Genomics and virulence factors of Fusobacterium necrophorum

Wright, K.

Genomics and virulence factors of

Fusobacterium necrophorum

Katie Louise Wright

A thesis submitted in partial fulfilment of the requirements of the University

of Westminster for the degree of Doctor of Philosophy

March 2016

Abstract

Fusobacterium necrophorum, a Gram negative, anaerobic bacterium, is a common cause of acute pharyngitis and tonsillitis and a rare cause of more severe infections of the head and neck. At the beginning of the project, there was no available genome sequence for *F. necrophorum*. The aim of this project was to sequence the *F. necrophorum* genome and identify and study its putative virulence factors contained using *in silico* and *in vitro* analysis. Type strains JCM 3718 and JCM 3724, *F. necrophorum* subspecies *necrophorum* (*Fnn*) and *funduliforme* (*Fnf*), respectively, and strain ARU 01 (*Fnf*), isolated from a patient with LS, were commercially sequenced by Roche 454 GS-FLX+ next generation sequencing and assembled into contigs using Roche GS Assembler. Sequence data was annotated semi-automatically, using the xBASE pipeline, BLASTp and Pfam. The *F. necrophorum* genome was determined to be approximately 2.1 – 2.3 Mb in size, with an estimated 1,950 ORFs and includes genes for a leukotoxin, ecotin, haemolysin, haemagglutinin, haemin receptor, adhesin and type Vb and Vc secretion systems. The prevalence of the *leukotoxin* gene was investigated in strains JCM 3718, JCM 3724 and ARU 01, as well as a clinical collection of 25 *Fnf* strains, identified using biochemical and molecular tests. The *leukotoxin* operon was found to be universal within the strain collection by PCR. HL-60 cells subjected to aliquots of concentrated high molecular weight culture supernatant, predicted to contain the secreted leukotoxins of strains JCM 3718, JCM 3724 and ARU 01, were killed in a dose-dependent manner. The cytotoxic effect of the leukotoxin against human donor white blood cells was also tested to validate the HL-60 assay. The differences in the results between the two assays were not statistically significant. Ecotin, a serine protease inhibitor, was found to be present in 100 % of the strain collection and had a highly conserved sequence with primary and secondary binding sites exposed on opposing sides of the protein. During enzyme inhibition studies, a purified recombinant *F. necrophorum* ecotin protein inhibited human neutrophil elastase, a protease that degrades bacteria at inflammation sites, and human plasma kallikrein, a component of the host clotting cascade. The recombinant ecotin also prolonged human plasma clotting times by up to 7-fold for the extrinsic pathway, and up to 40-fold for the intrinsic pathway. The genome sequence data provides important information about *F. necrophorum* type strains and enables comparative study between strains and subspecies. Results from the leukotoxin and ecotin assays can be used to build up an understanding of how the organism behaves during infection.

Acknowledgements

Dedicated to the memory of my Grandmother

Irene Jones

(1932 – 2015)

Author's declaration

I declare that the present work was carried out in accordance with the Guidelines and Regulations of the University of Westminster. The work is original except where indicated by special reference in the text.

The submission as a whole or part is not substantially the same as any that I previously or am currently making, whether in published or unpublished form, for a degree, diploma or similar qualification at any university or similar institution.

Until the outcome of the current application to the University of Westminster is known, the work will not be submitted for any such qualification at another university or similar institution. Any views expressed in this work are those of the author and in no way represent those of the University of Westminster.

Signed: Date: 30th November 2015

Contents

Supplementary materials

Additional Appendices are located on the attached USB flash drive.

Appendix A	JCM 3718 Raw FASTA sequence reads
Appendix B	JCM 3724 Raw FASTA sequence reads
Appendix C	ARU 01 Raw FASTA sequence reads
Appendix D	JCM 3718 Qual file
Appendix E	JCM 3724 Qual file
Appendix F	ARU 01 Qual file
Appendix G	JCM 3718 Assembly contigs
Appendix H	JCM 3724 Assembly contigs
Appendix I	ARU 01 Assembly contigs
Appendix J	JCM 3718 xBASE annotation
Appendix K	JCM 3724 xBASE annotation
Appendix L	ARU 01 xBASE annotation
Appendix M	File conversion script
Appendix N	JCM 3718 xBASE and BLASTp Annotation
Appendix O	JCM 3724 xBASE and BLASTp Annotation
Appendix P	ARU 01 xBASE and BLASTp Annotation
Appendix Q	ATCC 51357 Assembly contigs download
Appendix R	ATCC 51357-JCM 3724 Merged contigs

List of figures

List of tables

List of abbreviations

A	Adenine
ACT	Artemis Comparison Tool
AFC	7-Amino-4-(trifluoromethyl)coumarin
ANOVA	Analysis of variance
APS	Ammonium persulfate
APTT	Activated partial thromboplastin time
ARU	Anaerobe Reference Unit
BHI	Brain heart infusion
BLAST	Basic Local Alignment Search Tool
bp	Base pair
C	Cytosine
cas	CRISPR-associated
CT	Computed tomography
DIC	Disseminated intravascular coagulation
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
E	Enzyme
EDTA	Ethylenediaminetetraacetic acid
ENA	European Nucleotide Archive
E value	Expect value
FAA	Fastidious anaerobe agar
FHA	Filamentous haemagglutinin
Fnf	*Fusobacterium necrophorum* subspecies. *funduliforme*
Fnn	*Fusobacterium necrophorum* subspecies. *necrophorum*

G	Guanine
GAS	Group A streptococcus
HNE	Human neutrophil elastase
HPK	Human plasma kallikrein
I	Inhibitor
IMAC	Immobilised-metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
JCM	Japanese Collection of Microorganisms
kb	Kilo bases
kDa	Kilo Daltons
K_i	Inhibitory constant
K_m	Michaelis constant
lktA	leukotoxin
LPS	Lipopolysaccharide
M	Molar
Mb	Megabases
MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
NCBI	National Center for Biotechnology Information
OD	Optical density
OLC	Overlap layout consensus
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PI	Propidium iodide
PMNs	Polymorphonuclear cells
pNA	p-nitroanilide
POTRA	Polypeptide transport-associated

PSTS	Persistent sore throat syndrome
PT	Prothrombin time
PTA	Peritonsillar abscess
RFU	Relative fluorescence units
rpm	Revolutions per minute
rRNA	Ribosomal ribonucleic acid
S	Substrate
SDS PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
spp.	Species (plural)
subsp.	Subspecies
T	Thymine
TAA	Trimeric autotransporter adhesin
TAE	Tris base, acetic acid and EDTA
TEMED	N,N,N',N'-Tetramethylethylenediamine
TPS	Two-partner secretion
tRNA	Transfer ribonucleic acid
TT	Thrombin time
UV	Ultra violet
V_{max}	Maximum rate of reaction
WBC	White blood cell

Chapter 1:

Introduction

1.1 Introduction

1.1.1 The *Fusobacterium* genus

The Fusobacterium genus contains Gram negative, non-motile, non-sporulating, rod-shaped bacteria that are obligate anaerobes. They can be found in association with both humans and animals, particularly in mucous membranes. In humans they can be found in the oral cavity, gastrointestinal tract, female genital tract, periurethral areas and in necrotic lesions. In animals such as cattle, horses, sheep, pigs, cats and dogs, they can be found in the gastrointestinal tract, oral cavities and necrotic lesions. The genus contains three species clades: *F. nucleatum, F. periodonticum, F. russi, F. simiae* and *F. canifelinum,* which form a group of oral species; *F. necrophorum*, *F. equinum* and *F. gonidiaformans* which form a second clade which are genealogically related, as are species *F. varium, F. mortiferum, F. necrogenes,* and *F. ulcerans* from the third clade. The classifications are based on 16S rRNA homology and are supported by phylogenetic analysis of a group of conserved proteins (Hofstad, 1998; Gupta and Sethi, 2014).

1.1.2 *Fusobacterium* spp*.* and associated disease

F. necrophorum, the focus of this study, is known to cause oropharyngeal disease, such as pharyngitis and tonsillitis. It also causes more serious infections such as peritonsillar abscess and in rare cases a systemic infection known as Lemierre's syndrome, which includes symptoms such as septicaemia and metastatic abscesses. (Batty and Wren, 2005; Eaton and Swindells, 2014; Kuppalli *et al.*, 2012; Lyle *et al.*, 2011). *F. necrophorum* causes infections in both humans and animals (Langworth, 1977) and it is debated whether the organism is a commensal in the human upper

respiratory tract (Riordan, 2007). It is known to have several virulence factors that are implicated in disease states in humans and animals (section 1.8).

Fusobacterium nucleatum is a commensal of the human oral cavity and is implicated in various diseases, including periodontal infections, gastrointestinal disorders and adverse pregnancy outcomes. *F. nucleatum* is also a causative agent of Lemierre's syndrome (Han, 2015), although less commonly so than *F. necrophorum* (Ridgway *et al.*, 2010). *F. nucleatum* has been shown to be enriched in colorectal cancer (Castellarin *et al.*, 2012). Rubinstein *et al.* (2013) reported an association of the organism's FadA adhesin/invasin, a key virulence factor, with inflammatory and oncogenic responses. *F. nucleatum* has also been isolated from infected cat and dog bite wounds (Abrahamian and Goldstein, 2011).

F. equinum has been isolated from horses in both healthy and diseased oral cavities, lower respiratory tract infections, paraoral abscesses, necrotising pneumonia and pleurisy. It was designated as a new species by Dorsch *et al.* (2001) after previously being classified as *F. necrophorum*. Like *F. necrophorum*, *F. equinum* has been found to have a leukotoxin gene and exhibit leukotoxic activity (Tadepalli *et al.*, 2008; Zhou *et al.*, 2009b).

F. russi and *F. gonidiaformans* have also been isolated from cat and dog bite wounds (Abrahamian and Goldstein, 2011), as has *F. canifelinum* (Conrads *et al.*, 2004). *F. necrogenes* and *F. mortiferum* have been isolated from human faeces, while *F. russii* and *F. varium* have been isolated from both human and animal faeces. *F. periodonticum* may be found in patients with advanced periodontal disease and *F. gonidiaformans* is implicated in human infections involving the respiratory tract, genitourinary tract and gastrointestinal tract.

1.1.3 Historical review of *F. necrophorum*

The early literature of the organism now referred to as *F. necrophorum* is predominantly published in French and German languages. It has, however, been reviewed in English by Riordan (2007) who described how the organism was first reported in a case of calf diphtheria in 1884 by Löffler (1884) and named *Bacillus necrophorus* by Flügge (1886). It was first isolated and cultured by Bang (1890)

from a cattle liver abscess and the term nekrosebazillus (necrobacillosis) was coined, meaning septicaemic or necrotic infection from which the organism now referred to as *F. necrophorum* is isolated. In 1891 the first known human infections with the organism were described by Schmorl (1891), when the author and fellow researchers developed finger abscesses during work on rabbits with necrobacillosis. The first description of the organism *Bacillus funduliformis*, now referred to as *F. necrophorum* subsp. *funduliforme* was of that isolated from a female genital tract by Hallé (1898). In 1900 this was followed by what is thought to be the first description of human systemic infection with *F. necrophorum,* by Courmont and Cade (1900) although it was not until 1936 that Lemierre provided a clear clinical description of postanginal septicaemia associated with *F. necrophorum* (Lemierre, 1936). This condition is now commonly referred to as Lemierre's syndrome, however this was not until the late 1980s, and the terms postanginal septicaemia and necrobacillosis were previously used until then (Riordan, 2007).

There have been many other names used to describe *F. necrophorum,* including *Bacillus funduliformis, Bacteroides funduliformis, Necrobacterium funduliforme* and *Sphaerophorus necrophorus* (Duerden, 1990). The term *Fusobacterium* was introduced by Knorr (1923) and by 1974 all species of *Sphaerophorus* had been transferred to *Fusobacterium*, leading to the name *Fusobacterium necrophorum* being in use today (Hofstad, 1998).

1.1.4 Focus of introduction

F. necrophorum is the causative agent of approximately 10 % of sore throats, as well as the rare but severe systemic infection known as Lemierre's syndrome (Batty and Wren, 2005; Aliyu *et al.*, 2004). The organism is not consistently screened for in patient throat swabs and its genome and virulence factors have not yet been extensively studied.

This introduction focusses on *F. necrophorum* infections originating in the oropharynx of humans and also covers microbiology of the organism, along with the epidemiology, pathogenesis and virulence factors.

1.2 Description and definition of *F. necrophorum*

1.2.1 *F. necrophorum* subspecies

F. necrophorum is divided into two subspecies: *F. necrophorum* subsp. *necrophorum* (*Fnn*) and *F. necrophorum* subsp. *funduliforme* (*Fnf*). This division was proposed by Shinjo *et al.* (1991) based on the previous biovar classification. Biovar A, now referred to as *Fnn*, is more often associated with animal infections than human infections and is assumed to be more virulent due to a greater production of a leukotoxin and haemagglutinin than biovar B. Biovar B, now referred to as *Fnf*, is mostly associated with infections of humans and is less virulent (Smith *et al.*, 1990). The two biovars were divided based on haemagglutinating activity (see section 1.8.4). Using a glass slide method and chicken erythrocytes, biovar A was positive for haemagglutination and biovar B was not. DNA homology was also used to confirm the results with a membrane filter and measuring radioactivity. Percentage homologies were then assigned to strains in pairs, with strains of the same biovar having high levels of homology compared to strains of the other biovar (Shinjo *et al.*, 1991).

Smith and Thornton (1997) examined the biotypes of a collection of human and animal isolates using pathogenicity tests in mice. Fourteen strains, all of animal origin, were classified as biotype A, while 10 animal strains and 18 human strains were classified as biotype B. They determined that biotype A strains caused necrobacillosis in animals but not humans, whereas biotype B strains were found in both humans and animals. The evidence appears to support the view that *Fnn* strains are the predominant pathogen in animals and *Fnf* strains are predominant in humans.

1.2.2 Phenotypic and biochemical profile

F. necrophorum may be cultured on several different agars, such as Columbia blood agar, Soy agar and Brain-Heart infusion agar, enriched with yeast extract, vitamin K and hemin, or Brucella agar (Markey *et al.*, 2013; Brazier *et al.*, 1990). However it is the opinion of Batty and Wren (2005), Brazier (2006) and Syed *et al.* (2007) that *F. necrophorum* is best cultured on fastidious anaerobe agar (FAA), supplemented with 5 – 10 % defibrinated horse blood. This is due to *F. necrophorum* being a fastidious organism, and FAA provides a ready-made nutritious medium for the organism

to grow, with cysteine as a specific growth promoting agent. Colonial characteristics vary slightly between the subspecies, with *Fnf* colonies typically cream in colour, opaque, smooth and umbonate/raised. *Fnn* colonies are typically grey in colour, semi-opaque, mottled and umbonate with an irregular edge. *F. necrophorum* colonies are usually 1 – 4 mm in diameter, with *Fnn* at the larger end of the scale and *Fnf* at the lower end of the scale. *F. necrophorum* strains are usually haemolytic and exhibit a butyric acid odour (Hall *et al.*, 1997; Batty and Wren, 2005; Barrow and Feltham, 2004).

F. necrophorum is a Gram negative, obligately anaerobic, pleomorphically rod shaped bacterium which is non-motile and non-sporulating (Langworth, 1977; Batty and Wren, 2005). Cell morphology ranges from a coccobacillus shape to long filaments, and is dependent on the strain, the age of the culture and the media used (Langworth, 1977; Markey *et al.*, 2013). Commercial kits such as the API rapid ID 32A, the RapID-ANA II and API 20A strip tests can be used to identify *F. necrophorum,* although as many of the tests on each strip are negative, excluding the indole and alkaline phosphate tests, it is often better to use specific tests when *F. necrophorum* is suspected (Batty *et al.*, 2005; Riordan, 2007).

F. necrophorum is known to produce lipase, an enzyme involved in the breakdown of triglycerides, which is exhibited as an opalescent sheen on the surface of cultures grown on egg yolk agar (Batty *et al.*, 2005; Markey *et al.*, 2013). *F. necrophorum* also produces tryptophanase enzymes, which produce indole as a metabolite from tryptophan. This is examined using a spot indole reagent which turns a green colour if indole is present (Batty *et al.*, 2005). Under long wave ultra-violet (UV) light (365 nm), *F. necrophorum* colonies fluoresce a green-yellow colour, and where gas chromatography facilities are available, propionic acid and butyric acid can be detected (Barrow and Feltham, 2004).

1.2.3 Species and subspecies identification using molecular techniques

In addition to using biochemical techniques to identify *F. necrophorum* strains, molecular techniques have been developed that are faster and more sensitive than traditional culture techniques. This is discussed in detail in chapter 3. Briefly, PCR primers specific to the *16S rRNA*

gene, the *RNA polymerase (rpo) B* gene and the *gyrase (gyr) B* gene of *F. necrophorum* have been used to identify the species by Oikonomou *et al.* (2012), Aliyu *et al.* (2004) and Jensen *et al.* (2007), respectively. In order to identify to the subspecies level, real-time PCR assays using subspecies-specific TaqMan probes targeting the *rpoB* and *gyrB* genes have been designed and used by Aliyu *et al.* (2004) and Jensen *et al.* (2007), respectively.

1.3 Is *F. necrophorum* a commensal or pathogen?

It is a matter of debate whether *F. necrophorum* is a component of the normal flora of the upper respiratory tract in humans. In a review by Riordan (2007) it is discussed whether it can be considered normal flora based on a meta-analysis of all available relevant publications. It was noted that in some studies confirming the presence of *F. necrophorum* in healthy individuals, the primary evidence was unclear, by often referring to the isolation of Fusobacteria with no mention of which species. In other cases, such as the highly cited paper by Hallé (1898), it was concluded that *F. necrophorum* (then *Bacillus funduliformis*) is a commensal, despite not being isolated from any of the healthy subjects in the study (Riordan, 2007). Batty *et al.* (2005) commented that following their own extensive search of the literature, there appeared to be no evidence that *F. necrophorum* is a commensal in the human oral cavity.

Recent investigations into this issue have increasingly used molecular techniques. For example, as part of a study discussed further in section 1.4.1, Aliyu *et al.* (2004) used real-time PCR to investigate the presence of *F. necrophorum* in 100 healthy control throat swabs and found all were negative. However Jensen *et al.* (2007) carried out a similar investigation (also discussed further in section 1.4.1) using real-time PCR and examined throat swabs from 92 healthy controls for the presence of *F. necrophorum*. *Fnf* was found to be present in 21 %.

Jensen *et al.* (2007) and Aliyu *et al.* (2004) both used real-time PCR, although with different primers and probes to target different genes. The differences between the *F. necrophorum* isolation rates from the healthy control samples are quite striking. There is a possibility that the sampling methods

varied, with one method sampling deeper into the tissue than the other. Or perhaps there were problems with the primers such as non-specificity, or mismatches with the target gene. Jensen *et al.* (2007) also tried to culture *F. necrophorum* from the swabs, however all of the healthy control swabs were culture-negative. They attributed this to a poor standard of agar and to low numbers of *F. necrophorum* present on healthy tonsils. It was also suggested that the low numbers of *F. necrophorum* is likely to be due the organism being present predominantly deep in the tonsillar crypts of healthy individuals, as opposed to on the surface, which is likely to occur with inflamed and infected tonsils.

A large-scale study into the carriage of *F. necrophorum* in healthy individuals would be of great value in understanding whether *F. necrophorum* is a commensal of the human upper respiratory tract. If the organism is not a commensal, then it must be referred to as a pathogen. While *F. necrophorum* is generally referred to as a commensal, sometimes without reference, this is not always the case. Recently, Eaton and Swindells (2014) suggested that *F. necrophorum* should be regarded as a true pathogen rather than a commensal and Yusuf *et al.* (2015) also describe the organism as a primary pathogen and suggest that it is not a commensal in the oral flora, based on their own interpretations of the literature.

It has been suggested that food and water contaminated with animal faeces are sources of *F. necrophorum* infection in impoverished and malnourished children, particularly with poor oral hygiene (Enwonwu *et al.*, 1999). Perhaps close contact with animals such as livestock and pets could be a source of infection in developed countries. This is discussed further in section 1.7.1.

F. necrophorum is considered to be a commensal of the gastrointestinal, respiratory and genitourinary tracts of animals such as cattle (Narayanan *et al.*, 1997; Tadepalli *et al.*, 2008b). Animal infections are briefly discussed in section 1.5, however are mostly beyond the scope of this literature review.

1.4 *F. necrophorum* infectious disease

1.4.1 Pharyngitis and tonsillitis

Pharyngitis, also known as sore throat, and tonsillitis are the inflammation of the pharynx and tonsils, respectively (Public Health England, 2015; Thibodeau and Patton, 1997). The pharynx is divided into three sections: the nasopharynx, oropharynx and laryngopharynx (Figure 1). The oropharynx, situated below the nasopharynx and above the laryngopharynx, is the region at the back of the throat (Thibodeau and Patton, 1997) and is lined by stratified squamous epithelium (Perry and Whyte, 1998). The palatine tonsils are situated either side of the back of the throat and are comprised of mucosa-associated lymphoid tissue and contain crypts that greatly increase their surface area (Barnes, 2000; Perry and Whyte, 1998). They sit within a capsule that provides blood vessels and nerves (Galioto, 2008).
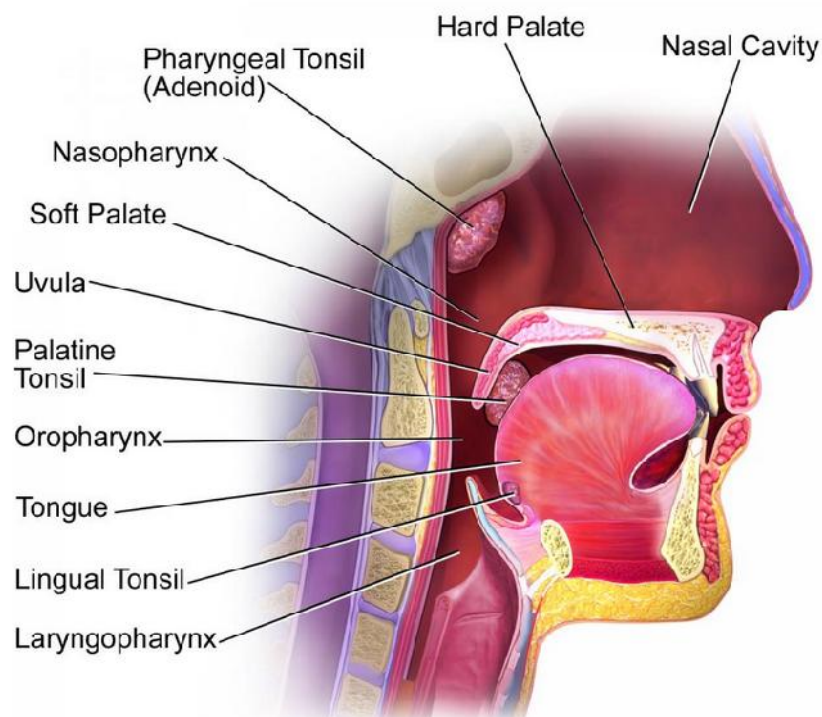
Figure 1: Sagittal view diagram of the location of the pharynx and tonsils. Source: (Merck Manuals Professional Edition, 2015).

Pharyngitis and tonsillitis can be caused by viruses and bacteria. Viruses involved include adenovirus, Epstein-Barr virus, and also common cold viruses such as rhinovirus and coronavirus (Barnes, 2000; Bird *et al.*, 2014). One of the most commonly implicated bacteria is *Streptococcus pyogenes*, an aerotolerant anaerobe, also known as group A β-haemolytic streptococcus. Aerobic bacteria isolated from cases of pharyngitis and tonsillitis include β-haemolytic streptococcus groups C and G, and *Corynebacterium diphtheriae* (Batty and Wren, 2005; Public Health England, 2015). Anaerobic bacteria are also implicated in throat infections. Increased numbers of anaerobes have been found to be present in chronic throat infections and anaerobes have been isolated from the cores of tonsils of children and adults with recurrent tonsillitis, in some cases with no aerobes present. Pigmented *Prevotella* and *Porphyromonas*, *Fusobacterium* and *Actinomyces spp.* are the anaerobic organisms that have been associated with tonsillitis and it has been suggested that they play a role in acute inflammatory processes in the tonsils (Brook, 2005).

Symptoms of pharyngitis include a sore throat, redness, difficulty swallowing, fever, malaise and a headache (Thibodeau and Patton, 1997; Public Health England, 2015). Symptoms of tonsillitis are similar, and may also include swelling of the tonsils, swollen lymph nodes in the neck, and exudate on the tonsils (Barnes, 2000). *F. necrophorum* is known to cause sore throats, persistent sore throat syndrome (PSTS), recurrent and persistent pharyngitis and tonsillitis (Batty and Wren, 2005; Eaton and Swindells, 2014). Incidence of *F. necrophorum* infections is discussed in section 1.6.1.

Batty and Wren (2005) investigated the presence of *F. necrophorum* in patient throat swabs using culture techniques. Two hundred and forty eight samples with clinical diagnoses of sore throat, tonsillitis and PSTS were tested; of those 24 (approximately 10 %) were positive for *F. necrophorum*. *F. necrophorum* was noted to be the most frequently isolated pathogen from a clinical diagnosis of PSTS. It was further noted that due to the lack of detail from the clinician, some cases may have been categorised as sore throat when they should have been in the PSTS group. *F. necrophorum* incidence in this cohort increased with age (section 1.6.2). It was suggested that this may be due to patients acquiring an infection that persists, and therefore leads to the prevalence experiencing a

cumulative effect. This study shows a strong link between *F. necrophorum* and PSTS, although no control group of healthy throat swabs were included.

Aliyu *et al.* (2004) investigated *F. necrophorum* as the cause of a simple sore throat using real-time PCR targeting the *RNA polymerase B* gene. One hundred throat swabs from a clinical diagnosis of pharyngitis were analysed and compared to 100 healthy control throat swabs (previously mentioned in section 1.3). Ten percent of the patient swabs were positive for *F. necrophorum* subspecies *funduliforme* while all of the healthy controls were negative. The results are a strong indication that *F. necrophorum* is implicated in some cases of pharyngitis. The 10 % incidence rate of *F. necrophorum* matches that found by Batty and Wren (2005), as discussed in the previous paragraph.

Jensen *et al.* (2007) examined throat swabs from 61 patients with non-streptococcal tonsillitis and 92 healthy controls (previously mentioned in section 1.3) for the presence of *F. necrophorum*. A real-time PCR assay was developed using TaqMan probes in order to detect both subspecies of *F. necrophorum* via the *gyrase B* gene. Forty eight percent of tonsillitis patients were positive for *F. necrophorum*, compared to 21 % of healthy controls. They concluded that tonsillitis and particularly recurrent tonsillitis may be caused by *F. necrophorum* present in large numbers. These prevalence rates are much higher than those found by Batty and Wren (2005) and Aliyu *et al.* (2004). The culture techniques are likely to be less sensitive than real-time PCR techniques. Culturing techniques require careful sampling, storage, transport and culturing conditions to meet the fastidious requirements of *F. necrophorum*. In contrast, PCR techniques do not require the organism to be viable and can detect small amounts present. The potential reasons for the difference in results between Jensen *et al.* (2007) and Aliyu *et al.* (2004) are discussed in section 1.3. It is unclear whether these *F. necrophorum* infections of pharyngitis and tonsillitis are due to the organism acting as a primary pathogen, or due to tissue damage allowing the organism to invade and spread more easily.

1.4.2 Peritonsillar abscesses

Peritonsillar abscess (PTA), also known as quinsy, is an acute infection that develops between the palatine tonsil and its capsule. In addition to an abscess in the peritonsillar region, symptoms of PTA may also include: sore throat, fever, malaise, uvula deviation, jaw muscle spasm, neck mass and tenderness in the neck (Hsiao *et al.*, 2012). It is thought that PTA may develop as a complication of an existing infection within the tonsils (Jousimies-Somer *et al.*, 1993), with an established infection spreading from the tonsil through local tissue resulting in further tissue damage and subsequent abscess formation (Powell *et al.*, 2013). An alternative theory states that damaged tissue in the peritonsillar region may be associated with abscess formation (Galioto, 2008; Powell *et al.*, 2013). Salivary glands, known as Weber glands, located in the supratonsillar space help to keep the tonsils and surrounding area clean (Passy, 1994; Powell *et al.*, 2013). The saliva excreted from the glands washes the area and contains antimicrobial peptides which help to control the oral flora (Ball *et al.*, 2007). It has been hypothesised that damage to these glands, resulting in scarring, inflammation and blockage, may lead to PTA (Passy, 1994). It has been suggested that the latter theory may supersede the former (Galioto, 2008). Powell *et al.* (2013) states that PTA and acute tonsillitis are distinct infections and describe how oral hygiene, smoking, antibiotic use and other host factors may play a role in triggering pathogenesis. Focussing on controlling these host factors may enable some degree of PTA prevention.

PTA are mostly polymicrobial with *Prevotella*, *Porphyromonas*, *Fusobacterium* and *Peptostreptococcus* spp. the most predominant anaerobic organisms. *Streptococcus pyogenes*, *Staphylococcus aureus* and *Haemophilus influenzae* are also predominant organisms within PTA. Anaerobes are almost ubiquitous in these abscesses (Brook, 2005). *F. necrophorum* is known to be a cause of tonsillar and peritonsillar abscesses (Batty and Wren, 2005).

Yusuf *et al.* (2015) carried out a 10 year epidemiology study by retrospectively reviewing patient records and associated microbiological data from 2004 to 2014. They found that of 27 peritonsillar abscesses, *F. necrophorum* was isolated as the major pathogen from 25 of them (92.6 %).

Klug *et al.* (2009) also conducted a retrospective study, investigating all patients at a Danish hospital with peritonsillar abscess from 2001 to 2006, totalling 847 patients. *F. necrophorum* was isolated from 23 % of cultures, making it the most frequently detected bacteria, above group A *Streptococci* at 17 %. Eighty one percent of the *F. necrophorum* cultures were pure. The isolation rate of 23 % is much smaller than the rate of 92.6 % reported by Yusuf *et al.* (2015), however the higher isolation rate came from a much smaller sample size.

Jousimies-Somer *et al.* (1993) investigated pus samples from 124 patients with peritonsillar abscesses and isolated *F. necrophorum* from 38 % of samples. Those infected with *F. necrophorum* had the highest rate of previous tonsillar or peritonsillar infections (52 %) and had the highest rate of recurrent infections (57 %). Of 15 cases that isolated only anaerobic cultures, 14 contained *F. necrophorum*, and three of these were pure cultures. This suggests that *F. necrophorum* is strongly associated with peritonsillar abscesses.

Riordan (2007) noted that PTA was previously seen as a precursor for Lemierre's syndrome (Lemierre, 1936) however Riordan found PTA in only 7 % of Lemierre's syndrome cases reviewed. There are several routes of infection *F. necrophorum* may spread (discussed in section 1.7.4), and it appears as if causing PTA is not a prerequisite for Lemierre's syndrome, as previously suggested.

1.4.3 Lemierre's syndrome

Lemierre's syndrome (LS) is a severe and life threatening infection in humans that is characterised by oropharyngeal infection, primarily in the palatine tonsils and pharynx (Chirinos *et al.*, 2002), thrombophlebitis of the internal jugular vein, metastatic abscesses and septicaemia. The metastatic abscesses are typically located in the lungs and large joints, however abscesses may also occur in the liver and spleen (Batty and Wren, 2005; Kuppalli *et al.*, 2012; Lyle *et al.*, 2011). A description of the spread of LS is provided in section 1.7.4. *F. necrophorum* is the organism most commonly isolated from cases of LS (Bennett and Eley, 1993; Golpe *et al.*, 1999; Ridgway *et al.*, 2010). Riordan (2007) reported that in 222 LS cases, identified by searching French and English language literature since 1970, an organism other than Fusobacteria was isolated in only 8 % of cases. The other species

implicated include those of *Bacteroides, Peptostreptococcus, Prevotella, Porphyromonas, Eikenella, Enterococcus, Proteus* and *Streptococcus*. LS typically affects children, adolescents and young adults who were previously healthy (Lyle *et al.*, 2011) and tends to occur in the winter months, during the season of sore throats, common colds and flu (see section 1.6 for epidemiology).

1.4.4 Other head and neck infections

Of the anaerobic infections that originate in the head and neck, approximately half involve *Fusobacterium spp.* (Brook, 2015). *F. necrophorum* infection is not required to begin in the pharynx or tonsils in order for a diagnosis of LS. Infections may also originate in the ears, mastoids and teeth (Hagelskjær *et al.*, 1998). Hagelskjær *et al.* (1998) reported that in 49 cases of *F. necrophorum* septicaemia, half were not related to the oropharynx, including eight originating in the gastrointestinal tract, five in the genitourinary tract and seven from skin infections.

1.4.4.1 Otogenic infection

Otitis media is an infection of the middle ear. In a study by Yusuf *et al.* (2015), *F. necrophorum* was found to be the major pathogen in 60 % of acute otitis cases. Mastoiditis is a suppurative infection of the mastoid bone. The bone sits behind the ear and is comprised of air spaces that help drain the middle ear. Mastoiditis can often follow otitis media (Yarden-Bilavsky *et al.*, 2013). Riordan (2007) reports that of the patients with *F. necrophorum* ear infections, 69 % of them also had mastoiditis.

1.4.4.2 Odontogenic infection

F. necrophorum does not typically cause dental infections, as it primarily infects the pharynx and tonsils, rather than dental cavities (Brook, 2005; Eaton and Swindells, 2014), therefore LS originating from a dental infection is rare (Riordan, 2007). However, there are cases such as that reported by Ali *et al.* (2003) where a patient received dental work three weeks prior to the onset of LS, with blood cultures positive for *F. necrophorum*.

Noma, also known as cancrum oris, is a form of gangrene affecting the face that causes facial deformity, oral stenosis, trismus and death in 70 – 90 % of untreated cases (Falkler *et al.*, 1999; Enwonwu *et al.*, 1999). *F. necrophorum*, thought to be acquired via animal faecal matter

contamination, is known to be implicated in noma infections and is thought to be involved in tissue-destruction that helps the lesion to spread. Noma infections typically only occur in children who are immunocompromised, malnourished and have poor oral hygiene. It is particularly prevalent in sub-Saharan Africa (Enwonwu *et al.*, 1999).

1.4.4.3 Deep neck space infection

Deep neck infection, or sepsis in the neck, may include symptoms such as fever, swelling and pain in the neck, sore throat, difficulty swallowing with pain associated, trismus and stridor (Lyle *et al.*, 2011). It commonly begins as odontogenic or oropharyngeal infection (Pinto *et al.*, 2008) and from there may spread throughout the fascial spaces of the neck (Lyle *et al.*, 2011). *F. necrophorum* is known to be capable of causing serious deep neck space infections, spreading through the fascial spaces in the head, neck and around the lungs, potentially also causing abscesses in these regions (Jousimies-Somer *et al.*, 1993; Chow, 1992).

1.4.5 Bacteraemia

It is possible for patients to have *F. necrophorum* bacteraemia without necessarily having LS. For a diagnosis of LS, thrombophlebitis of the internal jugular vein and metastatic abscesses would also need to be present (Kuppalli *et al.*, 2012; Batty and Wren, 2005). Centor *et al.* (2010) reports on cases of bacteraemic tonsillitis and describe the symptoms as including fevers, vomiting, rigors and sweating. It was noted that prompt admission to hospital and early antimicrobial treatment was linked with a lack of metastasis. Improved diagnosis techniques, leading to faster administration of appropriate antimicrobial treatment may therefore reduce the incidence of LS by limiting the progression of the infection at the onset of bacteraemia, if not before.

1.4.6 Metastatic abscesses, joint infections and complications of LS

Metastatic abscesses are a symptom of LS and may be located in the lungs, large joints, liver and spleen (Kuppalli *et al.*, 2012; Batty and Wren, 2005). In a review by Riordan (2007), where 222 cases of LS were studied (discussed previously in section 1.4.3), it was reported that pleuropulmonary lesions, the most common metastatic complication, occurred in 92 % of cases, while empyema, the

collection of pus and fluid particularly in the pleural cavity, occurred in 17 % of cases reviewed. Septic emboli have been described as typical lesions present in LS, and are often part of the inclusion criteria for diagnosis (Riordan, 2007; Lu *et al.*, 2009). Mild disseminated intravascular coagulation (DIC) with thrombocytopenia is another potential complication of LS, reportedly developing in up to 23 % of cases (Hagelskjaer Kristensen and Prag, 2000). Riordan (2007) agrees that thrombocytopenia is not uncommon, however states that DIC only occurred in 4 % of LS cases reviewed. Other reported rates were septic shock developing in 7 % of cases and renal failure in 2 %. The prevalence of abscesses in the joints, liver and spleen were 9.5 %, 5 % and 3.6 % of LS cases, respectively (Riordan, 2007).

1.5 *F. necrophorum* infections in animals

F. necrophorum is not only an infectious agent for humans, but is also the cause of necrobacillosis in animals (Langworth, 1977). It is implicated in liver abscesses and footrot in cattle and sheep (Zhou *et al.*, 2009a). Nagaraja *et al.* (2005) describe these infections, stating that liver abscesses are the most common *F. necrophorum* infection in cattle, and that infection tends to begin in the rumen and spread via septic emboli to the liver. Footrot, also known as interdigital necrobacillosis, is a necrotising infection of the soft tissues on the feet and surrounding skin which often leads to fever and lameness in cattle and other hooved livestock. Damp ground and existing injury to the interdigital skin region are predisposing factors. It is thought that the primary source of infection is faecal secretion. *F. necrophorum* also causes calf diphtheria, also referred to as necrotic laryngitis, and mastitis in cattle (Tan *et al.*, 1996).

These infections have a large economic effect on the farming industry and also concern the welfare of the animals. As a result of this, the majority of research carried out on *F. necrophorum* so far is centred on the bacterium being predominantly an animal pathogen. Riordan (2007) describes the bacterium as a more common and important pathogen in animals than in humans. However, Wright

et al. (2012) describe this as a historical view, highlighting the increasing importance of the infection in humans.

1.6 Epidemiology of *F. necrophorum*

1.6.1 Incidence

As discussed in section 1.4.1, Batty and Wren (2005) reported the incidence of *F. necrophorum* among patients diagnosed with sore throat, tonsillitis and PSTS as 10 %. Aliyu *et al.* (2004) also reported an incidence rate of 10 % among cases of pharyngitis compared to 0 % for healthy controls. Jensen *et al.* (2007), however, reported an incidence rate of 48 % for tonsillitis swabs, compared to 21 % of healthy controls and Centor *et al.* (2015) found a *F. necrophorum* detection rate of 20.5 % for patients compared to 9.4 % for asymptomatic controls using PCR methods. *F. necrophorum* is a significant cause of sore throat, tonsillitis and PSTS and should therefore be routinely screened for in laboratories receiving throat swabs (Batty and Wren, 2005).

The incidence of *F. necrophorum* in PTA appears to be higher, with Yusuf *et al.* (2015) reporting an incidence rate of 92.6 %, as discussed in section 1.4.2. However, Jousimies-Somer *et al.* (1993) reported an isolation rate of 38 % and Klug *et al.* (2009) reported an isolation rate of 23 %. The mean annual incidence in Denmark of PTA was calculated as 41 cases per 100,000 population (Klug *et al.*, 2009).

Lemierre's syndrome is rare, with a reported incidence of 0.9 cases per million inhabitants per year in the southwest of England between 1994 and 1999 (Jones *et al.*, 2001). Other reports give an estimate of between 0.6 and 2.3 per million inhabitants per year based reviewing medical literature (Syed *et al.*, 2007). Hagelskjær Kristensen and Prag (2008) reported the incidence in Denmark as 3.6 cases per million inhabitants per year based on a study between 1998 and 2001. When broken down by age group per year, the incidence was reported as 14.4 cases per million 15 – 24 year olds

for the study duration. It was estimated by Centor (2009) that 1 in 400 *F. necrophorum* pharyngitis cases will progress to LS.

Eaton and Swindells (2014) suggested that rural communities may experience lower incidences of *F. necrophorum* infection, following their observation that the *F. necrophorum* isolation rate of 5.6 % from sore throat swabs during their study was lower than rates observed in London by Batty and Wren (2005) (section 1.4.1). Karkos *et al.* (2009) reported that 10 % of acute sore throats and 21 % of recurring sore throats were attributed to *F. necrophorum*.

