATAATTAGCT TCTTGGGGTA TCTTTAAATA CTGTAGAAAA
     AGCAGAACAA TATTAATCGA AGAACCCCAT AGAAATTTAT GACATCTTTT

                       START kan
 851 GAGGAAGGAA ATAATAAATG GCTAAAATGA GAATATCACC GGAATTGAAA
     CTCCTTCCTT TATTATTTAC CGATTTTACT CTTATAGTGG CCTTAACTTT
                       M   A  K  M  R   I  S  P   E  L  K   kan

 901 AAACTGATCG AAAAATACCG CTGCGTAAAA GATACGGAAG GAATGTCTCC
     TTTGACTAGC TTTTTATGGC GACGCATTTT CTATGCCTTC CTTACAGAGG
     K  L  I  E   K  Y  R   C  V  K   D  T  E  G   M  S  P  kan

 951 TGCTAAGGTA TATAAGCTGG TGGGAGAAAA TGAAAACCTA TATTTAAAAA
     ACGATTCCAT ATATTCGACC ACCCTCTTTT ACTTTTGGAT ATAAATTTTT
      A  K  V   Y  K  L  V   G  E  N   E  N  L   Y  L  K  M kan

1001 TGACGGACAG CCGGTATAAA GGGACCACCT ATGATGTGGA ACGGGAAAAG
     ACTGCCTGTC GGCCATATTT CCCTGGTGGA TACTACACCT TGCCCTTTTC
       T  D  S   R  Y  K   G  T  T  Y   D  V  E   R  E  K   kan

1051 GACATGATGC TATGGCTGGA AGGAAAGCTG CCTGTTCCAA AGGTCCTGCA
     CTGTACTACG ATACCGACCT TCCTTTCGAC GGACAAGGTT TCCAGGACGT
     D  M  M  L   W  L  E   G  K  L   P  V  P  K   V  L  H  kan

1101 CTTTGAACGG CATGATGGCT GGAGCAATCT GCTCATGAGT GAGGCCGATG
     GAAACTTGCC GTACTACCGA CCTCGTTAGA CGAGTACTCA CTCCGGCTAC
      F  E  R   H  D  G  W   S  N  L   L  M  S   E  A  D  G kan

1151 GCGTCCTTTG CTCGGAAGAG TATGAAGATG AACAAAGCCC TGAAAAGATT
     CGCAGGAAAC GAGCCTTCTC ATACTTCTAC TTGTTTCGGG ACTTTTCTAA
       V  L  C   S  E  E   Y  E  D  E   Q  S  P   E  K  I   kan
```

```
1201 ATCGAGCTGT ATGCGGAGTG CATCAGGCTC TTTCACTCCA TCGACATATC
     TAGCTCGACA TACGCCTCAC GTAGTCCGAG AAAGTGAGGT AGCTGTATAG
     I  E  L  Y   A  E  C   I  R  L   F  H  S  I   D  I  S kan

1251 GGATTGTCCC TATACGAATA GCTTAGACAG CCGCTTAGCC GAATTGGATT
     CCTAACAGGG ATATGCTTAT CGAATCTGTC GGCGAATCGG CTTAACCTAA
      D  C  P   Y  T  N  S   L  D  S   R  L  A   E  L  D  Y kan

1301 ACTTACTGAA TAACGATCTG GCCGATGTGG ATTGCGAAAA CTGGGAAGAA
     TGAATGACTT ATTGCTAGAC CGGCTACACC TAACGCTTTT GACCCTTCTT
       L  L  N   N  D  L   A  D  V   D  C  E   N  W  E  E   kan

1351 GACACTCCAT TTAAAGATCC GCGCGAGCTG TATGATTTTT TAAAGACGGA
     CTGTGAGGTA AATTTCTAGG CGCGCTCGAC ATACTAAAAA ATTTCTGCCT
     D  T  P  F   K  D  P   R  E  L   Y  D  F  L   K  T  E kan

1401 AAAGCCCGAA GAGGAACTTG TCTTTTCCCA CGGCGACCTG GGAGACAGCA
     TTTCGGGCTT CTCCTTGAAC AGAAAAGGGT GCCGCTGGAC CCTCTGTCGT
      K  P  E   E  E  L  V   F  S  H   G  D  L   G  D  S  N kan

1451 ACATCTTTGT GAAAGATGGC AAAGTAAGTG GCTTTATTGA TCTTGGGAGA
     TGTAGAAACA CTTTCTACCG TTTCATTCAC CGAAATAACT AGAACCCTCT
       I  F  V   K  D  G   K  V  S   G  F  I   D  L  G  R   kan

1501 AGCGGCAGGG CGGACAAGTG GTATGACATT GCCTTCTGCG TCCGGTCGAT
     TCGCCGTCCC GCCTGTTCAC CATACTGTAA CGGAAGACGC AGGCCAGCTA
     S  G  R  A   D  K  W   Y  D  I   A  F  C  V   R  S  I kan

1551 CAGGGAGGAT ATCGGGGAAG AACAGTATGT CGAGCTATTT TTTGACTTAC
     GTCCCTCCTA TAGCCCCTTC TTGTCATACA GCTCGATAAA AAACTGAATG
      R  E  D   I  G  E  E   Q  Y  V   E  L  F   F  D  L  L kan

1601 TGGGGATCAA GCCTGATTGG GAGAAAATAA AATATTATAT TTTACTGGAT
     ACCCCTAGTT CGGACTAACC CTCTTTTATT TTATAATATA AAATGACCTA
       G  I  K   P  D  W   E  K  I   K  Y  Y   I  L  L  D   kan

1651 GAATTGTTTT AGTACCTAGA TTTAGATGTC TAAAAAGCTT GATATCGAAT
     CTTAACAAAA TCATGGATCT AAATCTACAG ATTTTTCGAA CTATAGCTTA
     E  L  F  *  kan

1701 TCCTGCAGCC CGGGGGATCC ACTAGTTCTA GAGCGGCCGC CACCGCGGTG
     AGGACGTCGG GCCCCCTAGG TGATCAAGAT CTCGCCGGCG GTGGCGCCAC

1751 GAGCTCCAGC TTTTGTTCCC TTTAGTGAGG GTTAATTCCG AGCAGATCTT
     CTCGAGGTCG AAAACAAGGG AAATCACTCC CAATTAAGGC TCGTCTAGAA

1801 CAAATGAAGA CACAAAAGCA GAAATAGCCT ATGCCTTTGA AAATACAGCT
     GTTTACTTCT GTGTTTTCGT CTTTATCGGA TACGGAAACT TTTATGTCGA

1851 TGTGATCATA TCATGGATAA ATTAGAAAAA ATTTTTAATC TTTATAAATT
     ACACTAGTAT AGTACCTATT TAATCTTTTT TAAAAATTAG AAATATTTAA
                  M  D  K   L  E  K   I  F  N  L   Y  K  F Cj1344c

1901 TAAAAATTTT GGCGTTGTAG GTGGAGCTAG TGCAAATCTT AACTTGCGTT
     ATTTTTAAAA CCGCAACATC CACCTCGATC ACGTTTAGAA TTGAACGCAA
      K  N  F   G  V  V  G   G  A  S   A  N  L   N  L  R  S Cj1344c

1951 CGCGTTTGCA AAATTTATGT CAAAAATATA ATGCAAATTT AAAACTAGCT
     GCGCAAACGT TTTAAATACA GTTTTTATAT TACGTTTAAA TTTTGATCGA
       R  L  Q   N  L  C   Q  K  Y   N  A  N   L  K  L  A   Cj1344c
```

```
2001 CCTTTAAAAT TCTGCTCTGA TAATGCTTTG ATGATAGCAA GAGCCGCAGT
     GGAAATTTTA AGACGAGACT ATTACGAAAC TACTATCGTT CTCGGCGTCA
     P  L  K  F   C  S  D   N  A  L   M  I  A  R   A  A  V  Cj1344c

2051 TGATGCTTAT GAAAAAAAGG AATTTGTAAG TGTAGAAGAA GATATTTTAA
     ACTACGAATA CTTTTTTTCC TTAAACATTC ACATCTTCTT CTATAAAATT
      D  A  Y   E  K  K  E   F  V  S   V  E  E   D  I  L  S Cj1344c

2101 GCCCTAAAAA TAAAAATTTT TCAAGGATAT AGATGAAAAA AGCTCGAGAA
     CGGGATTTTT ATTTTTAAAA AGTTCCTATA TCTACTTTTT TCGAGCTCTT
       P  K  N   K  N  F   S  R  I  *    Cj1344c

2151 TCACTAGTGA ATTCGCGGCC GCCTGCAGGT CGACCATATG GGAGAGCTCC
     AGTGATCACT TAAGCGCCGG CGGACGTCCA GCTGGTATAC CCTCTCGAGG

2201 CAACGCGTTG GATGCATAGC TTGAGTATTC TATAGTGTCA CCTAAATAGC
     GTTGCGCAAC CTACGTATCG AACTCATAAG ATATCACAGT GGATTTATCG

2251 TTGGCGTAAT CATGGTCATA GCTGTTTCCT GTGTGAAATT GTTATCCGCT
     AACCGCATTA GTACCAGTAT CGACAAAGGA CACACTTTAA CAATAGGCGA

2301 CACAATTCCA CACAACATAC GAGCCGGAAG CATAAAGTGT AAAGCCTGGG
     GTGTTAAGGT GTGTTGTATG CTCGGCCTTC GTATTTCACA TTTCGGACCC

2351 GTGCCTAATG AGTGAGCTAA CTCACATTAA TTGCGTTGCG CTCACTGCCC
     CACGGATTAC TCACTCGATT GAGTGTAATT AACGCAACGC GAGTGACGGG

2401 GCTTTCCAGT CGGGAAACCT GTCGTGCCAG CTGCATTAAT GAATCGGCCA
     CGAAAGGTCA GCCCTTTGGA CAGCACGGTC GACGTAATTA CTTAGCCGGT

2451 ACGCGCGGGG AGAGGCGGTT TGCGTATTGG GCGCTCTTCC GCTTCCTCGC
     TGCGCGCCCC TCTCCGCCAA ACGCATAACC CGCGAGAAGG CGAAGGAGCG

2501 TCACTGACTC GCTGCGCTCG GTCGTTCGGC TGCGGCGAGC GGTATCAGCT
     AGTGACTGAG CGACGCGAGC CAGCAAGCCG ACGCCGCTCG CCATAGTCGA

2551 CACTCAAAGG CGGTAATACG GTTATCCACA GAATCAGGGG ATAACGCAGG
     GTGAGTTTCC GCCATTATGC CAATAGGTGT CTTAGTCCCC TATTGCGTCC

2601 AAAGAACATG TGAGCAAAAG GCCAGCAAAA GGCCAGGAAC CGTAAAAAGG
     TTTCTTGTAC ACTCGTTTTC CGGTCGTTTT CCGGTCCTTG GCATTTTTCC

2651 CCGCGTTGCT GGCGTTTTTC CATAGGCTCC GCCCCCCTGA CGAGCATCAC
     GGCGCAACGA CCGCAAAAAG GTATCCGAGG CGGGGGGACT GCTCGTAGTG

2701 AAAAATCGAC GCTCAAGTCA GAGGTGGCGA AACCCGACAG GACTATAAAG
     TTTTTAGCTG CGAGTTCAGT CTCCACCGCT TTGGGCTGTC CTGATATTTC

2751 ATACCAGGCG TTTCCCCCTG GAAGCTCCCT CGTGCGCTCT CCTGTTCCGA
     TATGGTCCGC AAAGGGGGAC CTTCGAGGGA GCACGCGAGA GGACAAGGCT

2801 CCCTGCCGCT TACCGGATAC CTGTCCGCCT TTCTCCCTTC GGGAAGCGTG
     GGGACGGCGA ATGGCCTATG GACAGGCGGA AAGAGGGAAG CCCTTCGCAC

2851 GCGCTTTCTC ATAGCTCACG CTGTAGGTAT CTCAGTTCGG TGTAGGTCGT
     CGCGAAAGAG TATCGAGTGC GACATCCATA GAGTCAAGCC ACATCCAGCA

2901 TCGCTCCAAG CTGGGCTGTG TGCACGAACC CCCCGTTCAG CCCGACCGCT
     AGCGAGGTTC GACCCGACAC ACGTGCTTGG GGGGCAAGTC GGGCTGGCGA
```

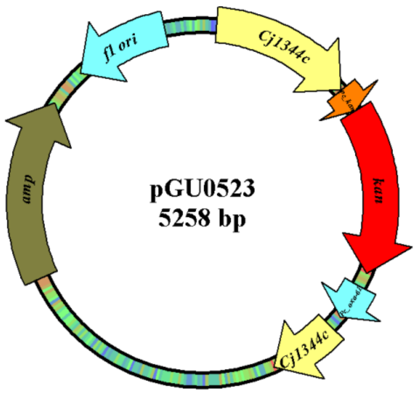
```
2951 GCGCCTTATC CGGTAACTAT CGTCTTGAGT CCAACCCGGT AAGACACGAC
     CGCGGAATAG GCCATTGATA GCAGAACTCA GGTTGGGCCA TTCTGTGCTG