1.6.2 Age distribution

Batty and Wren (2005) plotted, in 10 year increments, the patient age distribution from all throat swabs containing *F. necrophorum* or group A streptococcus (GAS). They showed a steady increase in *F. necrophorum* incidence with age up to 40 years, starting with an isolation rate of <4 % for the 0 – 10 years category, increasing to almost 18 % for the 31 – 40 years category. The GAS prevalence remained fairly stable throughout, between 10 and 15 %. For the 41 – 50 age group, the prevalence of both fell to approximately 4 %. In the 10 year epidemiological study by Yusuf *et al.* (2015), the median age of all *F. necrophorum* infections was 19. Broken down into groups, the median age for acute tonsillitis was 30, peritonsillar abscess was 27 and acute otitis was 3 years of age.

LS typically affects previously healthy infants and adolescents (Lyle *et al.*, 2011). Riordan (2007) noted that among LS patients, 89 % were aged between 10 and 35 years and the median age was 19 years. Ridgway *et al.* (2010) describe that 70 % of LS cases are in the 16 – 25 year old group. This is in agreement with Brazier *et al.* (2002) who noted the peak age range for *F. necrophorum* bacteraemia is between 16 and 23 years old. They also describe those over 65 as having the next highest incidence rate. Those over the age of 65 are more likely to have a weaker immune system which could explain the peak in incidence for that age group. As for the peak age range between 16 and 23 years, this could be related to the similar peak age range for Epstein-Barr virus infections (Brazier, 2006). There may be symbiotic relationship between the virus and bacterium, or it could

be that simply close mouth-to-mouth contact between individuals in that age group leads to increased transmission of the bacterium and therefore higher rates of infection.

1.6.3 Sex ratio

F. necrophorum infections tend to be more common in males than females, according to Brazier *et al.* (2002), who showed that 68 % of *F. necrophorum* bacteraemia cases in England and Wales were male. Hagelskjær *et al.* (1998) demonstrated that, in Denmark, two thirds of Lemierre's syndrome cases with tonsillitis as the primary focus were male. Riordan (2007) also discusses the male predominance, although suggests no explanation. This does not always appear to be the case, however, with a strong female predominance of 74 % in a Swedish study by Björk *et al.* (2015).

1.6.4 Seasonal and geographical distribution

A seasonal variation in the number of cases of *F. necrophorum* infections have been reported on a number of occasions. Brazier *et al.* (2002) noted a peak in incidence in England and Wales from January to March. Eaton and Swindells (2014) reported that during their UK study the peak in incidence was between April and June. A slight seasonal variation was reported by Hagelskjær Kristensen and Prag (2008) in Denmark, with peaks occurring during late winter and early autumn months. Riordan (2007) states that reports of infections are mostly from Europe and North America, suggesting ethnic group and geographical area play a role in the incidence. This could be due to genetic differences between ethnic groups that affect the bacteria's ability to cause an infection. These may be similar factors to those pertaining to greater numbers of infections in males compared to females. Alternatively, differing methods of detecting and reporting cases may be the cause of this geographical variation in incidence.

1.6.5 Factors causing increase in incidence

Lemierre's syndrome had become the forgotten disease (Hagelskjaer Kristensen and Prag, 2000) following the introduction of penicillin treatment in the 1940s, however an increase in cases has been reported in several publications more recently. Batty and Wren (2005) noted the increase in isolation rate of *F. necrophorum* from patients in the accident and emergency department. This rise

is supported by Brazier (2006) who reported an increase in the number of *F. necrophorum* bacteraemia referrals to the Anaerobe Reference Unit. Hagelskjær *et al.* (1998) also reported the increase in the incidence of cases over time. A review of medical literature containing Lemierre's syndrome as the key word by Karkos *et al.* (2009) found an increase in the number of relevant articles over time, from six between 1980 and 1990, 50 between 1991 and 2000, and 121 between 2001 and 2008. Several articles discuss the reasons for this alleged increase in incidence, suggesting reasons such as a reduction in the number of tonsillectomies and a reduction in antibiotic prescriptions for cases of sore throats (Kuppalli *et al.*, 2012; Wright *et al.*, 2012; Brook, 2015). Additional suggestions were an increase in awareness of *F. necrophorum* infections and population changes in areas that have seen an increase (Wright *et al.*, 2012). An improvement in laboratory techniques regarding methods for blood culture, isolation and identification of anaerobic organisms, and the use of more sensitive detection methods, such as polymerase chain reaction were also suggested as reasons for an increase in reported incidence (Brook, 2015). There is a possibility that the importance of *F. necrophorum* as a pathogen has been greatly underestimated for some time and the improvement of techniques and reporting has led to the perceived increase in incidence. Alternatively, incidence could indeed be increasing. Surveillance of the organism would be beneficial in order to track incidence and changes to the organism.

1.6.6 Mortality

Lemierre's syndrome is associated with high morbidity and mortality, often due to delays in diagnosis (Lyle *et al.*, 2011). For patients who do not receive treatment for LS the mortality rate is up to 90 % (Wright *et al.*, 2012). However for those who do receive treatment (section 1.10) there is usually a full recovery (Brazier *et al.*, 2002). Karkos *et al.* (2009) reported a mortality rate of 5 % among patients who had received treatment for LS, while Centor (2009) reviewed several case series and reported a rate of 4.6 %. Hagelskjær Kristensen and Prag (2008) reported a mortality rate of 9 % for LS, however the mortality rate for disseminated *F. necrophorum* infections was 26 %. The median age for these two groups was 20 and 66, respectively. The variation in mortality rates between these groups may be due to the group with disseminated infection containing

predominantly elderly patients, who may have weakened immune systems, leading to a higher risk of death from the organism.

1.7 *F. necrophorum* Pathogenesis

1.7.1 Source of infection

The source of *F. necrophorum* infection is unknown as it is unclear whether the organism is a commensal or primary pathogen (section 1.3) (Riordan, 2007). If a commensal, then *F. necrophorum* is present in the oropharynx but may potentially cause infection as a primary pathogen (Chirinos *et al.*, 2002), or opportunistically as a secondary invader (section 1.7.3). In either case the mucosal tissue may previously be compromised due to other factors. If *F. necrophorum* is not a commensal then the source of infection is less clear. Human-to-human transmission is thought to be possible, following close contact between a carrier and susceptible individuals (Batty *et al.*, 2005). Aside from intimate contact such as kissing, sharing a toothbrush may also potentially transmit the organism between individuals (Riordan, 2007). It is known that *F. necrophorum* is present in the gastrointestinal tracts of animals, such as cattle (Narayanan *et al.*, 1997), and therefore it may be possible that infection may occur via the faecal-oral route due to contamination from animal faeces, as previously mentioned in section 1.3.

1.7.2 Host factors

Certain host factors may enable *F. necrophorum* to invade more easily, such as the age and gender of an individual, as discussed in section 1.6.2 and 1.6.3. The health of the oral mucosa is likely to be important. If the mucosal barrier is compromised or damaged due to an existing or previous infection, or trauma from an injury or surgery, then it will be easier for *F. necrophorum* to invade and establish an infection (Brook, 2015; Syed *et al.*, 2007; Riordan, 2007). Oral/dental hygiene is important in order to maintain a healthy microflora and to prevent periodontal disease. The oral microflora may be disrupted by the use of antibiotics and subsequently allow colonisation by a

pathogen, such as *F. necrophorum* (Powell *et al.*, 2013). Smoking has also been linked to an increased risk of respiratory infections by damaging the normal oral microflora and mucosal barrier, acting synergistically with bacterial toxins and compromising the inflammatory response and the functioning of leukocytes (Bagaitkar *et al.*, 2008). Variations in the immune systems and genetics between individuals may affect the level of protection against *F. necrophorum* infection (Powell *et al.*, 2013).

1.7.3 Coinfection

Brook (2015) speculates that adhesion and invasion are possibly dependent on co-infection with viruses such as Epstein-Barr. Eaton and Swindells (2014) also suggest that viral co-infection may be a precursor for invasive *F. necrophorum* disease. This is supported by the overlap in peak ages of LS and Epstein Barr virus (Brazier, 2006), as discussed in section 1.6.2, and the peak in cases during the winter months coinciding with the common cold and influenza season, which could suggest more individuals are immunocompromised and therefore prone to infection, or that *F. necrophorum* is capable of coinfection with viruses present in the throat. Batty *et al.* (2005) suggest that the bacteria is acquired simultaneously with infectious mononucleosis. There have also been reports regarding coinfection and synergy between group A β-haemolytic streptococcus and other aerobic and anaerobic bacteria involved in mixed tonsillitis infections (Brook and Gillmore, 1996).

1.7.4 Spread from tonsil to vein

F. necrophorum infections will most commonly cause only a primary infection of pharyngitis or tonsillitis, however these infections may be persistent or recurrent (Batty and Wren, 2005). In rare cases, *F. necrophorum* infection may progress to LS, which is thought to occur in a series of stages (Figure 2) most commonly beginning in the tonsils and pharynx (Chirinos *et al.*, 2002). Following initial infection the organism may invade the lateral pharyngeal space, the peritonsillar vein and the internal jugular vein (Ridgway *et al.*, 2010; Chirinos *et al.*, 2002). The organism may progress to cause bacteraemia, septic emboli and subsequently metastatic complications. Complications may be pertaining to cardiovascular, neurological, pulmonary, musculoskeletal, liver and renal systems

(Ridgway *et al.*, 2010). The exact pathogenic mechanisms of *F. necrophorum* attachment, invasion, proliferation and host evasion, along with host inflammation and coagulation initiation mechanisms have not yet been explained. Inflammation may cause disruption to local tissue that promotes invasion (Wright *et al.*, 2012) whilst damage to the mucocutaneous barrier may also allow for Fusobacterial invasion (Brook, 2015).
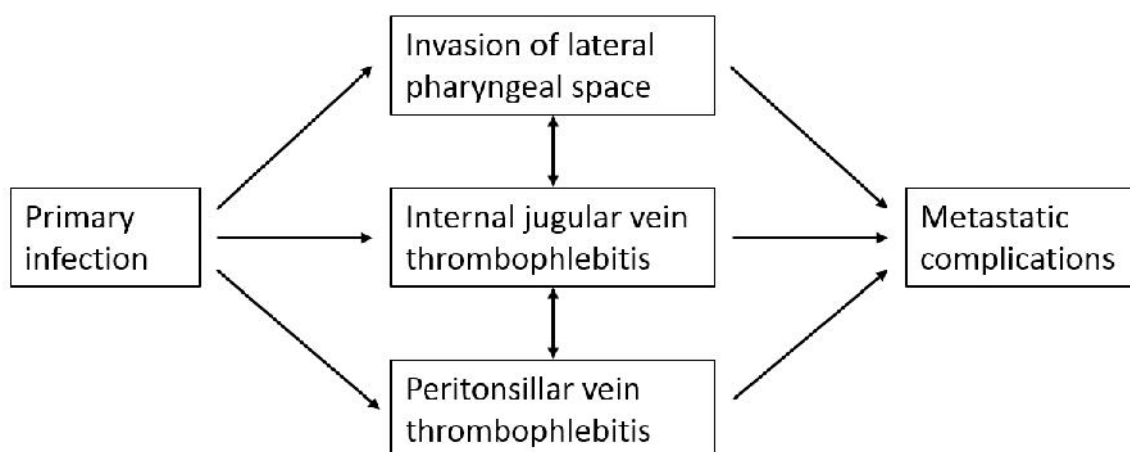
Figure 2: Clinical progression of Lemierre's syndrome following primary infection of the throat and/or tonsils with F. necrophorum.

1.8 Virulence factors of *F. necrophorum*

The human body has evolved innate and adaptive host defence mechanisms to protect itself from invading pathogens. Epithelial cells contributing to the oral mucosa provide a physical barrier, while also producing and secreting inflammatory cytokines and antimicrobial peptides. Mucous provides an extra barrier, assisting in the prevention of bacterial adherence and penetration into the tissue and bloodstream. Mucous contains lysozyme, which digests peptidoglycan found in bacterial cell walls, and lactoferrin, which prevents the growth of bacteria by sequestering iron (Wilson *et al.*, 2011). Immunoglobulins are also present and can identify and neutralise potentially pathogenic material, while saliva is also thought to contribute to the defence process by containing receptors to lipopolysaccharide present on the bacterium (Sugawara *et al.*, 2002). There are also a variety of

immune cells which defend against invading bacteria by phagocytosis. These include polymorphonuclear cells (PMNs), monocytes, macrophages and dentritic cells and they ingest and kill bacteria. They do this with the help of complement proteins and cytokines. Complement proteins label bacteria for phagocytosis and also attract PMNs to the site of infection. Cytokines also provide cell signalling to PMNs (Wilson *et al.*, 2011). There are several putative virulence factors which may be utilised by *F. necrophorum* to enable it to colonise more effectively and to avoid these host defence mechanisms.

1.8.1 Leukotoxin

Many bacteria produce and secrete soluble proteins known as exotoxins that enter host cells and alter the physiology to cause pathological effects (Barbieri, 2009). A type of toxin, referred to as a leukotoxin is produced by *F. necrophorum* (Tadepalli *et al.*, 2008a) and is toxic against leukocytes, particularly PMNs (Narayanan *et al.*, 2002a). It is believed to be the main virulence factor of this species (Tan *et al.*, 1994a). The killing of these PMNs prevents the invading bacteria from being engulfed and therefore allows them the opportunity to establish an infection. The leukotoxin is discussed further in chapter 5.

1.8.2 Endotoxin

Endotoxic lipopolysaccharide (LPS) is a bacterial membrane component that consists of a polysaccharide chain, known as the O-antigen, and a toxic lipid segment, known as Lipid A, that is released upon bacterial lysis (Caroff and Karibian, 2003). The lipid segment interacts with host macrophages and neutrophils and promotes the uncontrolled secretion of proinflammatory cytokines and nitric oxide resulting in endotoxic shock (Wilson *et al.*, 2011). LPS is present in the outer membrane of *F. necrophorum* and has been shown to be lethal to mice and rabbits (Tan *et al.*, 1996). It is thought to induce disseminated intravascular coagulation and be involved in abscess formation by creating an anaerobic microenvironment (Nagaraja *et al.*, 2005). According to Garcia *et al.* (1998), *Fnn* and *Fnf* contain endotoxin lipid A in a 4:1 ratio, respectively, which may have some relevance in relation to levels of virulence, considering that lipid A is toxic.

1.8.3 Haemolysin

Haemolysins are a type of pore-forming toxin that damage the membrane of the host cell (Wilson *et al.*, 2011). They are cell-associated, extracellular proteins that lyse erythrocytes (Nagaraja *et al.*, 2005). *F. necrophorum* is known to produce a haemolysin that is thought to help the organism to increase iron acquisition from the host, which stimulates bacterial growth. It has also been suggested that it may also reduce the transportation of oxygen to the infection site, thus making the environment more anaerobic (Tan *et al.*, 1996; Nagaraja *et al.*, 2005). When *F. necrophorum* is grown on agar, β-haemolysis can usually be observed around the colonies (Batty and Wren, 2005).

1.8.4 Haemagglutinin

Haemagglutinin causes the clumping of red blood cells. It is produced by *F. necrophorum* and is thought to be a cell wall associated protein (Nagaraja *et al.*, 2005; Kanoe *et al.*, 1997). Tan *et al.* (1996) discuss the purification and characterisation of the *F. necrophorum* haemagglutinin by Nagai *et al.* (1984) and describe it as filamentous and having a molecular weight of 19 kDa. PCR primers specific for the haemagglutinin gene in *Fnn* have been used to subspeciate strains of *F. necrophorum* (Kumar *et al.*, 2013; Aliyu *et al.*, 2004). It is unclear whether this method was chosen due to an understanding that *Fnf* strains lack the gene or due to strict specificity of the primers. However, a haemagglutination test using chicken red blood cells *in vitro* has been used to subspeciate strains also, with only *Fnn* causing haemagglutination (Narongwanichgarn *et al.*, 2001). This is suggestive of *Fnf* strains either lacking the gene, or lacking sufficient haemagglutinin on the cell surface to cause clumping. Haemagglutinin also mediates attachment to host cells (Nagaraja *et al.*, 2005) which is likely to assist the organism in establishing an infection.

1.8.5 Adhesins and invasins

Forming an attachment to the mucosal surface of the host is key to pathogenesis (Riordan, 2007). As mentioned in section 1.8.4, the *F. necrophorum* haemagglutinin is thought to be one mechanism that mediates *F. necrophorum* attachment to host cells (Nagaraja *et al.*, 2005). In addition to this,

it has been suggested that the organism possesses fimbriae, also known as pili, in order to aid attachment (Riordan, 2007; Miyazato *et al.*, 1978).

Brook and Walker (1986) used electronmicroscopy to show that *F. necrophorum* possesses a mucopolysaccharide capsule. This protects the organism from harmful immune secretions and helps it avoid detection by the host. Polysaccharide capsules prevent complement proteins from binding to the bacteria and therefore help to avoid opsonisation and evade phagocytosis (Wilson *et al.*, 2011; Brook and Walker, 1986). Polysaccharide capsules may also be involved in adhesion (Wilson *et al.*, 2011).

F. nucleatum, F. periodonticum and *F. simiae*, all oral species, are known to express a conserved adhesin known as FadA which binds to surface proteins of oral mucosa KB cells (Han *et al.*, 2005). It was not revealed whether *F. necrophorum* also expresses this adhesin and it would therefore be of interest to search for this within *F. necrophorum* genome data and by PCR techniques.

It has been demonstrated that *F. necrophorum* is capable of adhering to ruminal cells by Takayama *et al.* (2000), who suggest that collagen is responsible for mediating this attachment. It is also thought that using collagenase may be a method that enables *F. necrophorum* to invade the host. Rather than the organism expressing its own collagenase, a type of endopeptidase, it appears that *F. necrophorum* and *F. nucleatum* are both capable of inducing Collagenase 3 mRNA production by the host, via activation of multiple cell signalling systems (Uitto *et al.*, 2005).

Proteolytic enzymes, or proteases, cause damage to host tissue and therefore promote invasion of the host mucosa and spreading of the infection (Wilson *et al.*, 2011). The presence of a *F. necrophorum* protease is also discussed by Tan *et al.* (1996).

1.8.6 Platelet aggregation

The aggregation of platelets, leading to thrombophlebitis and disseminated intravascular coagulation, is one of the main symptoms of Lemierre's syndrome and is therefore of great relevance (Tan *et al.*, 1996; Riordan, 2007). Kanoe and Yamanaka (1989) investigated *F. necrophorum* bovine platelet aggregation and demonstrated that it is mediated by haemagglutinin.

When *F. necrophorum* cultures were pre-treated with haemagglutinin antiserum, platelet aggregation was strongly inhibited in a dose-dependent manner up to 64 %. Forrester *et al.* (1985) showed that 13 out of 16 biovar A strains could aggregate human platelets, whereas none of the biovar B strains could. This demonstrates the greater virulence of biovar A and the need to further understand how the less virulent strains maintain infection.

1.9 Diagnosis of *F. necrophorum* infections

1.9.1 Clinical diagnosis

Factors contributing to a diagnosis of LS are pharyngitis within the preceding 4 weeks, thrombophlebitis of the jugular vein, evidence of metastatic lesions and isolation of *F. necrophorum* or other bacteria from a typically sterile body site, such as the blood (Wright *et al.*, 2012). A high fever may often be the only indicator during the initial stage of LS (Kuppalli *et al.*, 2012) as the pharyngeal symptoms may have resolved by this stage (Wright *et al.*, 2012). The diagnosis of LS is most often determined through laboratory investigations rather than clinical observations due to the rarity of the disease. This unfortunately leads to delays in diagnosis which allow for further disease progression (Ridgway *et al.*, 2010).

1.9.2 Radiology/imaging

Diagnostic imaging can be used in order to confirm the presence of thrombosis in the internal jugular vein, emboli in the lungs and abscesses (Lyle *et al.*, 2011; Chirinos *et al.*, 2002; Kuppalli *et al.*, 2012). Ultrasonography is considered to be a cost effective, minimally invasive, readily available method, therefore helping to provide a rapid diagnosis. However, it provides poor detection of newly formed clots and in areas underneath the jaw or collarbone (Kuppalli *et al.*, 2012; Syed *et al.*, 2007). Computed tomography (CT) is the best method for detecting thrombosis in the internal jugular vein, according to Ridgway *et al.* (2010) and Kuppalli *et al.* (2012). It does involve the use of X-rays, although it is more sensitive (Syed *et al.*, 2007). Magnetic resonance imaging (MRI) is the

most accurate and does not require exposure to X-rays, however it is very expensive and not widely available and therefore has limited involvement in diagnosing LS (Syed *et al.*, 2007; Kuppalli *et al.*, 2012).

1.9.3 Laboratory diagnosis

Throat swabs with subsequent laboratory analysis by culture and/or PCR are the methods to determine the cause of bacterial pharyngitis, tonsillitis and PTA. In cases of persistent sore throat or peritonsillar abscess, the current Public Health England guidelines are to test pus aspirate or throat swabs for *F. necrophorum* (Public Health England, 2015). Batty and Wren (2005) suggest that *F. necrophorum* should be routinely screened for, however it is unknown how widely this advice has been adhered to.

As discussed in section 1.2.2, *F. necrophorum* is best cultured on fastidious anaerobe agar (FAA) and grown under anaerobic conditions at 37 °C for 48 hours (Syed *et al.*, 2007; Batty and Wren, 2005). The organism can then be purified if needed and identified using Gram staining and biochemical tests such as indole production, lipase production and alkaline phosphatase testing, or commercial API testing (Batty *et al.*, 2005). PCR techniques can be used for rapid identification of *F. necrophorum*, as discussed in section 1.2.3.

1.10 Treatment of *F. necrophorum* infections

1.10.1 Antibiotics

According to Centor *et al.* (2015), the United States do not have any guidelines for *F. necrophorum* pharyngitis treatment, although they do support antibiotic prescriptions for GAS pharyngitis. In the UK the National Institute for Health and Care excellence, also known as NICE, published guidelines detailing three strategies for what it describes as self-limiting respiratory tract infections, including acute cases of otitis media, sore throat, pharyngitis and tonsillitis. The strategies are no antibiotic prescribing, delayed antibiotic prescribing and immediate prescribing (Tan *et al.*, 2008). As

discussed in section 1.6.5, a decrease in antibiotic prescribing for sore throats has been linked with the reported increase of *F. necrophorum* infections (Brook, 2015). The NICE guidelines emphasise the natural resolution time of each infection, with acute otitis media resolving in four days, and acute sore throat, acute pharyngitis and acute tonsillitis usually resolving in one week (Tan *et al.*, 2008).

Brazier *et al.* (2002) reported antimicrobial susceptibility results of 100 *F. necrophorum* clinical isolates and showed 2 % resistance to penicillin and 15 % resistance to erythromycin. In an American study, 22.7 % of 22 *F. necrophorum* isolates were determined to be β-lactamase producers, based on nitrocefin disk testing (Appelbaum *et al.*, 1990), suggesting that penicillin was not such an appropriate choice in that region unless in combination with a β-lactamase inhibitor. Brazier *et al.* (2002) also reported that there was no resistance to metronidazole, amoxicillin/clavulanate, cefoxitin, chloramphenicol, clindamycin or imipenem. However *F. necrophorum* is resistant to gentamicin and quinolones, and tetracyclines are relatively ineffective (Riordan, 2007).

Antimicrobial therapy for LS is usually prolonged (approximately 3 – 6 weeks) with high dosage due to the inaccessible nature of septic emboli and abscesses (Ridgway *et al.*, 2010) and also to prevent local and systemic spread (Brook, 2015). Kuppalli *et al.* (2012) recommend a combination of 500 mg intravenous metronidazole every 8 hours and 2 g intravenous ceftriaxone every 24 hours for adults. Brazier *et al.* (2002) recommend penicillin and metronidazole as appropriate treatments. For antibiotic treatment of PTA it is recommended that both GAS and oral anaerobes should be targeted (Galioto, 2008). A review by Wright *et al.* (2012) discusses the factors contributing to metronidazole being the first choice antimicrobial, stating a lack of resistance and excellent bioavailability and tissue penetration.

1.10.2 Surgery

In severe cases of *F. necrophorum* infection, in addition to antimicrobial treatment, patients may require tonsillectomy (Björk *et al.*, 2015) or abscess drainage (Riordan, 2007; Brook, 2015). Kuppalli

et al. (2012) describe the drainage of pus and debridement of necrotic tissue as essential. In a retrospective study on PTA patients in Denmark, 847 patients were included. Of these, 726 patients had a tonsillectomy and 110 patients had pus aspirated. *F. necrophorum* was isolated from 191 of these cases and GAS from 141 (Klug *et al.*, 2009).

1.10.3 Anticoagulants

In cases of LS, anticoagulation therapy such as heparin may also be given to patients. It has been controversial due to the argument that it is problematic if the patient requires surgery (Hagelskjær Kristensen and Prag, 2008) and that clots generally resolve on their own. However, it has been suggested that a faster resolution of septic emboli is clinically important and that anticoagulation therapy should therefore be used where thrombosis occurs (Ridgway *et al.*, 2010).

1.11 Genomics of *F. necrophorum*

At the beginning of this project, in September 2011, the only available *Fusobacterium* genome sequence was that of *F. nucleatum*. The genome sequencing and metabolic analysis was completed by Kapatral *et al.* (2002) and deposited in GenBank under accession number AE009951. The genome features a single circular chromosome, containing 2.17 Mb, a GC content of 27 % and 2,067 predicted open reading frames. This annotated genome sequence was used to make predictions regarding the *F. necrophorum* genome during this project, such as the expected size of the genome and genes that it may contain. *F. necrophorum* and *F. nucleatum* occupy similar niches, with *F. nucleatum* in the oral cavity (Han, 2015) and *F. necrophorum* in the oropharynx (Brazier *et al.*, 2002), therefore it seems plausible that they may share some virulence mechanisms.

Several months into the project, the first *F. necrophorum* genome became available on the GenBank database. This has since been followed by the addition of a collection of *F. necrophorum* genomes in 2014, leading to a total of eleven. The available genomes are strains isolated from bovine, deer

and human origin and are approximately 2 Mb in size, as was expected based on the *F. nucleatum* genome. This topic is discussed further in chapter 4.

1.12 Project aims

F. necrophorum appears to be a common cause of acute pharyngitis and tonsillitis and a rare cause of more severe infections of the head and neck. Several virulence factors appear to be involved in pathogenesis, however there is a limited understanding of them. The *F. necrophorum* genome is also largely understudied.

This project intended to increase the understanding of *F. necrophorum* by using next generation sequencing to analyse the genome and follow up laboratory analysis on virulence determinants of interest. The project aims were as follows:

1. Collect clinical strains of *F. necrophorum* from human origin and a reference strain of both *F. necrophorum* subspecies *necrophorum* and *F. necrophorum* subspecies *funduliforme*. Grow and characterise strains using a combination of conventional microbiological techniques and molecular techniques.

2. Use next generation sequencing technologies to *de novo* sequence the reference strains of both subspecies *Fnn* and *Fnf*, along with a clinical strain known to have caused Lemierre's syndrome in a patient.

3. Assemble sequence data and annotate with gene products using *F. nucleatum* as a reference. Use manual curation alongside automatic annotation in order to increase the likelihood of novel and interesting characteristics of *F. necrophorum* being discovered.

4. Search databases for additional *Fusobacterial* sequences, particularly those related to virulence, such as the leukotoxin. Perform sequence alignments to look for similarities and conserved regions between strains.

5. Use PCR to determine the presence of putative virulence factors in the strain collection, such as adhesins and proteases.

6. Carry out *in vitro* assays on virulence determinants of interest.

Upon completion of the genome annotation work, the virulence determinants chosen for *in vitro* analysis were the leukotoxin, the organism's primary virulence factor, and ecotin, a serine protease inhibitor not previously described in *F. necrophorum*. The carriage and sequence of the *leukotoxin* gene was investigated within the three sequenced strains and the larger collection of *F. necrophorum* clinical strains. The cytotoxic effects on the HL-60 cell line and human white blood cells were also investigated. Ecotin was investigated for inhibitory effects against human plasma clotting and individual protease targets, following the production and purification of a recombinant protein.

Chapter 2:

Materials and methods

2.1 Bacterial stocks

Bacterial strains used in this study were *Fusobacterium necrophorum* and chemically competent *Escherichia coli*. *F. necrophorum* reference strains JCM 3718 and JCM 3724 from the Japan Collection of Microorganisms, and clinical strain ARU 01, were provided by the Anaerobe Reference Unit, Cardiff (Table 1). Strains JCM 3718 and JCM 3724, subspecies *necrophorum* and *funduliforme*, respectively, are both of bovine liver abscess origin, and strain ARU 01 is known to have caused Lemierre's syndrome in a patient. A collection of 26 human clinical isolates, provided by Antonia Batty, University College Hospital London (Table 2) were also used in this study. No identifying patient data was provided with the strains. Chemically competent *E. coli* cells were Top10 and BL21(DE3) strains (Table 3).

Table 1: Reference strains and clinical strain provided by the Anaerobe Reference Unit, Cardiff, UK.

Strain/isolate	Subspecies
JCM 3718	*Fusobacterium necrophorum* subspecies *necrophorum*
JCM 3724	*Fusobacterium necrophorum* subspecies *funduliforme*
ARU 01	*Fusobacterium necrophorum* subspecies *funduliforme*

Table 2: Clinical strain collection provided by Antonia Batty, University College Hospital London, UK.

Collection	isolate numbers
Fusobacterium necrophorum clinical strains	1, 5, 11, 21, 24, 30, 39, 40, 41, 42, 52, 59, 62, 70, 80, 82, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95

Table 3: Chemically competent strains used for cloning.

Strain	Species	Genotype	Source
Top10	*Escherichia coli*	F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lacZ*ΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	Ms Emma Bentley, University of Westminster
BL21(DE3)	*Escherichia coli*	F- ompT hsdSB (rBmB-) gal dcm (DE3)	Dr Anatoliy Markiv, University of Westminster

2.2 Plasmid stocks

A pET-16b plasmid was provided by Dr Anatoliy Markiv, University of Westminster, for use in section 2.14.

2.3 Cell line stocks

The HL-60 cell line (ATCC CCL-240) was purchased from Sigma, Gillingham. The cell line was derived from peripheral blood leukocytes from a 36 year old Caucasian female with acute promyelocytic leukaemia.

2.4 Reagents and solutions

Solutions were prepared using deionised water from a Purite Select deioniser. Sterilisation conditions were by autoclaving at 121 °C for 15 minutes except where alternative conditions are stated. Acetic acid, hydrochloric acid and Dimethyl sulfoxide (DMSO) were purchased from Sigma,

Gillingham. Ethanol and isopropanol were purchased from VWR BDH PROLABO Chemicals, Leicestershire.

Methods for preparing ethylenediaminetetraacetic acid (EDTA), 50X Tris base, acetic acid and EDTA (TAE) buffer, 6X Gel electrophoresis loading dye, 6X Laemmli buffer, 10X Tris-Glycine buffer, Bradford reagent and Phosphate buffered saline (PBS) are listed in Appendix 1.

2.5 Media and antibiotics

Methods for preparing LB agar and broth, fastidious anaerobe agar (FAA), fastidious anaerobe broth (FAB), brain heart infusion (BHI) broth, complete cell culture medium, SOC media, 100 mg/ml ampicillin and 100X penicillin-streptomycin solution are listed in Appendix 2.

2.6. Bacterial culture and storage

F. necrophorum strains were cultured on fastidious anaerobe agar (LabM, Lancashire) supplemented with 5 % defibrinated horse blood (TCS Bioscience, Buckingham) and incubated in a sealed chamber for 48 hours under anaerobic conditions at 37 °C using AnaeroGen sachets (Oxoid, Basingstoke). *F. necrophorum* strains are usually 1 – 3 mm in diameter, cream-grey in colour with an irregular or smooth edge, exhibiting beta-haemolysis and a butyric acid odour. For long term storage, cultures were frozen at -80 °C in a solution of sterilised 15 % glycerol (Sigma, Gillingham).

2.7 Identification using biochemical tests

Tests were used to confirm the biochemical profiles of the strain collection matched that of *F. necrophorum*. These tests consisted of Gram stains, oxidase and catalase tests and a test for indole production using the Bactidrop Spot Indole test (Remel, Dartford). Expected results were Gram negative pleomorphic rods, often with some long filaments present. Oxidase and catalase tests

should be negative with a positive result for indole production. Methods for carrying out a Gram stain, oxidase test, catalase test and indole production test are detailed in Appendix 3.

2.8 Nucleic acid analysis

2.8.1 DNA extraction

DNA was extracted using a Qiagen DNA mini kit. Several loopfuls of colonies from a 48 hour anaerobic culture were suspended in 1 ml sterile deionised water (SDW) and a 200 µl sample was aliquoted into a microfuge tube. 200 µl buffer ATL was added and the mixture vortexed, followed by adding 20 µl 20 mg/ml proteinase K and vortexing again. The mixture was incubated at 56 °C for 10 minutes, vortexed every 2 minutes, and then briefly centrifuged in an Eppendorf 5415 D benchtop centrifuge to remove drops from the inside of the lid. 200 µl buffer AL was then added and pulse vortexed for 15 seconds, followed by incubating at 70 °C for 10 minutes and briefly centrifuging again. 200 µl ethanol (96 – 100 %) was added to the sample, which was mixed again by pulse vortexing for 15 seconds, and briefly centrifuged. The mixture was carefully applied to a spin column in a 2 ml collection tube and centrifuged at 8000 rpm for 1 minute. The filtrate was discarded. 500 µl buffer AW1 was added to the column before centrifuging at 8000 rpm for 1 minute and discarding the filtrate. 500 µl buffer AW2 was added to column. This was centrifuged at 13,000 rpm for 3 minutes and the filtrate was discarded before an additional spin at 13,000 rpm for 1 minute. The collection tube was discarded and the column placed in a new microfuge tube. 200 µl buffer AE was added and the sample was incubated at room temperature for 5 minutes. The DNA was collected by centrifuging at 8000 rpm for 1 minute. The extracted DNA was stored in 1.5 ml tubes at -20 °C. Aliquots of the extracted DNA were diluted 1 in 20 and stored at -20 °C in 100 µl aliquots for use in PCR.

2.8.2 Polymerase chain reaction (PCR)

PCR master mixes were prepared using the components in Table 4. The components, excluding the DNA template, were added to a single 1.5 ml microfuge tube in sufficient quantities for multiple reactions, and then dispensed in 20 µl aliquots into 0.2 ml PCR tubes. DNA template (diluted 1 in 20 from stock) and a molecular grade water (Fisher Scientific, Loughborough) negative control were added to the tubes, labelled with the corresponding sample name or number.

PCR was performed using a BioRad DNA Engine Peltier Thermal Cycler. Conditions were an initial denaturation step of 95 °C for 5 minutes, then 35 cycles of denaturing at 95 °C for 30 seconds, annealing at 55 °C for 30 seconds then extension at 72 °C for 1 minute. This was followed by a final extension step of 72 °C for 5 minutes before holding at 8 °C.

Primer pairs used in PCR are listed in Table 20, Table 30, Table 42 and Appendices 10, 11 and 14.

Table 4: PCR components for each tube

Component	Source	Quantity
Taq PCR master mix kit	Qiagen, Crawley	12.5 µl
Forward primer (10 µM)	Synthesised by Eurofins MWG Operon	2.5 µl
Reverse primer (10 µM)	Synthesised by Eurofins MWG Operon	2.5 µl
Molecular grade water	Fisher Scientific, Loughborough	2.5 µl
DNA template		5 µl

2.8.3 Gel electrophoresis

Gel electrophoresis was used to visualise amplified PCR products. Dehydrated, molecular grade agarose was obtained from Lonza, Rockland, USA. 0.8 % (w/v) or 1 % (w/v) solutions were prepared in 1X TAE buffer. Solutions were boiled in a microwave until fully dissolved and then cooled to 50 °C – 60 °C. 10,000X GelRed nucleic acid gel stain (Biotum, Hayward, CA, USA) was added to the cooled gel solution as a 1 in 10,000 dilution, which was then swirled and poured into a gel cast.

Plasmids were electrophoresed on 0.8 % (w/v) agarose gels with a GeneRuler 1 kb DNA ladder and 100 bp DNA ladder (Thermo Fisher Scientific, Hertfordshire), while PCR fragments were electrophoresed on 1 % (w/v) agarose gels with a 100 bp ladder. DNA products were mixed with 6X loading dye before loading into the agarose gel well. All gels were run with a negative control sample. The electrophoresis running buffer was 1X TAE buffer. Gels were visualised using a UV imaging system, UVIPRO (UVITEC, Cambridge). Plasmids and PCR fragment sizes were determined by comparison to the DNA ladder standards run in parallel.

2.8.4 Real-time PCR using TaqMan probes

TaqMan probes were used in conjunction with *gyrB* primers in a RotorGene Q PCR machine in order to subspeciate the *F. necrophorum* isolates. The *gyrB* primers (Table 20) and probes used were taken from the paper by Jensen *et al.* (2007) and synthesised by Eurofins MWG Operon. The *gyrB* primers were diluted to a working concentration of 4 µM and the probes were diluted to a working concentration of 1 µM. See Table 5 for 5' – 3' probe sequences. The *Fnn* probe had a FAM fluorophore (absorbance: 495 nm; emission 520 nm) and the *Fnf* probe had a JOE fluorophore (absorbance: 520 nm; emission: 548 nm). Both probes used black hole quencher (BHQ) 1 (quenching range: 480 – 580 nm).

Master mixes were prepared using the components in Table 6. The components, excluding the DNA template, were added to a single 1.5 ml microfuge tube in sufficient quantities for multiple reactions, and then dispensed in 20 µl aliquots into 0.2 ml PCR tubes. DNA template (diluted 1 in 20 from stock) and a molecular grade water (Fisher Scientific, Loughborough) negative control were added to the tubes, labelled with the corresponding sample name or number. The probes were tested in separate reaction tubes.

Cycle conditions were an initial denaturation step of 95 °C for 5 minutes, then 40 cycles of 95 °C for 30 seconds followed by 60 °C for 60 seconds while acquiring on the green and yellow channels. The subspecies *necrophorum* probe was detected on the green channel and the subspecies

funduliforme was detected on the yellow channel. The RotorGene Q software was used to analyse results.

Table 5: TaqMan probes used in conjunction with gyrB PCR primers to subspeciate F. necrophorum strains using real-time PCR.

Probe	5' – 3' probe sequence
Subsp. *necrophorum*-specific *gyrB* probe	FAM-TCTACTTTGGAGGTTGGAGAAACAAC-BHQ 1
Subsp. *funduliforme*-specific *gyrB* probe	JOE-TCCGCTTTAGAGGCTGGAGAAACGAC-BHQ 1

Table 6: Reaction components for each real-time PCR tube

Component	Source	Quantity
RotorGene probe PCR kit	Qiagen, Crawley	12.5 µl
Forward primer (4 µM)	Synthesised by Eurofins MWG Operon	2.5 µl
Reverse primer (4 µM)	Synthesised by Eurofins MWG Operon	2.5 µl
Fnn- or *Fnf*-specific probe (1 µM)	Synthesised by Eurofins MWG Operon	2.5 µl
DNA template		5 µl

2.8.5 Gel extraction of nucleic acid product

The GenElute Gel extraction kit was used for purification of DNA fragments from agarose gels according to the manufacturer's instructions. See Appendix 4 for full details.

2.8.6 Quantification of DNA using Nanodrop

Concentrations of purified PCR fragments, plasmids, and other DNA samples were measured using a Nanodrop2000 spectrophotometer (Thermo Scientific, Loughborough). The optical density of 1 µl of sample DNA was measured at 260 nm and the concentration calculated by the Nanodrop2000 software.

2.8.7 Sanger sequencing purified DNA fragment

Primers designed for closing gaps in the leukotoxin sequence data and those designed for sequencing the full leukotoxin operon are detailed in Appendices 10 and 11, respectively. The PCR products amplified using these primer pairs were sequenced using the Sanger method. Primer sets used for detection of the *lktA* gene are detailed in Appendix 14. PCR products amplified using these primer pairs were not sequenced, with the exception of the first 12 strains (1 – 59) for Ludlam_LT1. A small selection of the Ludlam_LT1 amplicons were sequenced to determine variations within that region, which were highlighted by non-amplification using lktA primers, targeting the same region.

A primer pair designed to amplify the *ecotin* gene (detailed in Table 42) was used to sequence strains 1 – 59, JCM 3718, JCM 3724 and ARU 01. This was to enable analysis of the JCM 3718, JCM 3724 and ARU 01 sequence data and to design a *F. necrophorum* ecotin plasmid insert. The remainder of the ecotin amplicons were sequenced to assess the level of conservation between the strains. Products amplified using primer pairs *gyrB* (Table 20) and *FadA* (Table 30) were not sequenced. The *gyrB* amplicons were not sequenced as verification of the strain identities as TaqMan probes were also used for identification work. There was no amplification of product within *F. necrophorum* strains using the *FadA* primers, so these primers were not used for sequencing.

For Sanger sequencing, prepaid barcodes were used to send purified plasmids and PCR fragments to GATC Biotech (Constance, Germany). 20 µl of purified PCR fragments (10 – 50 ng/µl) or purified plasmid (30 – 100 ng/µl) were sent in 1.5 ml microfuge tubes and 20 µl of the forward and reverse primers (10 pmol/µl), in separate 1.5 ml microfuge tubes, were sent via a London collection point.

2.9 *de novo* genome sequencing, assembly and annotation

2.9.1 Next generation sequencing and assembly

Two *F. necrophorum* reference strains and a LS clinical isolate were commercially sequenced by GATC Biotech (Konstanz, Germany) using Roche 454 GS-FLX+ next generation sequencing technology. The raw sequence reads of the genomes were assembled into larger contigs semi-automatically using Roche GS Assembler, also known as Newbler.

2.9.2 Genome annotation

The assembled contigs were uploaded into the xBASE pipeline (Chaudhuri *et al.*, 2008) in FASTA format for prediction of open reading frames and gene product annotations. The pipeline combines several programs used for annotation purposes: Glimmer predicted the open reading frames, tRNAScan-SE searched for tRNA genes, RNAmmer searched for ribosomal RNA genes, and BLASTp searched translated coding sequences against the selected reference sequence (*Fusobacterium nucleatum* in this case). The best result for each search was selected as the annotation product, providing the criteria in Table 7 was satisfied:

Table 7: Criteria required for BLASTp assignment of product annotation by xBASE pipeline.

Minimum gene length:	90 base pairs
Maximum gene overlap:	50 base pairs
BLAST E-value cut-off:	1e-10

For BLASTp searches that yield no acceptable result, the product identity field was left blank.

2.9.3 File conversion script

xBASE output files are produced in GenBank format. To convert them to a user-friendly Microsoft Excel file, a program was created using an OpenVMS operating system to extract the information of interest. The GenBank file was the input and a comma-separated value file was the output. The resulting Excel spreadsheet had data under the headings: assembly contig, base location, orientation, product and protein sequence.

2.9.4 BLASTp annotation

For manual annotation using a BLASTp search, the translated protein sequence was entered into the query sequence field and then searched against the non-redundant protein sequence database using the default settings. The Broad Institute prokaryotic annotation pipeline standard operating procedure (2009) was used to define the acceptable parameters for BLAST annotation (Table 8). For each query sequence the top BLAST result that satisfies the Broad Institute standard operating procedure conditions was accepted as the gene naming product. Products labelled as hypothetical protein were disregarded unless no other acceptable result was available. The results were added to the Microsoft Excel file containing the xBASE annotation results.

Table 8: Acceptable parameters for BLAST annotation, as defined by the Broad Institute.

Blast Database:	Non-redundant (bacteria)
Maximum E value:	10^{-10}
Minimum identity:	30 %
Minimum query coverage:	30 %

2.9.5 Contig merging using Minimus2

Genomic data sets from the same strain were sent to Source Bioscience, Nottingham, to be merged into a single set of contigs using the program Minimus2, in order to reduce the overall number of contigs. This was carried out as a commercial package.

2.10 Web-based sequence tools and open source software

2.10.1 Primer3

Primer3 (Untergasser *et al.*, 2012) is a PCR primer design and analysis tool that was used for all primer design during this study. The DNA sequence to be targeted, and a region of flanking

sequence were entered into the text box with regions to flank or avoid marked. PCR primers were generated using default settings.

2.10.2 Clustal Omega

To compare multiple relatively short sequences of DNA or protein, Clustal Omega (Sievers *et al.*, 2011) was used (Clustal Omega, 2015). Sequence data for a single gene or operon were entered in FASTA format into the input box from each of the different strains. The sequences were aligned against one another with an asterisk below identical nucleotides or amino acids. Variation between the strains is indicated by a lack of asterisk symbols.

2.10.3 ExPASy translate

To translate DNA sequence into protein sequence, the ExPASy translate tool was used (Expasy - Translate Tool, 2015). DNA sequence was entered into the query text box and an output format selected. The tool then translated the DNA sequence into the six reading frames.

2.10.4 Hydophobicity plots

To create a hydrophobicity plot of protein sequence data, ExPASy ProtScale was used (Expasy - Protscale, 2015). Protein sequence was entered into the query text box and the Kyte and Doolittle hydrophobicity plot option was selected. Potential transmembrane regions are found by peaks with a score of 1.6 or greater when a recommended window size of 19 is used. The window size represents the number of surrounding amino acids used at each amino acid position to generate the average score that predicts hydrophobicity at that position.

2.10.5 Sequence alignments using WebACT

A web-based version of the Artemis Comparison Tool program (Abbott *et al.*, 2005; Webact, 2015) was used to generate comparison files of the leukotoxin operons of the three sequenced genomes. Sequences were uploaded in FASTA format under the generate tab. The results from each query were saved and viewed within the Artemis Comparison Tool program. Regions of similarity between

sequences are shown by red blocks, reversed sections of similarity are shown by blue blocks and white sections (or breaks) indicate unique regions.

2.10.6 Seaview and PhyML

Phylogenetic trees were constructed using programs Seaview and PhyML within a Linux operating system. DNA sequences of interest were first loaded into Seaview, before selecting the view as protein option. Once the DNA sequence had been translated, the align function was used. The aligned sequence was then converted back into DNA sequence with optimised spacing between nucleotides for a better alignment. To make the tree, the PhyML option was selected, with 100 bootstrap replicates set in options. The scale bar shown on the resulting tree represents the number of inferred substitutions per site.

2.11 Cell line culture

All sterile work involving the HL-60 cell line was carried out in a Bioair Safeflow microbiological safety cabinet. Methods for propagating from frozen, cell counting, serial passage and storage are detailed in Appendix 5.

2.12 Donor blood collection and associated ethics approval

2.12.1 Ethics approval

Blood was taken from informed, consenting, healthy volunteers after obtaining ethics approval from the University of Westminster. Application numbers were: 12_13_04 for blood collection for use in the cytotoxicity assay with *F. necrophorum* leukotoxin and VRE1314-1070 for blood collection for use in coagulation assays with *F. necrophorum* ecotin. For both assays, no participant information is linked to the results. All samples were discarded by the end of the day and no cellular

or DNA containing samples were retained. See Appendices 6 and 7 for associated participant information sheets, consent form templates and ethics approval letters.

2.12.2 Venipuncture and blood collection

Samples of blood were collected from participants by venipuncture. This was carried out by an experienced phlebotomist, and 20 ml of blood was collected from each donor in four 5 ml vacutainers. EDTA vacutainers were used where white blood cells were to be separated by histopaque method, and buffered sodium citrate vacutainers were used where platelet-poor plasma was required.

2.12.3 Histopaque separation of blood components

White blood cells were prepared for cytotoxicity assays by Histopaque method. 3 ml of room temperature Histopaque-1119 was added to a 15 ml conical centrifuge tube and 3 ml of room temperature Histopaque-1077 was carefully layered on top. 6 ml of whole blood, diluted 1:1 with 1X PBS, was layered onto the upper gradient of the tube. The tube was centrifuged at 700 x g for 30 minutes at room temperature, resulting in two distinct opaque layers. The plasma was aspirated to within 0.5 cm of the lower layer and discarded. Cells from this layer were transferred to a tube labelled granulocytes. The cells were washed by the addition of 10 ml of 1X PBS to the tubes, followed by centrifuging for 10 minutes at 300 x g. The supernatant was discarded. Wash steps were repeated twice, with cells resuspended by gently drawing in and out of a Pasteur pipette. Cells were then resuspended in complete RPMI 1640 culture medium (Life Technologies, Paisley). The concentration of viable cells was determined by a 0.4 % trypan blue dye exclusion assay.

2.12.4 Collecting platelet-poor plasma by centrifugation

Platelet-poor plasma was prepared for clotting time assays by centrifugation method. Whole blood was separated by centrifugation at 2000 x g for 15 minutes at room temperature and the plasma layer was retained, taking care not to disturb the buffy coat layer. The plasma from one tube was kept at 4 °C for activated partial thromboplastin time tests and the remainder were kept at room temperature.

2.13 Cytotoxicity assay

2.13.1 Collection of high molecular weight protein for cytotoxicity assay

JCM 3718, JCM 3724 and ARU 01 strains were grown on FAA media at 37 °C for 48 hours under anaerobic conditions. They were subsequently cultured into BHI broth and incubated at 37 °C and 150 rpm until they reached an OD_{600} of 0.7 – 0.8. The cultures were centrifuged at 3000 x g for 30 minutes at 4 °C and the supernatant was sterile filtered using 0.2 μm filters (Millipore, Hertfordshire) and 75X concentrated with 100 kDa molecular weight cut off filters (Millipore, Hertfordshire) also centrifuged at 3000 x g and 4 °C for 30 minutes. The resulting high molecular weight samples were aliquoted and stored at -20 °C for up to two weeks. Samples were not defrosted more than once and were analysed for the presence of high molecular weight protein by SDS PAGE.

2.13.2 Quantification of protein sample by Bradford assay

The concentrations of the high molecular weight protein fractions from section 2.13.1, hypothesised to contain leukotoxin, were determined by Bradford assay (Bradford, 1976) in 96 well clear flat-bottomed plates. A standard curve was constructed using 9, 11, 13, 15, 17 and 19 μg/ml human serum albumin (HSA) in triplicate, in 200 μl final volumes containing 100 μl per well Bradford reagent. Putative leukotoxin samples were diluted 1 in 50. Absorbance was measured on a Versamax microplate reader at 595 nm, and the equation of the graph and R squared value were recorded. Sample concentrations were adjusted for the dilution factor before reported.

2.13.3 Analysis of protein sample by SDS PAGE

Concentrated samples of high molecular weight protein (from section 2.13.1) were analysed using a sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE) method introduced by Laemmli (1970). Protogel buffers were purchased from Fisher (Loughborough) and Ammonium persulfate (APS) and N,N,N',N'-Tetramethylethylenediamine (TEMED) were purchased from Sigma (Gillingham). 0.75 mm polyacrylamide gels were made using Mini Protean Tetra Cell apparatus (BioRad, Hertfordshire, UK). 10 % polyacrylamide running gels were made by mixing the

components in Table 9 and poured between glass plates with a 4 % polyacrylamide stacking gel, made by mixing the components in

Table 10, poured on top, as per the manufacturer's instructions. The APS solution, made by adding 0.1 g to 1 ml deionised water, and TEMED were added quickly before pouring. 6X Laemmli reducing buffer and non-reducing buffer were added to the samples separately, which were then boiled for 5 minutes before being loaded into the gel. Gels were run with either blue protein standard (broad range) or colour protein standard (broad range) protein ladders, purchased from New England Biolabs (Hitchin). Gels were run in a tank of 1X Tris-Glycine buffer (pH 8.3) at 110 V for 90 minutes. Gels were stained with 0.5 % Coomassie blue dye for 1 hour, destained overnight in 7 % (v/v) acetic acid and then viewed on a GS-800 scanner using the Quantity one program.

Table 9: Composition of 10 % running gel.

Component	Volume
Protogel, containing 30 % (w/v) acrylamide/methylene bis-acrylamide solution (37.5:1 ratio)	3.3 ml
Resolving buffer	2.6 ml
Deionised water	3.96 ml
Ammonium persulfate 10 % (w/v)	100 µl
TEMED	20 µl

Table 10: Composition of 4 % stacking gel.

Component	Volume
Protogel, containing 30 % (w/v) acrylamide/methylene bis-acrylamide solution (37.5:1 ratio)	1.3 ml
Stacking buffer	2.5 ml
Deionised water	6.1 ml
Ammonium persulfate 10 % (w/v)	50 µl

TEMED	10 μl

2.13.4 Cytotoxicity assay using HL-60 cell line

HL-60 cells were passaged as described in Appendix 5 in order to be used in a cytotoxicity assay, with methods based on those of Tadepalli *et al.* (2008a). Following cell passage, cells were resuspended in an appropriate volume to result in 1 x 10^6 viable cells/ml. 12 ml centrifuge tubes were labelled with the sample treatments and 100 μl of cell suspension was added to each to result in 1 x 10^5 viable cells per tube. Cell treatments included: 1X PBS as a negative control, ethanol as positive control at a final concentration of 7.5 %, concentrated high molecular weight culture supernatant collected from JCM 3718, JCM 3724 and ARU 01 at final concentrations 125 μg/ml, 150 μg/ml, and 175 μg/ml, concentrated high molecular weight culture supernatant collected from JCM 3718 and ARU 01 at a final concentration of 250 μg/ml, and filter flow through from the three strains. Concentrated high molecular weight culture supernatant, filter flow through, PBS or ethanol were added to each tube and the volume made up to 300 μl with 1X PBS. Treated cells were incubated for 45 minutes at 37 °C in 5 % CO_2. Cells were washed twice with 1X PBS and centrifuging at 300 x g, with cells resuspended by gently drawing in and out of a Pasteur pipette. Cells were then resuspended in 300 μl complete culture medium (Appendix 2.4) and stained with 10 μl propidium iodide (50 μg/ml) in the dark for 5 minutes before being processed on a flow cytometer. Experiments were repeated over three separate days for triplicate data sets.

2.13.5 Cytotoxicity assay using human donor white blood cells

The cytotoxicity assay from section 2.13.4 using HL-60 cells was partially repeated using white blood cells separated from human donor blood by Histopaque method (section 2.12.3). Following a trypan blue dye exclusion assay, as described in Appendix 5.2, cells were adjusted to a concentration of 1 x 10^6 viable cells/ml by addition of RPMI 1640 culture medium (Life Technologies, Paisley). 12 ml centrifuge tubes were labelled with the sample treatments and 100 μl of cell suspension was added to each to result in 1 x 10^5 viable cells per tube. Cell treatments included: 1X PBS as a negative control, ethanol as positive control at a final concentration of 7.5 %, and toxin collected from JCM

3718, JCM 3724 and ARU 01 at a final concentration of 150 µg/ml. Treated cells were incubated for 45 minutes at 37 °C in 5 % CO_2. Cells were washed twice with 1X PBS and centrifuging at 300 x g, with cells resuspended by gently drawing in and out of a Pasteur pipette. Cells were then resuspended in 300 µl RPMI 1640 culture medium and stained with 10 µl propidium iodide (50 µg/ml) in the dark for 5 minutes before being processed on a flow cytometer. Experiments were repeated over three separate days for triplicate data sets.

2.13.6 Flow cytometric analysis

For each sample 10,000 cells were analysed on the Cyan ADP Flow Cytometer (DakoCytomation) using the Summit V4.3 software. Unstained, untreated samples of cell lines and human white blood cells were used to adjust the lasers and set gates for cell size and granularity before samples stained with propidium iodide were measured for viability. The percentage viabilities for the negative controls were set to match those of the trypan blue dye exclusion assay in order to obtain the most accurate gate settings. Cells were therefore gated according to approximately 98 % viability.

2.14 Cloning

2.14.1 Preparation of competent cells using $CaCl_2$ method

A few colonies from an overnight culture of Top10 or BL21(DE3) *E. coli* were inoculated into 10 ml of LB broth in a sterile tube and incubated for 16 hours at 37 °C and 330 rpm. 50 ml of pre-warmed LB broth in a conical flask was inoculated with 1 ml of the 16 hour culture and grown at 37 °C and 220 rpm until OD_{600} reached 0.3 – 0.4. The culture was transferred to a pre-chilled 50 ml tube and chilled on ice for 20 minutes before being centrifuged at 3000 rpm for 5 minutes at 4 °C. The supernatant was discarded and the cells gently resuspended on ice in 5 ml of filter sterilised 0.1 M $CaCl_2$ pre-chilled to 4 °C. After 30 minutes on ice, the cells were centrifuged again at 3000 rpm for 5 minutes at 4 °C and the supernatant was discarded. The cells were gently resuspended in 2 ml

pre-chilled, filter sterilised 0.1 M CaCl$_2$ containing 15 % glycerol. Cells were left on ice for 30 minutes before being aliquoted into pre-chilled 1.5 ml tubes and stored at -80 °C.

2.14.2 Plasmid design

A plasmid insert containing the *F. necrophorum ecotin* gene was designed to include an NcoI restriction site, a deca-histidine tag, an enterokinase cleavage site, an NdeI restriction site, followed by the *ecotin* gene sequence including the stop codon, and restriction site XhoI. It was ensured that the *ecotin* gene start codon would be in frame. The plasmid was commercially synthesised by GeneArt (Regensburg, Germany) and was supplied as 5 µg of lyophilised plasmid DNA. The DNA sequence was verified as 100 % homologous to the design by Sanger sequencing.

2.14.3 Resuspending lyophilised plasmid DNA

Lyophilised plasmid DNA was pulse centrifuged to ensure all DNA collected at the bottom of the tube. 50 µl of molecular grade water (Fisher Scientific, Loughborough) was added to 5 µg plasmid DNA. This was mixed by gentle flicking of the tube and pulse centrifuged to incorporated all DNA and incubated for 1 hour at room temperature. 1 µl of the resuspended DNA was transformed into chemically competent *E. coli* cells (section 2.14.4) and the remainder stored in aliquots at -20 °C.

2.14.4 Transformation of plasmid into chemically competent Top10 cells using heat shock method

The synthesised plasmid and the pET-16b plasmid provided by Dr Anatoliy Markiv, University of Westminster, were both separately transformed into chemically competent Top10 *E. coli* cells in order to amplify the plasmids. In each case, 30 µl Top10 cells were transferred to a 1.5 ml microfuge tube on ice. 10 ng of plasmid was added to the cells which were incubated on ice for 30 minutes. The cells were heat shocked at 42 °C for 40 seconds then placed back on ice for 2 minutes. 200 µl SOC medium was added and the culture was placed in a shaking incubator for 45 minutes at 37 °C at 330 rpm. 100 µl competent bacteria and plasmid were inoculated onto an LB agar plate with 100 µg/ml ampicillin and incubated overnight at 37 °C. One colony was selected and grown in 6 ml LB

broth with 100 μg/ml ampicillin for 16 hours at 37 °C, shaking at 250 rpm. An aliquot was saved in 50 % glycerol and a plasmid miniprep kit from Life Technologies, Paisley (section 2.14.5) was used to extract plasmid DNA, following the manufacturer's instructions. Plasmid DNA was quantified using a Nanodrop.

2.14.5 Plasmid extraction

Plasmid DNA was purified using a PureLink Quick plasmid miniprep kit, purchased from Invitrogen, by Life Technologies (Paisley), and used as per the manufacturer's instructions. 5 ml of an overnight broth culture of *E. coli* cells transformed with plasmid was centrifuged at 12,000 x g and the supernatant discarded. Cells were resuspended in 250 μl Resuspension buffer with RNase A and mixed until homogenous. 250 μl Lysis buffer was added and mixed gently, then incubated at room temperature for 5 minutes. 350 μl Precipitation buffer was added and mixed immediately by inverting the tube until the mixture was homogenous. The sample was then centrifuged at 12,000 x g for 10 minutes. The supernatant was loaded into a spin column in a 2 ml wash tube and was centrifuged for 12,000 x g for 1 minute and the flow through discarded. 700 μl Wash buffer with ethanol was added to the column which was then centrifuged at 12,000 x g for 1 minute and the flow through was discarded. The centrifugation step was repeated and the spin column placed in a clean 1.5 ml recovery tube. 75 μl TE buffer, preheated to 65 – 70 °C, was added to the centre of the column and incubated for 1 minute at room temperature. The column was centrifuged at 12,000 x g for 2 minutes and the eluted, purified plasmid was stored in 25 μl aliquots at -20 °C.

2.14.6 Plasmid digestion

Purified plasmids from Top10 cells were digested using restriction enzymes to cleave the DNA at the NcoI and XhoI cloning sites. Digestions were carried out simultaneously to provide the pET-16b plasmid backbone and the *ecotin* gene insert. The reaction components listed in Table 11 were added and the tubes were incubated at 37 °C for 30 minutes before being run in a 0.8 % agarose gel. The resulting bands were excised with a scalpel and the DNA extracted using a DNA gel

purification kit (Sigma, Gillingham) (section 2.8.5). The nucleic acid was then quantified using a

Nanodrop (section 2.8.6) for both the plasmid backbone and insert.

Table 11: Reaction components of plasmid digestion for both pET-16b plasmid and ecotin synthesised plasmid.

Reaction component	Reaction volume
Plasmid (pET-16b and Ecotin, separately)	28 μl
Digestion buffer (Fermentas, Hertfordshire)	3.5 μl
Molecular grade water (Fisher Scientific, Loughborough)	0.5 μl
NcoI (Fermentas, Hertfordshire)	1.5 μl
XhoI (Fermentas, Hertfordshire)	1.5 μl
Total	**35 μl**

2.14.7 Plasmid ligation

The digested ecotin insert was ligated into the pET-16b plasmid backbone by adding the plasmid

components in a 1:3 molar ratio (1 part backbone – 3 parts insert). The reaction components from

Table 12 were added and the mixture was incubated at 4 °C overnight. The ligation mix was heat

inactivated at 70 °C for 10 minutes before 10 ng was transformed into Top10 cells using the method

described in section 2.14.4. One colony was selected and grown in 6 ml LB broth with 100 μg/ml

ampicillin for 16 hours at 37 °C, shaking at 250 rpm. An aliquot was then saved in glycerol and a

plasmid miniprep (section 2.14.5) was carried out to extract plasmid DNA. A second digest (section

2.14.6) was carried out on the purified plasmid and the resulting digested plasmid run on a 0.8 %

agarose gel (section 2.8.3) to check for the presence of the insert. This was indicated by two bands,

one being the plasmid backbone and the other being the insert.

Table 12: Reaction components of plasmid ligation of pET-16b plasmid backbone and ecotin insert.

Reaction component	Reaction volume
Ecotin insert	1.6 µl
pET-16b plasmid backbone	4 µl
10X ligation buffer (New England Biolabs, Hitchin)	1 µl
Molecular grade water (Fisher Scientific, Loughborough)	2.4 µl
T4 DNA ligase (New England Biolabs, Hitchin)	1 µl
Total	**10 µl**

2.14.8 Transformation of plasmid into chemically competent BL21(DE3) cells using heat shock method

Purified pET-16b plasmid containing the ecotin insert was transformed into BL21(DE3) cells for expression. 50 µl of chemically competent BL21(DE3) *E. coli* cells were transferred to a 1.5 ml microfuge tube on ice. 10 ng of plasmid DNA was added to the cells and incubated on ice for 30 minutes. The cells were heat shocked by placing the tube into a 42 °C heat block for 10 seconds and then placed on ice for 5 minutes. 950 µl SOC medium was added and the culture was incubated at 37 °C for 1 hour, shaking at 250 rpm. 100 µl was inoculated onto pre-warmed LB agar plates with 100 µg/ml ampicillin and incubated overnight at 37 °C. One colony was selected and grown in 6 ml LB broth with 100 µg/ml ampicillin for 16 hours at 37 °C, shaking at 250 rpm. An aliquot was then saved in 15 % glycerol and a plasmid miniprep (section 2.14.5) was carried out to extract plasmid DNA, which was subsequently quantified using a Nanodrop (section 2.8.6). 500 ng of plasmid was

sent for commercial Sanger sequencing (section 2.8.7) in order to verify the final plasmid contained the correct sequence.

2.15 Histidine tag expression and purification

2.15.1 Expression of recombinant protein

BL21(DE3) expression strain, transformed with pET-16b/ecotin, was grown on LB agar containing 100 µg/ml ampicillin and incubated overnight at 37 °C. A BL21(DE3) expression strain containing no plasmid was grown in parallel on media without ampicillin to be used as a negative control. The following day, one colony from each culture was inoculated into 10 ml LB broth, containing 100 µg/ml ampicillin for the plasmid culture only, and incubated at 37 °C and 250 rpm for 16 hours. 5 ml of each starter culture was used to inoculate a 1 L conical flask containing 200 ml LB, with 100 µg/ml ampicillin for the plasmid-containing culture. The cultures were incubated at 37 °C and 250 rpm until the OD_{600} reached between 0.3 and 0.4. The flasks were placed on ice for 30 minutes then protein production was induced with IPTG at 0.5 mM for 3 hours at 37 °C and 250 rpm. 1 ml samples were taken from each culture at IPTG induction (0 hours) and 1, 2 and 3 hours post-induction for analysis by Tricine method SDS PAGE (section 2.15.3). The flask cultures were pelleted in 50 ml aliquots by centrifugation at 3000 x g and stored overnight at -20 °C.

2.15.2 Purification of histidine-tagged protein by immobilised metal ion chromatography (IMAC) under native conditions

The histidine-tagged recombinant protein was purified by immobilised-metal affinity chromatography (IMAC) using buffers listed in Table 13. Pellets were resuspended in 5 ml His-Binding buffer with 1.5 µl rLysozyme solution added (Merck Millipore, Watford). Resuspended cultures were mechanically lysed with 5 cycles of 10 seconds of sonicating followed by a 20 second rest period. 20 µl of 'Total' fraction was collected for Tricine method SDS PAGE analysis (section 2.15.3) and stored as a 1:1 ratio in 6 M Urea. Lysate was centrifuged at 12,000 x g for 10 minutes at

4 °C and the supernatant was collected. 20 µl of the 'Soluble' fraction was collected for SDS PAGE analysis. IMAC was carried out using a nickel-nitrilotriacetic acid matrix. A 0.25 ml bed volume of IMAC Sepharose 6 Fast Flow Resin (GE Healthcare, Amersham) was primed with 0.2 M NiSO₄.6H₂O and equilibrated with His-Binding buffer. The lysate supernatant containing the His-tagged protein was run through the column by gravity filtration. 20 µl of 'Flow through' was collected for SDS PAGE analysis. The column was washed with 5 ml His-Wash buffer and 20 µl 'Wash' fractions were collected for SDS PAGE analysis. After each wash step, a 2 µl aliquot was analysed on a Nanodrop2000 using the A280 setting to quantify protein concentration. Wash steps were repeated until the concentration had reduced to 0 mg/ml. His-tagged protein was eluted from the metal ion matrix four times in 250 µl aliquots with His-Elution buffer. 40 µl of each 'Elution' fraction was collected for SDS PAGE (section 2.15.3) and Bradford assay analysis (section 2.15.4). All purification steps were carried out on ice with buffers at 4 °C and all protein aliquots were stored at -20 °C.

Table 13: Protein purification buffers used in IMAC under native conditions. Buffers were made up in water and filter sterilised.

Buffer	pH	Composition
His-Binding buffer	8.0	50 mM HEPES, 500 mM NaCl, 5 mM Imidazole, 5% Glycerol
His-Wash buffer	8.0	50 mM HEPES, 500 mM NaCl, 30 mM Imidazole, 5% Glycerol
His-Elution buffer	8.0	50 mM HEPES, 500 mM NaCl, 250 mM Imidazole, 5% Glycerol

2.15.3 Analysis of protein expression and purification by Tricine SDS PAGE

Samples taken during protein expression and purification were analysed using the Tricine-SDS PAGE method by Schagger and Jagow (1987) for separation of proteins in the range from 1 to 100 kDa. Ammonium persulphate (APS) and N,N,N',N'-Tetramethylethylenediamine (TEMED) were purchased from Sigma (Gillingham). Buffers and 0.75 mm polyacrylamide gels were made as per Table 14 and Mini Protean Tetra Cell apparatus was used (BioRad, Hertfordshire, UK). 15 % polyacrylamide running gels were made and poured between the glass plates with a 5 % polyacrylamide stacking gel poured on top. The APS solution, made by adding 0.1 g to 1 ml

deionised water, and TEMED were added quickly before pouring the gels. 6X Laemmli reducing buffer was added to the samples, which were then boiled for 5 minutes before being loaded into the gel. Gels were run with either blue protein standard (broad range) or colour protein standard (broad range) protein ladders, purchased from New England Biolabs (Hitchin). Gels were run with 1X cathode buffer between the glass plates and 1X anode buffer in the surrounding tank and at 110 V for 140 minutes. Gels were stained with 0.5 % Coomassie blue dye for 1 hour, destained overnight in 7 % (v/v) acetic acid and then viewed on a GS-800 scanner using the Quantity one program.

Table 14: Composition of gels and buffers used in Tricine SDS PAGE. All were prepared in deionised water.

Buffer/Gel	pH	Composition
Gel buffer	8.45	3 M Tris base, 0.3 % SDS
Running gel	NT	15 % Acrylamide/bisacrylamide, 1 M gel buffer, 13.3 % glycerol, 0.1 % APS, 2 µl/ml TEMED
Stacking gel	NT	5 % Acrylamide/bisacrylamide, 0.75 M gel buffer, 0.1 % APS, 2 µl/ml TEMED
Anode buffer	8.9	0.2 M Tris base
Cathode buffer	8.25	0.1 M Tris base, 0.1 M Tricine, 0.1 % SDS

NT = Not tested

2.15.4 Bradford assay to obtain purified recombinant protein concentration

The concentration of purified His-tagged protein was determined by Bradford assay (Bradford, 1976) as described in section 2.13.2, with a standard curve constructed using human serum albumin (HSA) concentrations of 0, 2, 4, 6, 8 and 10 µg/ml. Elution samples were first tested on a Nanodrop using the default option within the A280 setting and results were used to approximate the necessary dilution required to fit the concentration within the range of standards.

2.16 Enzyme inhibition assays

Enzyme inhibition assays were conducted in 96 well flat-bottomed microtitre plate formats. Changes in relative fluorescence units (RFU) were monitored on a Glomax Mulit+ E8032 for human plasma kallikrein (HPK) assays and black microtitre plates were used. Excitation wavelength was set at 405 nm and emission wavelength at 495 – 505 nm. Absorbance changes were monitored on a Versamax microplate reader at 410 nm for human neutrophil elastase (HNE) assays and clear microtitre plates were used.

Reaction buffers A and B were made according to Table 15 and were autoclaved, stored at 4 °C and equilibrated to room temperature before use. Ecotin, the inhibitor, was stored in aliquots at -20 °C, and was diluted to a working concentration of 500 nM in the corresponding enzyme buffer (A or B) before use. Aliquots of enzymes, substrates and inhibitor were not defrosted more than once and were used on the same day. For each set of enzyme assays, the standards were characterised on day 1 and the inhibition was tested on day 2.

Table 15: Composition of buffers used for enzyme inhibition assays. Buffer A was used with human plasma kallikrein and Buffer B was used with human neutrophil elastase.

	Composition
Buffer A	50 mM Tris pH 7.5, 100 mM NaCl, 2 mM $CaCl_2$, 0.005% Triton X-100
Buffer B	0.1 M Tris pH 7.5, 0.5 M NaCl

50 µg human plasma kallikrein was purchased from Merck Millipore (Watford) in liquid form in NaCl and Tris-HCl buffer at a concentration of 6.28 µM. This was stored in 6 µl aliquots at -20 °C. A working concentration of 0.1 µM was obtained by diluting in Buffer A. 5 mg fluorogenic HPK substrate (H-D-Val-Leu-Arg-AFC.2HCl) was also obtained from Merck Millipore (Watford) and was dissolved in DMSO at a concentration of 2 mM and stored in 405 µl aliquots at -20 °C. A working concentration of 1 mM was obtained by diluting in Buffer A.

50 μg of lyophilised human neutrophil elastase was purchased from Merck Millipore (Watford). This was reconstituted in a buffer containing 50 mM sodium acetate pH 5.5 and 200 mM sodium chloride to provide an enzyme stock concentration of 10 μM which was stored in 12 μl aliquots at -20 °C. A working concentration of 0.1 μM was obtained by diluting in Buffer B. 50 mg chromogenic HNE substrate (Methoxysuccinyl-Ala-Ala-Pro-Val-p-nitroanilide) was also obtained from Merck Millipore (Watford) and was dissolved in DMSO at a concentration of 8 mM and stored in 500 μl aliquots at -20 °C. A working concentration of 2 mM was obtained by diluting in Buffer B.

A range of substrate standards were tested to establish normal, uninhibited enzyme kinetics. HPK was added to a microtitre plate in preparation for a final concentration of 10 nM in a final volume of 100 μl. 1 mM HPK substrate was then added to produce final concentrations of 0.015, 0.03, 0.06, 0.125, 0.25 and 0.5 mM. All substrate concentrations were tested side by side in triplicate (see Table 16 for template), with the substrate added quickly using a multichannel pipette. The plate was pre-blanked with Buffer A and fluorescence was monitored every minute for 30 minutes.

Table 16: Microtitre plate template for HPK substrate standards. [E] = enzyme concentration, [S] = substrate concentration and [I] = inhibitor concentration. Final volume was 100 μl.

Location	Contents
A1-A3	Buffer A
B1-B3	[E] = 10 nM, [S] = 0.015 mM, [I] = 0 nM
C1-C3	[E] = 10 nM, [S] = 0.03 mM, [I] = 0 nM
D1-D3	[E] = 10 nM, [S] = 0.06 mM, [I] = 0 nM
E1-E3	[E] = 10 nM, [S] = 0.125 mM, [I] = 0 nM
F1-F3	[E] = 10 nM, [S] = 0.25 mM, [I] = 0 nM
G1-G3	[E] = 10 nM, [S] = 0.5 mM, [I] = 0 nM

A calibration curve was produced for 7-Amino-4-(trifluoromethyl)coumarin (AFC), the fluorophore cleaved from HPK substrate. Ten concentrations in 0.05 mM increments in the range of 0 – 0.5 mM

were tested in triplicate and a graph of AFC concentration against mean relative fluorescent units (RFU) was plotted and the equation of the line determined using linear regression as part of the GraphPad Prism package.

To establish the rate of reaction for HPK at each substrate concentration, the changes in RFU were recorded at 5 minutes. These values were converted into concentration of product using the equation from the AFC calibration curve. The concentration values were adjusted to take into account the volume of liquid (0.1 ml) in order to establish a quantity of product. The rate of reaction was calculated by dividing this value by the number of minutes (5) in order to determine the rate of reaction.

HNE standards were measured as follows: HNE was added to a microtitre plate in preparation for a final concentration of 17 nM in a final volume of 200 µl. 2 mM HNE substrate was then added to produce final concentrations of 0.015, 0.03, 0.06, 0.125, 0.25 and 0.5 mM. All substrate concentrations were tested side by side in triplicate (see Table 17 for template), with the substrate added quickly using a multichannel pipette. The plate was pre-blanked with Buffer B and absorbance was monitored every minute for 30 minutes.

Table 17: Microtitre plate template for HNE substrate standards. [E] = enzyme concentration, [S] = substrate concentration and [I] = inhibitor concentration. Final volume was 200 µl

Location	Contents
A1-A3	Buffer B
B1-B3	[E] = 17 nM, [S] = 0.015 mM, [I] = 0 nM
C1-C3	[E] = 17 nM, [S] = 0.03 mM, [I] = 0 nM
D1-D3	[E] = 17 nM, [S] = 0.06 mM, [I] = 0 nM
E1-E3	[E] = 17 nM, [S] = 0.125 mM, [I] = 0 nM
F1-F3	[E] = 17 nM, [S] = 0.25 mM, [I] = 0 nM
G1-G3	[E] = 17 nM, [S] = 0.5 mM, [I] = 0 nM

In order to calculate the rates of reaction from the data for Michaelis-Menten kinetics, the changes in absorbance were recorded at 5 minutes for each substrate concentration. Beer-Lambert Law (Equation 1) was used to establish the concentration of product from these values, where A = absorbance, ε = molar extinction coefficient and L = light path (cm). p-nitroanilide (pNA), the chromophore product cleaved from HNE substrate, has a known molar extinction coefficient of 8800 M^{-1} cm^{-1}, as published by Merck-Millipore. The light path was 0.7 cm.

Equation 1: Beer-Lambert Law equation. A = absorbance, ε = molar extinction coefficient and L = light path (cm).

$$A = \varepsilon\, C\, L$$

The concentration values were adjusted to take into account the volume of liquid (0.2 ml) in order to establish a quantity of product. The rate of reaction was calculated by dividing this value by the number of minutes (5) in order to determine the rate of reaction.

For each set of data, substrate concentration was plotted against rate of reaction and Michaelis-Menten non-linear regression was applied as part of the GraphPad Prism package to calculate the V_{max} and K_m values. Both HPK and HNE experiments were repeated using the templates from Table 16 and Table 17 with the enzyme excluded in order to test for autolysis of the substrate.

HPK inhibition was tested one substrate concentration per microtitre plate. In triplicate, each plate contained a blank well of Buffer A and the full range of ecotin concentrations to be tested, using a final volume of 100 µl. HPK was added into each well, excluding the blank, for a final concentration of 10 nM. Ecotin was added at 0, 12.5, 25, 50 and 100 nM (See Table 18 for template). The plate was incubated at room temperature for 1 hour to equilibrate. At 1 hour, a multichannel pipette was used to quickly add HPK substrate to rows B – F, resulting in a final concentration of substrate that was the same for all rows. Fluorescence was then monitored every minute for 30 minutes. Row F, containing 0 nM ecotin, was compared to the substrate controls from the previous day to test for reproducibility.

Table 18: Microtitre plate template for range of ecotin concentrations against HPK and each substrate concentration. [E] = enzyme concentration, [I] = inhibitor concentration, [S] = substrate concentration. Final volume was 100 μl.

Location	Contents
A1-A3	Buffer A
B1-B3	[E] = 10 nM, [I] = 100 nM, [S] = *v*
C1-C3	[E] = 10 nM, [I] = 50 nM, [S] = *v*
D1-D3	[E] = 10 nM, [I] = 25 nM, [S] = *v*
E1-E3	[E] = 10 nM, [I] = 12.5 nM, [S] = *v*
F1-F3	[E] = 10 nM, [I] = 0 nM, [S] = *v*

v = variable from plate to plate

HNE inhibition was also tested one substrate concentration per microtitre plate. In triplicate, each plate contained a blank well of Buffer B and the full range of ecotin concentrations to be tested, using a final volume of 200 μl. HNE was added into each well, excluding the blank, for a final concentration of 17 nM. Ecotin was added at 0, 12.5, 25, 50 and 100 nM (See Table 19 for template). The plate was incubated at room temperature for 1 hour to equilibrate. At 1 hour, a multichannel pipette was used to quickly add HNE substrate to rows B – F, resulting in a final concentration of substrate that was the same for all rows. Absorbance was then monitored every minute for 30 minutes. Row F, containing 0 nM ecotin, was compared to the substrate controls from the previous day to test for reproducibility.

Table 19: Microtitre plate template for range of ecotin concentrations against HNE and each substrate concentration. [E] = enzyme concentration, [I] = inhibitor concentration, [S] = substrate concentration. Final volume was 200 µl.

Location	Contents
A1-A3	Buffer B
B1-B3	[E] = 17 nM, [I] = 100 nM, [S] = *v*
C1-C3	[E] = 17 nM, [I] = 50 nM, [S] = *v*
D1-D3	[E] = 17 nM, [I] = 25 nM, [S] = *v*
E1-E3	[E] = 17 nM, [I] = 12.5 nM, [S] = *v*
F1-F3	[E] = 17 nM, [I] = 0 nM, [S] = *v*

v = variable from plate to plate

For each concentration of ecotin, XY plots of time against RFU were plotted for HPK assays, and time against absorbance for HNE assays. Plots were also made for the inhibitor concentration against the ratio of inhibited to uninhibited rate of reaction for each substrate concentration. Morrison K_i nonlinear regression was applied as part of the GraphPad Prism package in order to establish the inhibition constants (K_i) for ecotin against each enzyme.

2.17 Clotting assays

Blood was taken from three different donors on separate days. The experiments with each blood sample were completed within 6 hours of the sample being taken. Platelet-poor plasma was prepared as described in section 2.12.4 and normal clotting times were established. A waterbath was set to 37 °C and maintained with 0.5 °C. All tests were carried out in 12 x 75 mm glass test tubes. Clotting was observed by visual assessment.