3001 TTATCGCCAC TGGCAGCAGC CACTGGTAAC AGGATTAGCA GAGCGAGGTA
     AATAGCGGTG ACCGTCGTCG GTGACCATTG TCCTAATCGT CTCGCTCCAT

3051 TGTAGGCGGT GCTACAGAGT TCTTGAAGTG GTGGCCTAAC TACGGCTACA
     ACATCCGCCA CGATGTCTCA AGAACTTCAC CACCGGATTG ATGCCGATGT

3101 CTAGAAGAAC AGTATTTGGT ATCTGCGCTC TGCTGAAGCC AGTTACCTTC
     GATCTTCTTG TCATAAACCA TAGACGCGAG ACGACTTCGG TCAATGGAAG

3151 GGAAAAAGAG TTGGTAGCTC TTGATCCGGC AAACAAACCA CCGCTGGTAG
     CCTTTTTCTC AACCATCGAG AACTAGGCCG TTTGTTTGGT GGCGACCATC

3201 CGGTGGTTTT TTTGTTTGCA AGCAGCAGAT TACGCGCAGA AAAAAAGGAT
     GCCACCAAAA AAACAAACGT TCGTCGTCTA ATGCGCGTCT TTTTTTCCTA

3251 CTCAAGAAGA TCCTTTGATC TTTTCTACGG GGTCTGACGC TCAGTGGAAC
     GAGTTCTTCT AGGAAACTAG AAAAGATGCC CCAGACTGCG AGTCACCTTG

3301 GAAAACTCAC GTTAAGGGAT TTTGGTCATG AGATTATCAA AAAGGATCTT
     CTTTTGAGTG CAATTCCCTA AAACCAGTAC TCTAATAGTT TTTCCTAGAA

3351 CACCTAGATC CTTTTAAATT AAAAATGAAG TTTTAAATCA ATCTAAAGTA
     GTGGATCTAG GAAAATTTAA TTTTTACTTC AAAATTTAGT TAGATTTCAT

3401 TATATGAGTA AACTTGGTCT GACAGTTACC AATGCTTAAT CAGTGAGGCA
     ATATACTCAT TTGAACCAGA CTGTCAATGG TTACGAATTA GTCACTCCGT

3451 CCTATCTCAG CGATCTGTCT ATTTCGTTCA TCCATAGTTG CCTGACTCCC
     GGATAGAGTC GCTAGACAGA TAAAGCAAGT AGGTATCAAC GGACTGAGGG

3501 CGTCGTGTAG ATAACTACGA TACGGGAGGG CTTACCATCT GGCCCCAGTG
     GCAGCACATC TATTGATGCT ATGCCCTCCC GAATGGTAGA CCGGGGTCAC

3551 CTGCAATGAT ACCGCGAGAC CCACGCTCAC CGGCTCCAGA TTTATCAGCA
     GACGTTACTA TGGCGCTCTG GGTGCGAGTG GCCGAGGTCT AAATAGTCGT

3601 ATAAACCAGC CAGCCGGAAG GGCCGAGCGC AGAAGTGGTC CTGCAACTTT
     TATTTGGTCG GTCGGCCTTC CCGGCTCGCG TCTTCACCAG GACGTTGAAA

3651 ATCCGCCTCC ATCCAGTCTA TTAATTGTTG CCGGGAAGCT AGAGTAAGTA
     TAGGCGGAGG TAGGTCAGAT AATTAACAAC GGCCCTTCGA TCTCATTCAT

3701 GTTCGCCAGT TAATAGTTTG CGCAACGTTG TTGCCATTGC TACAGGCATC
     CAAGCGGTCA ATTATCAAAC GCGTTGCAAC AACGGTAACG ATGTCCGTAG

3751 GTGGTGTCAC GCTCGTCGTT TGGTATGGCT TCATTCAGCT CCGGTTCCCA
     CACCACAGTG CGAGCAGCAA ACCATACCGA AGTAAGTCGA GGCCAAGGGT

3801 ACGATCAAGG CGAGTTACAT GATCCCCCAT GTTGTGCAAA AAAGCGGTTA
     TGCTAGTTCC GCTCAATGTA CTAGGGGGTA CAACACGTTT TTTCGCCAAT

3851 GCTCCTTCGG TCCTCCGATC GTTGTCAGAA GTAAGTTGGC CGCAGTGTTA
     CGAGGAAGCC AGGAGGCTAG CAACAGTCTT CATTCAACCG GCGTCACAAT

3901 TCACTCATGG TTATGGCAGC ACTGCATAAT TCTCTTACTG TCATGCCATC
     AGTGAGTACC AATACCGTCG TGACGTATTA AGAGAATGAC AGTACGGTAG
```

```
3951 CGTAAGATGC TTTTCTGTGA CTGGTGAGTA CTCAACCAAG TCATTCTGAG
     GCATTCTACG AAAAGACACT GACCACTCAT GAGTTGGTTC AGTAAGACTC

4001 AATAGTGTAT GCGGCGACCG AGTTGCTCTT GCCCGGCGTC AATACGGGAT
     TTATCACATA CGCCGCTGGC TCAACGAGAA CGGGCCGCAG TTATGCCCTA

4051 AATACCGCGC CACATAGCAG AACTTTAAAA GTGCTCATCA TTGGAAAACG
     TTATGGCGCG GTGTATCGTC TTGAAATTTT CACGAGTAGT AACCTTTTGC

4101 TTCTTCGGGG CGAAAACTCT CAAGGATCTT ACCGCTGTTG AGATCCAGTT
     AAGAAGCCCC GCTTTTGAGA GTTCCTAGAA TGGCGACAAC TCTAGGTCAA

4151 CGATGTAACC CACTCGTGCA CCCAACTGAT CTTCAGCATC TTTTACTTTC
     GCTACATTGG GTGAGCACGT GGGTTGACTA GAAGTCGTAG AAAATGAAAG

4201 ACCAGCGTTT CTGGGTGAGC AAAAACAGGA AGGCAAAATG CCGCAAAAAA
     TGGTCGCAAA GACCCACTCG TTTTTGTCCT TCCGTTTTAC GGCGTTTTTT

4251 GGGAATAAGG GCGACACGGA AATGTTGAAT ACTCATACTC TTCCTTTTTC
     CCCTTATTCC CGCTGTGCCT TTACAACTTA TGAGTATGAG AAGGAAAAAG

4301 AATATTATTG AAGCATTTAT CAGGGTTATT GTCTCATGAG CGGATACATA
     TTATAATAAC TTCGTAAATA GTCCCAATAA CAGAGTACTC GCCTATGTAT

4351 TTTGAATGTA TTTAGAAAAA TAAACAAATA GGGGTTCCGC GCACATTTCC
     AAACTTACAT AAATCTTTTT ATTTGTTTAT CCCCAAGGCG CGTGTAAAGG

4401 CCGAAAAGTG CCACCTGATG CGGTGTGAAA TACCGCACAG ATGCGTAAGG
     GGCTTTTCAC GGTGGACTAC GCCACACTTT ATGGCGTGTC TACGCATTCC

4451 AGAAAATACC GCATCAGGAA ATTGTAAGCG TTAATATTTT GTTAAAATTC
     TCTTTTATGG CGTAGTCCTT TAACATTCGC AATTATAAAA CAATTTTAAG

4501 GCGTTAAATT TTTGTTAAAT CAGCTCATTT TTTAACCAAT AGGCCGAAAT
     CGCAATTTAA AAACAATTTA GTCGAGTAAA AAATTGGTTA TCCGGCTTTA

4551 CGGCAAAATC CCTTATAAAT CAAAAGAATA GACCGAGATA GGGTTGAGTG
     GCCGTTTTAG GGAATATTTA GTTTTCTTAT CTGGCTCTAT CCCAACTCAC

4601 TTGTTCCAGT TTGGAACAAG AGTCCACTAT TAAAGAACGT GGACTCCAAC
     AACAAGGTCA AACCTTGTTC TCAGGTGATA ATTTCTTGCA CCTGAGGTTG

4651 GTCAAAGGGC GAAAAACCGT CTATCAGGGC GATGGCCCAC TACGTGAACC
     CAGTTTCCCG CTTTTTGGCA GATAGTCCCG CTACCGGGTG ATGCACTTGG

4701 ATCACCCTAA TCAAGTTTTT TGGGGTCGAG GTGCCGTAAA GCACTAAATC
     TAGTGGGATT AGTTCAAAAA ACCCCAGCTC CACGGCATTT CGTGATTTAG

4751 GGAACCCTAA AGGGAGCCCC CGATTTAGAG CTTGACGGGG AAAGCCGGCG
     CCTTGGGATT TCCCTCGGGG GCTAAATCTC GAACTGCCCC TTTCGGCCGC

4801 AACGTGGCGA GAAAGGAAGG GAAGAAAGCG AAAGGAGCGG GCGCTAGGGC
     TTGCACCGCT CTTTCCTTCC CTTCTTTCGC TTTCCTCGCC CGCGATCCCG

4851 GCTGGCAAGT GTAGCGGTCA CGCTGCGCGT AACCACCACA CCCGCCGCGC
     CGACCGTTCA CATCGCCAGT GCGACGCGCA TTGGTGGTGT GGGCGGCGCG

4901 TTAATGCGCC GCTACAGGGC GCGTCCATTC GCCATTCAGG CTGCGCAACT
     AATTACGCGG CGATGTCCCG CGCAGGTAAG CGGTAAGTCC GACGCGTTGA
```

```
4951 GTTGGGAAGG GCGATCGGTG CGGGCCTCTT CGCTATTACG CCAGCTGGCG
     CAACCCTTCC CGCTAGCCAC GCCCGGAGAA GCGATAATGC GGTCGACCGC

5001 AAAGGGGGAT GTGCTGCAAG GCGATTAAGT TGGGTAACGC CAGGGTTTTC
     TTTCCCCCTA CACGACGTTC CGCTAATTCA ACCCATTGCG GTCCCAAAAG

5051 CCAGTCACGA CGTTGTAAAA CGACGGCCAG TGAATTGTAA TACGACTCAC
     GGTCAGTGCT GCAACATTTT GCTGCCGGTC ACTTAACATT ATGCTGAGTG

5101 TATA
     ATAT
```

Appendix I

pGU0523 (pGU0501ΔCj1344c::Pc_{kan}_kan_Pc_{oxa-61})




```
  1 GGGCGAATTG GGCCCGACGT CGCATGCTCC CGGCCGCCAT GGCGGCCGCG
    CCCGCTTAAC CCGGGCTGCA GCGTACGAGG GCCGGCGGTA CCGCCGGCGC