2.17.1 Thrombin time (TT)

Bovine thrombin (Diagnostic Reagents, Thame) was purchased in a lyophilised preparation for use in TT assays and was reconstituted in 1 ml of sterile deionised water and used within 2 hours. Thrombin was diluted in barbitone buffered saline (TCS Biosciences, Buckingham) to a working solution of 7.5 units/ml.

200 µl platelet-poor plasma was warmed to 37 °C in a waterbath. 50 µl His-Elution buffer (see Table 13) was added, to be used as the negative control, and the mixture was incubated for a further 2 minutes. 100 µl bovine thrombin was added and a stopwatch was used to record the clotting time.

2.17.2 Prothrombin time (PT)

Calcium rabbit brain thromboplastin (Diagnostic Reagents, Thame) was purchased in a lyophilised preparation for use in PT assays and was reconstituted in 4 ml of sterile deionised water and used within 2 hours.

100 µl platelet-poor plasma was warmed to 37 °C in a waterbath. 50 µl His-Elution buffer was added, to be used as the negative control, and the mixture was incubated for a further 2 minutes. 200 µl warmed thromboplastin was added and a stopwatch was used to record the clotting time. Tubes were gently tilted at regular intervals.

2.17.3 Activated partial thromboplastin time (APTT)

Kaolin platelet substitute mixture (Diagnostic Reagents, Thame) was purchased in a lyophilised preparation for use in APTT assays and was reconstituted in 5 ml of sterile deionised water and used within 4 hours.

200 µl kaolin platelet substitute mixture was warmed to 37 °C in a waterbath. 50 µl His-Elution buffer was added, to be used as the negative control, and 100 µl platelet-poor plasma. The mixture was incubated for 2 minutes. 100 µl of prewarmed 25 mM calcium chloride was added and a stopwatch was used to record the clotting time. Tubes were gently tilted at regular intervals.

2.17.4 Effect of ecotin on TT, PT and APTT assays

Once the clotting time of the negative controls (0 µM ecotin) had been recorded for TT, PT and APTT assays, a range of ecotin concentrations were tested. The ecotin concentrations tested were 0, 0.25, 0.5, 0.75, 1 and 2 µM for TT and PT tests and 0, 0.06, 0.125, 0.25, 0.5, 0.75, 1 and 2 µM for APTT tests. Following inhibition tests, uninhibited clotting was repeated to confirm that the prolongation of clotting was not due to time bias. All experiments were run in duplicate and mean clotting times were plotted as ecotin concentration against fold prolongation.

Chapter 3:

F. necrophorum strain collection and identification

3.1 Introduction and aims

Strain type collections include those of the Japan Collection of Microorganisms (JCM) (Microbe Division (Jcm)(Riken Brc), 2015), the American Type Culture Collection (ATCC) (Atcc: The Global Bioresource Center, 2015) and the National Collection of Type Cultures (NCTC) (Culture Collections, 2015). These collections represent bacterial strains, cell lines, viruses, fungi and DNA material that are preserved and characterised, and can be used as standards and controls during experiments, such as antimicrobial testing, or biochemical tests.

F. necrophorum can be identified biochemically, by molecular techniques, or a combination of the two; see section 1.2.2 for biochemical profile. Narongwanichgarn *et al.* (2001) used the API 20A test for identification of anaerobes to confirm strains as *F. necrophorum*, based on the key test of indole production. A haemagglutination test was used to subspeciate them, with strains causing haemagglutination of washed chicken red blood cells classed as subspecies *necrophorum*, and those not causing haemagglutination were classed as subspecies *funduliforme*. All 19 strains demonstrated β-haemolysis on agar plates. They trialled the use of Random Amplified Polymorphic DNA PCR (RAPD PCR) as a new test for subspeciation and found one primer (W1L-2) to be effective, causing amplification in those identified as *Fnn* and no amplification in those identified as *Fnf*. This work was followed up by using two new primer pairs in a one-step duplex PCR reaction. One of the primer pairs amplified a 250 bp product from both subspecies, while the other a 900 bp from *Fnn* only. These fragments were identified by sequence data as the *rpoB* and haemagglutinin-related

protein genes, respectively (Narongwanichgarn *et al.*, 2003), suggesting that these genes are appropriate targets for molecular subspeciation assays.

Jensen *et al.* (2007) developed an assay to detect the subspecies of *F. necrophorum* using real-time PCR. Colonies on agar plates were initially identified based on colony morphology, odour, green fluorescence when irradiated with UV light, Gram stain, antimicrobial susceptibility testing and β-haemolysis on blood agar plates to be confirmed as *F. necrophorum*. PCR primers were then used to target the gyraseB (*gyrB*) gene of both subspecies, amplifying a region of 306 bp. Subspecies-specific TaqMan probes were designed to bind within the amplified region. The PCR results were followed up by using species-specific RNA polymerase β-subunit (*rpoB*) PCR primers and a single probe specific to *F. necrophorum*. Both sets of primers amplified product in 100 % of *F. necrophorum* strains. All isolates were assigned the correct subspecies by the *gyrB* probes, based on previous records of the strains.

Aliyu *et al.* (2004) also used real-time PCR targeting the *rpoB* gene. Two TaqMan probes were used that were both specific to *F. necrophorum*, but did not subspeciate the strains. Strains were subspeciated using PCR primers targeting the haemagglutinin-related protein gene in a SYBR-green assay. Strains showing amplification in the SYBR-green assay were identified as *Fnn*, and those that did not show amplification were identified as *Fnf*.

Other studies have used 16S rRNA gene sequences to identify (Oikonomou *et al.*, 2012) or diagnose infection with (Sanmillán *et al.*, 2013) *F. necrophorum*. Farooq *et al.* (2015) used PCR primers specific to the leukotoxinA (*lktA*) gene alone to identify *F. necrophorum* strains, suggesting confidence that this gene is universal among *F. necrophorum* strains. Antiabong *et al.* (2013) used PCR primers targeting the *rpoB* and haemagglutinin-related protein genes in order to identify and subspeciate strains. *16S rRNA* and *lktA* primers were also tested.

A collection of *F. necrophorum* strains was required to carry out experimental work for this project. Two laboratories were approached and strains were donated for use in this study. Type strains of *F. necrophorum* subspecies *necrophorum* and *F. necrophorum* subspecies *funduliforme* and a

collection of clinical isolates were acquired. All isolates for use in this project were identified to species and subspecies levels, using a mixture of biochemical and molecular techniques, and type strains as positive controls. All of the strains received had been identified as *F. necrophorum* prior to arrival at the University of Westminster. It was therefore hypothesised that all strains would be identified as *F. necrophorum* during identification tests for this project. It was also hypothesised that all of the clinical strains would be identified as subspecies *funduliforme*, as human infections are most often this subspecies (Smith *et al.*, 1990; Riordan, 2007).

3.2 Results

3.2.1 *Fusobacterium necrophorum* strain collection

F. necrophorum strains JCM 3718 and JCM 3724 were kindly donated by the Anaerobe Reference Unit (ARU), Cardiff, for use in this study. The strains are from the Japan Collection of Microorganisms, although also feature in the American Type Culture Collection, as strains ATCC 25286 and ATCC 51357 for JCM 3718 and JCM 3724, respectively. JCM 3718 has been identified as being subspecies *necrophorum* and JCM 3724 as subspecies *funduliforme*. Both strains were isolated from bovine liver abscesses and added to the JCM in 1985. The two subspecies were proposed and named by Shinjo *et al.* (1991) based on haemagglutination tests and DNA homology (section 1.2.1) and the type strains JCM 3718 and JCM 3724 were selected.

The ARU also provided a clinical strain, isolated in 2002 from a blood sample of a patient with Lemierre's syndrome. For this study the isolate was named ARU 01. A collection of *F. necrophorum* human clinical isolates were also provided by Antonia Batty, University College Hospital London. Twenty six of these were used for this study.

3.2.2 Identification of strains using biochemical tests

Biochemical profiles of the human clinical isolates were characterised to ensure they matched that of *F. necrophorum* (methods section 2.7). Twenty six strains were subcultured and confirmed as

Gram negative rods, with negative results for oxidase and catalase tests and a positive result for indole production. All strains had a typical cream-grey appearance and butyric acid odour characteristic of *F. necrophorum*, with the exception of strain number 88 which had a white appearance on the agar plate, which was noted as atypical.

The three strains provided by the ARU were also tested for biochemical profiles. JCM 3718, JCM 3724 and ARU 01 all matched the expected profile and appearance.

3.2.3 Identification of strains using *F. necrophorum*-specific *gyraseB* PCR primers

The 26 strains from the clinical collection and the three strains from the ARU were tested with *F. necrophorum*-specific gyrB PCR primers (Jensen et al., 2007) (Table 20) to confirm the identity of the strains as *F. necrophorum* (methods section 2.8.2 and 2.8.3). Twenty five out of the 26 strains from the clinical collection produced amplified products of the expected size (306 bp). Strain number 88, which was previously noted as having an atypical appearance had no amplification of product. The strain was therefore identified as not being *F. necrophorum* and was excluded from the collection. JCM 3718, JCM 3724 and ARU 01 all had amplified products of the expected size. See Figure 3 for a gel electrophoresis image of amplified *gyrB* products.

Table 20: PCR primers used to amplify F. necrophorum gyrB genes.

Primer	Oligonucleotide sequence (5' → 3')
gyrB_F	AGGATTGCATGGAGTAGGAA
gyrB_R	CCTATTTCATTTCGACAATCCA
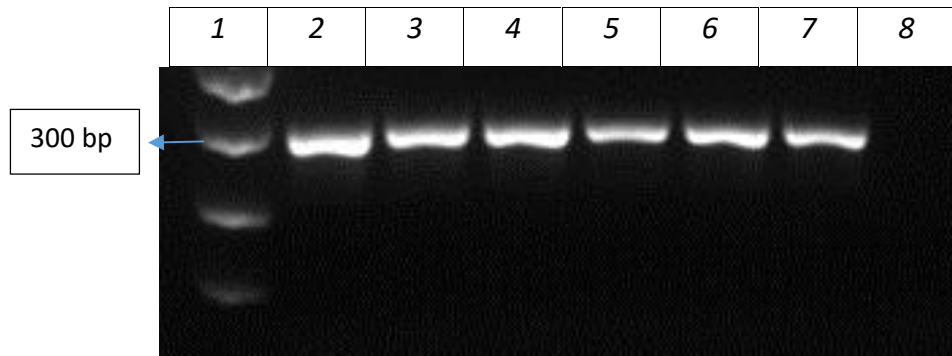
Figure 3: Analysis of PCR products from F. necrophorum-specific gyrB primers on a 1% agarose gel. Lane 1: 100bp ladder; lanes 2 – 7: strains 1, 5, 11, 21, 24, 30, respectively; lane 8: negative control (molecular grade water).

3.2.4 Subspeciation using TaqMan probes

The strains were subspeciated using *Fnn* and *Fnf*-specific TaqMan probes (Jensen *et al.*, 2007) with the *F. necrophorum* specific *gyrB* primers; with the probes tested in separate PCR tubes (see methods 2.8.4). Fluorescence from *Fnf*-specific probes was measured on the yellow channel of the RotorGene Q PCR machine, while fluorescence from *Fnn*-specific probes was measured on the green channel. Samples showing an increase in fluorescence sufficient to exceed the threshold were considered as having a positive result for the associated probe. Thresholds were set in the early logarithmic phase of the amplification curves.

JCM 3718, JCM 3724 and ARU 01 were first tested along with five of the clinical strains in order to validate the probes. Figure 4 (and sample key in Table 21) displays the result of these strains tested with the subspecies *funduliforme* probe. JCM 3724 and ARU 01 had a positive result with the *Fnf* probe while JCM 3718 and the negative control showed no increase in fluorescence. The five clinical strains from the collection were also expected to be subspecies *funduliforme*, due to their human origin, and all showed a positive result with the *Fnf* probe.
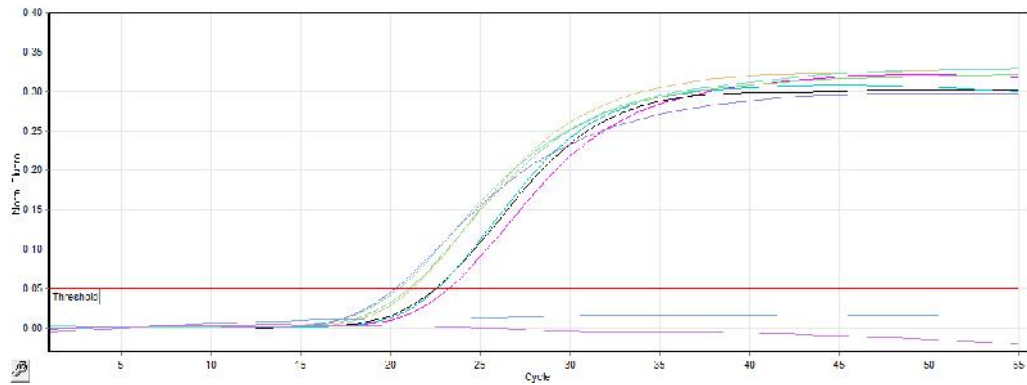
Figure 4: Fluorescence from Fnf-specific TaqMan probes measured on the yellow channel of a RotorGene Q PCR machine.

See Table 21 for sample key.

Table 21: Sample key for Figure 4. All tubes contained Fnf probe.

Colour	Sample name	Threshold exceeded?
	11	+
	21	+
	39	+
	42	+
	59	+
	ARU 01	+
	JCM 3718	-
	JCM 3724	+
	Negative control	-

The same strains were tested with the subspecies *necrophorum* probe (Figure 5 and Table 22), which measured a positive result for JCM 3718 only.
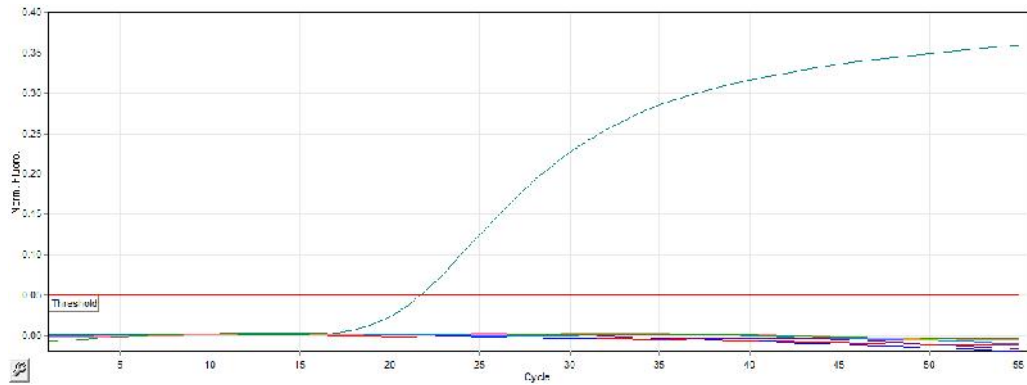
Figure 5: Fluorescence from Fnn-specific TaqMan probes measured on the green channel of a RotorGene Q PCR machine. See Table 22 for sample key.

Table 22: Sample key for Figure 5. All tubes contained Fnn probe.

Colour	Sample name	Threshold exceeded?
	11	-
	21	-
	39	-
	42	-
	59	-
	ARU 01	-
	JCM 3718	+
	JCM 3724	-
	Negative control	-

The probes had the expected results for the reference strains, resulting in a confirmed subspecies of *funduliforme* for JCM 3724 and *necrophorum* for JCM 3718. For the remaining samples sets, the number of cycles was reduced from 55 to 40 and JCM 3724 was used as a positive control. Figure 6 and associated sample key Table 23, along with Figure 7 and Table 24 show the remainder of the clinical strain collection testing positive for subspecies *funduliforme*.
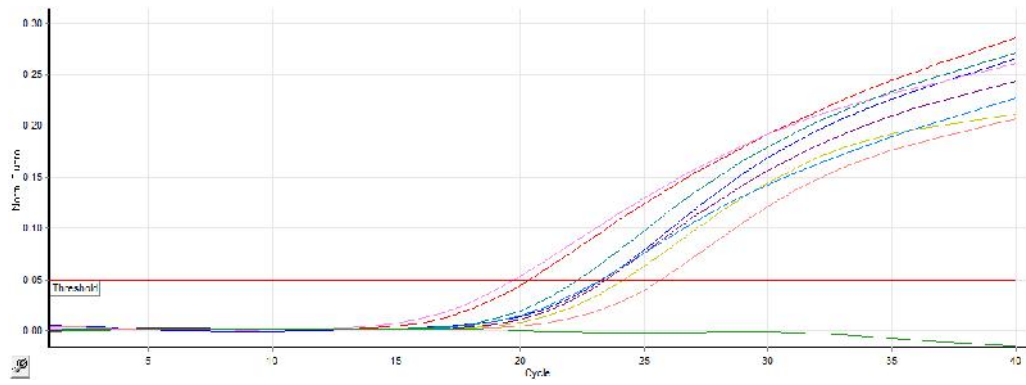
Figure 6: Fluorescence from Fnf-specific TaqMan probes measured on the yellow channel. See Table 23 for sample key.

Table 23: Sample key for Figure 6. All tubes contained Fnf probe.

Colour	Sample name	Threshold exceeded?
■	1	+
■	5	+
■	24	+
■	30	+
■	40	+
■	41	+
■	52	+
■	JCM 3724	+
■	Negative control	-
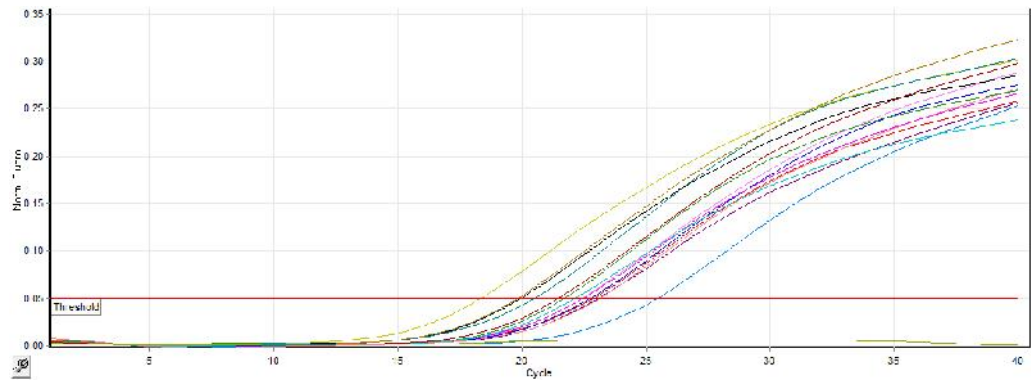
Figure 7: Fluorescence from Fnf-specific TaqMan probes measured on the yellow channel. See Table 24 for sample key.

Table 24: Sample key for Figure 7. All tubes contained Fnf probe.

Colour	Sample name	Threshold exceeded?
	70	+
	80	+
	82	+
	86	+
	87	+
	89	+
	90	+
	91	+
	92	+
	93	+
	94	+
	95	+
	62	+
	JCM 3724	+
	Negative control	-

3.3 Discussion

It had been hypothesised that all strains would identify as *F. necrophorum* during identification tests for this project, due to prior identification of the strains. The isolates received from Antonia Batty, University College Hospital London, had previously been confirmed as *F. necrophorum* as part of previously published work (Batty *et al.*, 2005; Batty and Wren, 2005). The strains received from the ARU consisted of two type strains that had been well characterised and identified to the subspecies level and a clinical isolate that had previously been identified as subspecies *funduliforme*. The biochemical tests that were chosen for identification purposes were therefore not exhaustive. The most important tests were determined to be a Gram stain and the indole test (Barrow and Feltham, 2004).

The *gyrB* TaqMan probe method developed by Jensen *et al.* (2007) was chosen as the method of subspeciation due to its simplicity. The haemagglutinin PCR primer method was avoided as the presence of the haemagglutinin gene in *F. necrophorum* subspecies has not been thoroughly studied in a large sample set. In addition, the type strains ATCC 25286 and ATCC 51357, which are the same strains as JCM 3718 and JCM 3724, respectively, had been tested with the *gyrB* primers and TaqMan probes as part of the Jensen *et al.* (2007) publication, and therefore served as both positive and negative controls for their corresponding/opposing probes.

The strains were first tested with the *gyrB* primers under standard PCR conditions and viewed on a gel to confirm the presence of one band of the expected size before the real-time PCR method was used. In future, it would be enough to test strains using the real-time PCR alone, as detection by the probes is dependent on amplification of the *gyrB* product.

Strain number 88 was determined not to be *F. necrophorum* based on a lack of amplification by the *gyrB* primers. This therefore demonstrated the hypothesis that all strains would prove to be *F. necrophorum* was invalid. Had the molecular techniques not have been used, additional biochemical tests would have been required to ensure an accurate identification. It was, however,

noted that the appearance and smell of strain 88 were not as expected, and more emphasis would have been placed on this, if not for the *gyrB* PCR and TaqMan probe work.

It had also been hypothesised that all clinical strains would be identified as *Fnf.* Upon completion of this section of the work, there were 28 *F. necrophorum* strains available for use in this study. 25 were *Fnf* strains from the UCLH collection, one *Fnf* clinical strain from the ARU collection, known to have caused Lemierre's syndrome in a patient, and one type strain of each subspecies. Of the clinical strains that were identified as *F. necrophorum*, 100 % were identified as *Fnf*, which is in agreement with the hypothesis.

Chapter 4:

Genomics of *F. necrophorum*

4.1 Introduction and aims

DNA sequencing has progressed a great deal since Sanger *et al.* (1977a) sequenced the first DNA genome, that of bacteriophage φX174, using the 'plus and minus' method. This method was soon developed into what is now commonly referred to as the Sanger sequencing method (Sanger *et al.*, 1977b). In 1995 the first bacterial genome, that of *Haemophilus influenzae*, was sequenced in full using a whole-genome, random sequencing method (Fleischmann *et al.*, 1995). Large sequencing projects have contributed a great deal to the advancement of DNA sequencing techniques and subsequent data analysis, particularly The Human Genome Project, which was launched in 1990. The initial draft and analysis of the genome was published in 2001, by which time the cost of sequencing had dropped 100-fold (Lander *et al.*, 2001), making DNA sequencing a far more accessible tool for researchers. More recently, large numbers of bacterial genomes have been sequenced as part of projects such as the Human Microbiome Project (HMP) and the Genomic Encyclopedia of Bacteria and Archaea. The HMP began in 2008 with the aim of sequencing microbial samples obtained from multiple body sites of healthy volunteers in order to study the correlation between human health and the microbiome. A metagenomic sequencing approach was used and in excess of 12,000 DNA samples were sequenced (Aagaard *et al.*, 2013). As there had been some debate over whether *F. necrophorum* is a commensal or strictly a pathogen, as reviewed by Riordan (2007) and discussed in section 1.3, it was unclear whether *F. necrophorum* genomes would be sequenced as part of the project. Also, if they were to be sequenced, there was uncertainty regarding the reliability of metagenomic data due to the nature of sequencing from mixed DNA

samples. A reliable genome sequence of an *F. necrophorum* type strain was required if the organism and its pathogenic mechanisms were to be better understood.

Bacterial genomes are comprised of double-stranded DNA as single circular chromosomes. They tend to have a high density of genes, with about 90 % of the DNA present encoding proteins. Plasmids may also be present as extrachromosomal DNA (Madigan and Martinko, 2006).

There are a variety of techniques used for bacterial genome sequencing, assembly and annotation, which change as new technologies become available. Next generation sequencing, also known as high-throughput sequencing, includes technologies such as Roche 454, Illumina, Ion Torrent and Pacific Biosciences (summarised in Table 25). The choice of technology is dependent on the project type and available funding. At the time the current project was planned, 454 sequencing was the primary choice for *de novo* sequencing projects, i.e. where there is no similar genome available to be used as a reference sequence.

454 sequencing uses pyrosequencing, which involves pyrophosphate being released each time a nucleotide is incorporated by DNA polymerase which subsequently leads to the production of light by the enzyme luciferase. The amount of light produced is proportional to the number of nucleotides incorporated. 454 sequencing produces longer sequence reads than Illumina, allowing for easier assembly, however the data is susceptible to errors in homopolymer regions due to the nature of simultaneously adding multiple nucleotides (Mardis, 2008).

Illumina technologies tend to be more cost effective than 454 sequencing and use a method resembling traditional Sanger sequencing (Hodkinson and Grice, 2015). At each base position, reversible terminators with removable fluorophores are incorporated and identified, then the fluorophore removed ready for the next base incorporation (Bentley *et al.*, 2008). Sequencing with Illumina technologies such as HiSeq or MiSeq are often used where a reference genome is already available, or to resequence a genome in order to correct errors, and improve coverage and assembly by providing additional sequence data. PCR and Sanger sequencing, which are not appropriate for acquiring high throughput data, may then be used to complete the genomes and

bridge gaps in the sequence. Due to financial constraints, resequencing and gap closing steps are often excluded. Errors in Illumina sequencing tend to be substitution errors rather than frame shift errors as nucleotides are added one by one (Hodkinson and Grice, 2015). Frame shift errors affect annotation to a much greater extent than substitution errors, as with a substitution a protein can still be annotated in full, whereas with a frameshift the remaining protein sequence will be read incorrectly or may be abruptly terminated.

Ion Torrent machines, which were not widely available at the time the current project was in the planning stages, were designed to be inexpensive and more accessible to research laboratories. Here, the technology monitors pH as nucleotides are added during the sequencing reaction, detecting H+ ions that are released as a by-product. As with 454 technology, Ion Torrent sequence data is also susceptible to errors in homopolymer regions.

A different type of technology is used by Pacific Biosciences, whose machines utilise a nanotechnology chip with DNA polymerases attached to detectors. Nucleotides with phospholinked dye-labels are incorporated and imaged in real time during the synthesis of DNA strands. The technology has a high base call error rate, however each DNA fragment is sequenced multiple times which corrects most errors and leads to a high level of accuracy (Mardis, 2013; Hodkinson and Grice, 2015).

Table 25: Summary of sequencing technologies and associated advantages and disadvantages of each.

Platform	Chemistry	Highest average read length	Advantages	Disadvantages
Roche 454	Pyrosequencing	700 bp	Longer read lengths are better for de novo assemblies	Errors in homopolymer regions Not cost effective
Illumina	Reversible dye terminator	300 bp	Cost effective Errors are usually substitution errors rather than frame shift errors	Short read length
Ion Torrent	Proton detection	400 bp	Inexpensive	Errors in homopolymer regions
Pacific Biosciences	Phospholinked fluorescent nucleotides	8,500 bp	Very long read lengths High level of accuracy	Expensive

Sequence data is created in multiple formats; FASTA format is a simple text-based format in which DNA or protein sequence is displayed using single letter nucleotide or amino acid code. Each sequence begins with a greater than symbol (>) and an identifying sequence header, containing the name of the sequence and/or a description. FASTQ format contains FASTA format along with corresponding data quality scores and is also a text based format. 454 sequence data uses Qual format instead of FASTQ. The difference being that in Qual files the FASTA data is not present, as it

is with FASTQ files, but quality scores are in the same order as the corresponding sequence file containing FASTA format data.

Once genomes have been sequenced the raw output data must be assembled by looking for overlapping sequence regions. Contiguous sequence data, known as contigs, are formed in order to create a putative reconstruction of the genome that is as complete as possible. Two common methods used in assemblers are Overlap Layout Consensus methods, which rely on overlap graphs, and de Bruijn Graph Methods, which use K-mer graphs. K-mers are sequences of base calls of length K, which can be any positive integer. Sequence reads that share K-mers in overlapping regions are connected into a graph which generates connected sequences (Miller *et al.*, 2010). Many sequence assembly algorithms are based on de Bruijn graphs, such as Velvet (Zerbino and Birney, 2008), ABYSS (Simpson *et al.*, 2009) and AllPaths (Butler *et al.*, 2008). K-mers are ideal for reads with short lengths and high coverage. They are less computationally expensive as redundant data is compressed (Miller *et al.*, 2010). Overlap Layout Consensus (OLC) methods are used for *de novo* assemblies and have three steps. Firstly overlaps are found by comparing every sequence against every other sequence, known as an all-against-all alignment, in the forward and reverse orientations. This is a very computationally expensive step and creates a graph of overlaps which are subsequently simplified during the layout step into contigs. The consensus stage uses the multiple sequence alignments to exclude sequencing errors. Higher sequencing coverage results in better error correction (Li *et al.*, 2012). Newbler (Margulies *et al.*, 2005) is a genome sequence assembler designed by 454 specifically for 454 data and uses the OLC method (Miller *et al.*, 2010).

Statistics of interest for genome sequence data are the coverage, number of contigs, and contig sizes, including N50 and N90 sizes. Coverage is determined by the total length of all sequenced bases in a genome divided by the total number of expected bases. It equates to the average number of reads that each base position is present on. The number of contigs indicates how close to completion the genome is. The N50 and N90 values are the sequence lengths at which 50 % and 90 % of the total assembly residues are contained within contigs of this size or larger.

Once a genome is assembled into contigs the next aim is to annotate the genes it contains. Open reading frames are defined by coding sequence with an initiation codon and stop codon in the same frame, and exceed a pre-defined base pair length. The number of genes are recorded along with their function. Non-protein encoding genes such as tRNAs and rRNAs are often also annotated. Automated web-based tools such as xBASE (Chaudhuri *et al.*, 2008), BASys (Van Domselaar *et al.*, 2005) and RAST (Aziz *et al.*, 2008) can be used for gene prediction and annotation. As part of the genome annotation process, virulence genes are often characterised based on sequence homology to those that have been previously characterised by laboratory analysis.

The Basic Local Alignment Search Tool (BLAST) (Blast: Basic Local Alignment Search Tool, 2015) can be used as part of automatic or manual annotation. BLAST finds regions of similarity between sequences by comparing nucleotide or protein input sequences to sequences in a database, and calculates the statistical significance of the matches. Expect (E) values are reported and correspond to the likelihood of the match occurring by chance. The lower the E value, the more significant the match is. BLAST can be used to identify conserved domains and members of gene families as well as providing clues to functionality and evolutionary history relationships between sequences (Altschul *et al.*, 1997). BLAST has several different programs, including BLASTn for searching a nucleotide database with a nucleotide input sequence and BLASTp for searching a protein database with a protein input sequence. Pfam (Pfam, 2015) is a protein database used for identifying protein families, domains and inferring protein function (Finn *et al.*, 2014), complementing BLAST query search results well. Most commonly, an automated, often web-based, tool is used for annotating genome data. The results may subsequently be manually curated. The annotation results are based on *in silico* predictions and are therefore hypothetical protein annotations, unless they have also been experimentally shown.

Different strains of the same species can show variation in the genes they contain and within the sequences of their genes, affecting metabolic pathways, virulence mechanisms, antimicrobial resistance mechanisms and genes that help the bacteria adapt to their environment. Bioinformatic tools can be used to compare genomes in order to identify these differences (Sangal *et al.*, 2014).

Programs such as Artemis Comparison Tool (Carver *et al.*, 2005), Mauve (Darling *et al.*, 2004) and CGView (Stothard and Wishart, 2005) are all available as open source software for comparative work.

When this project started, in September 2011, there was no available genome sequence for *F. necrophorum*. Genomic data was becoming increasingly important and providing new methods for analysing organisms and the genes that they contain. The aim of this chapter was to provide the first genome sequence for both subspecies of *F. necrophorum*, complete with annotation. This was to identify genes in *F. necrophorum* that were previously unknown. It was also to increase the sequence data available for individual genes within *F. necrophorum* that have already been sequenced and studied, while also providing additional information such as the surrounding genes and operons. Type strains JCM 3718 and JCM 3724, alongside ARU 01, a clinical strain of human origin associated with Lemierre's syndrome (section 3.2.1) were sequenced. It was hypothesised that the genomes would be approximately 2.2 Mb, based on the sequenced genome of *F. nucleatum* (Kapatral *et al.*, 2002); the most closely related organism with a completed genome. It was also hypothesised that once annotated the genome data would contain virulence genes such as the leukotoxin, which had previously been described in *F. necrophorum*, and additional virulence genes that had not previously been described in *F. necrophorum*. An additional aim of the chapter was to investigate whether differences in virulence between the two subspecies could be explained by genomic differences, testing the hypothesis that there was a simple explanation for the difference in virulence between *Fnn* and *Fnf*.

The best available method was used for *de novo* sequencing and assembly, which at the time was 454 sequencing and Newbler assembly. The genomes were annotated using the xBASE pipeline (Chaudhuri *et al.*, 2008) and BLASTp, and were subsequently mined in order to search for genes reported to be associated with virulence in other organisms, with the intention of further characterising some of these. Once the genomes had been sequenced, the first *F. necrophorum* genome became available online and was subsequently followed by additional genomes. The initial genome sequencing and analysis was carried out in 2012. In 2015, an update on the genomic data

available was incorporated to compare the data from other projects to the data generated as part of this project.

4.2 Results

4.2.1 Genome sequencing and assembly

JCM 3718, JCM 3724 and ARU 01 were commercially sequenced by GATC Biotech (Konstanz, Germany) using Roche 454 GS-FLX+ next generation sequencing. The raw sequence reads, which had mean sizes of 301, 299 and 296 nucleotides for JCM 3718, JCM 3724 and ARU 01 respectively, were assembled into larger contigs using Roche GS Assembler, also known as Newbler. Table 26 summarises the resulting data. Genome coverage was calculated using Equation 2 and an initial estimate of 2.17 Mb for the expected number of base pairs in each genome, based on the genome size of *F. nucleatum* (Kapatral *et al.*, 2002). Sequence information that later became available (first sequence shown in Table 28) led to this figure being revised to 1.96 Mb, which lead to the value 1,960,000 being used in the calculations in Table 26.

Equation 2: Equation to calculate the coverage of a genome sequence.

$$\text{\textit{Coverage} = \frac{\textit{\underline{(Number of raw reads x mean size of raw reads)}}}{\textit{Expected number of base pairs in genome}}}$$

Files containing the raw sequence reads, the associated quality (Qual) files and the FASTA format assembly contigs are located in Appendix files A – I.

	JCM 3718 (*Fnn*)	JCM 3724 (*Fnf*)	ARU 01 (*Fnf*)
Number of raw reads	85,434	64,685	90,274
Mean size of raw reads	301	299	296
Genome coverage	13.12X	9.87X	13.63X
Number of assembled contigs	812	629	436
Average contig size	2,835	3,097	4,715
Largest contig	28,070	40,328	41,389
Total number of bases in contigs	2,301,868	1,947,814	2,055,836

4.2.2 Contig annotation using xBASE and BLAST

The assembled contigs were uploaded into the xBASE pipeline (Chaudhuri *et al.*, 2008) in FASTA format for prediction of open reading frames and gene product annotations, using *F. nucleatum* as a reference (methods section 2.9.2). Table 27 summarises the xBASE annotation data. Files containing the xBASE annotation data in Genbank format are located in Appendix files J – L.

Table 27: Summary of xBASE annotation showing the number of predicted open reading frames (ORFs) for each genome, and the proportion that were assigned a gene product. Those unlabelled or labelled as hypothetical were later annotated using BLAST.

	JCM 3718 (*Fnn*)	JCM 3724 (*Fnf*)	ARU 01 (*Fnf*)
Total number of ORFs predicted by Glimmer	1991	1754	1918
Number assigned gene product	1194 (60 %)	1112 (63 %)	1167 (61 %)
Number unlabelled or labelled as hypothetical	797 (40 %)	642 (37 %)	751 (39 %)

xBASE output files are produced in GenBank format. To convert them to a user-friendly comma-separated value file to be used with Microsoft Excel, a program was created using an OpenVMS operating system to extract the information of interest (Program code in Appendix M). The GenBank file was the input and a comma-separated value file was the output. The resulting Excel spreadsheet had data under the headings: assembly contig, base location, orientation, product and protein sequence.

The predicted open reading frames that were unlabelled or had been labelled as hypothetical were then manually annotated using BLASTp searches using the methods described in section 2.9.4. The results were added to the Microsoft Excel file containing the xBASE annotation results (Appendix files N – P). See section 4.2.6 for genes of interest.

4.2.3 Comparison of additional *F. necrophorum* draft genomes

Following the genome sequencing from section 4.2.1 of this project, the *F. necrophorum subsp. funduliforme* ATCC 51357 genome sequence was made available online as part of the Human Microbiome Project (HMP). This strain from the American Type Culture Collection is also known as

JCM 3724, from the Japan Collection of Microorganisms and hence should be identical to the DNA sequenced as part of the current project. This resulted in two sets of available genome data for this strain.

The genome sequence for strain *Fnf* 1_1_36S was the first *F. necrophorum* genome available online, labelled as the reference genome for the Human Microbiome Project. This genome sequence was later labelled as misassembled and was therefore excluded from the following comparisons.

Additional draft *F. necrophorum* genomes have also been recently sequenced and made available on the NCBI online database (National Center for Biotechnology Information, 2015) and are listed in Table 28. One of these, *F. necrophorum* subspecies *funduliforme* strain B35, has been published (Calcutt *et al.*, 2014). See section 4.2.6 for genes of interest that were found.

The strains that have been sequenced are mostly of animal origin therefore there is still a need for genome sequences of *F. necrophorum* isolated from humans, particularly those associated with disease.

Table 28: F. necrophorum genomes downloaded from the NCBI database.

Strain	Origin	NCBI assembly number	Sequencing technology	Assembly method	Coverage	Size (Mb)	Contigs/ Scaffolds	Genes	Date submitted
D12	Human	GCA_000158295.2	454	Newbler	15x	1.96	17	1942	27/07/2011
ATCC 51357*f*	Bovine liver abscess	GCA_000262225.1	Illumina	Newbler	23.27x	2.11	45	2026	10/05/2012
Fnf 1007*f*	Human	GCA_000292975.1	Illumina	Newbler	30x	2.17	87	2145	27/08/2012
B35*f*	Bovine liver abscess	GCA_000600355.1	454	Newbler	59x	2.09	40	1978	26/03/2014
HUN048	Rumen microbiome	GCA_000622045.1	Illumina HiSeq	Allpaths	Unknown	2.03	17	1945	08/04/2014
BL	Bovine liver	GCA_000691645.1	Illumina MiSeq	Velvet	22.26x	2.46	235	2309	15/05/2014
DAB	Deer jaw abscess	GCA_000691705.1	Illumina MiSeq	Velvet	17.92x	2.52	254	2347	15/05/2014
BFTR-1	Bovine footrot	GCA_000691685.1	Illumina MiSeq	Velvet	15.77x	2.53	304	2412	15/05/2014
BFTR-2	Bovine footrot	GCA_000691725.1	Illumina MiSeq	Velvet	11.96x	2.61	389	2599	15/05/2014
DJ-1	Deer Jaw	GCA_000691665.1	Illumina MiSeq	Velvet	11.47x	2.46	370	2393	15/05/2014
DJ-2	Deer Jaw	GCA_000691745.1	Illumina MiSeq	Velvet	25.98x	2.52	226	2382	15/05/2014

f = subspecies *funduliforme*

The average sizes of the contigs in Table 28 were calculated by dividing the genome size by the number of contigs. These values were then plotted against the number of genes present in each genome (Figure 8). Gene numbers were retrieved by Batch Entrez (Batch Entrez, 2015). Nonlinear regression was applied using GraphPad Prism and a strong trend, with an R squared value of 0.94, seemed to show that smaller average contig sizes are linked with higher numbers of genes. Smaller average contig sizes are a result of higher contig numbers and/or poor assembly. The graph does not take into consideration the varying overall sizes of the genomes but still has a strong pattern, suggesting that all *F. necrophorum* genomes may have very similar numbers of genes present and be approximately the same size. The graph was estimated to plateau at 1,950 genes. The majority of the variation in total gene numbers may be due to the assembly and annotation. Gaps in the genome may cause a gene to be split over two contigs, being counted as two genes instead of one. The more gaps in a sequence there are, the more false genes that will be counted.
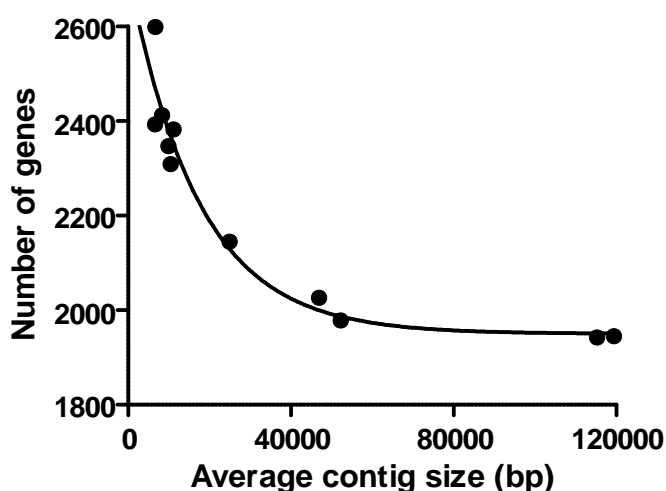
Figure 8: XY plot of average contig size against the number of overall genes annotated in the genomes listed in Table 28.

The majority of *F. necrophorum* genomes available in Table 28 have not been subspeciated. Isolates ATCC 51357, *Fnf* 1007 and B35 have been previously identified as subspecies *funduliforme*. As the strains were not available for laboratory based work, *in silico* analysis of the genome data was the only option for subspeciating the strains. The DNA sequences of the *gyrase B* PCR primers and TaqMan probes used in section 3.2.4 and the *gyrB* sequences from each of the genomes were downloaded from the European Nucleotide Archive website (European Nucleotide Archive, 2015).

The sequences were trimmed to exclude sequence upstream of the forward primer and downstream of the reverse primer and a phylogenetic tree was made using Seaview and PhyML (methods section 2.10.6). Figure 9 shows two distinct groups of strains. The bottom group contains strains of bovine and deer origin and they appear to have identical sequences in this region. JCM 3718 is known to be subspecies *necrophorum*. The top group contains strains that were previously known to be subspecies *funduliforme*, with the exception of D12 and HUN048 which were not subspeciated. They are noted as being from human origin, so subspecies *funduliforme* seems likely, suggesting that those in the top group are *Fnf*, while those in the bottom group are *Fnn*.
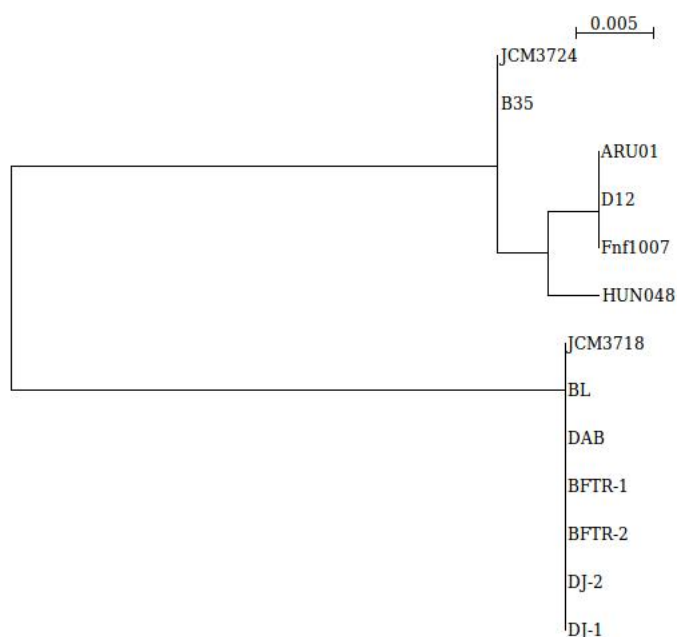
Figure 9: Phylogenetic tree of gyrB sequences to subspeciate strains, constructed using Seaview and PhyML. The scale represents the average number of base changes per position for the length shown. Longer branch lengths therefore represent a greater number of base changes.

In addition to the phylogenetic tree in Figure 9, a Clustal alignment was made of the top group sequences against the subspecies *funduliforme* TaqMan probe sequence, and the bottom group sequences against the subspecies *necrophorum* TaqMan probe (Appendix 8). In both cases, the sequences were all 100 % congruent with the expected matching probe.

4.2.4 Contig merging JCM 3724 data for assembly improvement

As there were two available genome data sets for strain JCM 3724/ATCC 51357, the genome assembly was improved by merging the two sets of data in order to bridge gaps between the contigs. The sequence data for ATCC 51357 was produced by Illumina sequencing and had been assembled using Newbler. The 45 assembled contigs were downloaded from the European Nucleotide Archive (Data contained in Appendix Q).

The Illumina set of contigs from the HMP and the 454 set of contigs from this project were sent to Source Bioscience to be merged into one set of contigs using the program Minimus2, in order to reduce the number of contigs (Resulting FASTA file in Appendix R).

Statistical analysis of the FASTA contigs from the Illumina contigs, 454 contigs and the Minimus2 merged data was carried out using the program Stats, within a Linux operating system. The minimum, maximum and average lengths of the contigs were calculated, along with the N50, N80 and N90 values, the number of 'N's within each set of contigs, the number of contigs and the total residues for each set of data (Table 29). The N50 is the length at which 50 % of the total assembly residues are contained within contigs of this size or larger. The N80 and N90 correspond to 80 % and 90 %, respectively.

The Minimus2 data had a larger average contig size and smaller number of contigs. It also had a larger number of total residues, leaving fewer bases to be sequenced if the gaps are to be filled in future work. The Minimus2 data is a substantial improvement on the 454 sequence data of JCM 3724. This is reflected in the marked increase in the N50, N80 and N90 statistics in Table 29.

Sequence lengths:	JCM 3724 454 data	ATCC 51357 Illumina data	Minimus2 output
Minimum	112	530	116
Maximum	40,328	183,913	305,638
Average	3,096.68	46,906.71	60,714.86
N50	5,720	85,366	95,412
N80	2,294	53,256	66,012
N90	1,308	38,220	41,642
'N' count	9	0	0
Number of contigs	629	45	35
Total residues	1,947,814	2,110,802	2,125,020

4.2.5 BASys statistics

In addition to the statistics calculated using the Stats program in Table 29, the BASys annotation pipeline was used to compare the gene lengths of the JCM 3724 genome before and after it was merged with the sequence data available under the strain name ATCC 51357. Figure 10 shows the changes that occurred to the predicted genes as a result of having a more complete assembly. The JCM 3724 sequence data was 1,947,814 bases in length and was predicted to include 2,521 genes. After the data was merged with the additional sequence data available as ATCC 51357, the number of bases increased to 2,125,020 and the number of genes predicted by BASys reduced to 2,368. This supports the suggestion that a higher number of contigs increases the number of false gene predictions. It appears that the additional gaps in the contigs cause more genes predicted of between 100 bp and 200 bp. The reduced number of contigs causes a more even spread of gene sizes. Graphs for strains JCM 3718 and ARU 01 are shown in Appendix 9. They show that JCM 3718, with 812 contigs, and ARU 01, which has 436 contigs also have a large proportion of small genes

that are likely to be overestimated. The predicted number of genes is higher than the number

predicted by xBASE, however as xBASE is no longer operational, the merged data cannot be

analysed with this annotation pipeline and BASys was used in its place.
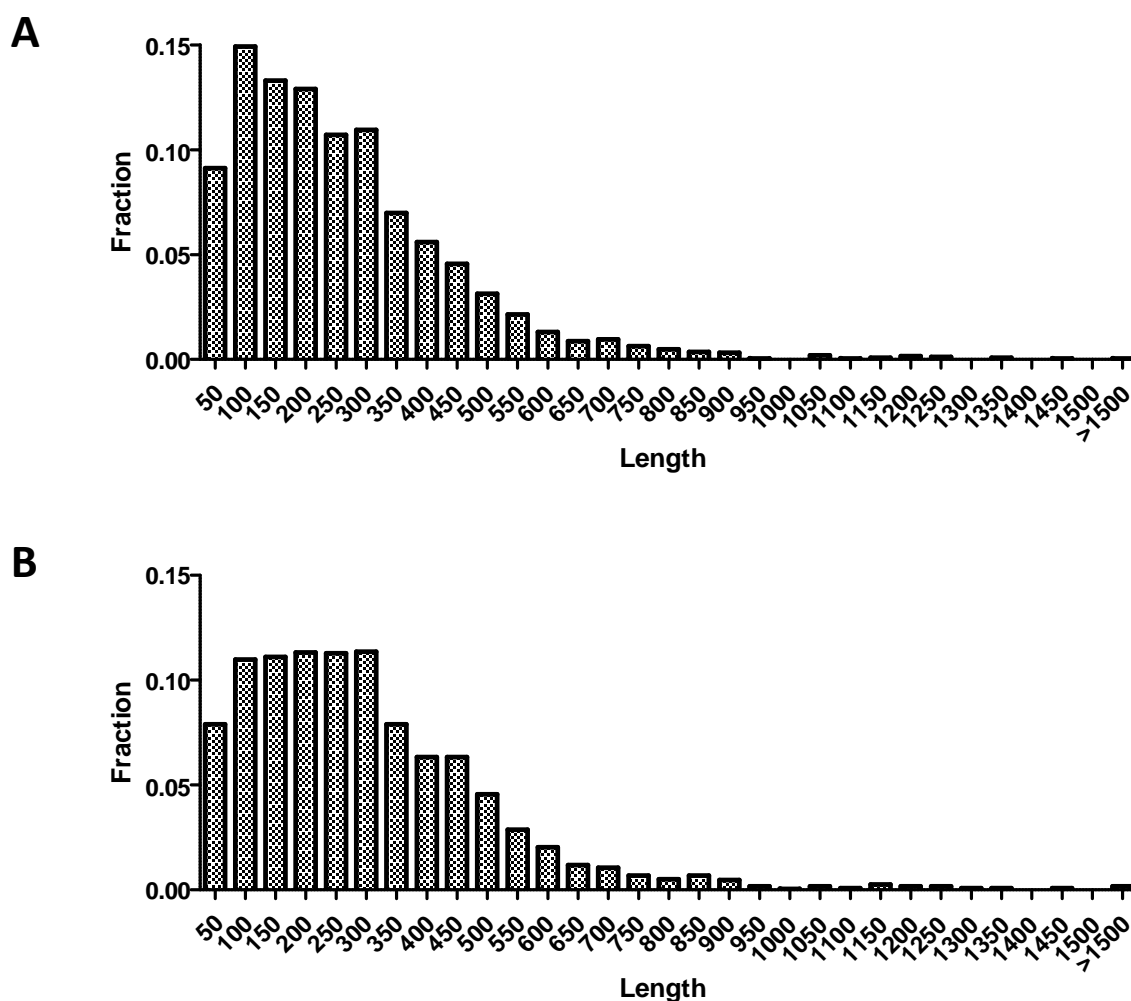
Figure 10: Gene lengths predicted by BASys in A) JCM 3724 genome data, and B) genome data sets merged using Minimus2.

4.2.6 Putative virulence genes

The annotation data from section 4.2.2 was mined for genes hypothesised to be related to virulence

of the organism, based on characterisation within other organisms. The results were compared to

those found in strain B35 (Calcutt *et al.*, 2014) and also to the annotation data present on the BioCyc

database (Biocyc, 2015) for strains ATCC 51357 and D12. Calcutt *et al.* (2014) discussed finding

unique genes linked to lipopolysaccharide biosynthesis, an adhesin-filamentous haemagglutinin

and four CRISPR loci. A selection of the virulence genes found in the genome data of JCM 3718, JCM 3724 and ARU 01 are discussed below.

4.2.6.1 Leukotoxin

The annotation for B35 on the ENA database (European Nucleotide Archive, 2015) includes the leukotoxin, which is thought to be the primary virulence factor of the organism (Tan *et al.*, 1994a). The leukotoxin operon, containing a total of three genes, was found in the genome annotations of JCM 3718, JCM 3724 and ARU 01 (annotation data in Appendix files N – P). Strains D12 and ATCC 51357 are annotated on the BioCyc database as containing the leukotoxin operon. While this is clearly annotated and named for strain D12, the ATCC 51357 strain includes it as a hypothetical protein and does not feature the label leukotoxin. The leukotoxin is investigated in further detail in chapter 5.

4.2.6.2 Ecotin

Strains JCM 3718, JCM 3724 and ARU 01 were all found to contain a serine protease inhibitor, known as ecotin. This was found by manual BLASTp search and not by the xBASE automatic annotation pipeline. Strains ATCC 51357 and D12 include ecotin in their annotation data, and searching the ENA database for ecotin within *F. necrophorum* shows that it is also present in strains *Fnf*1007, B35, BFTR-1, BFTR-2, DAB, BL, DJ-1 and DJ-2. This protein from *F. necrophorum* has not previously been characterised in laboratory experiments and is investigated in chapter 6.

4.2.6.3 *FadA* adhesin

As discussed in section 1.8.5, the FadA adhesin is a known virulence determinant in *F. nucleatum* and other oral *Fusobacterium* spp., *F. periodonticum* and *F. simiae*. The xBASE and BLASTp genome annotation data (section 4.2.2) was mined for the FadA gene and no copies of the gene were initially found. In order to further test for the presence of the gene within *F. necrophorum* genomes, PCR primers were used, designed to amplify a product of 359 bp based on the *F. nucleatum FadA* gene (Han *et al.*, 2005) (Table 30). Late in the project, when additional genomes had become available online, the BLAST searches were repeated for the FadA adhesin, and copies of the gene were found

in strains DJ-1, DJ-2, BFTR-1, BFTR-2, BL, DAB and JCM 3718. There was not sufficient time or resource left to investigate this further as part of this project, but this would be studied as part of future work.

As part of the strain identification work in chapter 3, strain number 88 was excluded from the *F. necrophorum* clinical collection due to its atypical morphology and non-amplification of the *F. necrophorum gyraseB* gene. The biochemical profile and morphology (section 3.2.2) indicate a possibility that it could be a *F. nucleatum* strain. The strain was therefore determined to have potential for amplification of FadA and providing an indication of how a positive control may appear for this gene. Strains JCM 3718, JCM 3724, ARU 01 and all 25 of the clinical *F. necrophorum* isolates were tested and resulted in non-amplification of FadA. See Figure 11 for example gel. Due to non-amplification of the gene product among the *F. necrophorum* strains, the identification of strain 88 was not confirmed and the FadA gene was deemed unlikely to be present within any of the strains in the collection that were identified as *F. necrophorum*.

Table 30: PCR primers used to amplify FadA gene.

Primer	Oligonucleotide sequence (5' → 3')
FadA_F	TTAGCTGTTTCTGCTTCAGC
FadA_R	TTACCAGCTCTTAAAGCTTG
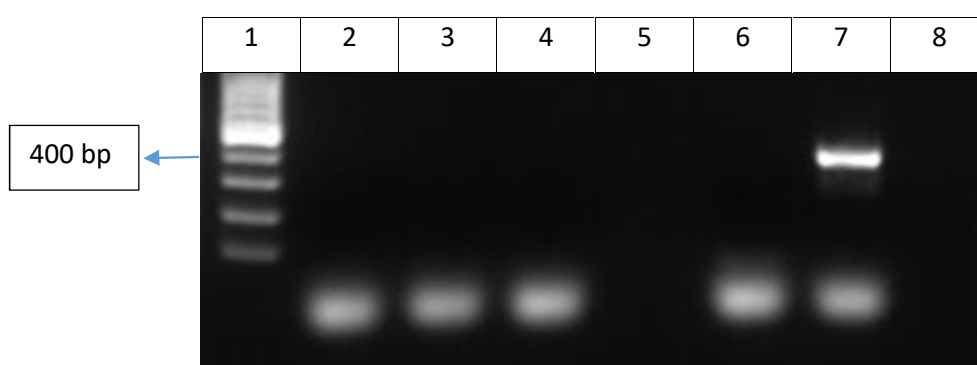
Figure 11: Analysis of PCR products from FadA primers on a 1% agarose gel. Lane 1: 100bp ladder; lanes 2 – 4: strains JCM 3718, JCM 3724, ARU 01, respectively; lane 5: blank; lane 6: negative control (molecular grade water); lane 7: previously excluded strain 88; lane 8: blank.

4.2.6.4 Type V secretion systems

There are three subtypes of type V secretion systems: Type Va are autotransporters, type Vb are two-partner secretion (TPS) systems, and type Vc are trimeric autotransporter adhesins. TPS systems are comprised of a secreted TpsA protein and a TpsB transporter protein. The TpsA protein is an exoprotein that also includes a signal peptide and a TPS domain that makes it specific to the TpsB protein. The TpsB protein is a β-barrel transporter and also contains a signal peptide and a polypeptide transport-associated domain (POTRA) (Figure 12). Each TpsB protein is specific for the TpsA protein that it secretes (Hodak and Jacob-Dubuisson, 2007).
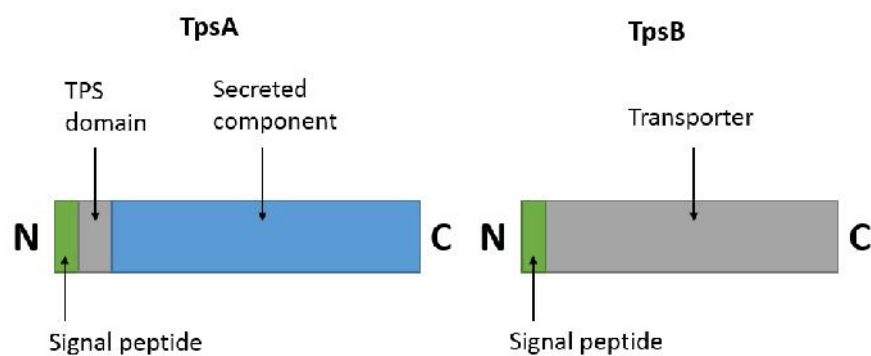
Figure 12: Protein organisation of TPS secretion system.

TPS systems are often used for secreting large virulence proteins, such as adhesins. Examples include the filamentous haemagglutinin (FHA) of *Bordetella pertussis* and haemolysin of *Serratia marcescens*. They contain conserved motifs with amino acid sequences NPNL and NPNGI within the TPS domain, although some TPS systems may contain slight variations (Jacob-Dubuisson *et al.*, 2001).

The BLASTp annotation data of strains JCM 3718 and JCM 3724 was found to contain TPS secretion domains, with motifs NPNL and NENGI, 35 residues apart within a haemolysin protein. Figure 13 shows a section of a protein sequence Clustal alignment of *B. pertussis* FHA and haemolysins from JCM 3718 and JCM 3724 (xBASE protein annotation numbers 1,601 and 1,234, respectively). ARU 01 contained a haemolysin gene with NDNGI 39 bases upstream of NKNL. The TPS motifs were not found in any other proteins.

```
fhaB       NP-GVVFNNGLTDGVSRIGGALTKNPNLTRQASAILAEVTDTSPSRLAGTLEVYGKGADL 172
JCM3718    DKNNVIFNNSQKNGTSVTGGEVSANPNLTNSASVILNEIQGNSASELNGGLEVFGKRADL 118
JCM3724    DKNNVIFNNSQKNGTSVTGGEVSANPNLTNSASVILNEIQGNSASELNGGLEVFGKRADL 151
           :   *:***. .:*.*  ** :: *****..**.** *:  .* *.* * ***:** ***

fhaB       IIANPNGISVNGLSTLNASNLTLTTGRPSVNGGRIGLDVQQGTVTIE--RGGVNATGLGY 230
JCM3718    VIANENGINVNGARFINTSALTLSTGKVSVDNKKISFNTATNNAKIAVKEKG-IETDSDY 177
JCM3724    VIANENGINVNGARFINTSALTLSTGKVSVDNKKISFNTATNNAKIAVKETG-MKTDSDY 210
           :*** ***.***   :*:* ***:**: **: :*.::.  ...*  . *   *   *
```

Figure 13: Clustal Omega protein alignment of B. pertussis FHA, a well characterised TPS system, and haemolysins from JCM 3718 and JCM 3724. The alignment is trimmed to focus on the TPS domain motifs, highlighted in yellow.

Autotransporters differ from the TPS system in that the gene is coded in a single coding sequence. There is one signal peptide, followed by the sequence for both the secreted protein and the transporter (Figure 14). The secreted protein, also known as the passenger domain, is passed through the outer membrane by the transporter, also known as the translocation domain. The two regions remain connected by a linker region (Hodak and Jacob-Dubuisson, 2007). Type Va and Vc secretion systems are both autotransporters, however trimeric autotransporter adhesins are formed by three monomers that each include head, stalk and anchor regions (Linke *et al.*, 2006).
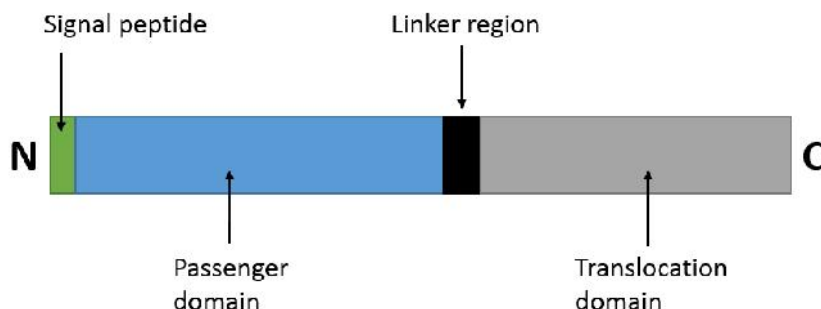
Figure 14: Protein organisation of autotransporter secretion system.

The BLASTp annotation data of strains JCM 3718, JCM 3724 and ARU 01 was found to include head, stalk and anchor regions of a trimeric autotransporter adhesin (TAA) similar to the YadA adhesin in *Yersinia spp*. YadA-like adhesins have been found in other organisms, such as *Moraxella catarrhalis*, and have been linked with autoagglutination and binding to host cells (Hoiczyk *et al.*, 2000).

4.2.6.5 CRISPR-associated genes

As mentioned in the B35 strain publication by Calcutt *et al.* (2014), *F. necrophorum* contains CRISPR-associated genes. These were also found within strains JCM 3718, JCM 3724 and ARU 01. The CRISPR-associated genes annotated by the manual BLASTp search are listed in Table 31.

Table 31: CRISPR-associated genes

JCM 3718 (*Fnn*)	JCM 3724 (*Fnf*)	ARU 01 (*Fnf*)
cas1	cas1	cas1
cas5	cas5	cas5
csx8	csx8	csx8
DevR	DevR	DevR
cst1		

When the ATCC 51357 database was searched on BioCyc, the operon in Figure 15 was detailed. CRISPR-associated genes have recently been shown to be effective tools for genome editing (Liu and Fan, 2014; Pennisi, 2013) and therefore they have been noted here as being present within the genomes, however they are not investigated further as part of this project.

Figure 15: CRISPR-associated (cas) genes present in ATCC 51357 database on BioCyc.

4.2.6.6 Additional putative virulence genes

The JCM 3718, JCM 3724 and ARU 01 genomes were also found to contain genes associated with phage, hemin receptors and iron acquisition, a filamentous haemagglutinin and other genes of interest listed in Table 32. These are annotated in Appendix files N – P but were not selected for further investigation within this project.

4.2.6.7 Prevalence of virulence genes within *Fusobacterium spp*. and other organisms and rationale for further work

Of the gene products covered within this chapter and in Table 32 below, many of these are prevalent within other organisms, while others are more specific to certain *Fusobacterium spp.*

The leukotoxin gene, *lktA*, has been found in *F. necrophorum spp*. (Tan *et al.*, 1994a) and *F. equinum (Tadepalli et al., 2008)*, but no other *Fusobacterium spp*. (Oelke *et al.*, 2005). Other leukotoxins have also been found in *Staphylococcus aureus* (Yoong and Torres, 2013), *Mannheimia haemolytica* (Highlander et al., 1989) and *Actinobacillus actinomycetemcomitans* (Karakelian et al., 1998), although there is very little similarity between the *lktA* gene and any other leukotoxins when compared by DNA or protein sequence alignments. This suggests that the *lktA* gene is unique to *F. necrophorum* and *F. equinum*.

The presence of ecotin has been shown in *Escherichia coli* (Chung *et al.*, 1983) and other Gram negative bacteria, including *Pseudomonas aeruginosa* and *Yersinia Pestis*, (Gillmor *et al.*, 2000). Searching online databases such as NCBI reveals additional ecotin sequence data for Gram negative bacteria. Prior to this project, ecotin had not been described in *Fusobacterium spp*. There are now sequences available on the NCBI database for *F. necrophorum*, *F. equinum* and *F. gonidiaformans*. No sequence data was found for ecotin in *F. nucleatum* despite a completed genome being available.

The FadA adhesin is unique to oral *Fusobacteria spp*., and was found to be absent from non-oral species, although *F. necrophorum* was not tested (Han *et al.*, 2005). Upon initial searches of the genomes of JCM 3718, JCM 3724 and ARU 01 during this project, no copies of the *FadA* gene were found. When tested with *F. nucleatum FadA*-specific PCR primers all *F. necrophorum* strains had no amplification of product. However, very late into the project, the *FadA* gene was located in several *Fnn* genomes using BLAST programmes. These strains, identified as presumptive *Fnn* in section 4.2.3, were JCM 3718, DJ1, DJ2, DAB, BL, BFTR-1 and BFTR-2.

Type V secretion systems are only present in Gram negative bacteria, due to the presence of the outer membrane in these organisms but not Gram positive bacteria (Leo *et al.*, 2012). These secretion systems are often associated with virulence factors that can be either maintained on the surface of the cell or secreted into the cell's surroundings (Leo *et al.*, 2012; Desvaux *et al.*, 2005). Secretion systems Va, Vb and Vc have been found in *F. nucleatum* using bioinformatics techniques (Desvaux *et al.*, 2005) and it was therefore expected that *F. necrophorum* would contain some of these, to support the organism's need to adhere to host surfaces in order to be pathogenic. Type V secretion genes found are briefly covered in section 4.2.6.4.

CRISPR-associated genes are found in many bacterial species (Kirchner and Schneider, 2015). The CRISPR-Cas system is considered to be a prokaryotic immune system, protecting the organism from foreign invading genetic elements (Rath *et al.*, 2015). A search of *Fusobacterium* CRISPR genes on the BioCyc website returns sequences for *F. necrophorum, F. nucleatum, F. gonidiaformans, F. mortiferum, F. periodonticum, F. russii, F. ulcerans* and *F. varium*.

Overall, the leukotoxin and ecotin sections of work were chosen as they were considered to be achievable within the restricted budget and timeframe, the facilities and tools were available for the experiments required, and there was potential for novel findings. The leukotoxin has been described as the organism's primary virulence factor (Tan *et al.*, 1994a) and was therefore considered important to investigate. Ecotin was a protein that had not previously been described in *F. necrophorum* and therefore had potential for novel findings.

The *FadA* gene was discovered in the *Fnn* strains too late in the project to proceed with experimental work, however it would be interesting to investigate as part of future work. Assays to investigate the haemagglutinin and haemolysin as part of a study into type V secretion systems were considered in depth, however it was concluded that the facilities required for the assays were not available, therefore work on these proteins was not taken any further. CRISPR genes became very topical in scientific literature during the course of this project, however it would have been

unlikely to achieve any novel findings that were specifically linked with *F. necrophorum* pathogenesis, as opposed to CRISPR processes in general, therefore these genes were not studied as part of this project. Phage-associated proteins and hemin receptor and transport system proteins were considered of interest, but it was concluded that these proteins were not the most likely to yield novel results and they were not investigated further.

Virulence-associated protein E and virulence factor mviN are both claimed to be virulence related, however there was not enough information available on these proteins to make an assessment of the likelihood of producing any novel work. Drug resistance proteins are important in terms of virulence and infections *in vivo*, however were not intended as the focus of this project, so were not investigated further. The lipid A biosynthesis protein from Table 32 was of interest due to lipid A being a component of endotoxic lipopolysaccharide (Caroff and Karibian, 2003). The capsular polysaccharide biosynthesis protein was also of interest due to *F. necrophorum* possessing a mucopolysaccharide capsule (Brook and Walker, 1986). Investigating either of these two proteins was not considered likely to produce novel information. Neutrophil-activating protein A could play a role in virulence of *F. necrophorum*. *Helicobacter pylori* is known to produce its own neutrophil-activating protein which acts as a virulence factor, stimulating immune responses (Amedei *et al.*, 2006). Investigating this in *F. necrophorum* could potentially yield some important novel findings, however this gene was not located until late into the project and the facilities to investigate this were also not available.

Table 32: Summary table of annotated genes of interest

Gene product	Gene present in:		
	JCM 3718	JCM 3724	ARU 01
leukotoxin operon	✓	✓	✓
haemolysin	✓	✓	✓
haemagglutinin	✓	✓	✓
FadA	✓	NF	NF
YadA-like adhesin	✓	✓	✓
ecotin	✓	✓	✓
CRISPR/CRISPR-associated proteins	✓	✓	✓
haemin receptor and binding/transport systems	✓	✓	✓
Phage proteins	✓	✓	✓
5-nitroimidazole antibiotic resistance protein	✓	✓	✓
acriflavin resistance proteins	✓	✓	✓
multi-drug resistance proteins	✓	✓	✓
Na+ driven multidrug efflux pump	✓	✓	✓
penicillin-binding protein	✓	✓	✓
florfenicol resistance	✓	✓	✓
virulence factor mviN	✓	✓	✓
virulence-associated protein E	✓	✓	NF
capsular polysaccharide biosynthesis protein	✓	NF	✓
Lipid A biosynthesis protein	NF	✓	✓
neutrophil-activating protein A	✓	✓	✓

NF = Not found

4.3 Discussion

At the time of sequencing, in early 2012, the use of 454 sequencing and Newbler assembly was an appropriate choice for *de novo* genome sequencing. The Pacific Biosciences technology was too expensive and Ion Torrent was not widely available at the time. 454 was more suitable than Illumina for *de novo* sequencing due to producing longer mean read lengths.

Genome sequencing has dramatically increased in popularity, with lower costs making the technology more accessible to researchers. The increasing number of *F. necrophorum* genomes, listed in Table 28, demonstrates this with only one submitted to the NCBI in 2011, increasing to eight in 2014. Table 28 also shows how the choice of sequencing technology switched from 454 to Illumina. This could be due to the fact that there was no longer the need to carry out a *de novo* assembly once an initial *F. necrophorum* sequence was available and also that the Illumina sequencing technology was improving, producing longer reads and better value for money. There does not appear to be an overall improvement in levels of coverage of the genomes between 2011 and 2014. Rather than being representative of sequencing technology it is likely to be due to project planning and the coverage deemed necessary for each genome, considering that additional coverage comes at additional cost.

A more appropriate choice for *F. necrophorum de novo* sequencing would now be Pacific Biosciences technology, due to the long read lengths, high level of accuracy and lower costs than in 2012. Although considering the number of genomes already available, *de novo* sequencing is no longer required, and Illumina sequencing now is the most cost effective method to acquire *F. necrophorum* genome sequence data. 454 will not be an option for much longer, as according to Hodkinson and Grice (2015), Roche are no longer supporting the 454 sequencing platforms after 2016. GATC Biotech have already stopped commercially using the technology.

The switch from 454 to other sequencing platforms also signals the potential switch away from Newbler assembly, which as mentioned in section 4.1 is specifically designed for 454 data (Miller *et al.*, 2010). Velvet is becoming much more popular and more recent versions are allowing for larger

K-mers, which is beneficial as regions of sequence repeats are subsequently more likely to be spanned (Chikhi and Medvedev, 2014). In addition to this, the web-based annotation pipeline xBASE, which was used for this project is no longer available for use. The field of genomics has advanced a great deal over recent years and as new technologies and analytical software become available, older ones become redundant.

At the start of this project there was no available genome sequence data for *F. necrophorum*. With much unknown about the pathogenesis of the organism during infection, and the greater emphasis recently on the use of genomic data, it was clear that the bacterium needed to be sequenced. Despite the addition of eleven genomes to the NCBI database, there remains only one published article on the *F. necrophorum* genome. The article, published by Calcutt *et al.* (2014) covers a brief description of a draft genome. There is still clearly a great deal of the genome left to be investigated.

The genome data provided by the work in this chapter represents the most complete assembly for strain JCM 3724/ATCC 51357 by combining the 454 and Illumina sequencing data. This strain is important as it is the type strain for *F. necrophorum* subsp. *funduliforme* (Shinjo *et al.*, 1991). The genome data also provides the first genome sequence for strain JCM 3718, also known as ATCC 25286, which is the type strain for *F. necrophorum* subsp. *necrophorum* (Shinjo *et al.*, 1991). These type strains are used as references during various types of characterisation experiments and provide a method of external quality control. For example, during the TaqMan PCR work by Jensen *et al.* (2007) that was used in section 3.2.4. The third genome sequenced as part of this project is the first to be recognised as originating from human disease. This is important as there may be virulence genes present in strains that are capable of causing LS that could be lacking from other *F. necrophorum* strains. Therefore, if the genome is to be studied for virulence factors associated with disease, then isolates should be used from disease sources.

While there is still no complete genome for *F. necrophorum*, the data provided in this chapter, along with the genomes available on the NCBI database (listed in Table 28), provide a great deal of information that was lacking at the start of this project, and also opportunity to further study the organism. By mapping the different sets of genome contigs against one another there is also enough

data to predict the order of some of the contigs in data sets with small contig numbers and begin the process of gap closing by PCR and Sanger sequencing.

The coverage of the three sequenced genomes ranged between 9.87X and 13.63X (Table 26). The level of coverage expected was 13X – 24X, as predicted by the GATC Biotech sequencing specialist. Factors that may cause a decrease in coverage include biases in the sample preparation or the sequencing (Sims *et al.*, 2014). As mentioned in section 4.2.1, the coverage is determined using the values for the number of raw reads, the mean size of the raw reads, and the expected number of base pairs in the genome. The low level of coverage with the GATC Biotech data appears to be due to the read lengths, detailed in Table 26. The 454 data was expected to produce read lengths with a mean of 700 bp, however it produced mean read lengths of approximately 300 bp. This is more closely aligned with the mean read length of Illumina technology. It seems that the short read length is the likely cause of the coverage being lower than expected, although no explanation was provided by GATC Biotech. As a consequence of this, the number of contigs for each genome was quite large, which is a problem if trying to complete the genome by gap closing, but does not greatly hinder the study of the genes contained within the genome.

The total number of bases in each of the three sequenced genomes ranges from 1.95 Mb – 2.3 Mb. Once the JCM 3724 454 data was merged with the Illumina data, the total number of bases in the assembly increased from 1.95 Mb to 2.13 Mb, causing the new size range for the three genomes to be 2.06 Mb – 2.3 Mb. This closely matches the hypothesis in section 4.1, which stated that the sequenced *F. necrophorum* genomes were expected to be approximately 2.2 Mb, due to the size of the *F. nucleatum* genome sequenced by Kapatral *et al.* (2002). The genome sequence of strain B35, published in 2014, was reported to be 2.09 Mb (Calcutt *et al.*, 2014), and it is also within the range of genome sizes listed in Table 28: 1.96 Mb – 2.61 Mb. When the genome sizes from Table 28 are separated according to assembly software, the size range for Newbler is 1.96 – 2.17 Mb, whereas the size range for Velvet assemblies is 2.46 Mb – 2.61 Mb, suggesting the choice of assembly software has a large effect on the total number of bases. When the phylogenetic tree (Figure 9) used to subspeciate the strains is taken into account, it can be seen that the genomes assembled

with Newbler are all *Fnf*, while the Velvet assembled genomes are all *Fnn*, so perhaps the *Fnn* genome is larger than that of *Fnf*. Such large differences seem unlikely, however, and differences in assembly methods are the most probable cause.

The annotation of *F. nucleatum* by Kapatral *et al.* (2002) using the ERGO bioinformatics suite resulted in 67 % of ORFs assigned a function. Using xBASE, with *F. nucleatum* as a reference, between 60 – 63 % of the ORFs in the *F. necrophorum* genomes were annotated (Table 27). Both organisms are likely to contain some genes that are absent in the other, such as the leukotoxin in *F. necrophorum* but not *F. nucleatum*, therefore the 60 – 63 % ORF annotation rate was within the expected range. The remaining ORFs were analysed using BLASTp to annotate as much of the genomes as possible and provide some manual curation. The graph in Figure 8 and associated non-linear regression suggest that the number of genes in the *F. necrophorum* genome is approximately 1,950. This is similar in number to 2,067 genes present in *F. nucleatum* (Kapatral *et al.*, 2002). If the *Fnn* genome was in fact larger than that of *Fnf*, then the data points would perhaps have an R squared value far lower than 0.94.

While GenBank is a commonly used file format and is compatible with some viewer programs, such as Artemis, it is not easy to read by eye. The script written for conversion to excel file format provides a table that is simple to read, and also allows easy extraction of the protein sequence.

The additional genomes that have been sequenced and are available on the NCBI database have been assembled and include some annotation data, however no analysis of this has been published, with exception of strain *Fnf* B35 which has brief details published (described in section 4.2.6). The subspecies of the majority of these *F. necrophorum* strains was unknown. The use of a phylogenetic tree to subspeciate these genomes allows for a better understanding of the data provided and allows for comparisons to be made between the subspecies. It also supports the general understanding that animal infections are caused by *Fnn* and that *Fnf* can be found in both animals and humans. The phylogenetic tree (Figure 9) and the alternative method of using Clustal alignments with the sequence of the *gyrB* TaqMan probes (4.2.3 and Appendix 8) were both an effective use of bioinformatic tools where no laboratory option was available. The tree provides

some additional information in that the *gyrase B* sequences of *Fnn* are very well conserved whereas there is a small amount of variation between the *Fnf* isolates. Given that these strains infect a wider range of hosts, that is perhaps not so surprising.

The genomes for strains JCM 3718, JCM 3724 and ARU 01 were mined for genes previously reported to be associated with virulence in other organisms. The hypothesis in section 4.1 stated that the genome annotation data would contain the leukotoxin and additional virulence genes that had not previously been described in *F. necrophorum*. The leukotoxin, which has been described as the organisms primary virulence factor (Tan *et al.*, 1994a), was found in all three genomes. Ecotin, a serine protease inhibitor which had not previously been described in *F. necrophorum* was also found in all three. *E. coli* ecotin has been associated with increased clotting times in human blood (Ulmer *et al.*, 1995) and this protein was therefore determined to be of interest. The *FadA* adhesin gene, known to be present in *F. nucleatum* (Han *et al.*, 2005), was briefly investigated in *F. necrophorum* genomes and found in JCM 3718, as well as other *Fnn* genomes. Chapters 5 and 6 discuss the leukotoxin and ecotin in further detail.

Motifs were found within a haemolysin gene in strains JCM 3718 and JCM 3724 (Figure 13), which suggest the protein is secreted via a two-partner secretion system. This has not previously been described in *F. necrophorum*. The motifs within strain ARU 01 are less well conserved and also in the reverse order, however, not all TPS proteins need both motifs (Jacob-Dubuisson *et al.*, 2001), so secretion may be unaffected if this system is indeed used. A YadA-like adhesin that acts as an autotransporter was also found within the annotation data. This, like the TPS system, is another type V secretion system. The head, stalk and anchor regions were all located.

With so little information published on the *F. necrophorum* genome, the information in this chapter represents the first description of several *F. necrophorum* virulence factors.

It had been hypothesised that there was a simple explanation for the difference in virulence between *Fnn* and *Fnf*, such as genomic differences between the strains. After studying and comparing the annotated genomes it was concluded that the differences in virulence cannot be

explained by gross differences in the genomes and that variations in virulence between strains is much more complex. Host and environmental factors are likely to play a role as well as subtle gene differences which could make virulence proteins more or less effective.

Chapter 5:

Analysis of *F. necrophorum* leukotoxin sequence, prevalence and activity

5.1 Introduction and aims

Many bacteria are known to produce toxins, which may be either toxic proteins released extracellularly (exotoxins) or toxic lipopolysaccharides maintained as part of the outer layer of the cell envelope (endotoxins) (Madigan and Martinko, 2006). Bacterial toxins cause damage to human and animal target cells in a variety of ways, such as by forming pores in membranes, inhibiting protein synthesis, activating secondary messenger pathways, activating the immune system and by protease action (Schmitt *et al.*, 1999). Examples of bacterial toxins include the leukotoxins of *Staphylococcus aureus* (Yoong and Torres, 2013), *Mannheimia haemolytica* (Highlander *et al.*, 1989) and *Actinobacillus actinomycetemcomitans* (Karakelian *et al.*, 1998). Leukotoxins are a type of exotoxin that are toxic to leukocytes, particularly polymorphonuclear cells (PMNs) (Narayanan *et al.*, 2002a). *F. necrophorum* subspecies *necrophorum* (*Fnn*) and *F. necrophorum* subspecies *funduliforme* (*Fnf*) are both known to produce a leukotoxin (lktA) that is a water soluble, heat-labile, protein exotoxin with an affinity for ruminant and human leukocytes (Tan *et al.*, 1994a; Tadepalli *et al.*, 2008a).