                   START Cj1344c
 51 GGAATTCGAT TCATATGAAA AATCTTATCC TAGCTATAGA AAGTTCTTGT
    CCTTAAGCTA AGTATACTTT TTAGAATAGG ATCGATATCT TTCAAGAACA
                   M  K   N  L  I  L   A  I  E   S  S  C   Cj1344c

101 GATGATAGTT CTATAGCTAT CATTGATAAA AACACCTTAG AATGTAAATT
    CTACTATCAA GATATCGATA GTAACTATTT TTGTGGAATC TTACATTTAA
    D  D  S  S   I  A  I   I  D  K   N  T  L  E   C  K  F  Cj1344c

151 TCATAAAAAA ATTTCCCAAG AATTAGATCA TAGTATCTAT GGGGGAGTGG
    AGTATTTTTT TAAAGGGTTC TTAATCTAGT ATCATAGATA CCCCCTCACC
     H  K  K   I  S  Q  E   L  D  H   S  I  Y   G  G  V  V Cj1344c

201 TACCTGAACT TGCTGCAAGA CTTCATAGCG AGGCTTTACC AAAGATGCTT
    ATGGACTTGA ACGACGTTCT GAAGTATCGC TCCGAAATGG TTTCTACGAA
      P  E  L   A  A  R   L  H  S  E   A  L  P   K  M  L   Cj1344c

251 AAGCAATGCA AAGAGCATTT TAAAAATCTT TGTGCCATAG CTGTGACAAA
    TTCGTTACGT TTCTCGTAAA ATTTTTAGAA ACACGGTATC GACACTGTTT
    K  Q  C  K   E  H  F   K  N  L   C  A  I  A   V  T  N  Cj1344c

301 TGAACCTGGA CTTAGTGTTT CTTTGCTCAG TGGAATTTCT ATGGCAAAAA
    ACTTGGACCT GAATCACAAA GAAACGAGTC ACCTTAAAGA TACCGTTTTT
     E  P  G   L  S  V  S   L  L  S   G  I  S   M  A  K  T Cj1344c

351 CCTTAGCAAG TGCGCTAAAT TTACCCTTAA TCCCTATAAA TCATCTTAAA
    GGAATCGTTC ACGCGATTTA AATGGGAATT AGGGATATTT AGTAGAATTT
      L  A  S   A  L  N   L  P  L  I   P  I  N   H  L  K   Cj1344c
```

```
 401 GGTCATATTT ATAGTCTTTT TTTGGAAGAA AAAATTTCTT TAGATATGGG
     CCAGTATAAA TATCAGAAAA AAACCTTCTT TTTTAAAGAA ATCTATACCC
     G  H  I  Y   S  L  F   L  E  E   K  I  S  L   D  M  G  Cj1344c

 451 AATTTTGCTT GTTAGTGGTG GGCATACCAT GGTGCTTTAT CTTAAAGATG
     TTAAAACGAA CAATCACCAC CCGTATGGTA CCACGAAATA GAATTTCTAC
      I  L  L   V  S  G  G   H  T  M   V  L  Y   L  K  D  D Cj1344c

 501 ATGCAAGCTT AGAGCTTTTA GCAAGTACAA ATGATGATAG CTTTGGAGAA
     TACGTTCGAA TCTCGAAAAT CGTTCATGTT TACTACTATC GAAACCTCTT
       A  S  L   E  L  L   A  S  T  N   D  D  S   F  G  E   Cj1344c

 551 AGTTTTGATA AAGTGGCTAA AATGATGAAT TTAGGTTACC CTGGTGGGGT
     TCAAAACTAT TTCACCGATT TTACTACTTA AATCCAATGG GACCACCCCA
     S  F  D  K   V  A  K   M  M  N   L  G  Y  P   G  G  V  Cj1344c

 601 CATCATAGAA AATTTAGCAA AAAATGCCAA ACTTAAAAAT ATCTCTTTTA
     GTAGTATCTT TTAAATCGTT TTTTACGGTT TGAATTTTTA TAGAGAAAAT
      I  I  E   N  L  A  K   N  A  K   L  K  N   I  S  F  N Cj1344c

 651 ACACACCTTT AAAGCATTCT AAAGAACTCG CTTTCAGTTT TTCAGGGCTT
     TGTGTGGAAA TTTCGTAAGA TTTCTTGAGC GAAAGTCAAA AAGTCCCGAA
       T  P  L   K  H  S   K  E  L  A   F  S  F   S  G  L   Cj1344c

                                                  Pc_kan→
 701 AAAAATGCAG TGCGTTTGGA AATTTTAAAA CATGAAAATT TTCTGTTTTC
     TTTTTACGTC ACGCAAACCT TTAAAATTTT GTACTTTTAA AAGACAAAAG
     K  N  A  V   R  L  E   I  L  K   H  E  N  F

 751 TGGTATTTAA GGTTTTAGAA TGCAAGGAAC AGTGAATTGG AGTTCGTCTT
     ACCATAAATT CCAAAATCTT ACGTTCCTTG TCACTTAACC TCAAGCAGAA

 801 GTTATAATTA GCTTCTTGGG GTATCTTTAA ATACTGTAGA AAAGAGGAAG
     CAATATTAAT CGAAGAACCC CATAGAAATT TATGACATCT TTTCTCCTTC

                START kan
 851 GAAATAATAA ATGGCTAAAA TGAGAATATC ACCGGAATTG AAAAAACTGA
     CTTTATTATT TACCGATTTT ACTCTTATAG TGGCCTTAAC TTTTTTGACT
                M  A  K  M   R  I  S   P  E  L   K  K  L  I kan

 901 TCGAAAAATA CCGCTGCGTA AAAGATACGG AAGGAATGTC TCCTGCTAAG
     AGCTTTTTAT GGCGACGCAT TTTCTATGCC TTCCTTACAG AGGACGATTC
       E  K  Y   R  C  V   K  D  T  E   G  M  S   P  A  K   kan

 951 GTATATAAGC TGGTGGGAGA AAATGAAAAC CTATATTTAA AAATGACGGA
     CATATATTCG ACCACCCTCT TTTACTTTTG GATATAAATT TTTACTGCCT
     V  Y  K  L   V  G  E   N  E  N   L  Y  L  K   M  T  D  kan

1001 CAGCCGGTAT AAAGGGACCA CCTATGATGT GGAACGGGAA AAGGACATGA
     GTCGGCCATA TTTCCCTGGT GGATACTACA CCTTGCCCTT TTCCTGTACT
      S  R  Y   K  G  T  T   Y  D  V   E  R  E   K  D  M  M kan

1051 TGCTATGGCT GGAAGGAAAG CTGCCTGTTC CAAAGGTCCT GCACTTTGAA
     ACGATACCGA CCTTCCTTTC GACGGACAAG GTTTCCAGGA CGTGAAACTT
       L  W  L   E  G  K   L  P  V  P   K  V  L   H  F  E   kan

1101 CGGCATGATG GCTGGAGCAA TCTGCTCATG AGTGAGGCCG ATGGCGTCCT
     GCCGTACTAC CGACCTCGTT AGACGAGTAC TCACTCCGGC TACCGCAGGA
     R  H  D  G   W  S  N   L  L  M   S  E  A  D   G  V  L  kan
```

```
1151 TTGCTCGGAA GAGTATGAAG ATGAACAAAG CCCTGAAAAG ATTATCGAGC
     AACGAGCCTT CTCATACTTC TACTTGTTTC GGGACTTTTC TAATAGCTCG
      C  S  E   E  Y  E  D   E  Q  S   P  E  K   I  I  E  L kan

1201 TGTATGCGGA GTGCATCAGG CTCTTTCACT CCATCGACAT ATCGGATTGT
     ACATACGCCT CACGTAGTCC GAGAAAGTGA GGTAGCTGTA TAGCCTAACA
       Y  A  E   C  I  R   L  F  H  S   I  D  I   S  D  C   kan

1251 CCCTATACGA ATAGCTTAGA CAGCCGCTTA GCCGAATTGG ATTACTTACT
     GGGATATGCT TATCGAATCT GTCGGCGAAT CGGCTTAACC TAATGAATGA
     P  Y  T  N   S  L  D   S  R  L   A  E  L  D   Y  L  L  kan

1301 GAATAACGAT CTGGCCGATG TGGATTGCGA AAACTGGGAA GAAGACACTC
     CTTATTGCTA GACCGGCTAC ACCTAACGCT TTTGACCCTT CTTCTGTGAG
      N  N  D   L  A  D  V   D  C  E   N  W  E   E  D  T  P kan

1351 CATTTAAAGA TCCGCGCGAG CTGTATGATT TTTTAAAGAC GGAAAAGCCC
     GTAAATTTCT AGGCGCGCTC GACATACTAA AAAATTTCTG CCTTTTCGGG
       F  K  D   P  R  E   L  Y  D  F   L  K  T   E  K  P   kan

1401 GAAGAGGAAC TTGTCTTTTC CCACGGCGAC CTGGGAGACA GCAACATCTT
     CTTCTCCTTG AACAGAAAAG GGTGCCGCTG GACCCTCTGT CGTTGTAGAA
     E  E  E  L   V  F  S   H  G  D   L  G  D  S   N  I  F  kan

1451 TGTGAAAGAT GGCAAAGTAA GTGGCTTTAT TGATCTTGGG AGAAGCGGCA
     ACACTTTCTA CCGTTTCATT CACCGAAATA ACTAGAACCC TCTTCGCCGT
      V  K  D   G  K  V  S   G  F  I   D  L  G   R  S  G  R kan

1501 GGGCGGACAA GTGGTATGAC ATTGCCTTCT GCGTCCGGTC GATCAGGGAG
     CCCGCCTGTT CACCATACTG TAACGGAAGA CGCAGGCCAG CTAGTCCCTC
       A  D  K   W  Y  D   I  A  F  C   V  R  S   I  R  E   kan

1551 GATATCGGGG AAGAACAGTA TGTCGAGCTA TTTTTTGACT TACTGGGGAT
     CTATAGCCCC TTCTTGTCAT ACAGCTCGAT AAAAAACTGA ATGACCCCTA
     D  I  G  E   E  Q  Y   V  E  L   F  F  D  L   L  G  I   kan

1601 CAAGCCTGAT TGGGAGAAAA TAAAATATTA TATTTTACTG GATGAATTGT
     GTTCGGACTA ACCCTCTTTT ATTTTATAAT ATAAAATGAC CTACTTAACA
      K  P  D   W  E  K  I   K  Y  Y   I  L  L   D  E  L  F kan

1651 TTTAGTACCT AGATTTAGAT GTCTAAAAAG CTTGATATCG AATTCCTGCA
     AAATCATGGA TCTAAATCTA CAGATTTTTC GAACTATAGC TTAGGACGT
       *  kan
                          START Pc_oxa61
1701 GCCCGGGGGA TCCACTAGTT CTAGCTAGAC TTGATATCGA ATTCCTGCAG
     CGGGCCCCCT AGGTGATCAA GATCGATCTG AACTATAGCT TAAGGACGTC

1751 CCCGGGGGAT CCATCGATGG ATTGCTTTAA TGGTTACAAT TTTAGAAAAT
     GGGCCCCCTA GGTAGCTACC TAACGAAATT ACCAATGTTA AAATCTTTTA

1801 TTCAGTTCAT GTTAAAGGAT ATTTAAATAA AATAAAGCTT TAAAAAGTAT
     AAGTCAAGTA CAATTTCCTA TAAATTTATT TTATTTCGAA ATTTTTCATA

1851 TTTGTTTAAA ATTATTTAAA TAGAAAGATA TTTTGTGTCT AGAGCGGCCG
     AAACAAATTT TAATAAATTT ATCTTTCTAT AAAACACAGA TCTCGCCGGC

1901 CCACCGCGGT GGAGCTCCAG CTTTTGTTCC CTTTAGTGAG GGTTAATTCC
     GGTGGCGCCA CCTCGAGGTC GAAAACAAGG GAAATCACTC CCAATTAAGG
```

split *Cj1344c*

1951 GAGCA**AAA**TG AAGACACAAA AGCAGAAATA GCCTATGCCT TTGAAAATAC
CTCGTTTTAC TTCTGTGTTT TCGTCTTTAT CGGATACGGA AACTTTTATG

2001 AGCTTGTGAT CATATCATGG ATAAATTAGA AAAAATTTTT AATCTTTATA
TCGAACACTA GTATAGTACC TATTTAATCT TTTTTAAAAA TTAGAAATAT

2051 AATTTAAAAA TTTTGGCGTT GTAGGTGGAG CTAGTGCAAA TCTTAACTTG
TTAAATTTTT AAAACCGCAA CATCCACCTC GATCACGTTT AGAATTGAAC

2101 CGTTCGCGTT TGCAAAATTT ATGTCAAAAA TATAATGCAA ATTTAAAACT
GCAAGCGCAA ACGTTTTAAA TACAGTTTTT ATATTACGTT TAAATTTTGA

2151 AGCTCCTTTA AAATTCTGCT CTGATAATGC TTTGATGATA GCAAGAGCCG
TCGAGGAAAT TTTAAGACGA GACTATTACG AAACTACTAT CGTTCTCGGC

2201 CAGTTGATGC TTATGAAAAA AAGGAATTTG TAAGTGTAGA AGAAGATATT
GTCAACTACG AATACTTTTT TTCCTTAAAC ATTCACATCT TCTTCTATAA

2251 TTAAGCCCTA AAAATAAAAA TTTTTCAAGG ATATAGATGA AAAAAGCTCG
AATTCGGGAT TTTTATTTTT AAAAAGTTCC TATATCTACT TTTTTCGAGC

2301 AGAATCACTA GTGAATTCGC GGCCGCCTGC AGGTCGACCA TATGGGAGAG
TCTTAGTGAT CACTTAAGCG CCGGCGGACG TCCAGCTGGT ATACCCTCTC

2351 CTCCCAACGC GTTGGATGCA TAGCTTGAGT ATTCTATAGT GTCACCTAAA
GAGGGTTGCG CAACCTACGT ATCGAACTCA TAAGATATCA CAGTGGATTT

2401 TAGCTTGGCG TAATCATGGT CATAGCTGTT TCCTGTGTGA AATTGTTATC
ATCGAACCGC ATTAGTACCA GTATCGACAA AGGACACACT TTAACAATAG

2451 CGCTCACAAT TCCACACAAC ATACGAGCCG GAAGCATAAA GTGTAAAGCC
GCGAGTGTTA AGGTGTGTTG TATGCTCGGC CTTCGTATTT CACATTTCGG

2501 TGGGGTGCCT AATGAGTGAG CTAACTCACA TTAATTGCGT TGCGCTCACT
ACCCCACGGA TTACTCACTC GATTGAGTGT AATTAACGCA ACGCGAGTGA

2551 GCCCGCTTTC CAGTCGGGAA ACCTGTCGTG CCAGCTGCAT TAATGAATCG
CGGGCGAAAG GTCAGCCCTT TGGACAGCAC GGTCGACGTA ATTACTTAGC

2601 GCCAACGCGC GGGGAGAGGC GGTTTGCGTA TTGGGCGCTC TTCCGCTTCC
CGGTTGCGCG CCCCTCTCCG CCAAACGCAT AACCCGCGAG AAGGCGAAGG

2651 TCGCTCACTG ACTCGCTGCG CTCGGTCGTT CGGCTGCGGC GAGCGGTATC
AGCGAGTGAC TGAGCGACGC GAGCCAGCAA GCCGACGCCG CTCGCCATAG

2701 AGCTCACTCA AAGGCGGTAA TACGGTTATC CACAGAATCA GGGGATAACG
TCGAGTGAGT TTCCGCCATT ATGCCAATAG GTGTCTTAGT CCCCTATTGC

2751 CAGGAAAGAA CATGTGAGCA AAAGGCCAGC AAAAGGCCAG GAACCGTAAA
GTCCTTTCTT GTACACTCGT TTTCCGGTCG TTTTCCGGTC CTTGGCATTT

2801 AAGGCCGCGT TGCTGGCGTT TTTCCATAGG CTCCGCCCCC CTGACGAGCA
TTCCGGCGCA ACGACCGCAA AAAGGTATCC GAGGCGGGGG GACTGCTCGT

2851 TCACAAAAAT CGACGCTCAA GTCAGAGGTG GCGAAACCCG ACAGGACTAT
AGTGTTTTTA GCTGCGAGTT CAGTCTCCAC CGCTTTGGGC TGTCCTGATA

2901 AAAGATACCA GGCGTTTCCC CCTGGAAGCT CCCTCGTGCG CTCTCCTGTT
TTTCTATGGT CCGCAAAGGG GGACCTTCGA GGGAGCACGC GAGAGGACAA

```
2951 CCGACCCTGC CGCTTACCGG ATACCTGTCC GCCTTTCTCC CTTCGGGAAG
     GGCTGGGACG GCGAATGGCC TATGGACAGG CGGAAAGAGG GAAGCCCTTC

3001 CGTGGCGCTT TCTCATAGCT CACGCTGTAG GTATCTCAGT TCGGTGTAGG
     GCACCGCGAA AGAGTATCGA GTGCGACATC CATAGAGTCA AGCCACATCC

3051 TCGTTCGCTC CAAGCTGGGC TGTGTGCACG AACCCCCCGT TCAGCCCGAC
     AGCAAGCGAG GTTCGACCCG ACACACGTGC TTGGGGGGCA AGTCGGGCTG

3101 CGCTGCGCCT TATCCGGTAA CTATCGTCTT GAGTCCAACC CGGTAAGACA
     GCGACGCGGA ATAGGCCATT GATAGCAGAA CTCAGGTTGG GCCATTCTGT

3151 CGACTTATCG CCACTGGCAG CAGCCACTGG TAACAGGATT AGCAGAGCGA
     GCTGAATAGC GGTGACCGTC GTCGGTGACC ATTGTCCTAA TCGTCTCGCT

3201 GGTATGTAGG CGGTGCTACA GAGTTCTTGA AGTGGTGGCC TAACTACGGC
     CCATACATCC GCCACGATGT CTCAAGAACT TCACCACCGG ATTGATGCCG

3251 TACACTAGAA GAACAGTATT TGGTATCTGC GCTCTGCTGA AGCCAGTTAC
     ATGTGATCTT CTTGTCATAA ACCATAGACG CGAGACGACT TCGGTCAATG

3301 CTTCGGAAAA AGAGTTGGTA GCTCTTGATC CGGCAAACAA ACCACCGCTG
     GAAGCCTTTT TCTCAACCAT CGAGAACTAG GCCGTTTGTT TGGTGGCGAC

3351 GTAGCGGTGG TTTTTTTGTT TGCAAGCAGC AGATTACGCG CAGAAAAAAA
     CATCGCCACC AAAAAAACAA ACGTTCGTCG TCTAATGCGC GTCTTTTTTT

3401 GGATCTCAAG AAGATCCTTT GATCTTTTCT ACGGGGTCTG ACGCTCAGTG
     CCTAGAGTTC TTCTAGGAAA CTAGAAAAGA TGCCCCAGAC TGCGAGTCAC

3451 GAACGAAAAC TCACGTTAAG GGATTTTGGT CATGAGATTA TCAAAAAGGA
     CTTGCTTTTG AGTGCAATTC CCTAAAACCA GTACTCTAAT AGTTTTTCCT

3501 TCTTCACCTA GATCCTTTTA AATTAAAAAT GAAGTTTTAA ATCAATCTAA
     AGAAGTGGAT CTAGGAAAAT TTAATTTTTA CTTCAAAATT TAGTTAGATT

3551 AGTATATATG AGTAAACTTG GTCTGACAGT TACCAATGCT TAATCAGTGA
     TCATATATAC TCATTTGAAC CAGACTGTCA ATGGTTACGA ATTAGTCACT

3601 GGCACCTATC TCAGCGATCT GTCTATTTCG TTCATCCATA GTTGCCTGAC
     CCGTGGATAG AGTCGCTAGA CAGATAAAGC AAGTAGGTAT CAACGGACTG

3651 TCCCCGTCGT GTAGATAACT ACGATACGGG AGGGCTTACC ATCTGGCCCC
     AGGGGCAGCA CATCTATTGA TGCTATGCCC TCCCGAATGG TAGACCGGGG

3701 AGTGCTGCAA TGATACCGCG AGACCCACGC TCACCGGCTC CAGATTTATC
     TCACGACGTT ACTATGGCGC TCTGGGTGCG AGTGGCCGAG GTCTAAATAG

3751 AGCAATAAAC CAGCCAGCCG GAAGGGCCGA GCGCAGAAGT GGTCCTGCAA
     TCGTTATTTG GTCGGTCGGC CTTCCCGGCT CGCGTCTTCA CCAGGACGTT

3801 CTTTATCCGC CTCCATCCAG TCTATTAATT GTTGCCGGGA AGCTAGAGTA
     GAAATAGGCG GAGGTAGGTC AGATAATTAA CAACGGCCCT TCGATCTCAT

3851 AGTAGTTCGC CAGTTAATAG TTTGCGCAAC GTTGTTGCCA TTGCTACAGG
     TCATCAAGCG GTCAATTATC AAACGCGTTG CAACAACGGT AACGATGTCC

3901 CATCGTGGTG TCACGCTCGT CGTTTGGTAT GGCTTCATTC AGCTCCGGTT
     GTAGCACCAC AGTGCGAGCA GCAAACCATA CCGAAGTAAG TCGAGGCCAA
```

```
3951 CCCAACGATC AAGGCGAGTT ACATGATCCC CCATGTTGTG CAAAAAAGCG
     GGGTTGCTAG TTCCGCTCAA TGTACTAGGG GGTACAACAC GTTTTTTCGC