During *F. necrophorum* infection, the leukotoxin appears to be the primary virulence factor (Tan *et al.*, 1994a), with a correlation between the amount of toxin produced by strains and their ability to cause abscesses in laboratory animals demonstrated by Emery *et al.* (1986). The toxin is known to target blood cells. When tested, PMNs from cattle and sheep were the most susceptible to the leukotoxic supernatant, whereas pig and rabbit PMNs were unaffected (Tan *et al.*, 1994a).

The leukotoxin, which can be found in the culture supernatant of the two subspecies of *F. necrophorum* (Tan *et al.*, 1994a), has been described as larger than other bacterial leukotoxins, with a molecular mass of 336 kDa (Narayanan *et al.*, 2001b). The leukotoxin of *Mannheimia haemolytica* is 105 kDa (Chang *et al.*, 1987), *Actinobacillus actinomycetemcomitans* is 114 kDa (Kachlany *et al.*, 2002), and *Staphylococcus aureus* is 37 kDa (Marshall *et al.*, 2000). The *F. necrophorum* leukotoxin has also been described as potentially novel, due to its lack of sequence similarity to other known bacterial leukotoxins (Narayanan *et al.*, 2001b). More recently, BLASTp searches of lktA protein sequences have revealed homology of approximately 25 – 30 % to adhesin and haemagglutinin proteins. Oelke *et al.* (2005) investigated the presence of the leukotoxin in *Fusobacterium spp.* using Southern blotting and Western blotting for leukotoxin DNA and protein, respectively. They reported a positive finding for *F. necrophorum* and negative findings for *F. nucleatum, F. gonidiaformans, F. mortiferum, F. necrogenes, F. simiae, F. ulcerans* and *F. varium.* The leukotoxin has also been found in *F. equinum*, an organism closely related to *F. necrophorum* that causes equine necrotic infections. The lktA protein was shown to be secreted, but lktB and lktC, the remaining proteins coded for by the leukotoxin operon in *F. necrophorum* were not found (Tadepalli *et al.*, 2008). This finding was supported by Zhou *et al.* (2009b) who confirmed the presence of the *F. equinum* leukotoxin but reported difficulties in identifying the DNA sequence data, resulting in only a short fragment of the *lktA* gene being sequenced. It would be of great interest to compare the leukotoxins of *F. necrophorum* and *F. equinum*, however due to the lack of available *F. equinum* sequence data and lack of access to *F. equinum* strains, the *F. equinum* leukotoxin was not studied as part of this project.

The mode of action of the *F. necrophorum* leukotoxin is unknown. A limited amount of research has been carried out on the leukotoxin, predominantly by a research group based in Kansas State University, Kansas, USA. The purification of the leukotoxin and determining the DNA sequence of the *lktA* gene have been particularly valuable contributions, enabling in depth analysis of purified leukotoxin, cloning experiments, molecular detection work and bioinformatic analysis.

In order to be able to carry out affinity purification of *F. necrophorum* leukotoxin, Tan *et al.* (1994b) produced monoclonal antibodies (mAbs) of the IgG class by injecting mice with semipurified leukotoxin. Antibodies were harvested, purified by affinity chromatography and labelled with biotin. The group showed that two of the mAbs partially neutralised the activity of the leukotoxin, but only when they used polyclonal serum could they completely neutralise the leukotoxin activity. This may suggest that the epitopes of the mAbs may be inaccessible when the leukotoxin is interacting with target cells, or in a region that does not largely affect toxin activity. The mAbs remain useful for affinity purification purposes.

Narayanan *et al.* (2001b) produced a library of bovine strain *F. necrophorum* subsp. *necrophorum* A25 genomic DNA fragments between 10 and 12 kb in length into Lambda Zap Express vectors for cloning and expression. Anti-leukotoxin polyclonal serum was used to determine which of the clones contained the leukotoxin. The recombinant DNA from the leukotoxin clone was sequenced and inverse PCR was used to extend the sequence to encompass the entire leukotoxin gene. The open reading frame was shown to be 9,726 bp, encoding for a protein of 3,241 amino acids and a molecular weight of 335,956 Da. The secreted protein was described as having substantial hydrophobicity, with 14 potentially membrane spanning regions, which could be very important in determining a potential mode of action. Pore-forming toxins insert transmembrane pores in the target cell to disrupt ions across the plasma membrane (Schmitt *et al.*, 1999). If the *F. necrophorum* leukotoxin is a pore-forming toxin it would require hydrophobic regions to create the pores.

E. coli host cells containing the first 3.5 kb of the cloned *lktA* gene lysed upon IPTG induction of the protein. This may have been due to the tertiary folding pattern of the truncated leukotoxin leaving an active region more exposed. Narayanan *et al.* (2001b) therefore constructed five smaller truncated polypeptides of the *lktA* gene (P1 – P5), which were purified and analysed by Western blotting for reactivity against polyclonal and monoclonal antibodies raised against native leukotoxin (Figure 16). The polyclonal serum reacted strongly with peptides one, two and five, which equated to bases 1 – 1,107, 919 – 3,696 and 7,405 – 9,723 of the *lktA* gene and weakly against peptides three and four (bases 3,553 – 5,691 and 5,626 – 7,509, respectively). The monoclonal antibody

reacted against the first peptide only. Antisera raised against peptides one and three were shown to have neutralising activity against the native leukotoxin, as determined by a microculture tetrazolium assay. Antibodies raised against peptides two, four and five did not have neutralising activity, suggesting they may be concealed by the folding pattern of the native leukotoxin. This analysis of truncated polypeptides provides some insight into which regions are immunogenic and potentially active toxin regions and may be of use during future sequence analysis. P1 reacted with monoclonal antibodies and strongly with polyclonal serum. Antisera raised against this truncated polypeptide had neutralising activity against native leukotoxin, suggesting that the P1 region of the *lktA* gene may be important to the structure and function of the leukotoxin. The 3.5 kb truncated peptide (which had the approximate length of truncated peptides 1 and 2 combined) was toxic to *E. coli* host cells, however full length recombinant leukotoxin was not. The toxicity of the 3.5 kb truncated peptide to *E. coli* cells further supports the suggestion that the active region may be within region P1 and/or P2. P2 also reacted strongly with polyclonal serum. The full length recombinant protein was found to be extremely unstable, with many breakdown products visible on Western blots. This was also observed with native purified leukotoxin. This will be a factor for consideration during future leukotoxin assays. Recombinant full length leukotoxin was toxic to bovine neutrophils (Narayanan *et al.*, 2001b).
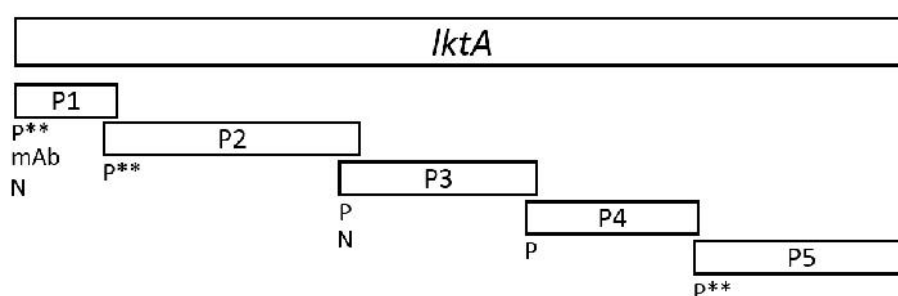
Figure 16: Diagram of five truncated polypeptides of the cloned leukotoxin gene and their respective reactivity with polyclonal serum and monoclonal antibodies. The lktA box represents the 9,726 bp leukotoxin gene and boxes P1 – P5 represent the truncated polypeptides. Labels P**, P, mAb and N represent a strong reaction with polyclonal serum, a weak reaction with polyclonal serum, a reaction with monoclonal antibodies and neutralising activity of native toxin by antisera raised against this truncated polypeptide, respectively.

In order to investigate the mechanism of toxicity induced by the leukotoxin, Narayanan *et al.* (2002b) exposed bovine peripheral leukocytes suspended in RPMI medium to various concentrations of immunoaffinity purified leukotoxin. After 45 minutes at 37 °C in an atmosphere of 5 % CO_2, the cells were washed and resuspended in Hank's balanced salt solution and stained with propidium iodide to evaluate cell viability. Flow cytometry and scanning and transmission electron microscopy were used to analyse the cells post treatment. The flow cytometry data showed that upon leukotoxin treatment the mononuclear cells had a dose-dependent decrease in cell size and an increase in granularity, which are both characteristics of apoptosis. The PMNs had decreased in size but had no significant increase in granularity. The reduction in cell size was suggestive of apoptosis being the mechanism of leukotoxic action, as opposed to necrosis. Numbers of cells decreased significantly at very high concentrations of toxin (>1,250 U/ml) which was suggestive of complete cell lysis. The amount of leukotoxin present in the host during infections is unknown. *F. necrophorum* is known to cause necrotic infections (Riordan, 2007; Hagelskjaer Kristensen and Prag, 2000), which suggests that the organism utilises toxic mechanisms potent enough to cause tissue necrosis to occur.

PMNs viewed with scanning electron microscopy appeared to clump, decrease in size and show lesions resembling pores after treatment with leukotoxin. At very high concentrations they showed large craters and were agglutinated. Indicators of apoptosis were plasma membrane blebbing and apoptotic bodies. Transmission electron microscopy of treated PMNs showed signs of apoptosis, including nuclear collapse, condensed chromatin, condensed cytoplasmic organelles, and a decrease in the cytoplasmic layer. The PMNs also showed signs of activation, such as translocation of granules to the cytoplasm periphery. The study concluded that at low concentrations of leukotoxin, apoptosis and activation were induced and at higher concentrations necrosis occurred. For PMNs, 0.02 – 0.2 U/ml leukotoxin induced apoptosis, 20 U/ml enhanced activation, and >200 U/ml induced necrosis. For monocytes, 2 – 20 U/ml induced apoptosis and >600 U/ml induced necrosis. They speculated that while membrane damage did occur to PMNs, pore formation was an

unlikely cause as the cells would have increased rather than decreased in size due to the uncontrolled entry of water into the cell (Narayanan *et al.*, 2002b).

Tadepalli *et al.* (2008a) tested for the leukotoxin gene and leukotoxic activity in four strains of *F. necrophorum* subsp. *funduliforme* of human origin. Analysis by PCR confirmed that all four strains contained the *lktA* gene and promoter, and using Western blotting the secretion of leukotoxin was confirmed. Human peripheral PMNs were treated with culture supernatants for 45 minutes at 37 °C and 5 % CO_2. Cells were washed twice and resuspended in 0.01 M PBS and stained with propidium iodide. Samples were processed on a flow cytometer to test for viability. All of the culture supernatants from the human strains exhibited toxicity to human PMNs. Bovine PMNs were not tested. The data showed strain to strain variation in the level of cytotoxicity. When culture supernatants were pretreated with proteases, leukotoxic activity decreased. It was therefore suggested that the strain variation in levels of cytotoxicity may be partly related to the levels of proteolytic enzymes secreted by the strains, cleaving the leukotoxin.

Secreted toxins from *Fnn* and *Fnf* had the same banding pattern on a Western blot (Tadepalli *et al.*, 2008b). This may represent consistency with the breakdown of products, or it may be specific proteolytic products. It was suggested by Tadepalli *et al.* (2008b) that these products are breakdown products, as the leukotoxin is known to be unstable (Tan *et al.*, 1994a; Narayanan *et al.*, 2001b), and that the differences in toxicity are therefore due to differing quantities of expression rather than differing stabilities. Recurring intense bands for *F. necrophorum* subsp. *necrophorum* were at ~250, 150, 130 and 110 kDa, amongst other less intense bands. *F. necrophorum* subsp. *funduliforme* produced less intense bands overall, although this may have been due to the polyclonal serum being raised against leukotoxin from a *Fnn* strain (Tadepalli *et al.*, 2008a). These banding patterns will provide a useful comparison for any protein analysis carried out as part of this project.

F. necrophorum leukotoxin has been shown to be very sensitive to changes in temperature and pH. Leukotoxic activity of culture supernatant was retained at 4 °C and 25 °C for up to 4 hours, however incubating at 56 °C or boiling for 5 minutes removed leukotoxic activity. At 37 °C leukotoxic activity

was decreased by 19 %. As this is a physiological temperature it has implications during infections. However, during infections leukotoxin may be continuously produced and therefore degraded toxin would be replaced. Maximal leukotoxic activity occurred between pH 6.6 and 7.8. At pH <6 the leukotoxin was inactivated (Tan *et al.*, 1994a). It is therefore of great importance to carefully store and handle the leukotoxin.

The *Fnn* leukotoxin gene is considered to be part of an operon of three genes (Narayanan *et al.*, 2001b). The leukotoxin operon of *F. necrophorum* subsp. *funduliforme* strain B35 was sequenced in full to include the *lktB, lktA* and *lktC* genes, concluding that both *F. necrophorum* subspecies contained these genes. When the genes were compared, strains A25 (*Fnn*) and B35 (*Fnf*) had similarities of 88% for *lktB*, 90% for *lktA* and 96% for *lktC*. The sequence lengths for *Fnf* B35 were 1,650 bp for *lktB*, 9,732 bp for *lktA* and 438 for *lktC*, as represented in Figure 17 (Tadepalli *et al.*, 2008b).
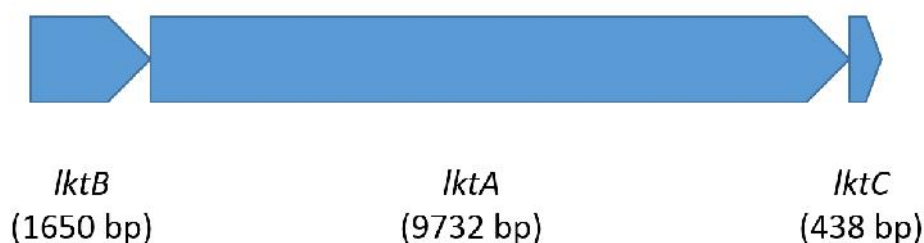
Figure 17: Representation of the leukotoxin operon of F. necrophorum subsp. funduliforme strain B35, showing genes lktB, lktA and lktC along with corresponding gene lengths.

It was found that *lktB* contained a POTRA_2 polypeptide transport associated domain which is highly suggestive of this protein being involved in the secretion of the lktA protein. Most of the differences between the two *lktA* genes were noted to be in the N-terminal and middle section of the gene. The *lktC* gene was described as highly conserved but the function was not hypothesised (Tadepalli *et al.*, 2008b).

When the titre of leukotoxin in the culture supernatant of both *Fnn* and *Fnf* strains was tested, the amount of leukotoxin present in the *Fnn* supernatant was 15-fold higher than the *Fnf* supernatant

(Tan *et al.*, 1994a). Tadepalli *et al.* (2008b) used qPCR to measure the relative expressions of *lktA* and found that for both subspecies the mid-log phase expressed the highest concentration of leukotoxin messenger RNA. When the two subspecies were compared they found that subspecies *necrophorum* had a 21.1 fold higher transcript level of *lktA* during the mid-log phase peak. It was concluded that the *necrophorum* subspecies is more virulent than the *funduliforme* subspecies, due to the likelihood that a higher transcript level of *lktA* would result in a higher level of leukotoxin secretion, and therefore cause a greater amount of cell death to leukocytes during infections. A comparison of the promoter regions of the two subspecies showed variations between the sequences, despite appearing to be conserved within each subspecies. The differing sequences for each subspecies may therefore provide a target for subspeciating strains. When assayed using *E. coli* cells, the *Fnn* promoter activity was approximately four-fold higher than that of the *Fnf* (Zhang *et al.*, 2006). This appears to support the hypothesis that *Fnn* strains produce more leukotoxin mRNA, and subsequently more leukotoxin protein, and are therefore more virulent.

In order to determine the prevalence of the *lktA* gene, a collection of *F. necrophorum* isolates from humans (43) and animals (57 – including mostly cows and sheep) were tested with a panel of custom designed leukotoxin primers. Two *lktA* gene sequences with 99.75% similarity were used to design three sets of primers targeting different regions of the gene. Each strain was tested against all three primer pairs and it was concluded that the *lktA* gene was present in 47 out of 100 strains studied. The *lktA* gene was found to be more prevalent in bovine strains and subspecies *necrophorum* strains than those of human origin, or subspecies *funduliforme* (Ludlam *et al.*, 2009a). This was later disputed by Bennett *et al.* (2010) who observed the dangers of declaring a gene to be absent based on non-amplification of a PCR product. They also highlighted how the two sequences used to design the primers were both of bovine origin and were identical. Several publications by one research group have since reported variations among *lktA* genes (Zhou *et al.*, 2009a; Zhou *et al.*, 2010; Zhou *et al.*, 2011b; Zhou *et al.*, 2011a). In order to more reliably test for the presence of the leukotoxin gene within a collection of isolates, PCR primers should be designed using a range of dissimilar sequences to take a greater level of variation into account.

F. necrophorum type strains JCM 3718 and JCM 3724 have been used by other research groups as control strains in *F. necrophorum* leukotoxin experiments. As previously mentioned in Chapter 4, prior to the start of this project the genomes of these strains had not been sequenced. This project provided the opportunity to sequence the genomes of these two strains, with the addition of a third strain of human clinical origin. This was to provide high quality sequence data of the leukotoxin operon and to carry out *in vitro* tests for cytotoxic activity analysis alongside this, looking at activity in human derived cells. The *F. necrophorum* leukotoxin has been studied previously, but there is still a great deal to be understood regarding its mode of action and the level of cytotoxic activity of the two subspecies. High quality sequence data can greatly contribute to the understanding of the leukotoxin operon, therefore the leukotoxin was considered worth investigating.

For this chapter, the sequence of the *F. necrophorum* leukotoxin operon was investigated within the two type strains JCM 3718 (*Fnn*) and JCM 3724 (*Fnf*) and the human clinical strain ARU 01. The majority of the work carried out on the leukotoxin previously has been by two research groups which investigate from a veterinary perspective, whereas human clinical isolates were also investigated here. The prevalence of the leukotoxin operon within a clinical collection of human isolates was also studied, using PCR primers from other studies, as well as custom primers, in order to critique the previously published primers. It was hypothesised that the leukotoxin would be found to be universal within the *F. necrophorum* strain collection, with the three sequenced genomes containing the *leukotoxin* operon in full. It was also predicted that when the sequences were aligned, those of the same subspecies would demonstrate greater similarity to each other than those of different subspecies, and that variations in activity could be explained by variations in the sequences. To confirm cytotoxic activity of the sequenced strains, concentrated culture supernatants were tested against the HL-60 human promyelocytic leukaemia cell line and human donor white blood cells. It was hypothesised that the concentrated, high molecular weight fraction of culture supernatant would demonstrate a cytotoxic effect in a dose-dependent manner. The antibodies described above have not been made commercially available and therefore were unavailable for use in this project.

5.2 Results

5.2.1 Locating the leukotoxin operons within the genome sequence data of JCM 3718, JCM 3724 and ARU 01.

The *Fnf* strain 1_1_36S draft genome (previously introduced in section 4.2.3; Genbank accession numbers ADLZ01000001-ADLZ01000040) was searched for the leukotoxin operon. Each of the 40 contigs was entered in FASTA format into a nucleotide BLAST search on the NCBI website (Blast: Basic Local Alignment Search Tool, 2015). Search settings were as follows: Database: Nucleotide collection (nr/nt). Organism: *Fusobacterium necrophorum*. One copy of the leukotoxin operon was found on accession number ADLZ01000005 from base pair 80957 to 93479.

As part of the commercial sequencing package, the raw reads from JCM 3718, JCM 3724 and ARU 01 were mapped to accession numbers ADLZ01000001-ADLZ01000040 using GS Reference Mapper. For reads that mapped within bases 80957 to 93479 on accession number ADLZ01000005, the corresponding assembly contigs were recorded in Table 33.

Table 33: Assembly contigs containing a region of the leukotoxin operon.

JCM 3718 (Fnn)	JCM 3724 (Fnf)	ARU 01 (Fnf)
contig00415	contig00225 (complement)	contig00305
contig00592 (complement)	contig00226 (complement)	contig00051 (complement)
contig00005	contig00177 (complement)	
	contig00205	
	contig00132	

Contigs were reverse complemented using a sequence editor (Fr33.Net, 2015) in the case of raw reads mapping to the complement strand. Each set of leukotoxin operon containing contigs were then concatenated into one FASTA file and aligned against bases 80957 to 93479 of accession

number ADLZ01000005 using Clustal Omega. The sequences were trimmed to exclude any bases outside of the region aligning to the ADLZ01000005 leukotoxin sequence. The locations of these operons were confirmed by the genomic annotation in chapter 4.

5.2.2 Sanger sequencing the leukotoxin operon in JCM 3718, JCM 3724 and ARU 01 to close gaps and confirm next generation sequencing data

All three of the leukotoxin operon sequences contained gaps in the next generation sequence data where the leukotoxin sequence had spanned multiple contigs. This was resolved by designing PCR primers to amplify these regions and using Sanger sequencing to obtain the missing bases. The primers were designed using the methods in section 2.10.1 and primer sequences were recorded in Appendix 10. Sanger sequencing was as described in methods section 2.8.7. The resulting sequence data was inserted into the appropriate gaps and a Clustal Omega alignment was carried out between the ADLZ01000005 leukotoxin sequence and the JCM 3718, JCM 3724 and ARU 01 leukotoxin sequences. This ensured the full leukotoxin operon was present in the new sequence data. As discussed in section 4.2.3, assembly contig accession numbers ADLZ01000001-ADLZ01000040 were misassembled. The results of the leukotoxin work were not adversely affected by this due to the leukotoxin sequences being confirmed by Sanger sequencing and aligned to other the leukotoxin genes from other *F. necrophorum* strains.

Next generation pyrosequencing is prone to errors in homopolymer regions (Hodkinson and Grice, 2015) which can lead to frame shift errors. In order to be confident about the leukotoxin operon sequence data a series of PCR primers were designed to Sanger sequence each of the three operons. A total of 37 PCR primers were designed (see Appendix 11 for primer sequences) with an average of 17 primer pairs being used for each strain. PCR, gel electrophoresis, gel extraction, and Sanger sequencing were carried out as described in section 2.8.7. Resulting Sanger sequence data, sequenced in both the forward and reverse orientations, were assembled using Clustal Omega. The next generation sequence data (sequence locations detailed in Table 33) and Sanger sequence data for the leukotoxin operon from each strain was aligned again using Clustal Omega to compare the

data sets. In each discrepancy between the two sets of data, the Sanger sequence data was taken as accurate. All three strains contained frame shift errors in the pyrosequencing data that were corrected by the Sanger sequencing data. The final Sanger sequence data for the three strains can be found as an alignment in Appendix 12.

5.2.3 Bioinformatic analysis of leukotoxin sequence

The leukotoxin operon sequences were translated into protein sequence using ExPASy Translate (methods section 2.10.3). For each operon sequence, the genes for *lktB* and *lktA* were present in full within the Sanger sequence data. There were, however, difficulties with designing primers for the *lktC* sequence due to the DNA sequence following the *lktC* gene being unknown at the time. The *lktC* sequence data is therefore taken from the next generation sequence data.

5.2.3.1 lktB

Figure 18 shows a Clustal Omega alignment of the three lktB protein sequences. The two *Fnf* lktB sequences are both 447 amino acids long and show a 99.8 % similarity. The only difference is at position 174, with a leucine present in the JCM 3724 strain and a valine present in the ARU 01 strain. Both amino acids are non-polar and hydrophobic. The JCM 3718 lktB is 336 amino acids long.

```
3718_lktB      ------------------------------------------------------------      0
3724_lktB      MERIWESYIEKEITITELYTIVQKINELYQEKGYLVCRAVLPAQKIQNGIVNILLIEGKT     60
ARU01_lktB     MERIWESYIEKEITITELYTIVQKINELYQEKGYLVCRAVLPAQKIQNGIVNILLIEGKT     60


3718_lktB      -------------------------------------------------MASGKVPGT        9
3724_lktB      GDITIQGNHSTREKYIKERIPLEKGKISNFKELDRSLTRFNLTNDSPIQVNMTSGKVLGT    120
ARU01_lktB     GDITIQGNHSTREKYIKERIPLEKGKISNFKELDRSLTRFNLTNDSPIQVNMTSGKVLGT    120
                                                                *:**** **

3718_lktB      TDYFVQIYEPKRQQFFVFADNLGQKNTGELRWGLNYINNSVTGNRDQLSLTSLVTEGTAS     69
3724_lktB      TDYFLQIYEPKRQQFFTFADNLGQKNTGELRWGISYINNSVTGNRDQLSLTSLLTEGTAS    180
ARU01_lktB     TDYFLQIYEPKRQQFFTFADNLGQKNTGELRWGISYINNSVTGNRDQLSLTSLVTEGTAS    180
                ****:*********** **************: *******************:******

3718_lktB      LSSFYTFPVSKKGTKISLQHSVGKLKHIQGALKHKITGNSYSYGVGIVHPILVHEKNKVE    129
3724_lktB      LSSIYTFPVSKKGTKVSLQHSLGKLKHIQGALKHKITGNSYSYGIGIVHPVLVDEKNKIE    240
ARU01_lktB     LSSIYTFPVSKKGTKVSLQHSLGKLKHIQGALKHKITGNSYSYGIGIVHPVLVDEKNKIE    240
                ***:*************:*****:***********************:*****:** .****:*

3718_lktB      LSLDWVKQRTVTDLLKLKWVNNRLSKYTAGIGISHYEEDSVFYTKQNITKGKFIPISGDA    189
3724_lktB      LSLDWGRQRTVTDLLKLNWVNNRLSKYTAGIGISHYEEDSIFYTKQNITKGKFIPISGDE    300
ARU01_lktB     LSLDWGRQRTVTDLLKLNWVNNRLSKYTAGIGISHYEEDSIFYTKQNITKGKFIPISGDE    300
                ***** :***********:*******************:******************

3718_lktB      RNYTKYDMFLIYQKNLKYNTLVTLKMAGQYSLSKKLPSVEQIYAGGAYNVRGYPENFMGA    249
3724_lktB      KKYTKYDMFLMYQKNLKYHTLATLRMTGQYSLSKKLPSVEQIYAGGAYNVRGYPESFMGA    360
ARU01_lktB     KKYTKYDMFLMYQKNLKYHTLATLRMTGQYSLSKKLPSVEQIYAGGAYNVRGYPESFMGA    360
                :.:*********:******* **.**:*:*******************.****

3718_lktB      EHGVFFNAELSKLVENKGEFFVFLDGASLHGESAWQENRIFSSGFGYKIRFLEKNNIAVS    309
3724_lktB      EHGIFFNIELSKLVENKGEFFVFLDGASLHGESAWQENRIFSSGFGYRLRFLEKNNIAVS    420
ARU01_lktB     EHGIFFNIELSKLVENKGEFFVFLDGASLHGESAWQENRIFSSGFGYRLRFLEKNNIAVS    420
                ***:*** ****************************************::***********

3718_lktB      MAFPWKKKINSISVDSNRIYITINHEF                336
3724_lktB      MAFPWKKTINSISVDSNRIYITINHEF                447
ARU01_lktB     MAFPWKKTINSISVDSNRIYITINHEF                447
                *******.******************
```

Figure 18: Clustal Omega alignment of the three lktB *protein sequences from JCM 3718, JCM 3724 and ARU 01.*

When entered into a protein BLAST search all three sequences resulted in a match of *Fusobacterium necrophorum* lktB with E values of 0. Table 34 lists the conserved domain hits found by the BLASTp search.

119

Table 34: Conserved domains found within the lktB protein by a BLASTp search.

Conserved domain hit	Description	E value
Polypeptide transport associated (POTRA) domain	Found towards the N-terminus of ShlB family proteins and important in the secretion and activation of the haemolysin ShlA. Hypothesised to have a chaperone-like function over ShlA.	6.36e-14
Surface antigen variable number repeat	Found primarily in bacterial surface antigens, usually as variable number repeats at the N-terminus.	6.36e-14
Surface antigen	This domain is found in outer membrane proteins from *H.influenzae*, *P.multocida*, *N.meningitidis* and *N.gonorrhoeae*.	1.39e-05
Haemolysin activation/secretion protein	Involved in intracellular trafficking, secretion, and vesicular transport.	4.29e-52
outer membrane protein assembly complex, YaeT	Members of this protein family are for assembling proteins into the outer membrane of Gram-negative bacteria. They typically have five tandem copies of a surface antigen variable number repeat, followed by an outer membrane beta-barrel domain.	6.02e-06

The lktB sequences were also entered into a Pfam search to look for conserved domains (Table 35).

Table 35: Conserved domains found within the lktB protein by a Pfam search.

Family	Description	E value
POTRA_2	POTRA domain, ShlB-type	8.3e-13
Bac_surface_Ag	Surface antigen	0.0002

The conserved domains found within the lktB proteins are suggestive of the protein having a role in secretion. The POTRA domain was found during both BLASTp and Pfam searches and is associated

with secretion. Several other domains are linked to the outer membrane of the bacteria and secretion.

5.2.3.2 lktA

A Clustal Omega alignment of the lktA protein sequences of JCM 3718, JCM 3724 and ARU 01 was made to compare the three sequences (Appendix 13). JCM 3724 was demonstrated to be 88.7 % similar to JCM 3718 and 75.3 % similar to ARU 01, while ARU 01 and JCM 3718 were 75.9 % similar to one another. A Clustal Omega alignment was also made using the DNA sequence for the *lktA* region. The resulting similarities were 90.0 % for JCM 3724 and JCM 3718, 78.5 % for JCM 3724 and ARU 01, and 79.1 % for JCM 3718 and ARU 01. This protein appears to be less conserved than the lktB protein. This could be due to sections of the lktA protein not being particularly important to the function of the leukotoxin and therefore not needing to be so well conserved. The lktB protein is considerably smaller and therefore is less likely to have regions that are of low importance.

Due to the large size of the protein, at approximately 3,200 amino acids, the sequence alignment is displayed here graphically using Artemis Comparison Tool (ACT) (Figure 19). The DNA sequences of the three *lktA* genes were submitted for comparison via the WebACT tool (methods section 2.10.5). The grey bars represent the three *lktA* sequences and the red bars represent areas of similarity. White spaces, or breaks, represent unique regions. The five truncated peptide regions from Figure 16 are overlaid onto this diagram. This enables the reactivity of these regions, based on experimental studies by Narayanan *et al.* (2001b), to be compared to how conserved the sequences of these regions are between the strains used in this project. P5 is the only truncated peptide region conserved between all three strains. Peptide regions P1 and P2, which both had a strong reaction with polyclonal serum in leukotoxin experimental studies by Narayanan *et al.* (2001b), show some sequence variation is present. P3 which was associated with neutralising activity shows a great deal of variation between the three strains.
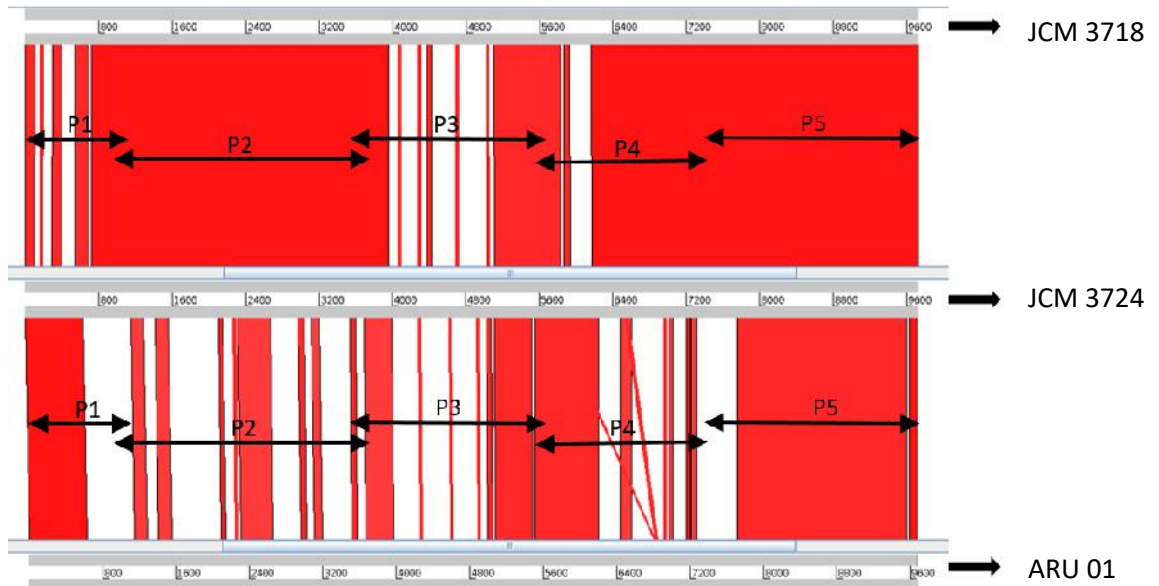
Figure 19: Artemis Comparison Tool alignment of the lktA genes from JCM 3718, JCM 3724 and ARU 01. Grey bars represent the three lktA sequences, red bars represent areas of similarity and white spaces represent unique regions. The five truncated peptides from Figure 16 are overlaid onto this diagram. In Figure 16, region P1 was associated with a strong reaction with polyclonal serum, a reaction with monoclonal antibodies and to neutralising activity of native toxin by antisera raised against truncated P1 peptide. Region P2 was associated with a strong reaction with polyclonal serum only. P3 was associated with neutralising activity of native toxin by antisera raised against truncated P3 peptide and P5 was associated with a strong reaction with polyclonal serum.

A phylogenetic tree was created in order to investigate the relationship between isolate source, DNA sequence and subspecies (Figure 20). DNA sequences of *F. necrophorum lktA* genes were taken from the NCBI website (National Center for Biotechnology Information, 2015) using the nucleotide database, and included with the Sanger sequence data of strains JCM 3718, JCM 3724 and ARU 01. The strains from which the DNA sequences were collated from are detailed in Table 36. The resulting tree shows no tight clustering between *Fnn* and *Fnf* strains. The top five strains are all *Fnf*; the top two are human strains, grouped with a rumen microbiome strain. Strains JCM 3724 and B35 are both of bovine liver abscess origin, and group more closely with the animal *Fnn* strains, although do not cluster tightly with them.

Table 36: Isolates used for collection of lktA gene sequence data.

Isolate name	Source	Subspecies (if known)
ARU 01	Human	*funduliforme*
D12	Human	*funduliforme**
HUN 048	Rumen microbiome	*funduliforme**
JCM 3724	Bovine liver abscess	*funduliforme*
B35	Bovine liver abscess	*funduliforme*
JCM 3718	Bovine liver abscess	*necrophorum*
H05	Bovine footrot	
BFTR1	Bovine footrot	*necrophorum**
DJ-2	Deer jaw	*necrophorum**
DAB	Deer jaw abscess	*necrophorum**
A25	Bovine liver abscess	*necrophorum*
BFTR-2	Bovine footrot	*necrophorum**

* = subspeciation result of phylogenetic analysis in section 4.2.3 (Figure 9).
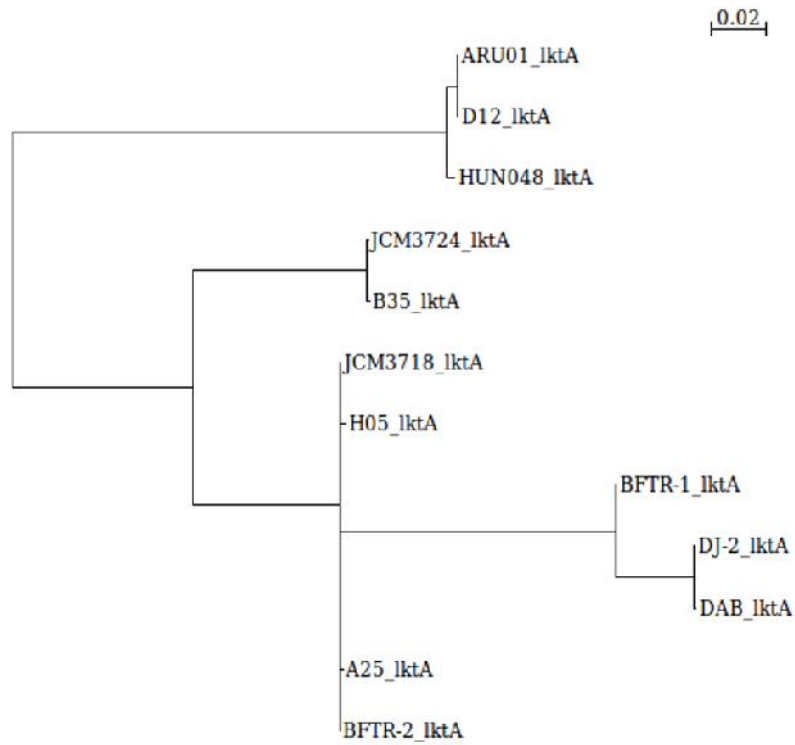
Figure 20: Phylogenetic tree of lktA genes of sequenced strains of F. necrophorum. Alignments were made in Seaview and maximum likelihood trees were made using PhyML. The scale represents the average number of base changes per position for the length shown. Longer branch lengths therefore represent a greater number of base changes.

Kyte and Doolittle hydrophobicity plots were made for the three lktA proteins in order to further characterise them. Large sections of hydrophobicity could indicate that the protein is membrane spanning, which is a requirement for pore-forming toxins. Due to the high level of similarity between the plots, only one is shown, characterising the hydrophobicity of the JCM 3718 lktA protein sequence (Figure 21). Potential transmembrane regions are found by peaks with a score of 1.6 or greater when a recommended window size of 19 is used. The window size represents the number of surrounding amino acids used at each amino acid position to generate the average score that predicts hydrophobicity. Figure 21 shows only two peaks reaching a score of 1.6, which disputes the suggestion of the protein being membrane spanning.
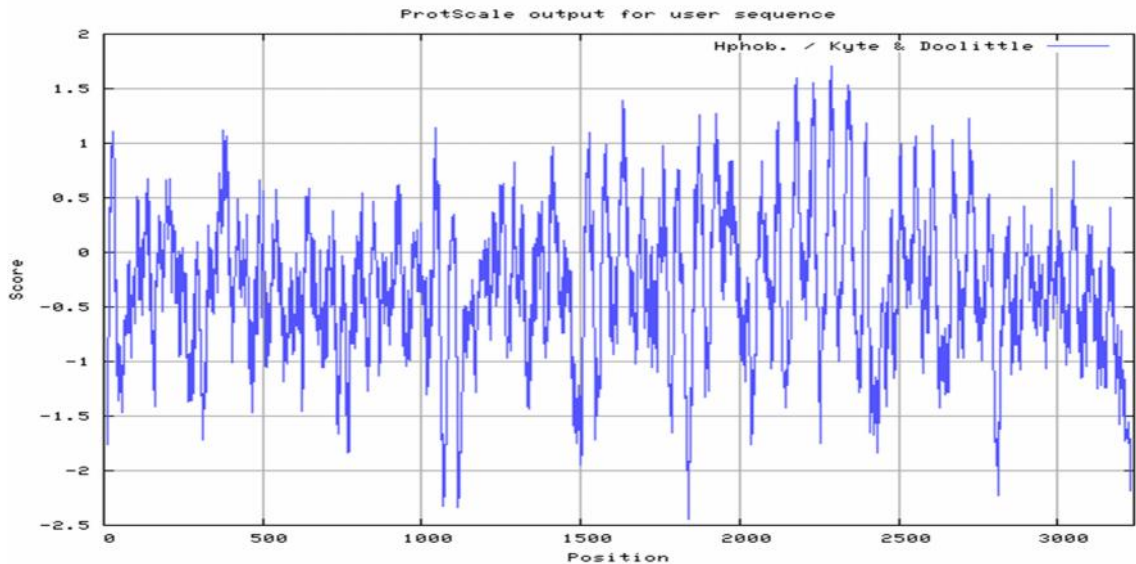
Figure 21: Kyte and Doolittle hydrophobicity plot of the JCM 3718 lktA gene. An amino acid window size of 19 was used.