4001 GTTAGCTCCT TCGGTCCTCC GATCGTTGTC AGAAGTAAGT TGGCCGCAGT
     CAATCGAGGA AGCCAGGAGG CTAGCAACAG TCTTCATTCA ACCGGCGTCA

4051 GTTATCACTC ATGGTTATGG CAGCACTGCA TAATTCTCTT ACTGTCATGC
     CAATAGTGAG TACCAATACC GTCGTGACGT ATTAAGAGAA TGACAGTACG

4101 CATCCGTAAG ATGCTTTTCT GTGACTGGTG AGTACTCAAC CAAGTCATTC
     GTAGGCATTC TACGAAAAGA CACTGACCAC TCATGAGTTG GTTCAGTAAG

4151 TGAGAATAGT GTATGCGGCG ACCGAGTTGC TCTTGCCCGG CGTCAATACG
     ACTCTTATCA CATACGCCGC TGGCTCAACG AGAACGGGCC GCAGTTATGC

4201 GGATAATACC GCGCCACATA GCAGAACTTT AAAAGTGCTC ATCATTGGAA
     CCTATTATGG CGCGGTGTAT CGTCTTGAAA TTTTCACGAG TAGTAACCTT

4251 AACGTTCTTC GGGGCGAAAA CTCTCAAGGA TCTTACCGCT GTTGAGATCC
     TTGCAAGAAG CCCCGCTTTT GAGAGTTCCT AGAATGGCGA CAACTCTAGG

4301 AGTTCGATGT AACCCACTCG TGCACCCAAC TGATCTTCAG CATCTTTTAC
     TCAAGCTACA TTGGGTGAGC ACGTGGGTTG ACTAGAAGTC GTAGAAAATG

4351 TTTCACCAGC GTTTCTGGGT GAGCAAAAAC AGGAAGGCAA AATGCCGCAA
     AAAGTGGTCG CAAAGACCCA CTCGTTTTTG TCCTTCCGTT TTACGGCGTT

4401 AAAAGGGAAT AAGGGCGACA CGGAAATGTT GAATACTCAT ACTCTTCCTT
     TTTTCCCTTA TTCCCGCTGT GCCTTTACAA CTTATGAGTA TGAGAAGGAA

4451 TTTCAATATT ATTGAAGCAT TTATCAGGGT TATTGTCTCA TGAGCGGATA
     AAAGTTATAA TAACTTCGTA AATAGTCCCA ATAACAGAGT ACTCGCCTAT

4501 CATATTTGAA TGTATTTAGA AAAATAAACA AATAGGGGTT CCGCGCACAT
     GTATAAACTT ACATAAATCT TTTTATTTGT TTATCCCCAA GGCGCGTGTA

4551 TTCCCCGAAA AGTGCCACCT GATGCGGTGT GAAATACCGC ACAGATGCGT
     AAGGGGCTTT TCACGGTGGA CTACGCCACA CTTTATGGCG TGTCTACGCA

4601 AAGGAGAAAA TACCGCATCA GGAAATTGTA AGCGTTAATA TTTTGTTAAA
     TTCCTCTTTT ATGGCGTAGT CCTTTAACAT TCGCAATTAT AAAACAATTT

4651 ATTCGCGTTA AATTTTTGTT AAATCAGCTC ATTTTTTAAC CAATAGGCCG
     TAAGCGCAAT TTAAAAACAA TTTAGTCGAG TAAAAAATTG GTTATCCGGC

4701 AAATCGGCAA AATCCCTTAT AAATCAAAAG AATAGACCGA GATAGGGTTG
     TTTAGCCGTT TTAGGGAATA TTTAGTTTTC TTATCTGGCT CTATCCCAAC

4751 AGTGTTGTTC CAGTTTGGAA CAAGAGTCCA CTATTAAAGA ACGTGGACTC
     TCACAACAAG GTCAAACCTT GTTCTCAGGT GATAATTTCT TGCACCTGAG

4801 CAACGTCAAA GGGCGAAAAA CCGTCTATCA GGGCGATGGC CCACTACGTG
     GTTGCAGTTT CCCGCTTTTT GGCAGATAGT CCCGCTACCG GGTGATGCAC

4851 AACCATCACC CTAATCAAGT TTTTTGGGGT CGAGGTGCCG TAAAGCACTA
     TTGGTAGTGG GATTAGTTCA AAAAACCCCA GCTCCACGGC ATTTCGTGAT

4901 AATCGGAACC CTAAAGGGAG CCCCCGATTT AGAGCTTGAC GGGGAAAGCC
     TTAGCCTTGG GATTTCCCTC GGGGGCTAAA TCTCGAACTG CCCCTTTCGG
```

```
4951 GGCGAACGTG GCGAGAAAGG AAGGGAAGAA AGCGAAAGGA GCGGGCGCTA
     CCGCTTGCAC CGCTCTTTCC TTCCCTTCTT TCGCTTTCCT CGCCCGCGAT