The first 60 amino acids from the lktA protein from each of the three strains were entered into SignalP (Petersen *et al.*, 2011) for prediction of signal peptides. JCM 3724 and ARU 01 had identical sequences in this region. Figure 22 shows the output of the two *Fnf* sequences and the resulting prediction of a signal peptide and a cleavage site between positions 36 and 37. JCM 3718 had some sequence variation (as shown in the Clustal Omega alignment in Appendix 13), however the prediction was the same, with a signal peptide and a cleavage site between positions 36 and 37. A signal peptide at the N-terminus of the protein is highly indicative that the protein will be targeted to the secretory pathway of the bacterium. The *F. necrophorum* leukotoxin is found in the culture supernatant, but the method of secretion is not yet known.
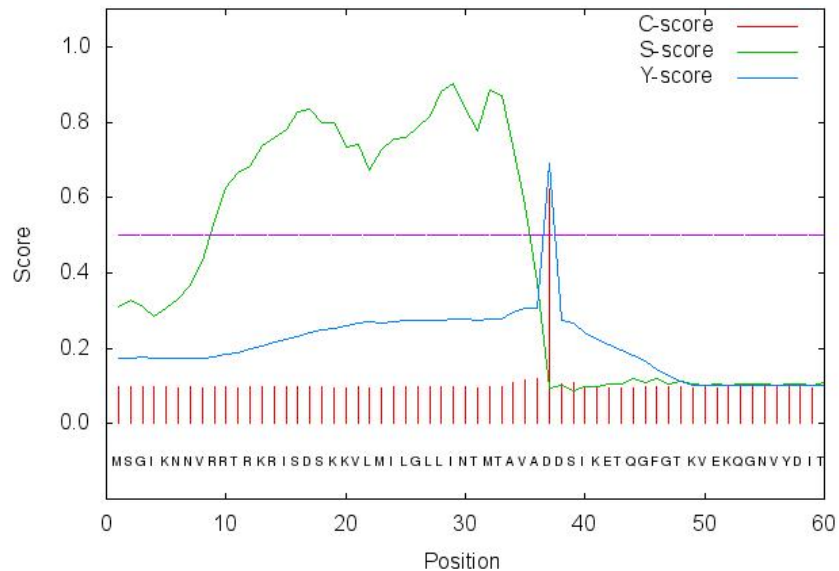
Figure 22: SignalP prediction of a signal peptide and a cleavage site at the start of the lktA protein sequence. C scores are high at the position immediately after the cleavage site, S scores distinguish between regions within the signal peptide and regions within the mature protein, and Y scores are a combination of the C score average and S score slopes. A D score is then used to make a weighted average of the mean S scores and the maximum Y scores which then determines whether a signal peptide is present.

When entered into a protein BLAST search, no conserved domain hits were found for the three lktA proteins. All three sequences did, however, match *Fusobacterium necrophorum* leukotoxin with E values of 0. As previously mentioned in section 5.1, recent BLASTp searches of these sequences have revealed homology of approximately 25 – 30 % to adhesin and haemagglutinin proteins. The lktA protein sequences were entered into a Pfam search to look for conserved domains. When the whole protein was entered no matches were found. However, when the protein was broken into smaller sequences and entered as individual searches, both of the *Fnf* strains had a haemagglutination activity domain with the first 180 bases, of 3,229 – 3,243 bases, with E values of 4e-06 for ARU 01 and 3.2e-06 for JCM 3724. This was not found for JCM 3718.

5.2.3.3 lktC

Figure 23 shows a Clustal Omega alignment of the three lktC protein sequences, which appear highly conserved. The JCM 3718 sequence contained a stop codon at the end, but for the two *Fnf* sequences, the remaining part of the sequence was unavailable.

When entered into a BLASTp search, the top hits for the three sequences resulted in a match of *Fusobacterium necrophorum* histidine kinase, E value 2e-99, for JCM 3718, *Fusobacterium necrophorum* histidine kinase, E value 3e-94 for JCM 3724 and sensory transduction regulator, E value 4e-96 for ARU 01.

A Pfam search of the three protein sequences resulted in a putative bacterial sensory transduction regulator domain, with E values of 7.6e-07, 1.9e-07 and 2.7e-07 for JCM 3718, JCM 3724 and ARU 01, respectively.

```
3718_lktC     MNLRESKFSEFLKNSNITCFEREEVKDELETVVYRSFMEVEGQNLPMVIVMDNSIYTNIR          60
3724_lktC     MNLRESKFSEFLKNSNITCFEREEVKDELETVVYRSFMEVEGQNLPMVIVMDNSIYTNIR          60
ARU01_lktC    MNLRESKFNDFLKNSNITCFEREEVKDELETVVYRSFMEVEGQNLPMVIVLDNSIYTNIR          60
              ********.:*****************************************:*********

3718_lktC     VQIAPKVIKDTNKEAVLSYINELNREYKVFKYYVTEDADVCLDSCVTSIAEEFNPEMVYT         120
3724_lktC     VQIAPKVIKDSNREAVLSYINELNREYKVFKYYVTEDADICLDSCITSIAEEFNPEMVYT         120
ARU01_lktC    VQIASKVIKDSNKEAVLSYINELNREYKVFKYYVTEDADVCLDSCVTSIAEEFNPEMVYT         120
              **** *****:*:**********************:*****:**************

3718_lktC     ILNVILEHITKHYSTFMKKIWSEEK 145
3724_lktC     ILNVILEHITEHYSTFMKKI----- 140
ARU01_lktC    ILNVILEHITEHYSTFMKKIW---- 141
              **********:*********
```

Figure 23: Clustal Omega alignment of the three lktC *protein sequences from JCM 3718, JCM 3724 and ARU 01.*

5.2.4 Testing the clinical strain collection with a range of leukotoxin-specific primers

A selection of published and custom designed leukotoxin primers were tested on the collection of 25 clinical strains, alongside strains JCM 3718, JCM 3724 and ARU 01, in order to test the prevalence of the gene. Ludlam_LT1 and Ludlam_LT2 primer pairs were designed by Ludlam *et al.* (2009a), primer pair lktA was designed by Tadepalli *et al.* (2008b) and lkt1 and lkt2 primer pairs were designed as part of this project. See Appendix 14 for primer sequences. All primers targeted the *lktA* gene and all results were recorded in duplicate over separate days.

The locations targeted by the primers are displayed in Table 37. Locations relate to JCM 3724 *lktA* gene. A summary of results is displayed in Table 38, with full results found in Appendix 15.

Table 37: Locations targeted by lktA primers. Base positions relate to the JCM 3724 lktA gene.

Primer set	5' start	3' end	Product size (bp)
Ludlam_LT1	2297	2668	372
lktA	2394	2589	196
Ludlam_LT2	6343	6540	198
lkt2	8181	8575	395
lkt1	8688	8854	167

Table 38: Summary of leukotoxin PCR results.

	Primer set				
	lktA	Ludlam_LT1	Ludlam_LT2	lkt1	lkt2
Number of isolates with a positive PCR result	12/28	28/28	25/28	28/28	28/28
Percentage of isolates with a positive PCR result	43	100	89	100	100

As shown in Table 37, the full length of the lktA product is within the forward and reverse primer of Ludlam_LT1 (see Appendix 14 for primer sequences). There was amplification of the Ludlam_LT1 product in 100 % of strains tested, compared to 43 % using primer pair lktA.

An example gel image of each of the five PCR primer sets is shown in Figure 24, showing the first six strains of the collection. Interestingly, the gel image of Ludlam_LT2 shows two different sized bands present. Two of the bands are of the expected size of 200 bp, and three bands in separate lanes are approximately 30 bp larger. Due to a positive result being recorded with the Ludlam_LT1 primers, this was not investigated further.

A Ludlam_LT1 (372 bp)
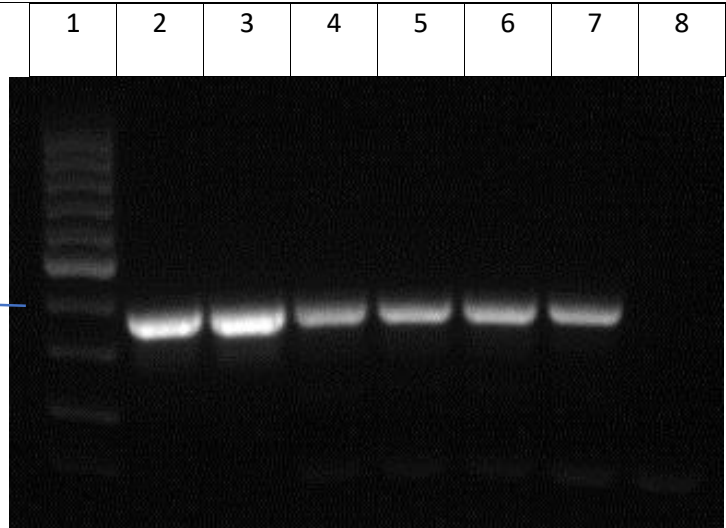
B lktA (196 bp)
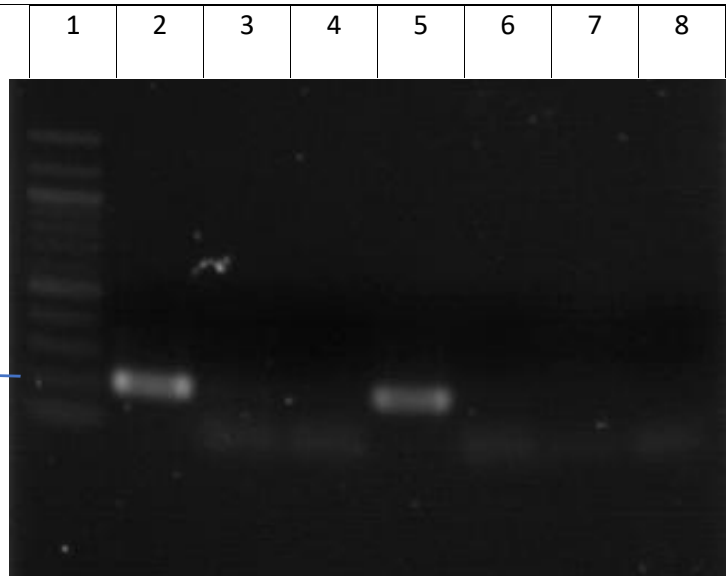
C Ludlam_LT2 (198 bp)
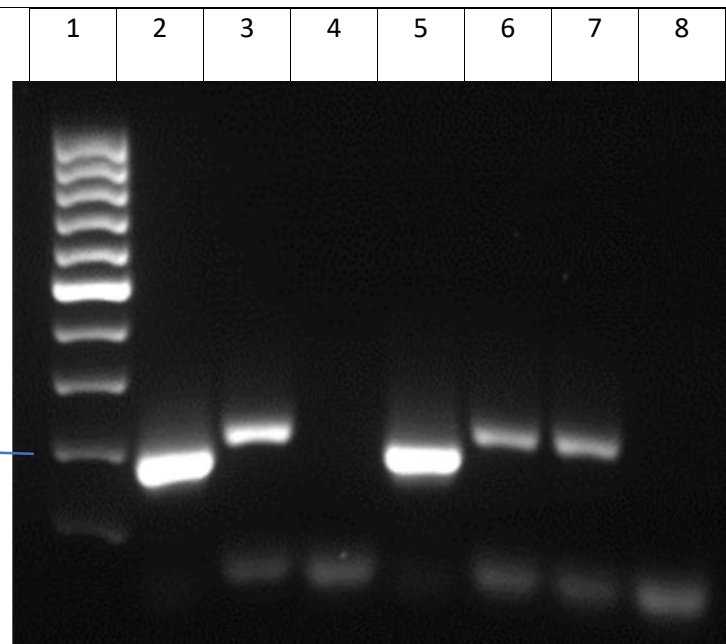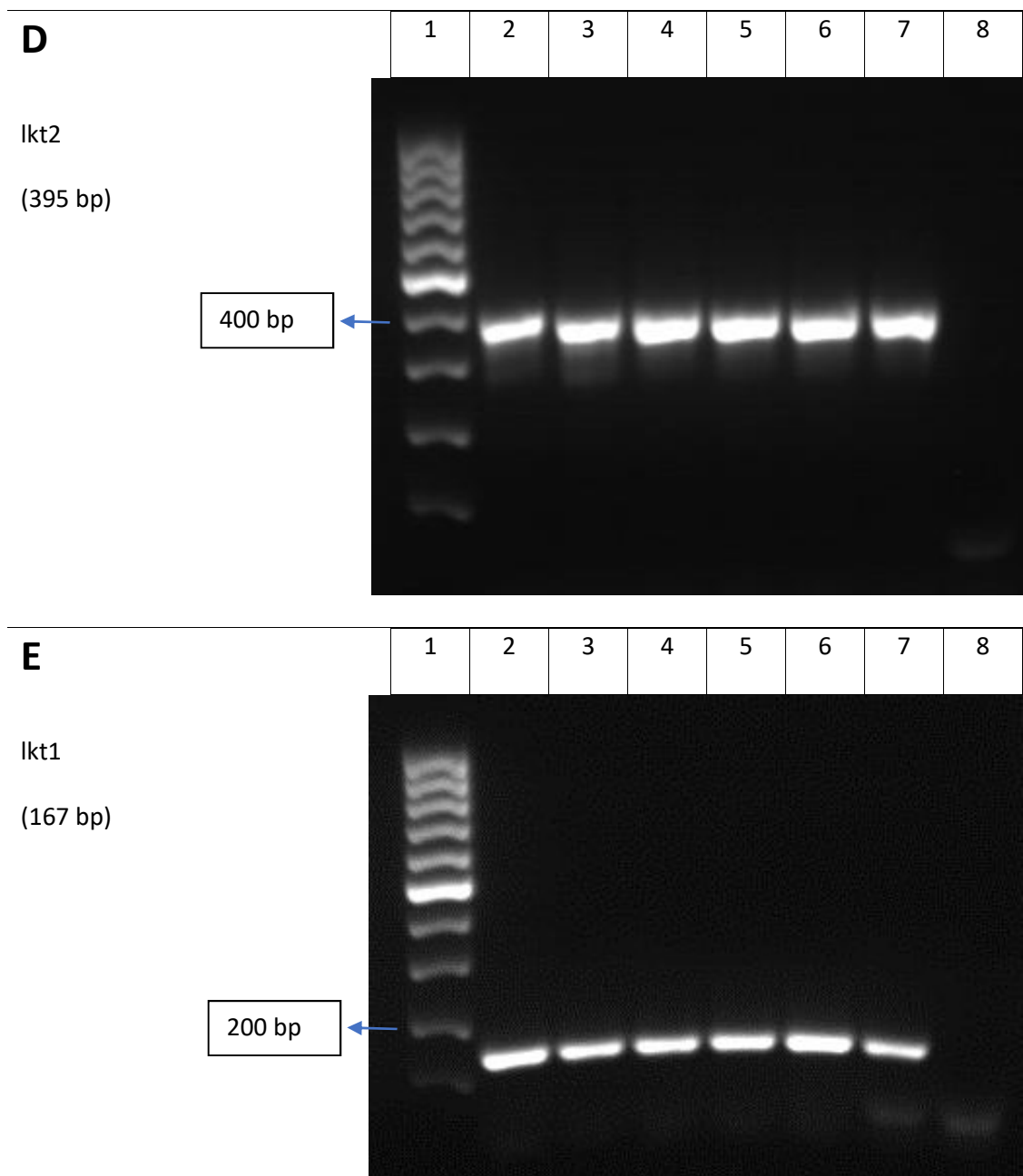
Figure 24: PCR products from the primer sets in Table 37 and Appendix 14. For each gel the lanes are as follows: 1) 100 bp marker; 2) F. necrophorum 1, 3) F. necrophorum 5; 4) F. necrophorum 11; 5) F. necrophorum 21; 6) F. necrophorum 24; 7) F. necrophorum 30; 8) Negative control (molecular grade water).

The first 12 strains tested with the Ludlam_LT1 primer set were sequenced by Sanger method in order to determine why there was no amplification for eight out of the first 12 strains tested with the lktA primers. The sequences were aligned using Clustal Omega (Appendix 16) and a phylogenetic tree was made (Figure 25). The lktA primer positions were highlighted on the Clustal alignment. All strains had a perfect match to the reverse primer, however only four strains matched the forward primer. The remaining strains only matched 14 out of 20 bases of the primer. The

phylogenetic tree shows two very distinct clusters that group into those that had amplification with lktA primers and those that did not.
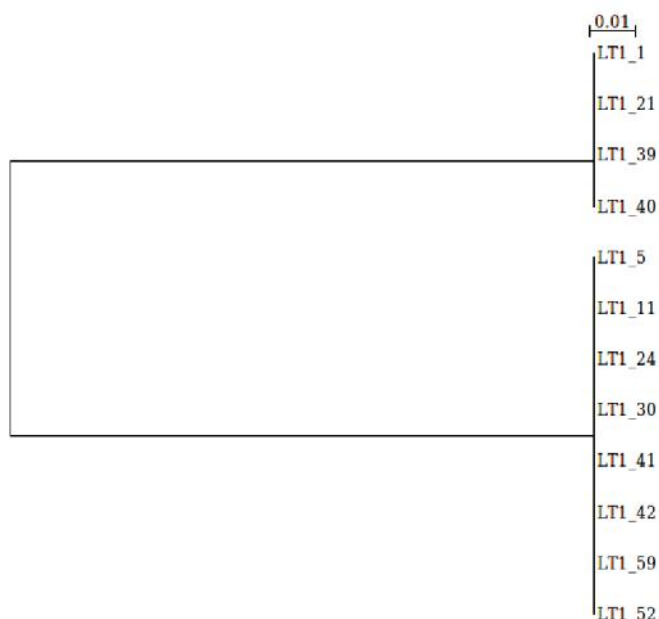
Figure 25: Phylogenetic tree of Ludlam_LT1 Sanger sequences. Produced using PhyML. The scale represents the average number of base changes per position for the length shown. Longer branch lengths therefore represent a greater number of base changes.

5.2.5 Collection of filtered culture supernatant of *F. necrophorum* strains

Aliquots of culture supernatant, hypothesised to contain a secreted leukotoxin, were collected from strains JCM 3718, JCM 3724 and ARU 01 to be used to check for cytotoxic activity against human white blood cells. Broth cultures were centrifuged and the supernatant was sterile filtered and then concentrated approximately 75X using 100 kDa molecular weight cut off filters (Millipore, Hertfordshire) (methods section 2.13.1). The aliquots of concentrated high molecular weight protein were frozen at -20 °C for up to a week and were defrosted one time only.

Bradford assays were used to determine the protein concentration of the high molecular weight fractions (methods section 2.13.2). A calibration curve was defined and samples were diluted 1 in 50 to measure within the range, using a Nanodrop as an initial indicator. Concentrations were then adjusted to account for the dilution factor and results were recorded in Table 39.

Table 39: Bradford assay results of high molecular weight protein samples.

Sample	Concentration (µg/ml)
JCM 3718	1267
JCM 3724	738
ARU 01	1161

The equation of the calibration curve was y=0.043+0.0183x with an R squared value of 0.992.

Samples were analysed using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE) under both reducing and non-reducing conditions (Figure 26 and Figure 27) (methods section 2.13.3). Samples were diluted 1:1 with loading buffer from neat concentrations.

Figure 26 shows the banding patterns of the high molecular weight fraction when boiled in reducing Laemmli buffer. The samples in Figure 27 were boiled in non-reducing buffer and do not show the same pattern. Some protein can be seen in the area of the well and has not migrated through the gel, possibly due to its large size. This protein is likely to be high molecular weight (>245 kDa) and is therefore expected to be the leukotoxin. There are very few additional bands so this indicates that the concentrated culture supernatant is not heavily contaminated with other proteins. Without using a monoclonal antibody for affinity purification, a low level of contamination is possible. The very high molecular weight protein in Figure 27 is missing from Figure 26 which increases the confidence that the bands present in Figure 26 are breakdown products of the leukotoxin. This is in agreement with the understanding that the leukotoxin protein is unstable (Tan *et al.*, 1994a; Tadepalli *et al.*, 2008a). No bands were present in the flow through lanes. This may be due to proteins running off the end of the gel, or because this fraction was not concentrated, unlike the leukotoxin-containing fraction, and therefore would have been a very dilute solution.
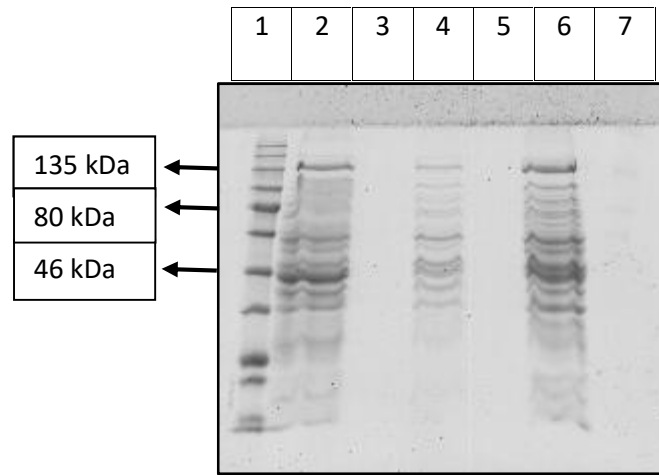
Figure 26: Leukotoxin samples boiled for 5 minutes in reducing buffer. Lane 1: NEB 245 kDa marker; 2: JCM 3718 putative toxin; 3: JCM 3718 filter flow through; 4: JCM 3724 putative toxin; 5: JCM 3724 filter flow through; 6: ARU 01 putative toxin; 7: ARU 01 filter flow through.
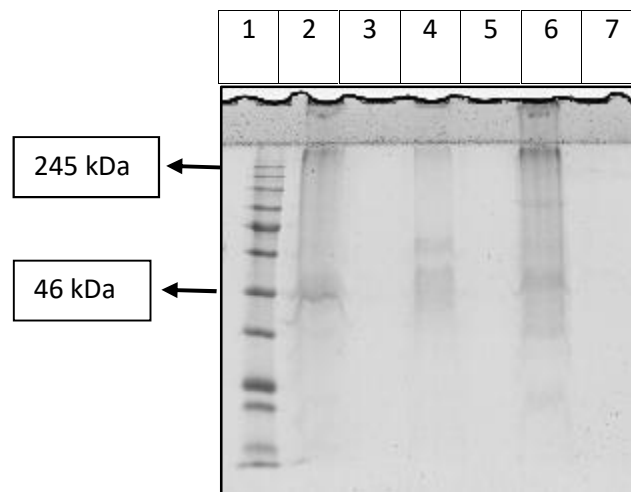
Figure 27: Leukotoxin samples boiled for 5 minutes in non-reducing buffer. Lane 1: NEB 245 kDa marker; 2: JCM 3718 putative toxin; 3: JCM 3718 filter flow through; 4: JCM 3724 putative toxin; 5: JCM 3724 filter flow through; 6: ARU 01 putative toxin; 7: ARU 01 filter flow through.

5.2.6 Assay to show cytotoxicity of filtered culture supernatant of *F. necrophorum* strains against HL-60 cells

Human promyelocytic leukaemia (HL-60, ATCC CCL-240) cells were used to represent human white blood cells for a cytotoxicity assay. HL-60 cells were treated with putative leukotoxin from strains JCM 3718, JCM 3724 and ARU 01 (methods section 2.13.4). To test for dose-dependent cell death,

concentrations were set at 125 µg/ml, 150 µg/ml and 175 µg/ml for JCM 3718, JCM 3724 and ARU 01, and 250 µg/ml as an additional concentration for JCM 3718 and ARU 01. JCM 3724 was excluded from the additional concentration due to limitations with concentrating the culture supernatant. 1X PBS was used as the negative control. After preliminary tests, 7.5 % ethanol was chosen as the positive control for cell death. Following treatment with the culture supernatant, cells were stained with propidium iodide (PI) and analysed on a flow cytometer (methods section 2.13.6) to assess the percentage viabilities of the cells. Uptake of PI is indicative of cell death. Results were recorded in triplicate over separate days, with 10,000 cells analysed per sample.

The concentrations for the cytotoxicity assay were chosen following a pilot study, where concentrations below 75 µg/ml appeared to have little effect. 125 µg/ml was chosen as the lowest concentration that had a notable effect on the cell viability. The top concentration was chosen to be 175 µg/ml for JCM 3724 due to limitations with the initial concentration of toxin obtained, although it was possible to use a higher concentration of 250 µg/ml for JCM 3718 and ARU 01. It was noted that once the viability of the cells fell below approximately 50 %, the structure of the cells were affected and the flow cytometer plots appeared to show severely disrupted cells. This resulted in dead cells not falling within the gated region and therefore being excluded from the results, which would skew the data. Care was taken to choose concentrations that were not high enough to cause this to happen.

Figure 28 demonstrates the creation of the HL-60 protocol within the Summit 4.3 program. Part A includes a gate created to exclude any cells that were not typical of the majority of cells. Each cell is represented as a red dot and is plotted according to forward scatter versus side scatter. Forward scatter is representative of cell size and side scatter relates to the granularity of the cell. This is independent of any staining. Part B excludes any cells external to the R1 gate in Part A. It plots the granularity of the cell against PI fluorescence, i.e. viability. A trypan blue exclusion assay predetermined the percentage viability of the healthy cells to typically be between 98 and 99% of cells present, and the threshold was therefore set on the plot to match this. The plot shows 98.13% cell viability (or 1.87% cell death) for this negative control sample. For the remainder of the samples

the gate and threshold remain the same, and cell death was determined as the number of cells recorded above the threshold, in the R4 region. Part C is a histogram of the PI data from plot B. The gated region is set at the same PE log value, measuring PI fluorescence, as in part B and therefore records cell death in the same way. Histograms can be overlaid to see the relationship between different sample treatments (see Figure 29).
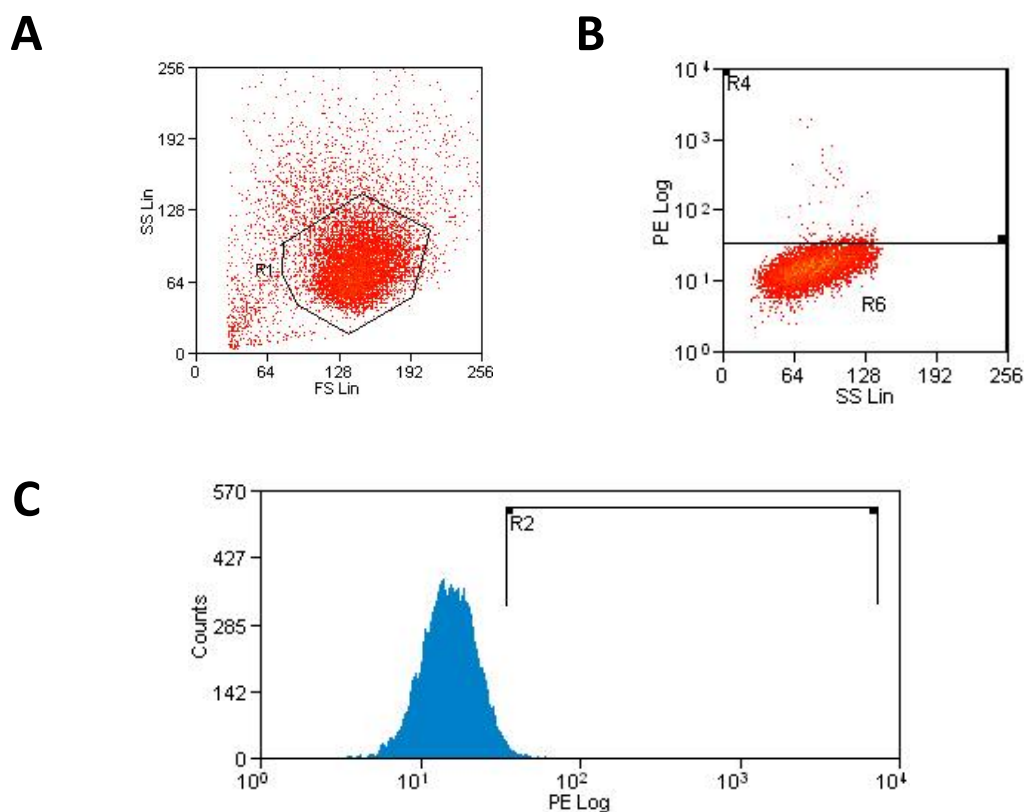
Figure 28: Gating of negative control HL-60 cells stained with propidium iodide (PI). (A) Plot of forward scatter versus side scatter. R1 gate denotes region that majority of cells presumed to be healthy fall within and contains 10,000 cell counts. (B) Plot of side scatter versus the log of PI fluorescence, containing 10,000 cell counts. Only cells from within R1 on plot A are analysed further on plot B. R6 contains cells presumed to be viable and those that fall into R4 are classed as not viable due to increased PI uptake. (C) PI fluorescence data from plot B displayed as a histogram.

Data represented as an overlay graph (Figure 29) is useful for analysing the relationship between putative leukotoxin concentration and cell viability, however it lacks clarity when displaying data in triplicate as the graph becomes crowded. Also the data cannot be normalised to the negative control. Therefore the data was instead plotted as bar graphs, with each set of results normalised

to the negative control. The mean of the three sets is displayed, with standard error bars included (Figure 30, Figure 31 and Figure 32).
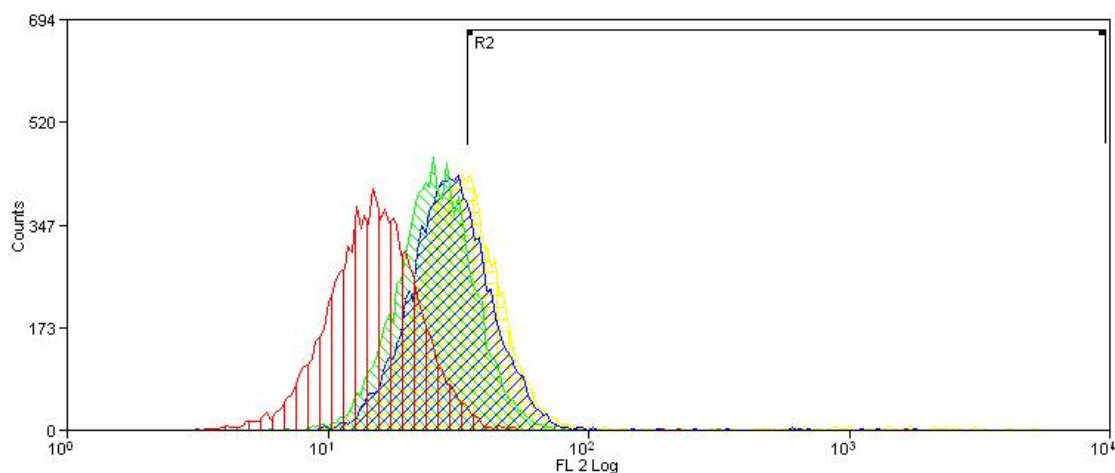
Figure 29: Overlay of a set of data from strain JCM 3724 showing dose-dependent cell death. PI fluorescence data is plotted against cell counts, as in Figure 28C. Red: Negative control; Green: 125 µg/ml putative leukotoxin; Blue: 150 µg/ml putative leukotoxin; Yellow: 175 µg/ml putative leukotoxin. Each is comprised of 10,000 cells. Cells within region R2 are considered not viable. See Table 40 for percentages.

Table 40: Percentage viability and percentage death data from samples shown in the overlay graph (Figure 29).

Symbol	Treatment group	Percentage viability	Percentage death
	Negative control	98.13	1.87
	125 µg/ml	78.11	21.89
	150 µg/ml	64.88	35.12
	175 µg/ml	54.70	45.30

Figure 30, Figure 31 and Figure 32 show a clear dose-dependent cytotoxic response of the high molecular weight fraction against the HL-60 cells. The higher the concentration of the putative toxin, the lower the percentage viability of the cells. See Appendix 17 for full results. The negative control was 1X PBS and the positive control was a final concentration of 7.5 % ethanol.
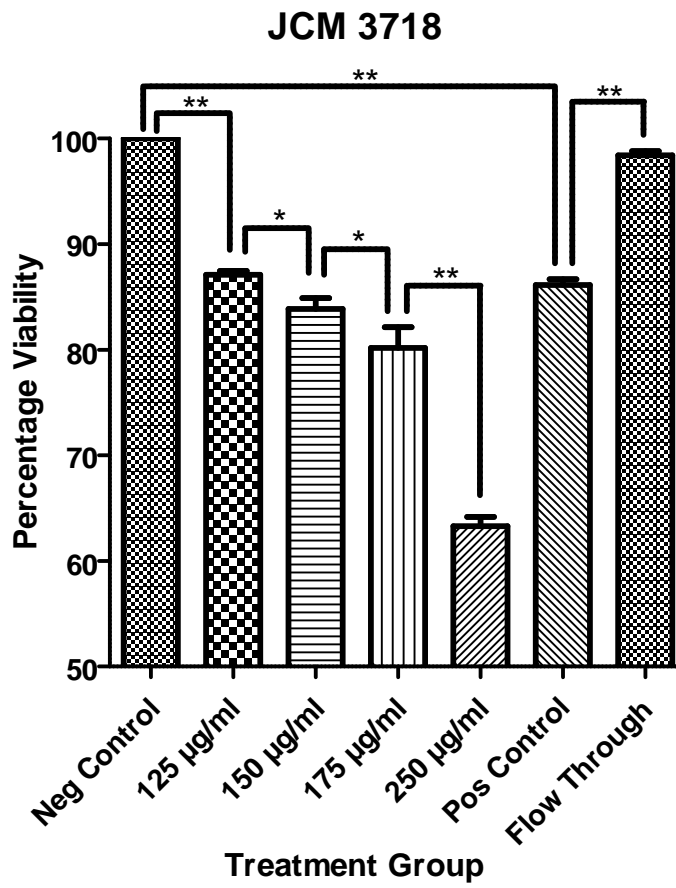
*Figure 30: Percentage viabilities of HL-60 cells following treatment with putative toxin from JCM 3718 at a range of concentrations. All values are the mean of triplicate readings and are normalised to the negative control. * = P <0.05 using Fisher's LSD test; ** = P<0.01.*
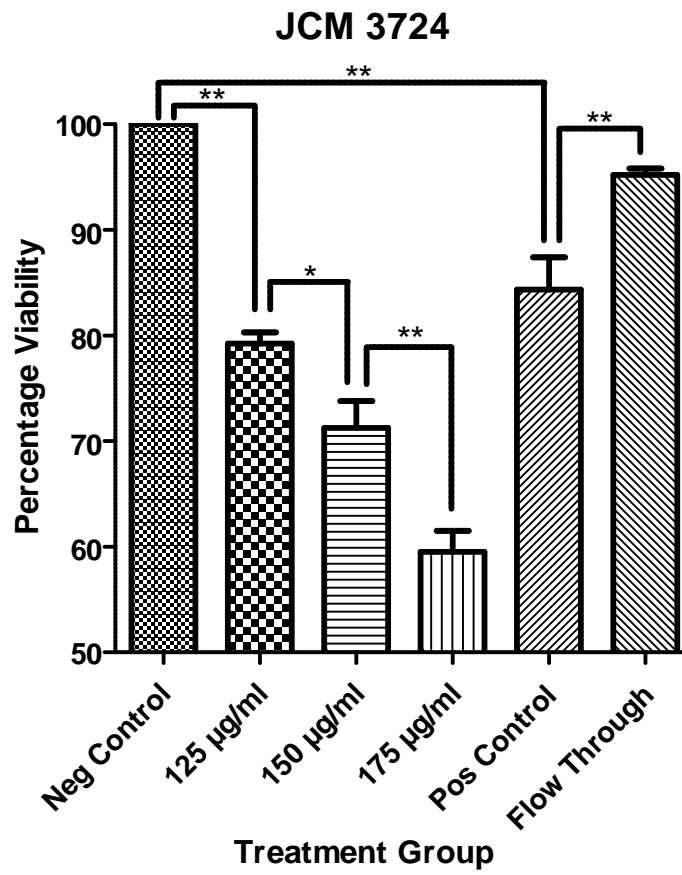
*Figure 31: Percentage viabilities of HL-60 cells following treatment with putative toxin from JCM 3724 at a range of concentrations. All values are the mean of triplicate readings and are normalised to the negative control. * = P <0.05 using Fisher's LSD test; ** = P<0.01.*
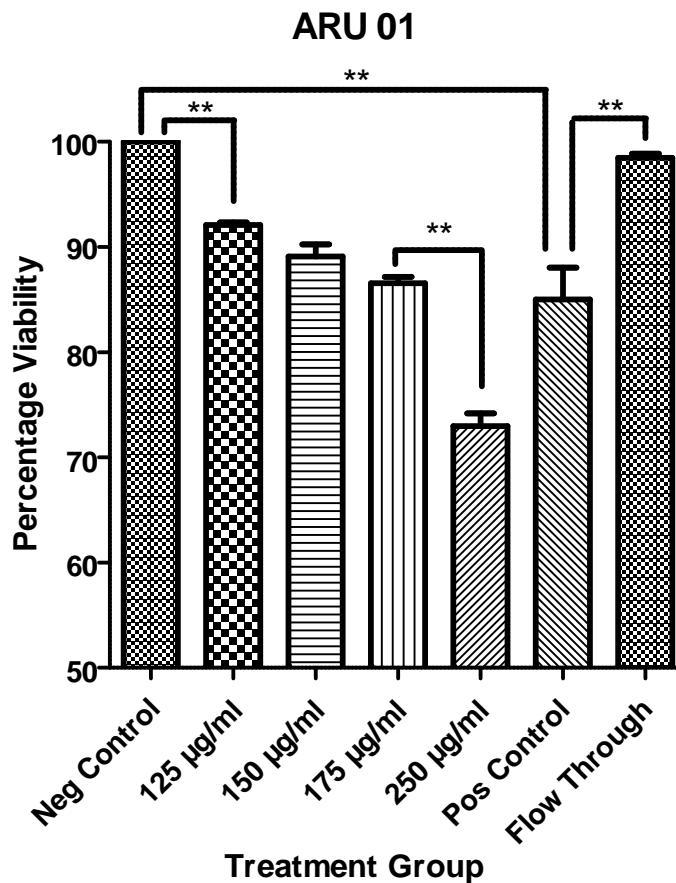
*Figure 32: Percentage viabilities of HL-60 cells following treatment with putative toxin from ARU 01 at a range of concentrations. All values are the mean of triplicate readings and are normalised to the negative control. * = P <0.05 using Fisher's LSD test; ** = P<0.01.*

5.2.7 Statistical analysis

A one way ANOVA carried out on the cell viability data from Figure 30, Figure 31 and Figure 32 resulted in a P value of <0.01 for each set of data, concluding that there is a significance between the treatment groups.

Fisher's Least Significant Difference (LSD) test was performed on the data to test for statistical significance between the treatment groups. The statistical analysis labels in Figure 30, Figure 31 and Figure 32 show a P value of <0.01 between the negative control and the lowest protein concentration of 125 µg/ml for each of the three sets of data. The level of statistical significance between the remainder of the concentrations varies between strains, with JCM 3724 having the strongest significance and ARU 01 having the weakest significance in the range of 125 µg/ml and

175 µg/ml. In each of the three sets of data the P value was <0.01 for the comparison of the negative control to the positive control, and for the positive control to the filter flow through. See Appendix 18 for full statistical analysis.