5001 GGGCGCTGGC AAGTGTAGCG GTCACGCTGC GCGTAACCAC CACACCCGCC
     CCCGCGACCG TTCACATCGC CAGTGCGACG CGCATTGGTG GTGTGGGCGG

5051 GCGCTTAATG CGCCGCTACA GGGCGCGTCC ATTCGCCATT CAGGCTGCGC
     CGCGAATTAC GCGGCGATGT CCCGCGCAGG TAAGCGGTAA GTCCGACGCG

5101 AACTGTTGGG AAGGGCGATC GGTGCGGGCC TCTTCGCTAT TACGCCAGCT
     TTGACAACCC TTCCCGCTAG CCACGCCCGG AGAAGCGATA ATGCGGTCGA

5151 GGCGAAAGGG GGATGTGCTG CAAGGCGATT AAGTTGGGTA ACGCCAGGGT
     CCGCTTTCCC CCTACACGAC GTTCCGCTAA TTCAACCCAT TGCGGTCCCA

5201 TTTCCCAGTC ACGACGTTGT AAAACGACGG CCAGTGAATT GTAATACGAC
     AAAGGGTCAG TGCTGCAACA TTTTGCTGCC GGTCACTTAA CATTATGCTG

5251 TCACTATA
     AGTGATAT
```

References

Abdullah, K. M., R. Y. C. Lo & A. Mellors, (1991) Cloning, Nucleotide Sequence, and Expression of the *Pasteurella haemolytica* A1 Glycoprotease Gene. *Journal of Bacteriology* **173**: 5597-5603.

Abdullah, K. M., E. A. Udoh, P. E. Shewen & A. Mellors, (1992) A neutral glycoprotease of *Pasteurella haemolytica* A1 specifically cleaves O-sialoglycoproteins. *Infection and Immunity* **60**: 56-62.

Adachi, JavierÂ A., L. Ostroskyâ€ • Zeichner, HerbertÂ L. DuPont & CharlesÂ D. Ericsson, (2000) Empirical Antimicrobial Therapy for Traveler's Diarrhea. *Clinical Infectious Diseases* **31**: 1079-1083.

Albert, M. J., A. S. G. Faruque, S. M. Faruque, R. B. Sack & D. Mahalanabis, (1999) Case-Control Study of Enteropathogens Associated with Childhood Diarrhea in Dhaka, Bangladesh. *Journal of Clinical Microbiology* **37**: 3458-3464.

Alfredson, D. A. & V. Korolik, (2003) Sequence analysis of a cryptic plasmid pCJ419 from *Campylobacter jejuni* and construction of an *Escherichia coli-Campylobacter* shuttle vector. *Plasmid* **50**: 152-160.

Alfredson, D. A. & V. Korolik, (2005) Isolation and Expression of a Novel Molecular Class D ß-Lactamase, OXA-61, from *Campylobacter jejuni*. *Antimicrob. Agents Chemother.* **49**: 2515-2518.

Allerberger, F., N. Al-Jazrawi, P. Kreidl, M. P. Dierich, G. Feierl, I. Hein & M. Wagner, (2003) Barbecued Chicken Causing a Multi-State Outbreak of *Campylobacter jejuni* Enteritis. *Infection* **31**: 19-23.

Altekruse, S. F., N. J. Stern, P. I. Fields & D. L. Swerdlow, (1999) *Campylobacter jejuni* - An Emerging Foodborne Pathogen. *Emerging Infectious Diseases* **5**: 28-35.

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

Ang, W. C., B. C. Jacobs & J. D. Laman, (2003) The Guillain-Barre syndrome: a true case of molecular mimicry.

Apicella, M. A., (2010) Personal communication.

Ashbolt, N. J., (2004) Microbial contamination of drinking water and disease outcomes in developing regions. *Toxicology* **198**: 229-238.

Baqar, S., L. A. Applebee & A. L. Bourgeois, (1995a) Immunogenicity and protective efficacy of a prototype *Campylobacter* killed whole-cell vaccine in mice. *Infect. Immun.* **63**: 3731-3735.

Baqar, S., L. A. Applebee, T. C. Gilliland, Jr., L. H. Lee, C. K. Porter & P. Guerry, (2008) Immunogenicity and Protective Efficacy of Recombinant *Campylobacter jejuni* Flagellum-Secreted Proteins in Mice. *Infect. Immun.* **76**: 3170-3175.

Baqar, S., A. L. Bourgeois, P. J. Schultheiss, R. I. Walker, D. M. Rollins, R. L. Haberberger & O. R. Pavlovskis, (1995b) Safety and immunogenicity of a prototype oral whole-cell killed *Campylobacter* vaccine administered with a mucosal adjuvant in non-human primates. *Vaccine* **13**: 22-28.

Biswas, D., K. Itoh & C. Sasakawa, (2003) Role of Microfilaments and Microtubules in the Invasion of INT 407 Cells by *Campylobacter jejuni*. *Microbiology and Immunology* **47**: 469-473.

Black, R., (1990) Epidemiology of travelers' diarrhea and relative importance of various pathogens. *Rev. Infect. Dis.* **12**: 73-79.

Black, R. E., M. M. Levine, M. L. Clements, T. P. Hughes & M. J. Blaser, (1988) Experimental Campylobacter jejuni infection in humans. *J. Infect. Dis.* **157**: 472-479.

Blaser, M. J., R. E. Black, D. J. Duncan & A. J, (1985) *Campylobacter jejuni*-specific serum antibodies are elevated in healthy Bangladeshi children. *J Clin Microbiol* **21**: 164–167.

Blaser, M. J. & J. Engberg, (2008) Clinical Aspects of *Campylobacter jejuni* and *Campylobacter coli* Infections. In: *Campylobacter,* 3rd ed. I. Nachamkin, C. Szymanski & M. J. Blaser (eds). Washington, DC: ASM Press, pp. 99-121.

Bleumink-Pluym, N., F. Verschoor, W. Gaastra, B. van der Zeijst & B. Fry, (1999) A novel approach for the construction of a *Campylobacter* mutant library. *Microbiology* **145**: 2145-2151.

Blixt, O., S. Head, T. Mondala, C. Scanlan, M. E. Huflejt, R. Alvarez, M. C. Bryan, F. Fazio, D. Calarese, J. Stevens, N. Razi, D. J. Stevens, J. J. Skehel, I. van Die, D. R. Burton, I. A. Wilson, R. Cummings, N. Bovin, C.-H. Wong & J. C. Paulson, (2004) Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 17033-17038.

Bolton, F. J., D. Coates, D. N. Hutchinson & A. F. Godfree, (1987) A study of thermophilic campylobacters in a river system. *Journal of Applied Microbiology* **62**: 167-176.

Bovill, R. A. & B. M. Mackey, (1997) Resuscitation of 'non-culturable' cells from aged cultures of *Campylobacter jejuni*. *Microbiology* **143**: 1575-1581.

Brieseman, M., (1990) A further study of the epidemiology of *Campylobacter jejuni* infections. *N Z Med J.* **103**: 207-209.

Buckley, A. M., J. Wang, D. L. Hudson, A. J. Grant, M. A. Jones, D. J. Maskell & M. P. Stevens, (2010) Evaluation of live-attenuated *Salmonella* vaccines expressing *Campylobacter* antigens for control of *C. jejuni* in poultry. *Vaccine* **28**: 1094-1105.

Burr, D. H., D. Rollins, L. H. Lee, D. L. Pattarini, S. S. Walz, J.-H. Tian, J. L. Pace, A. L. Bourgeois & R. I. Walker, (2005) Prevention of disease in ferrets fed an inactivated whole cell *Campylobacter jejuni* vaccine. *Vaccine* **23**: 4315-4321.

Butzler, J.-P., Y. Glupczynski & H. Goossens, (1992) *Campylobacter* and *Helicobacter* infections. *Current Opinion in Infectious Diseases* **5**: 80-87.

Butzler, J. & J. Oosterom, (1991) *Campylobacter*: pathogenicity and significance in foods. *International Journal of Food Microbiology* **12**: 1-8.

Butzler, J. P., (2004) *Campylobacter*, from obscurity to celebrity. *Clinical Microbiology and Infection* **10**: 868-876.

Castelli, F. & G. Carosi, (1995) Epidemiology of traveler's diarrhea. *Chemotherapy* **41**: 20-32.

Cladman, W. M., M.-A. V. Watt, J.-P. Dini & A. Mellors, (1996) The *Pasteurella haemolytica O*-Sialoglycoprotein Endopeptidase is Inhibited by Zinc Ions and Does Not Cleave Fetuin. *Biochemical and Biophysical Research Communications* **220**: 141-146.

Coburn, B., I. Sekirov & B. B. Finlay, (2007) Type III Secretion Systems and Disease. *Clin. Microbiol. Rev.* **20**: 535-549.

Coker, A. O., R. D. Isokpehi, B. N. Thomas, K. O. Amisu & L. C. Obi, (2002) Human Campylobacteriosis in Developing Countries. *Emerging Infectious Diseases* **8**: 237-243.

Crushell, E., S. Harty, F. Sharif & B. Bourke, (2004) Enteric Campylobacter: Purging Its Secrets? *Pediatr Res* **55**: 3-12.

Day, C., G. Tiralongo, R. Hartnell, C. Logue, J. Wilson, M. von Itzstein & V. Korolik, (2009) Differential carbohydrate recognition by *Campylobacter jejuni* strain 11168: Influences of temperature and growth conditions. *PLos One*.

De Melo, M. A. & J.-C. Pechère, (1988) Effect of mucin on Campylobacter jejuni association and invasion on HEp-2 cells. *Microbial Pathogenesis* **5**: 71-76.

Dekeyser, P., M. Gossuin-Detrain, J. Butzler & J. Sternon, (1972) Acute enteritis due to related vibrio: first positive stool cultures. *J Infect Dis* **125**: 390-392.

Dekker, J., J. W. A. Rossen, H. A. Büller & A. W. C. Einerhand, (2002) The MUC family: an obituary. *TRENDS in Biochemical Sciences* **27**: 126-131.

Dingle, K. E., N. van den Braak, F. M. Colles, L. J. Price, D. L. Woodward, F. M. Rodgers, H. P. Endtz, A. van Belkum & M. C. J. Maiden, (2001) Sequence Typing Confirms that Campylobacter jejuni Strains Associated with Guillain-Barre and Miller-Fisher Syndromes Are of Diverse Genetic Lineage, Serotype and Flagella Type. *Journal of Clinical Microbiology* **39**: 3346-3349.

Emanuelsson, O., S. Brunak, G. von Heijne & H. Nielsen, (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protocols* **2**: 953-971.

Ermolova, N. V., M. Ann Cushman, T. Taybi, S. A. Condon, J. C. Cushman & R. Chollet, (2003) Expression, purification, and initial characterization of a recombinant form of plant PEP-carboxylase kinase from CAM-induced *Mesembryanthemum crystallinum* with enhanced solubility in *Escherichia coli*. *Protein Expression and Purification* **29**: 123-131.

Everest, P. H., H. Goossens, J. P. Butzler, D. Lloyd, S. Knutton, J. M. Ketley & H. Williams P, (1992) Differentiated Caco-2 cells as a model for enteric invasion by *Campylobacter jejuni* and *C. coli*. *Journal of Medical Microbiology* **37**: 319-325.

Fouts, D. E., E. F. Mongodin, R. E. Mandrell, W. G. Miller, D. A. Rasko, J. Ravel, L. M. Brinkac, R. T. DeBoy, C. T. Parker, S. C. Daugherty, R. J. Dodson, A. S. Durkin, R. Madupu, S. A. Sullivan, J. U. Shetty, M. A. Ayodeji, A. Shvartsbeyn, M. C. Schatz, J. H. Badger, C. M. Fraser & K. E. Nelson, (2005) Major Structural Differences and Novel Potential Virulence Mechanisms from the Genomes of Multiple *Campylobacter* Species. *PLOS Biology* **3**: e15.

Friedman, C. R., J. Neimann, H. C. Wegner & R. V. Tauxe, (2000) Epidemiology of *Campylobacter jejuni* Infections in the United States and Other Industrialized Nations. In: Campylobacter. I. Nachamkin & M. J. Blaser (eds). Washington: ASM Press, pp. 121-138.

Fry, B., V. Korolik, J. ten Brinke, M. Pennings, R. Zalm, B. Teunis, P. Coloe & B. van der Zeijst, (1998) The lipopolysaccharide biosynthesis locus of *Campylobacter jejuni* 81116. *Microbiology* **144**: 2049-2061.

Fry, B. N., S. Feng, Y.-Y. Chen, D. G. Newell, P. J. Coloe & V. Korolik, (2000) The *galE* Gene of *Campylobacter jejuni* Is Involved in Lipopolysaccharide Synthesis and Virulence. *Infect. Immun.* **68**: 2594-2601.

Gaudreau, C. & H. Gilbert, (2003) Antimicrobial Resistance of *Campylobacter jejuni* subsp. *jejuni* Strains Isolated from Humans in 1998 to 2001 in Montreal, Canada. *Antimicrob. Agents Chemother.* **47**: 2027-2029.

Godschalk, P. C. R., A. P. Heikema, M. Gilbert, T. Komagamine, C. W. Ang, J. Glerum, D. Brochu, J. Li, N. Yuki, B. C. Jacobs, A. v. Belkum & H. P. Endtz, (2004) The crucial role of *Campylobacter jejuni* genes in anti-ganglioside antibody induction in Guillain-Barre syndrome. *The Journal of Clinical Investigation* **114**: 1659-1665.

Grant, C. C. R., M. E. Konkel, W. Cieplak & L. S. Tompkins, (1993) Role of Flagella in Adherence, Internalization, and Translocation of *Campylobacter jejuni* in Nonpolarized and Polarized Epithelial Cell Cultures. *Infection and Immunity* **61**: 1764-1771.

Grys, T. E., M. B. Siegel, W. W. Lathem & R. A. Welch, (2005) The StcE Protease Contributes to Intimate Adherence of Enterohemorrhagic *Escherichia coli* O157:H7 to Host Cells. *Infect. Immun.* **73**: 1295-1303.

Hall, G., M. D. Kirk, N. Becker, J. E. Gregory, L. Unicomb, G. Millard, R. Stafford, K. Lalor & t. O. W. Group, (2005) Estimating Foodborne Gastroenteritis, Australia. *Emerging Infectious Diseases* **11**: 1257-1264.

Hartley, L., N. Kaakoush, J. Ford, V. Korolik & G. Mendz, (2009) Characterisation of *Campylobacter jejuni* genes potentially involved in phosphonate degradation. *Gut Pathogens* **1**: 13.

Haselhorst, T., A.-C. Lamerz & M. v. Itzstein, (2009) Saturation Transfer Difference NMR Spectroscopy as a Technique to Investigate Protein–Carbohydrate Interactions in Solution. In: Glycomics. Humana Press, pp. 1-12.

Hendrixson, D. R. & V. J. DiRita, (2004) Identification of Campylobacter jejuni genes involved in commensal colonization of the chick gastrointestinal tract. *Molecular Microbiology* **52**: 471-484.

Holmen Larsson, J. M., H. Karlsson, H. Sjovall & G. C. Hansson, (2009) A complex, but uniform *O*-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MS^n. *Glycobiology* **19**: 756-766.

Hooper, N. M., (1994) Families of zinc metalloproteases. *FEBS Letters* **354**: 1-6.

Horrocks, S. M., R. C. Anderson, D. J. Nisbet & S. C. Ricke, Incidence and ecology of *Campylobacter jejuni* and *coli* in animals. *Anaerobe* **15**: 18-25.

Horvat, Š. & A. Jakas, (2004) Peptide and amino acid glycation: new insights into the Maillard reaction. *Journal of Peptide Science* **10**: 119-137.

Hu, L. & D. J. Kopecko, (1999) Campylobacter jejuni 81-176 associates with microtubules and dynein during invasion of human intestinal cells. *Infect Immun.* **67**: 4171-4182.

Hu, L. & D. J. Kopecko, (2000) Interactions of *Campylobacter* with eukaryotic cells: gut luminal colonization and mucosal invasion mechanisms. In: Campylobacter. I. Nachamkin & M. J. Blaser (eds). Washington, D.C.: ASM Press, pp. 191-215.

Hu, R.-H., A. Mellors & V. P. Bhavanaandan, (1994) Cleavage of Epitectin, a Mucin-Type Sialoglycoprotein, from the surface of Human Laryngeal Carcinoma Cells by a Glycoprotease from Pasteurella haemolytica. *Archives of Biochemistry and Biophysics* **310**: 300-309.

Hugdahl, M. B., J. T. Beery & M. P. Doyle, (1988) Chemotactic behavior of *Campylobacter jejuni*. *Infect. Immun.* **56**: 1560-1566.

Hughes, R., R. Hadden, N. Gregson & K. Smith, (1999) Pathogenesis of Guillain-Barre syndrome. *Journal of Neuroimmunology* **100**: 74-97.

Humphrey, T., S. O'Brien & M. Madsen, (2007) Campylobacters as zoonotic pathogens: A food production perspective. *International Journal of Food Microbiology* **117**: 237-257.

Inoue, H., H. Nojima & H. Okayama, (1990) High efficiency transformation of Escherichia coli with plasmids. *Gene* **96**: 23-28.

Jacobs, B. C., A. Van Belkum & H. P. Endtz, (2008) Guillain-Barré Syndrome and *Campylobacter* Infection. In: *Campylobacter,* 3rd ed. I. Nachamkin, C. Szymanski & M. J. Blaser (eds). Washington, DC: ASM Press, pp. 245-261.

Jagusztyn-Krynicka, E. K., P. Łaniewski & A. Wyszyńska, (2009) Update on *Campylobacter jejuni* vaccine development for preventing human campylobacteriosis. *Expert Review of Vaccines* **8**: 625-645.

Jain, D., K. N. Prasad, S. Sinha & N. Husain, (2008) Differences in virulence attributes between cytolethal distending toxin positive and negative *Campylobacter jejuni* strains. *J Med Microbiol* **57**: 267-272.

Kaino, K., H. Hayashidani, K. Kaneko & M. Ogawa, (1988) Intestinal colonization of *Campylobacter jejuni* in chickens. *Japanese Journal of Veterinary Science* **50**: 489-494.

Karandashova, I., I. Elanskaya, K. Marin, J. Vinnemeier & M. Hagemann, (2002) Identification of Genes Essential for Growth at High Salt Concentrations Using Salt-Sensitive Mutants of the Cyanobacterium *Synechocystis* sp. Strain PCC 6803. *Current Microbiology* **44**: 184-188.

Karin, A., H. M. V. Arnoud, G. K. Johannes, F. Monique, D. Jacob & E. Arie, (2003) Induced *Helicobacter pylori* vacuolating cytotoxin VacA expression after initial colonisation of human gastric epithelial cells. *FEMS Immunology and Medical Microbiology* **39**: 251-256.

Katz, C., I. Cohen-Or, U. Gophna & E. Z. Ron, (2010) The ubiquitous conserved glycopeptidase gcp prevents accumulation of toxic glycated proteins. *MBio* **1**: e00195-00110.

Ketley, J. M., (1997) Pathogenesis of enteric infection by *Campylobacter*. *Microbiology* **143**: 5-21.

Konkel, M., S. Garvis, S. Tipton, D. J. Anderson & W. J. Cieplak, (1997) Identification and molecular cloning of a gene encoding a fibronectin-binding protein (CadF) from Campylobacter jejuni. *Molecular Microbiology* **24**: 953-963.

Konkel, M. E., L. A. Joens & P. F. Mixter, (2000) Molecular characterization of *Campylobacter jejuni* virulence determinants. In: Campylobacter. I. Nachamkin & M. J. Blaser (eds). Washington, D.C.: ASM Press, pp. 217-240.

Konkel, M. E., J. D. Klena, V. Rivera-Amill, M. R. Monteville, D. Biswas, B. Raphael & J. Mickelson, (2004) Secretion of Virulence Proteins from *Campylobacter jejuni* Is Dependent on a Functional Flagellar Export Apparatus. *Journal of Bacteriology* **186**: 3296-3303.

Kopecko, D. J., L. Hu & K. J. M. Zaal, (2001) *Campylobacter jejuni* - microtubule-dependent invasion. *TRENDS in Microbiology* **9**: 389-396.

Korolik, V. & J. Ketley, (2008) Chemosensory Signal Transduction Pathway of *Campylobacter jejuni*. In: Campylobacter, 3rd ed. I. Nachamkin, C. Szymanski & M. J. Blaser (eds). Washington DC: ASM Press, pp. 351-366.

Lee, C., PE Shewen, WM Cladman, JA Conlon, A Mellors & R. Lo, (1994) Sialoglycoprotease of *Pasteurella haemolytica* A1: detection of antisialoglycoprotease antibodies in sera of calves. *Can J Vet Res* **58**: 93-98.

Lee, L. H., E. Burg, III, S. Baqar, A. L. Bourgeois, D. H. Burr, C. P. Ewing, T. J. Trust & P. Guerry, (1999) Evaluation of a Truncated Recombinant Flagellin Subunit Vaccine against *Campylobacter jejuni*. *Infect. Immun.* **67**: 5799-5805.

Liu, C., S. Stirzaker, D. Knuckey, K. Robinson, K. Knope, G. Fitzsimmons, J. Wall, K. Roper, N. Martin, A. Reynolds, R. Owen, A. Hii, C. Barry, P. Wright, L. Sanders, J. Fielding & N. A. R. W. Group, (2009) Australia's notifiable disease status, 2007: annual report of the National Notifiable Diseases Surveillance System. *Commun Dis Intell.* **33**: 89-154.

Lo, R., M. Watt, S. Gyroffy & A. Mellors, (1994) Preparation of recombinant glycoprotease of *Pasteurella haemolytica* A1 utilizing the *Escherichia coli* alpha-hemolysin secretion system. *FEMS Microbiol Lett* **116**: 2.

Lönnerdal, B. & S. Iyer, (1995) Lactoferrin: Molecular Structure and Biological Function. *Annual Review of Nutrition* **15**: 93-110.

Luechtefeld, N. A., M. J. Blaser, L. B. Reller & W. L. Wang, (1980) Isolation of *Campylobacter fetus* subsp. *jejuni* from migratory waterfowl. *J. Clin. Microbiol.* **12**: 406-408.

Marchler-Bauer, A. & S. H. Bryant, (2004) CD-Search: protein domain annotations on the fly. *Nucl. Acids Res.* **32**: W327-331.

Martin, P. M., J. Mathiot, J. Ipero, M. Kirimat, A. J. Georges & M. C. Georges-Courbot, (1989) Immune response to *Campylobacter jejuni* and *Campylobacter coli* in a cohort of children from birth to 2 years of age. *Infect. Immun.* **57**: 2542-2546.

McAuley, J., S. Linden, C. Png, R. King, H. Pennington, S. Gendler, T. Florin, G. Hill, V. Korolik & M. McGuckin, (2007) MUC1 cell surface mucin is a critical element of the mucosal barrier to infection. *Journal of Clinical Investigation* **117**: 2313-2324.

Mellors, A., (2008) Personal communication. In., pp.

Mellors, A. & P. Jiang, (1998) Membrane Protein Proteolysis Assayed by Fluorescence Quenching: Assay of *0*-Sialoglycoprotein Endopeptidase. *Analytical Biochemistry* **259**: 8-15.

Miller, J. F., W. J. Dower & L. S. Tompkins, (1988) High-voltage electroporation of bacteria: genetic transformation of *Campylobacter jejuni* with plasmid DNA. *Proceedings of the National Academy of Sciences of the United States of America* **85**: 856-860.

MMWR, (2005) Preliminary FoodNet Data on the Incidence of Infection with Pathogens Transmitted Commonly Through Food - 10 Sites, United States, 2004. *Morbidity and Mortality Weekly Report* **54**: 352-356.

Molina, J., Casin I, Hausfater P, Giretti E, Welker Y, Decazes J, Garrait V, Lagrange P & M. J., (1995) Campylobacter infections in HIV-infected patients: clinical and bacteriological features. *AIDS* **9**: 881-885.

Monteiro, M. A., S. Baqar, E. R. Hall, Y.-H. Chen, C. K. Porter, D. E. Bentzel, L. Applebee & P. Guerry, (2009) Capsule Polysaccharide Conjugate Vaccine against Diarrheal Disease Caused by *Campylobacter jejuni*. *Infect. Immun.* **77**: 1128-1136.

Monteville, M. R. & M. E. Konkel, (2002) Fibronectin-Facilitated Invasion of T84 Eukaryotic Cells by *Campylobacter jejuni* Occurs Preferentially at the Basolateral Cell Surface. *Infection and Immunity* **70**: 6666-6671.

Moore, J. E., M. D. Barton, I. S. Blair, D. Corcoran, J. S. G. Dooley, S. Fanning, I. Kempf, A. J. Lastovica, C. J. Lowery, M. Matsuda, D. A. McDowell, A.

McMahon, B. C. Millar, J. R. Rao, P. J. Rooney, B. S. Seal, W. J. Snelling & O. Tolba, (2006) The epidemiology of antibiotic resistance in *Campylobacter*. *Microbes and Infection* **8**: 1955-1966.

Murphy, G. J., P. Echeverria, L. Jackson, M. Arness, C. LeBron & C. Pitarangsi, (1996) Ciprofloxacin- and azithromycin-resistant *Campylobacter* causing traveler's diarrhea in U.S. troops deployed to Thailand in 1994. *Clin Infect Dis* **22**: 868-869.

Myers, J. D. & D. J. Kelly, (2005) A sulphite respiration system in the chemoheterotrophic human pathogen Campylobacter jejuni. *Microbiology* **151**: 233-242.

Nachamkin, I., B. M. Allos & T. Ho, (1998) *Campylobacter* Species and Guillain-Barre Syndrome. *Clinical Microbiology Reviews* **11**: 555-567.

Nachamkin, I., B. M. Allos & T. W. Ho, (2000a) Clinical aspects of *Campylobacter* infection. In: Campylobacter. I. Nachamkin & M. J. Blaser (eds). Washington, D.C.: ASM Press, pp. 155-175.

Nachamkin, I., J. Engberg & F. Aarestrup, (2000b) Diagnosis and antimicrobial susceptibility of *Campylobacter* species. In: *Campylobacter*, 2nd Edition. I. Nachamkin & M. J. Blaser (eds). Washington, DC: American Society for Microbiology, pp. 45-66.

Nachamkin, I., X. H. Yang & N. J. Stern, (1993a) Role of *Campylobacter-Jejuni* Flagella as Colonization Factors for 3-Day-Old Chicks - Analysis with Flagellar Mutants. *Applied and Environmental Microbiology* **59**: 1269-1273.

Nachamkin, I., X. H. Yang & N. J. Stern, (1993b) Role of *Campylobacter jejuni* flagella as colonization factors for three-day-old chicks: analysis with flagellar mutants. *Appl Environ Microbiol.* **59**: 1269–1273.

Nesin, M., Lupski JR, Svec P & Godson GN, (1987) Possible new genes as revealed by molecular analysis of a 5-kb *Escherichia coli* chromosomal region 5' to the rpsU-dnaG-rpoD macromolecular-synthesis operon. *Gene* **51**: 149-161.

Newell, D. G., (2001) Animal models of *Campylobacter jejuni* colonization and disease and the lessons to be learned from similar *Helicobacter pylori* models. *Journal of Applied Microbiology* **90**: 57S-67S.

Nichols, C. E., Johnson C, Lockyer M, Charles I. G., Lamb H. K, Hawkins A. R & Stammers D. K, (2006) Structural characterization of *Salmonella typhimurium* YeaZ, an M22 O-sialoglycoprotein endopeptidase homolog. *Proteins: Structure, Function, and Bioinformatics* **64**: 111-123.

NNDSS, (2003) National Notifiable Disease Surveillance System. In. Australia, pp.

Novagen, (2010) pET System Manual 11th Edition. *EMD Chemicals*.

Nuijten, P., L. Marquez-Magana & B. v. d. Zeijst, (1995) Analysis of flagellin gene expression in flagellar phase variants of *Campylobacter jejuni* 81116. *Antonie Van Leeuwenhoek.* **67**: 377-383.

Oberhelman, R., R. Gilman, P. Sheen, D. Taylor, R. Black, L. Cabrera, A. Lescano, R. Meza & G. Madico, (1999) A placebo-controlled trial of *Lactobacillus* GG to prevent diarrhea in undernourished Peruvian children. *Journal of Pediatrics* **134**: 15-20.

Oberhelman, R. A. & D. N. Taylor, (2000a) *Campylobacter* infections in developing countries. In: Campylobacter. I. Nachamkin & M. J. Blaser (eds). Washington, D.C.: ASM Press, pp.

Oberhelman, R. A. & D. N. Taylor, (2000b) *Campylobacter* infections in developing countries. In: *Campylobacter.* I. Nachamkin & M. J. Blaser (eds). Washington, D.C.: ASM Press, pp. 139-153.

Ogden, I. D., J. F. Dallas, M. MacRae, O. Rotariu, K. W. Reay, M. Leitch, A. P. Thomson, S. K. Sheppard, M. Maiden, K. J. Forbes & N. J. C. Strachan, (2009) *Campylobacter* Excreted into the Environment by Animal Sources: Prevalence, Concentration Shed, and Host Association. *Foodborne Pathogens and Disease* **6**: 1161-1170.

Otulakowski, G. L., P. E. Shewen, A. E. Udoh, A. Mellors & B. N. Wilkie, (1983) Proteolysis of sialoglycoprotein by *Pasteurella haemolytica* cytotoxic culture supernatant. *Infection and Immunity* **42**: 64-70.

Palyada, K., D. Threadgill & A. Stintzi, (2004) Iron acquisition and regulation in *Campylobacter jejuni*. *Journal of Bacteriology* **186**: 4714-4729.

Park, S. F., (2002) The physiology of *Campylobacter* species and its relevance to their role as foodborne pathogens. *International Journal of Food Microbiology* **74**: 177-178.

Parkhill, J., B. Wren, K. Mungall, J. M. Ketley, C. Churcher, D. Basham, T. Chillingworth, R. M. Davies, T. Feltwell, S. Holroyd, K. Jagels, A. V. Karlyshev, S. Moule, M. J. Pallen, C. W. Penn, M. A. Quail, M.-A. Rajandream, K. M. Rutherford, A. H. M. van Vilet, S. Whitehead & B. G. Barrell, (2000a) The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* **403**: 665-668.

Parkhill, J., B. W. Wren, K. Mungall, J. M. Ketley, C. Churcher, D. Basham, T. Chillingworth, R. M. Davies, T. Feltwell, S. Holroyd, K. Jagels, A. V. Karlyshev, S. Moule, M. J. Pallen, C. W. Penn, M. A. Quail, M.-A. Rajandream, K. M. Rutherford, A. H. M. v. Vliet, S. Whitehead & B. G. Barrell, (2000b) The genome sequence ofthe food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* **403**: 665-668.

Pei, Z. & M. Blaser, (1993) PEB1, the major cell-binding factor of *Campylobacter jejuni*, is a homolog of the binding component in gram-negative nutrient transport systems. *J. Biol. Chem.* **268**: 18717-18725.

Pei, Z., C. Burucoa, B. Grignon, S. Baqar, X.-Z. Huang, D. J. Kopecko, A. L. Bourgeois, J.-L. Fauchere & M. J. Blaser, (1998) Mutation in the *peb1A* Locus of *Campylobacter jejuni* Reduces Interactions with Epithelial Cells and Intestinal Colonization of Mice. *Infect. Immun.* **66**: 938-943.

Penn, C. W., (2001) Surface components of Campylobacter and Helicobacter. *Journal of Applied Microbiology* **90**: 25S-35S.

Pickett, C., E. Pesci, D. Cottle, G. Russell, A. Erdem & H. Zeytin, (1996) Prevalence of cytolethal distending toxin production in *Campylobacter jejuni* and relatedness of *Campylobacter* sp. *cdt*B gene. *Infect. Immun.* **64**: 2070-2078.

Piddock, L. J. V., V. Ricci, K. Stanley & K. Jones, (2000) Activity of antibiotics used in human medicine for *Campylobacter jejuni* isolated from farm animals and their environment in Lancashire, UK. *J. Antimicrob. Chemother.* **46**: 303-306.

Poly, F., C. Ewing, S. Goon, T. E. Hickey, D. Rockabrand, G. Majam, L. Lee, J. Phan, N. J. Savarino & P. Guerry, (2007) Heterogeneity of a *Campylobacter jejuni* Protein That Is Secreted through the Flagellar Filament. *Infect. Immun.* **75**: 3859-3867.

Pope, J. E., A. Krizova, A. X. Garg, H. Thiessen-Philbrook & J. M. Ouimet, (2007) *Campylobacter* Reactive Arthritis: A Systematic Review. *Seminars in Arthritis and Rheumatism* **37**: 48-55.

Pratt, A. & V. Korolik, (2005) Tetracycline resistance of Australian *Campylobacter jejuni* and *Campylobacter coli* isolates. *Journal of Antimicrobial Chemotherapy* **55**: 452-460.

Prokhorova, T. A., P. N. Nielsen, J. Petersen, T. Kofoed, J. S. Crawford, C. Morsczeck, A. Boysen & P. Schrotz-King, (2006) Novel surface polypeptides of *Campylobacter jejuni* as traveller's diarrhoea vaccine candidates discovered by proteomics. *Vaccine* **24**: 6446-6455.

Robbe, C., C. Capon, B. Coddeville & J.-C. Michalski, (2004) Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *Journal of Biochemistry* **384**: 307-316.

Sambrook, J. & D. W. Russel, (2001) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press,, Cold Spring Harbor, N.Y.

Scot, D. A., (1997) Vaccines against Campylobacter jejuni. *The Journal of Infectious Diseases* **176**: S183-188.

Shewen, P. E., C. W. Lee, A. Perets, D. C. Hodgins, K. Baldwin & R. Y. C. Lo, (2003) Efficacy of recombinant sialoglycoprotease in protection of cattle against pneumonic with *Mannheimia (Pasteurella) haemolytica* A1. *Vaccine* **21**: 1901-1906.

Simon, M., P. Bénit, A. Vassal, C. Dubois & F. G., (1994) Sequence of the PHO2-POL3 (CDC2) region of chromosome IV of *Saccharomyces cerevisiae*. *Yeast* **10**: 1653-1656.

Singh, R., A. Barden, T. Mori & L. Beilin, (2001) Advanced glycation end-products: a review. *Diabetologia* **44**: 129-146.

Sizemore, D. R., B. Warner, J. Lawrence, A. Jones & K. P. Killeen, (2006) Live, attenuated *Salmonella typhimurium* vectoring *Campylobacter* antigens. *Vaccine* **24**: 3793-3803.

Skirow, M. B. & M. J. Blaser, (2000) Clinical aspects of *Campylobacter* infection. In: Campylobacter. I. Nachamkin & M. J. Blaser (eds). Washington, D.C.: ASM Press, pp.

Skirrow, M., (1977) Campylobacter enteritis: a "new" disease. *British Medical Journal* **2**: 9-11.

Skirrow, M., (1991) Epidemiology of Campylobacter enteritis. *International Journal of Food Microbiology* **12**: 9-16.

Skirrow, M. B. & M. J. Blaser, (2000) Clinical Aspects of *Campylobacter* Infection. In: *Campylobacter*, 2nd ed. I. Nachamkin & M. J. Blaser (eds). Washington, DC: ASM Press, pp. 69-88.

Sorvillo, F., L. Lieb & S. Waterman, (1991) Incidence of campylobacteriosis among patients with AIDS in Los Angeles County. *Journal of Acquired Immune Defficiency Syndrome* **4**: 598-602.

Sutherland, D., K. Abdullah, P. Cyopick & A. Mellors, (1992) Cleavage of the cell-surface O-sialoglycoproteins CD34, CD43, CD44, and CD45 by a novel glycoprotease from Pasteurella haemolytica. *J Immunol* **148**: 1458-1464.

Sylvester, F. A., D. Philpott, B. Gold, A. Lastovica & J. F. Forstner, (1996) Adherence to lipids and intestinal mucin by a recently recognized human pathogen, *Campylobacter upsaliensis*. *Infect. Immun.* **64**: 4060-4066.

Takata, T., S. Fujimoto & K. Amako, (1992) Isolation of nonchemotactic mutants of *Campylobacter jejuni* and their colonization of the mouse intestinal tract. *Infect. Immun.* **60**: 3596–3600.

Tauxe, R., (2002) Emerging foodborne pathogens. *International Journal of Food Microbiology* **78**: 31-41.

Taylor, D. N., M. J. Blaser, P. Echeverria, C. Pitarangsi, L. Bodhidatta & W. L. Wang, (1987) Erythromycin-resistant *Campylobacter* infections in Thailand. *Antimicrob. Agents Chemother.* **31**: 438–442.

Taylor, D. N. & Y. Wang, (1990) Chloramphenicol resistance in *Campylobacter coli*: nucleotide sequence, expression, and cloning vector construction. *Gene* **94**: 23-28.

Tee, W. & A. Mijch, (1998) *Campylobacter jejuni* Bacteremia in Human Immunodeficiency Virus (HIV)-Infected and Non-HIV-Infected Patients: Comparison of Clinical Features and Review. *Clinical Infectious Diseases* **26**: 91-96.

Thompson, J. S., D. S. Hodge, D. E. Smith & Y. A. Yong, (1990) Use of tri-gas incubator for routine culture of *Campylobacter* species from fecal specimens. *J. Clin. Microbiol.* **28**: 2802-2803.

Toribara, N., Gum JR Jr, Culhane PJ, Lagace RE, Hicks JW, Petersen GM & Y. Kim, (1991) MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorphism. *J clin Invest.* **88**: 1005-1013.

Trotter, C. L., J. McVernon, M. E. Ramsay, C. G. Whitney, E. K. Mulholland, D. Goldblatt, J. Hombach & M.-P. Kieny, (2008) Optimising the use of conjugate vaccines to prevent disease caused by *Haemophilus influenzae* type b, *Neisseria meningitidis* and *Streptococcus pneumoniae*. *Vaccine* **26**: 4434-4445.

Tsang, R. S., (2002) The relationship of *Campylobacter jejuni* infection and the development of Guillain-Barre syndrome. *Current Opinion in Infectious Diseases* **15**: 221-228.

Tu, Q. V., M. A. McGuckin & G. L. Mendz, (2008) *Campylobacter jejuni* response to human mucin MUC2: modulation of colonization and pathogenicity determinants. *J Med Microbiol* **57**: 795-802.

Vandamme, P., (2000) Taxonomy of the Family *Campylobacteraceae*. In: Campylobacter. I. Nachamkin & M. J. Blaser (eds). Washington: ASM Press, pp. 3-26.

Vliet, A. H. M. v. & J. M. Ketley, (2001) Pathogenesis of enteric *Campylobacter* infection. *Journal of Applied Microbiology* **90**: 45S-56S.

Wallis, M. R., (1994) The pathogenesis of Campylobacter jejuni. *British Journal of Biomedical Science* **51**: 57-64.

Walsh, F. S., M. Cronin, S. Koblar, P. Doherty, J. Winer, A. Leon & R. A. C. Hughes, (1991) Association between glycoconjugate antibodies and *Campylobacter* infection in patients with Guillain-Barré syndrome. *Journal of Neuroimmunology* **34**: 43-51.

Wassenaar, T., N. Bleumink-Pluym & B. v. d. Zeijst, (1991) Inactivation of *Campylobacter jejuni* flagellin genes by homologous recombination demonstrates that flaA but not flaB is required for invasion. *EMBO* **10**: 2055-2061.

Wassenaar, T., B. Fry & B. v. d. Zeijst, (1993) Genetic manipulation of *Campylobacter:* evaluation of natural transformation and electro-transformation. *Gene* **132**: 131-135.

Wassenaar, T. M., (1997) Toxin production by *Campylobacter* spp. *Clinical Microbiology Reviews* **10**: 466-476.

Watt, M.-A. V., R. Y. C. Lo & A. Mellors, (1997a) Refolding of recombinant *Pasteurella haemolytica* A1 glycoprotease expressed in an *Escherichia coli* thioredoxin gene fusion system. *Cell Stress Chaperones.* **2**: 180-190.

Watt, M.-A. V., A. Mellors & R. Y. C. Lo, (1997b) Comparison of the recombinant and authentic forms of the *Pasteurella haemolytica* A1 glycoprotease. *FEMS Microbiology Letters* **147**: 37-43.

Whiteley, L., S. Maheswaran, D. Weiss, T. Ames & M. Kannan, (1992) *Pasteurella haemolytica* A1 and bovine respiratory disease: pathogenesis. *J Vet Intern Med.* **6**: 11-22.

Wiesner, R. S., D. R. Hendrixson & V. J. DiRita, (2003) Natural Transformation of *Campylobacter jejuni* Requires Components of a Type II Secretion System. *J. Bacteriol.* **185**: 5408-5418.

Wosten, M. M. S. M., M. Boeve, M. G. A. Koot, A. C. van Nuenen & B. A. M. van der Zeijst, (1998) Identification of *Campylobacter jejuni* Promoter Sequences. *J. Bacteriol.* **180**: 594-599.

Yao, R., R. A. Alm, T. J. Trust & P. Guerry, (1993) Construction of new Campylobacter cloning vectors and a new mutational cat cassette. *Gene* **130**: 127-130.

Yao, R., D. Burr, P. Doig, T. Trust, H. Niu & P. Guerry, (1994) Isolation of motile and non-motile insertional mutants of *Campylobacter jejuni*: the role of motility in adherence and invasion of eukaryotic cells. *Molecular Microbiology* **14**: 883-893.

Yuki, N., T. Taki, F. Inagaki, T. Kasama, M. Takahashi, K. Saito, S. Handa & T. Miyatake, (1993) A bacterium lipopolysaccharide that elicits Guillain-Barre syndrome has a GM1 ganglioside-like structure. *J. Exp. Med.* **178**: 1771-1775.

Zheng, L., J. Yang, C. Landwehr, F. Fan & Y. Ji, (2005) Identification of an essential glycoprotease in *Staphylococcus aureus*. *FEMS Microbiology Letters* **245**: 279-285.

Zheng, L., C. Yu, K. Bayles, I. Lasa & Y. Ji, (2007) Conditional Mutation of an Essential Putative Glycoprotease Eliminates Autolysis in *Staphylococcus aureus*. *J. Bacteriol.* **189**: 2734-2742.

Zuther, E., H. Schubert & M. Hagemann, (1998) Mutation of a Gene Encoding a Putative Glycoprotease Leads to Reduced Salt Tolerance, Altered Pigmentation, and Cyanophycin Accumulation in the Cyanobacterium *Synechocystis* sp. Strain PCC 6803. *Journal of Bacteriology* **180**: 1715-1722.