5.2.8 Cytotoxicity assay with human white blood cells

In order to confirm the suitability of HL-60 cells as a substitute for human white blood cells, the experiment was repeated using human white blood cells and a single concentration of 150 µg/ml leukotoxin, to match the middle of the three initial concentrations tested. The results were recorded in triplicate on three separate days, with 10,000 cells analysed per sample. Figure 33 shows the percentage viability of each treatment group, with HL-60 data in parallel with the donor blood data. See Appendix 19 for full results. Fisher's LSD test was performed on the data as before to test for statistical significance between each set of HL-60 data and that from the corresponding WBC data. In each case, despite the appearance of the HL-60 cells having a slightly lower percentage viability, there was no statistical significance (P>0.05). See Appendix 20 for full statistical analysis.
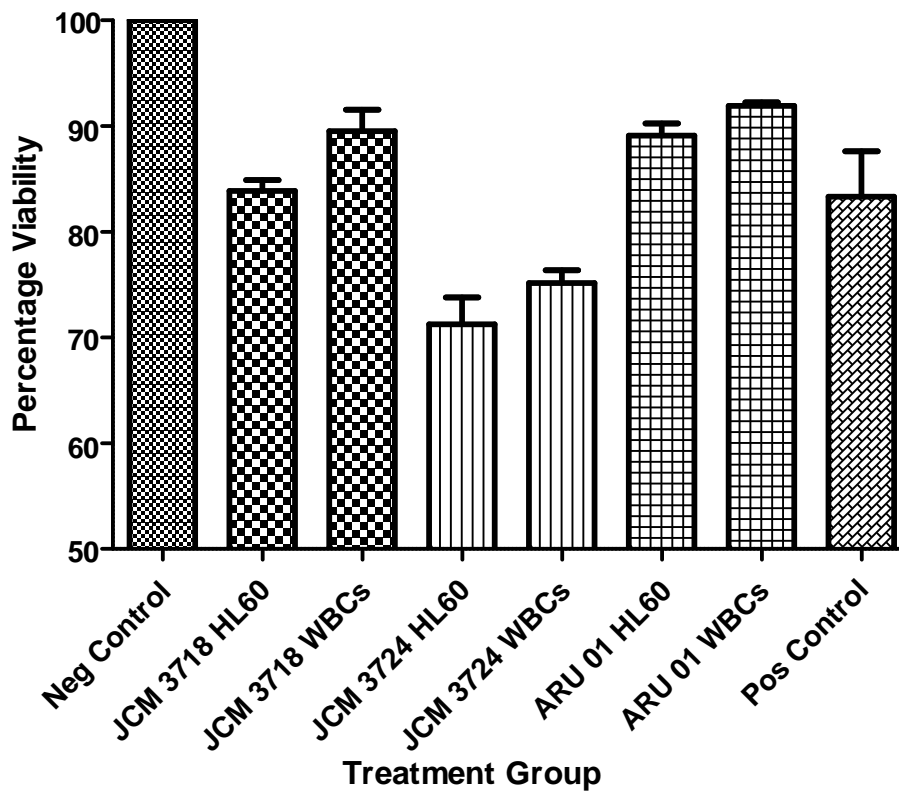
Figure 33: HL-60 results compared to those using human donor blood at a single concentration of 150 μg/ml. All values are the mean of triplicate readings and are normalised to the negative control.

5.3 Discussion

One copy of the leukotoxin operon was found in each of the three sequenced strains, which was as hypothesised, based on the prevalence of the gene according to the *F. necrophorum* literature and based on reports that it was the primary virulence factor of the bacterium (Tan *et al.*, 1994a). The size of the leukotoxin operon is in excess of 11,000 bp, and the average genome assembly contig size for JCM 3718, JCM 3724 and ARU 01 were all <5,000bp (chapter 4.2.1; Table 26), therefore it was likely that there would be gaps within the sequences. The majority of these could be bridged using PCR and Sanger sequencing, however, where the C terminus of the *lktC* gene was missing it was more difficult to design matching primers due to the neighbouring sequence not having been characterised previously in the literature at the time. More recently, *F. necrophorum* genomes have been deposited in online databases such as NCBI and BioCyc that now mean that this region of sequence could be located in other genomes and primers designed based on those sequences.

When the leukotoxin genes were translated into protein sequences it was noted that they began similarly and then lacked similarity later in the sequence. It was suspected that this was due to errors in the sequencing leading to frameshifts, most likely in homopolymer regions due to the use of pyrosequencing. It was therefore decided to use Sanger sequencing to confirm the sequence and remove the errors. Primers designed based on the JCM 3724 sequence were first tested against the three strains. Only four of the 17 primer sets resulted in amplification of product for ARU 01 and nine of the primer sets for JCM 3718, which indicates the level of variation between the three sequences. Once the sequences of the amplified products were assembled and aligned to the original sequence the presence of errors was confirmed and the new Sanger sequence data was used in its place. When these sequences were translated into protein sequence and aligned, some similarity was evident throughout the sequence, suggesting the frameshift errors had been removed. This illustrates that the Sanger sequence data was more reliable.

The Clustal Omega alignment of the lktB protein sequence (Figure 18) shows that it is highly conserved for the two *Fnf* strains, but has a different start codon location for the *Fnn* strain. Where

JCM 3724 and ARU 01 have the start codon (ATG), JCM 3718 has ATC instead, which codes for isoleucine rather than methionine. The result of this is that the JCM 3718 lktB protein sequence starts 112 amino acids later at the next methionine. The next generation sequence data and the Sanger sequence data of both the forward and reverse strand of the associated PCR product were all congruent for this section of the sequence, so it is highly unlikely to be a sequencing error. All three lktB sequences still matched *F. necrophorum* lktB in a BLASTp search. It had been hypothesised that sequences from the same subspecies would demonstrate greater similarity to each other than those of the other subspecies. This was therefore shown to be the case for lktB sequences in these strains with the two *Fnf* protein sequences almost identical and the *Fnn* sequence only three quarters of the length of the *Fnf* strains.

The Pfam search for conserved domains (Table 34) within the three lktB sequences used sequences both whole and also in smaller segments. The POTRA domain was found at the N terminus of the lktB sequence for JCM 3724 and ARU 01, and was missing from JCM 3718 due to the start codon appearing later in the sequence for that strain. It is this domain that has led to the speculation that the lktB protein is responsible for assisting the active toxin to be secreted (Tadepalli *et al.*, 2008b). Other related domains found towards the C terminus of the lktB sequence by BLASTp were those of a haemolysin activation/secretion protein and outer membrane protein assembly complex. A possible implication of the POTRA domain being absent in JCM 3718 could be the inability for lktA to be secreted, however the results of the cytotoxicity assay suggest this is not the case. Other possibilities include multiple secretion domains being present so that back up secretion systems are in place, or that the POTRA domain is redundant or only partial. For the *Fnf* strains, the first 20 of 76 amino acids are missing from the POTRA hidden Markov model alignment so it may be that the domain is not functional. The signal peptide found in the *Fnf* strains (Figure 22) and *Fnn* strain support the knowledge that this is a secreted protein (Tadepalli *et al.*, 2008b), although the mechanism of secretion remains unknown.

The lktA protein sequences aligned using Clustal Omega (Appendix 13) had a higher percentage similarity for the two bovine strains, JCM 3718 (*Fnn*) and JCM 3724 (*Fnf*), than for the two *Fnf*

strains, JCM 3724 and ARU 01. The ACT alignment (Figure 19) supports this and shows a greater level of similarity between JCM 3718 and JCM 3724 than between JCM 3724 and ARU 01. The first 500 bp appear to be more similar between the *Fnf* strains, with a similarity of 98 %. For the remainder of the alignment the JCM strains appear more similar, with similarities of 94 % between 1,000 bp and 4,000 bp, and 97 % between 6,100 bp and 9,700 bp. In comparison, the majority of regions of similarity between the two *Fnf* strains are between 80 % and 90 % similar. It had been hypothesised that the two *Fnf* strains would exhibit more similarity with each other than with the *Fnn* strain, but this has been shown not to be the case for these three *lktA* sequences. The reason for the unexpected result could be related to the source of the isolates. Both JCM 3718 and JCM 3724 were isolated from bovine liver abscesses, whereas ARU 01 was isolated from a blood sample of a human Lemierre's syndrome patient. The date of isolation may also be important. The JCM reference strains were isolated prior to 1985 (Shinjo *et al.*, 1991), whereas ARU 01 was isolated in 2002. The differences seen between the *Fnf* strains may be an indication of the evolution of the sequence over the years. As discussed below, when *lktA* genes from a number of additional genomes were investigated using phylogenetic analysis, *Fnf* strains were in distinct clusters from *Fnn* strains, which appears to support the hypothesis that strains of the same subspecies would demonstrate greater levels of similarity.

The implication from the work by Narayanan *et al.* (2001b) is that the active toxin region is within the first 3.5 kb of the *lktA* gene, as discussed in section 5.1. The ACT alignment displays the region as not highly conserved between the two *Fnf* strains, however it is well conserved between the two bovine strains, with similarities of 94 %.

The phylogenetic tree in Figure 20 was created using the data generated from the current project and that now available in databases in order to investigate the relationship between the isolate source, DNA sequence and subspecies, where known. The top cluster contains the two human *Fnf* isolates: ARU 01 and D12. HUN048, the third isolate, is also *Fnf* and presumably a commensal, as this was isolated from a rumen microbiome with no associated disease. These three isolates are tightly clustered as a result of having highly similar *lktA* sequences. The second cluster from the top

contains two *Fnf* strains, JCM 3724 and B35, both isolated from bovine liver abscesses. These first five isolates are all subspecies *funduliforme*, but the bovine liver abscess *Fnf* strains are clearly distinct from the remaining *Fnf* strains. The remaining clusters of isolates contain subspecies *necrophorum* strains and strain H05 that has an unknown subspecies identity due to a lack of sequence data for the *gyraseB* gene. Isolates DAB and DJ-2 are very tightly clustered and are both from deer jaw abscesses. The remaining isolates are bovine liver abscess or bovine footrot. There appears to be no separation of liver abscess or footrot source within the clusters, with a footrot isolate being clustered with a liver abscess isolate two separate times. The animal species from which the isolate is sourced appears to have a strong effect on the sequence, with bovine and deer strains distinctly separated from human strains, with the possible exception of the ruminant strain. It appears that both the subspecies and source of the isolate have a strong effect on sequence clustering.

According to Narayanan *et al.* (2001b), there are 14 regions within the lktA protein that have the potential to be membrane spanning. The Kyte and Doolittle hydrophobicity plot in Figure 21 disputes this and found only two potential membrane spanning regions for each of the three strains. Figure 21 was produced using the ExPASy ProtScale tool (Expasy - Protscale, 2015), whereas Narayanan *et al.* (2001b) used TMpred (Tmpred, 2015). When the JCM 3718, JCM 3724 and ARU 01 lktA protein sequences were entered into the TMpred online tool, it found 14, 19 and 15 potential transmembrane helices for the three strains, respectively. The peaks are very thin and are not sustained over many base pairs, and therefore these predictions seem less likely to be accurate.

The lktC protein is highly conserved between the strains, which suggests that it may play an important role. As all three sequences were so highly conserved, these results do not support the hypothesis from section 5.1, which predicted greater levels of similarity between strains of the same subspecies. Hypotheses regarding the function appear to be lacking from the literature, however, the lktC protein from strain *F. necrophorum* D12 is annotated on the NCBI website as a sensory transduction regulator (accession number: NZ_GL988012.1). This same protein from the same strain is annotated on the BioCyc genome database (Biocyc, 2015) as leukotoxin-activating

lysine-acyltransferase. The BLASTp and Pfam searches for protein product matches and conserved domains from section 5.2.3.3 resulted in hits to a histidine kinase and sensory transduction regulator. Histidine kinases are enzymes involved in signal-transduction and are very widespread in bacteria (Wolanin *et al.*, 2002). This could suggest that lktC is somehow involved in regulating leukotoxin production. Alternatively, the annotation of leukotoxin-activating lysine-acyltransferase may be more accurate. There are examples in the literature of lysine-acyltransferases activating other toxins, such as a haemolysin in *Escherichia coli* (Langston *et al.*, 2004) and a haemolysin in *Bordetella pertussis* (Basar *et al.*, 2001).

A range of PCR primers were needed to test for the prevalence of the leukotoxin within the strain collection; this has been disputed in the literature (Ludlam *et al.*, 2009a; Bennett *et al.*, 2010). It was determined that by using a range of published PCR primers and designing custom sets that an analysis of the PCR primers in the literature could be performed. The three sets of primers designed and tested by Ludlam *et al.* (2009a) were initially tested, all renamed to add a prefix of Ludlam, although primer set Ludlam_LT3 was later discarded due to non-specific binding. The DNA sequences used to design the PCR primers from this publication were strains *F. necrophorum* A25 (Oelke *et al.*, 2005) and H05 (Sun *et al.*, 2009), from bovine liver abscess and bovine footrot, respectively. All of the strains tested with primer set Ludlam_LT1 in this project produced amplicons, so it is difficult to speculate on the reasons for an overall amplification rate of 47 % in the strains tested by Ludlam *et al.* (2009a) after the use of three sets of PCR primers. When their results were separated by subspecies, only 33 % of *Fnf* strains had amplification with at least one of the primer sets. Of the strains tested in this project, all were subspecies *funduliforme*, with the exception of JCM 3718, and all had amplified product with Ludlam_LT1. There was a high rate of amplification with primer set Ludlam_LT2, at 89 %, although there were two different band sizes present. The expected band size was 198 bp. For the gel image in Figure 24C, two bands can be seen at the expected size and three bands can be seen that are slightly larger than expected at just above 200 bp. All bands were recorded as a positive result and this was not investigated further due to the positive result with Ludlam_LT1 for the strains. The lktA primers were found to be quite

ineffective for the strain collection used in this project, with an amplification rate of 43 %. The Sanger sequences of the first 12 Ludlam_LT1 products showed two very distinct groups. On closer analysis it was apparent that the sequences of strains 1, 21, 39 and 40 were identical to each other, as were the sequences of strains 5, 11, 24, 30, 41, 42, 52 and 59. There is a possibility that the same strain has infected multiple patients who contributed to the strain collection, or that these strains are simply closely related, possibly due to being isolated from the same geographic region. The amplicon sequences were only 291 bp long, which is not large enough to make any conclusions regarding strain similarity. The lkt1 and lkt2 PCR primer sets were designed based on a region that was believed to be highly conserved, based on the ACT alignment in Figure 19. They appeared to be very effective with 100 % amplification for each primer set for the strain collection tested. The hypothesis that the leukotoxin would be universal within the *F. necrophorum* strain collection is supported here, with three of the leukotoxin primer sets having an amplification rate of 100 %. For future testing, primer sets Ludlam_LT1, lkt1 and lkt2 would be considered. However, the Clustal Omega results in Figure 18 show that the lktB protein is more highly conserved. This may make it likely that the *lktB* gene sequence would be a better target for leukotoxin detection using PCR primers. The Clustal alignment in Appendix 12 shows the beginning of the *leukotoxin* operon, containing the *lktB* gene, as more conserved than the region containing the *lktA* gene, suggesting that this would be a good target for PCR primers.

F. necrophorum leukotoxin is known for being highly unstable (Tan *et al.*, 1994a; Tadepalli *et al.*, 2008a), therefore great care was taken during the steps to collect the leukotoxin fraction of the culture supernatants. The products analysed by SDS PAGE under denaturing conditions (Figure 26) showed a variety of bands that are likely to be breakdown products. Tadepalli *et al.* (2008a; 2008b) showed this with both subspecies of *F. necrophorum*. The three leukotoxin fractions, which were approximately 75X concentrated from the culture supernatant, were not diluted prior to mixing with the loading buffer and the intensity of the bands on the gel appear to correspond with the varying concentrations as measured by Bradford assay, with JCM 3724 having the faintest bands. The image in Figure 27 is the SDS PAGE under non-reducing conditions. It shows a band at

approximately 46 kDa, and one larger than 245 kDa for each of the three strains, as well as a few other very faint bands. There also appears to be some high molecular weight protein that failed to migrate into the gel and has remained at the gel interface. This has also been seen by Tan *et al.* (1994a; 1994b) in 7.5 % and 10 % SDS PAGE gels when imaging leukotoxin fractions. 10 % gels were used here.

For the cytotoxicity assay (section 5.2.6), HL-60 cells were used in place of human donor blood for the majority of the study in order to minimise the number of times blood samples would need to be taken. This cell line has been used in previous research to test the toxicity of *Actinobacillus actinomycetemcomitans* leukotoxin (Karakelian *et al.*, 1998; Lear *et al.*, 2000), which suggests that this cell line is an appropriate model to use in place of healthy human blood cells. Apoptotic effects on HL-60 cells have also been studied upon addition of cytotoxic cancer treatments (Willimott *et al.*, 2007). The inclusion of human donor blood later in the experiment helped to validate the HL-60 cell line results and demonstrate that they were an appropriate choice. HL-60 cells can be induced to differentiate into mature granulocyte types (Birnie, 1988). While this would potentially have given a more suitable model, this takes six days to occur, by which point the cells are less viable and give a more scattered plot on the flow cytometer. It was therefore decided to use the cells at two days after passage, as they were healthy and had not differentiated, giving a more uniform and healthy sample of cells.

It had been hypothesised that the concentrated, high molecular weight fraction of culture supernatant would demonstrate a cytotoxic effect in a dose-dependent manner. A dose-dependent cytotoxic response against HL-60 cells can be clearly seen in Figure 30, Figure 31 and Figure 32, which supports the hypothesis. For each of the three strains, they show high statistical significance between the negative control and the 125 µg/ml treatment group, and while not necessarily statistically significant between each step, as is the case with ARU 01, they do show further significance with higher treatment concentrations. The graphs also consistently show high statistical significance between the negative control and the positive control, as well as between the positive control and the filter flow through, which serves as a second negative control. There

was no statistical significance between the negative control and the filter flow through, demonstrating that this fraction was not toxic to the HL-60 cells.

Fisher's least significant difference test was chosen because the analysis was not an all against all comparison, so there was no correction for multiple comparisons as there would have been with a Bonferroni test. Specific comparisons were chosen in advance of the statistical test being carried out. While Fisher's LSD test is less vigorous, it does allow a better comparison between the three data sets as there is more variation with the statistical significance results. The test was chosen to for a better comparison between the strains rather than to test whether the cytotoxicity was significant or not. With the P values <0.01 for each of the 125 µg/ml treatment groups the cytotoxic effect was clear.

The expectation of the cytotoxicity results was that JCM 3724 would have the weakest cytotoxic effect, due to the *funduliforme* subspecies being known for being less virulent (Tadepalli *et al.*, 2008b); JCM 3718 and ARU 01 were expected to have a stronger cytotoxic effect, as JCM 3718 is a *Fnn* strain, and ARU 01 exhibited high virulence in a patient. The lower virulence of *Fnf*, however, is due to a lower expression of the toxin. For this assay the concentrations were standardised across the three strains and therefore the effect of the differing expression was negated. For the three sets of data it appears that a higher concentration of leukotoxin in the culture supernatant did not necessarily correlate with a highly virulent protein. The bovine strains were more toxic than the human strain, although the human strain had a higher concentration in the Bradford assay than JCM 3724. Strains that cause Lemierre's syndrome may therefore have more efficient leukotoxin promoters; investigating this as part of future work would be of interest. In order to confirm that the high molecular weight fractions were not contaminated with other non-toxic high molecular weight proteins that affected the protein concentrations measured by Bradford assay, anti-leukotoxin antibodies would be needed for affinity purification. As discussed previously in the aims of this chapter (section 5.1) these antibodies are not commercially available. It had been hypothesised that variations in activity could be explained by variations in the leukotoxin

sequences. From examination of the DNA and protein sequence data alongside the cytotoxicity results, it appears that there is no obvious link between these data sets.

The results of the cytotoxicity assay in section 5.2.8 using human donor blood confirm the results of the HL-60 experiments. The Fisher's LSD test was used again for consistency and specific comparisons were selected in advance of the test. In each case the corresponding pair of HL-60 and human donor blood data had no statistical significance between them, supporting the suggestion that HL-60 cells are a good model to use in place of human donor blood cells.

When the data from the cytotoxicity assay was compared to results from related publications there was a consensus that the leukotoxin is unstable (Tan *et al.*, 1994a), shows many breakdown products with SDS PAGE analysis under denaturing conditions (Tadepalli *et al.*, 2008a), and that some product fails to migrate into the gel (Tan *et al.*, 1994a; Tan *et al.*, 1994b). The *Fnn* strain was expected to produce a higher concentration of leukotoxin than the *Fnf* strains, due to the findings by Zhang *et al.* (2006) that showed the *Fnn* promoter activity was four-fold higher. The Bradford assay results in Table 39 show the *Fnn* strain had the highest concentration. The promoter region of the leukotoxin operon was not present in the sequence data and therefore these could not be compared, but it is likely that the *Fnf* promoters were weaker, due to lower concentrations of protein product in the high molecular weight fractions. Both the human and bovine strains were cytotoxic to human white blood cells, as expected based on the findings of Tadepalli *et al*. (2008a; 2008b). Similarly the statement applies that strains of both subspecies were cytotoxic. The three strains contained the full leukotoxin operon (*lktB, lktA* and *lktC*) within the genomes and a signal peptide at the beginning of the lktA protein. The two bovine strains, JCM 3718 and JCM 3724 had the most cytotoxic effect, compared to ARU 01, which had the highest cell viabilities following leukotoxin treatment for both HL-60 and human donor blood data (Figures 30 – 33) and therefore the least statistical significance. It was thought that this strain may be highly virulent as it was isolated from a patient with Lemierre's syndrome, however this was not the case and the bovine strains were more cytotoxic when concentrations were standardised. Virulence is likely to be

multifactorial, therefore additional factors such as coinfection with another bacteria or virus, or health of the host may determine *F. necrophorum* progression to Lemierre's syndrome.

Chapter 6:

Ecotin: a serine protease inhibitor

6.1 Introduction

Serine proteases are a widespread class of proteolytic enzymes, of which chymotrypsin/trypsin-like proteases are the most abundant type. They have functions involved with the immune response, blood coagulation, fibrinolysis, digestion and reproduction. Serine proteases are characterised by a nucleophilic serine residue at the active site of the enzyme (Hedstrom, 2002). Serine protease inhibitors (serpins), which have a critical role of regulating proteolytic activity, such as in clotting, thrombolytic and inflammatory pathways, are also widespread (Silverman *et al.*, 2010).

Neutrophils are abundant in normal, healthy adults, accounting for more than half of circulating white blood cells. They are the body's core defence against purulent, inflammatory bacterial infections (Bain *et al.*, 2012). They are classified as granulocytes due to their granule content within the cytoplasm and it is these granules that contain serine proteases: neutrophil elastase, proteinase 3 and cathepsin G, which are involved in the destruction of pathogens. The proteases digest microorganisms that have been engulfed into phagosomes and are also released at inflammatory sites in order to degrade microorganisms extracellular to the neutrophils (Korkmaz *et al.*, 2010). Human neutrophil elastase (HNE) is a 30 kDa trypsin-like serine protease that is able to degrade a variety of proteins in addition to elastin, including bacterial virulence proteins (Thusberg and Vihinen, 2006) such as outer membrane protein A (Belaaouaj *et al.*, 2000).

Human plasma kallikrein is involved in blood clotting as part of the contact activation system, otherwise known as the intrinsic pathway. Formed by the activation of prekallikrein, it is involved in the cleavage of high molecular weight kininogen and Factor XII, which also contribute to

coagulation. Kallikrein and Factor XIIa additionally attract neutrophils by chemotaxis (Bain *et al.*, 2012). Kallikrein consists of a heavy chain (52 kDa) and two light chains (33 – 36 kDa), linked by disulphide bonds (Colman *et al.*, 1985).

Ecotin is a serine protease inhibitor that forms dimers with a molecular weight of 32 kDa (Gaboriaud *et al.*, 2013). It was first isolated from the periplasm of *Escherichia coli* and was found to be a strong inhibitor of pancreatic chymotrypsin and elastase, rat mast cell chymase, and a less effective inhibitor of human plasma urokinase; it did not inhibit kallikrein, plasmin or thrombin. Ecotin was tested against all of the known *E. coli* proteases and was found to inhibit none of them, suggesting that ecotin plays a role in protecting the cell against external proteases; its location in the periplasm supports this (Chung *et al.*, 1983). Studies by Ulmer *et al.* (1995) and Castro *et al.* (2006) have since shown evidence that kallikrein and thrombin activity are affected by the presence of *E. coli* ecotin.

Ecotin forms homodimers via interactions at the C terminal and the ecotin dimer then forms a heterotetramer complex with two proteases that it competitively inhibits. The crystal structure of the tetramer reveals that each ecotin molecule forming the initial dimer interacts with one bound enzyme via its primary binding site, and the other enzyme in the tetramer via its secondary binding site, meaning that both ecotin monomers assist in the inhibition of both enzymes (Yang *et al.*, 1998).

Ecotin has also been found in other Gram negative bacterial genomes such as *Pseudomonas aeruginosa* and *Yersinia pestis* (Gillmor *et al.*, 2000) and there are now many sequences available in online databases. Using site directed mutagenesis it has been shown that the sequence of the primary substrate binding site can be altered and still result in a functional inhibitor (Yang *et al.*, 1998; Pál *et al.*, 1994).

E. coli ecotin has been shown by means of prothrombin time and activated partial thromboplastin time tests to be a potent anticoagulant. It was found to be a potent inhibitor of Factors Xa and XIIa, plasma kallikrein, human leukocyte elastase and bovine trypsin and chymotrypsin. It does not inhibit thrombin, VIIa, XIa, activated protein C, plasmin and tissue plasminogen activator (Seymour

et al., 1994; Ulmer *et al.*, 1995). The anticoagulant effect seen in clotting time tests has been explained as being due to the inhibition of Factor Xa, Factor XIIa and kallikrein (Ulmer *et al.*, 1995).

Ecotin has been extensively studied in *E. coli*, but has not previously been studied in *Fusobacterium spp*. *F. necrophorum* infections are known to spread via the bloodstream (Riordan, 2007), therefore the ecotin protein could be an important protective measure for the bacteria as it is likely to encounter neutrophils and their proteases while in the bloodstream. The aim of this chapter was to investigate whether *F. necrophorum* produces a functional ecotin protein, and to test its activity against host proteases.

It was hypothesised that *F. necrophorum* ecotin would be universal within the strain collection and that protein model predictions would demonstrate a structure similar to that of *E. coli*, with both primary and secondary binding sites present. PCR methods were used to determine the prevalence of the ecotin gene within the strain collection and an analysis of the ecotin DNA and translated protein sequences of the JCM 3718, JCM 3724 and ARU 01 strains was used to predict the protein structure.

It was also hypothesised that a recombinant *F. necrophorum* ecotin protein would inhibit human neutrophil elastase and human plasma kallikrein. This was tested using enzyme-substrate inhibition tests. To assess the effect of the recombinant ecotin on clotting time in normal human donor plasma, which could be affected due to the presence of serine proteases within the clotting cascade, thrombin time, prothrombin time and activated partial thromboplastin time assays were also measured.

6.2 Results

6.2.1 *Ecotin* gene locations

The *ecotin* gene was located by manually mining the annotations of the three genomes. In each genome annotation the *ecotin* gene appeared once, as part of the BLASTp manual annotation, and the locations were recorded in Table 41.

Table 41: Locations of ecotin genes found within the BLAST search results.

Isolate	Assembly contig	Base location	Strand
JCM 3718	Contig00021	8696..9100	Complement
JCM 3724	Contig00015	77..733	Complement
ARU 01	Contig00099	125..604	Complement

Using the DNA sequence from the assembly contigs, a pair of custom PCR primers were designed to amplify the putative gene in the three strains. The forward primer was located upstream of the predicted ecotin start codon and the reverse primer was located downstream of the stop codon (see Table 42).

Table 42: PCR primers used to amplify the ecotin genes in the three F. necrophorum genomes.

Primer	Oligonucleotide sequence (5' → 3')
Ecotin_F	GGCAACCAAAGACATGTAGGG
Ecotin_R	GTACCACGAAACATGCATACTT

The PCR primers in Table 42 were used to test for the presence of the *ecotin* gene within the clinical strain collection (methods section 2.8.2). Twenty five clinical strains, alongside strains JCM 3718, JCM 3724 and ARU 01, were tested with the ecotin primers. Of these, a product of approximately 700 bp was amplified in 100 % of strains. See Figure 34 for an example gel image.
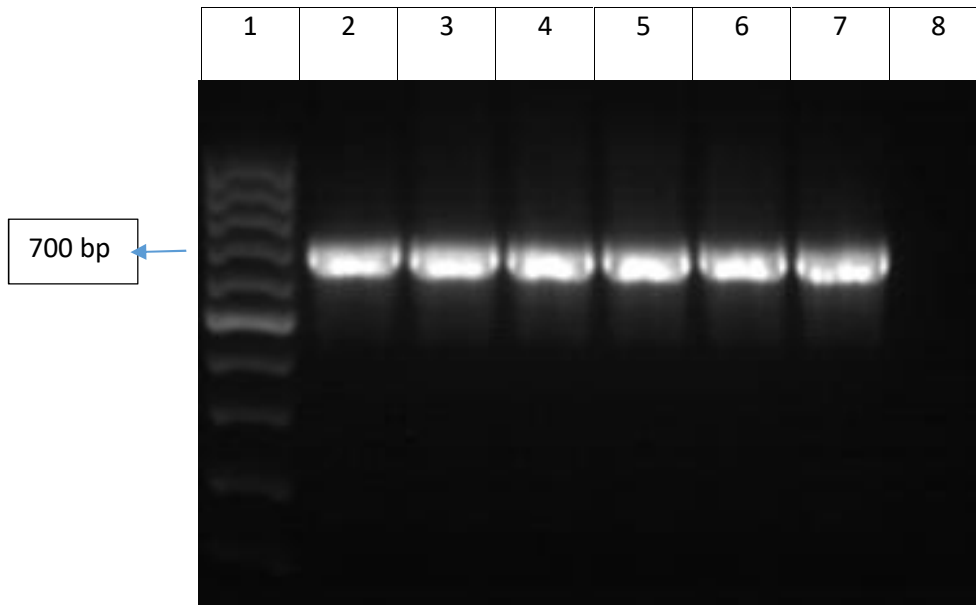
Figure 34: Analysis of PCR products from Ecotin primers on a 1% agarose gel. Lane 1: 100 bp marker; 2: F. necrophorum 1; 3: F. necrophorum 5; 4: F. necrophorum 11; 5: F. necrophorum 21; 6: F. necrophorum 24; 7: F. necrophorum 30; 8: Negative control (molecular grade water).

6.2.2 Gene sequence analysis

6.2.2.1 Clustal alignments

Amplified ecotin PCR products from strains JCM 3718, JCM 3724 and ARU 01 were run in a 1 % agarose gel (methods section 2.8.3), excised and purified (methods section 2.8.5) before being sent for commercial Sanger sequencing (methods section 2.8.7). The resulting sequences were aligned using Clustal Omega (Figure 35) and show the three sequences are very highly conserved. The sequences from the two *Fnf* strains are 100 % identical, however the sequence from the *Fnn* strain differs in two locations, resulting in a 99.58 % identity. At position 189 there is a cytosine in place of a thymine, and in position 250 there is an adenine in place of a guanine.

The DNA sequences of the first 12 strains from the clinical collection were also determined by Sanger sequencing method and a Clustal alignment was performed (Appendix 21). The sequences were shown to have percentage identities of 98.96 % or above.

```
JCM3718    ATGAAAAAATGTATTTATGCTATCGGCTTACTGTTTTCTTTTTCCGTCAGTGTTTTTGCA
JCM3724    ATGAAAAAATGTATTTATGCTATCGGCTTACTGTTTTCTTTTTCCGTCAGTGTTTTTGCA
ARU01      ATGAAAAAATGTATTTATGCTATCGGCTTACTGTTTTCTTTTTCCGTCAGTGTTTTTGCA
           ************************************************************

JCM3718    ATGCAGCATCCGGATATGAACTTGGAAATATATCCTAAGGCAAAACAAGGCATGAAGAAG
JCM3724    ATGCAGCATCCGGATATGAACTTGGAAATATATCCTAAGGCAAAACAAGGCATGAAGAAG
ARU01      ATGCAGCATCCGGATATGAACTTGGAAATATATCCTAAGGCAAAACAAGGCATGAAGAAG
           ************************************************************

JCM3718    GTTGTGTATCTTTTAGAGAAAAAGAAAAGAGAAGAGACTATAAATTGGAAATAAAATTT
JCM3724    GTTGTGTATCTTTTAGAGAAAAAGAAAAGAGAAGAGACTATAAATTGGAAATAAAATTT
ARU01      GTTGTGTATCTTTTAGAGAAAAAGAAAAGAGAAGAGACTATAAATTGGAAATAAAATTT
           ************************************************************

JCM3718    GGAAAAGACCTTGTTGTAGATGATAATCTTCATCACTTTTTAGGAGGAAAGCTGGAAGAG
JCM3724    GGAAAAGATCTTGTTGTAGATGATAATCTTCATCACTTTTTAGGAGGAAAGCTGGAAGAG
ARU01      GGAAAAGATCTTGTTGTAGATGATAATCTTCATCACTTTTTAGGAGGAAAGCTGGAAGAG
           ******** ***************************************************

JCM3718    AAAGATGTAAAAGGTTGGGGCTATCCTTATTACATTTTTTCAGGAGATTCTCAAATGGCA
JCM3724    AAAGATGTAGAAGGTTGGGGCTATCCTTATTACATTTTTTCAGGAGATTCTCAAATGGCA
ARU01      AAAGATGTAGAAGGTTGGGGCTATCCTTATTACATTTTTTCAGGAGATTCTCAAATGGCA
           ********* **************************************************

JCM3718    CAAACTTTAATGGCGTTTCCTTTAGGAAGTGAACGAGAAAAAAGAGTATATTATCCCACA
JCM3724    CAAACTTTAATGGCGTTTCCTTTAGGAAGTGAACGAGAAAAAAGAGTATATTATCCCACA
ARU01      CAAACTTTAATGGCGTTTCCTTTAGGAAGTGAACGAGAAAAAAGAGTATATTATCCCACA
           ************************************************************

JCM3718    GCTACGAAAATATTGCCTTATCATTCAAAACTTCCTTTGGTTTTATATGTTCCGGAAGAT
JCM3724    GCTACGAAAATATTGCCTTATCATTCAAAACTTCCTTTGGTTTTATATGTTCCGGAAGAT
ARU01      GCTACGAAAATATTGCCTTATCATTCAAAACTTCCTTTGGTTTTATATGTTCCGGAAGAT
           ************************************************************

JCM3718    GTGAAGGTAGAAGTATCTCTTTGGAATCGAATGCAGGAAATCAAAGAAGTTTCTCGTTAA
JCM3724    GTGAAGGTAGAAGTATCTCTTTGGAATCGAATGCAGGAAATCAAAGAAGTTTCTCGTTAA
ARU01      GTGAAGGTAGAAGTATCTCTTTGGAATCGAATGCAGGAAATCAAAGAAGTTTCTCGTTAA
           ************************************************************
```

Figure 35: Clustal Omega alignment of the DNA sequence results from Sanger sequencing the three ecotin genes. Locations where sequences are homologous are marked with '*'. Locations where bases do not match are marked in red.

The sequences were translated into protein using ExPASy translate (methods section 2.10.3), with the first frame of the 5' to 3' sequence representing an open reading frame. The predicted protein sequences were aligned using Clustal Omega (Figure 36). The protein alignment shows that of the two substitutions highlighted in Figure 35, only one resulted in an amino acid substitution. The difference at base number 189 caused no amino acid substitution as GAC and GAU both code for aspartic acid. However the difference at base 250 resulted in a prediction of a lysine (AAA) in the *Fnn* gene, compared to a glutamic acid (GAA) in the two *Fnf* genes. The location of this amino acid variation in relation to the active site of the protein is discussed in section 6.2.2.3.

157

```
JCM3718    MKKCIYAIGLLFSFSVSVFAMQHPDMNLEIYPKAKQGMKKVVYLLEKKEKEEDYKLEIKF
JCM3724    MKKCIYAIGLLFSFSVSVFAMQHPDMNLEIYPKAKQGMKKVVYLLEKKEKEEDYKLEIKF
ARU01      MKKCIYAIGLLFSFSVSVFAMQHPDMNLEIYPKAKQGMKKVVYLLEKKEKEEDYKLEIKF
           ************************************************************

JCM3718    GKDLVVDDNLHHFLGGKLEEKDVKGWGYPYYIFSGDSQMAQTLMAFPLGSEREKRVYYPT
JCM3724    GKDLVVDDNLHHFLGGKLEEKDVEGWGYPYYIFSGDSQMAQTLMAFPLGSEREKRVYYPT
ARU01      GKDLVVDDNLHHFLGGKLEEKDVEGWGYPYYIFSGDSQMAQTLMAFPLGSEREKRVYYPT
           **********************:**************************************

JCM3718    ATKILPYHSKLPLVLYVPEDVKVEVSLWNRMQEIKEVSR
JCM3724    ATKILPYHSKLPLVLYVPEDVKVEVSLWNRMQEIKEVSR
ARU01      ATKILPYHSKLPLVLYVPEDVKVEVSLWNRMQEIKEVSR
           ***************************************
```

Figure 36: Clustal Omega alignment of the predicted protein sequence results following the use of the ExPASy online translation tool.

The *ecotin* gene sequences from strains JCM 3724 and ARU 01 were identical, therefore this *Fnf* consensus sequence was used for the *in vitro* investigations as this subspecies is the one most associated with humans and is therefore most relevant.

6.2.2.2 Signal peptide prediction

The *Fnf* ecotin protein sequence was entered into the web-based tool SignalP (Petersen *et al.*, 2011) using Gram negative settings to predict the presence and location of signal peptide cleavage sites (Figure 37). The results showed a predicted cleavage site between amino acid positions 20 and 21, as shown by the first red and blue peak, suggesting that the mature protein begins on the second methionine of the sequence. The D value weighted average score (see figure legend for definition) was 0.777 with the cutoff being 0.570. The second, smaller peak in the red and blue lines was not considered significant by the algorithm. It is not uncommon for there to be multiple high peaking C scores when only one cleavage site is truly present (Petersen *et al.*, 2011).
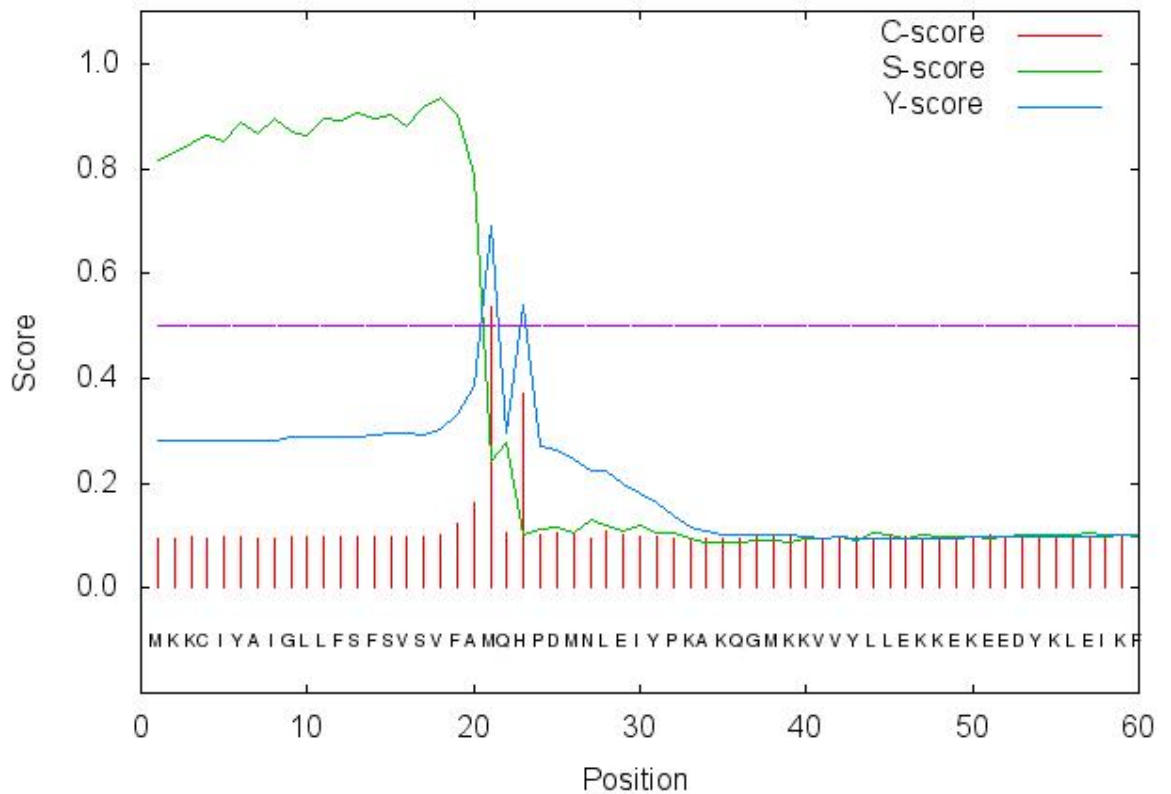
Figure 37: SignalP prediction of signal peptide cleavage site. C scores are high at the position immediately after the cleavage site, S scores distinguish between regions within the signal peptide and regions within the mature protein, and Y scores are a combination of the C score average and S score slopes. A D score is then used to make a weighted average of the mean S scores and the maximum Y scores which then determines whether a signal peptide is present.

6.2.2.3 Pfam search for conserved domains

The sequence was searched for conserved protein domains using Pfam, resulting in a match for ecotin with an E value of 1.3e-32. A Clustal Omega alignment was carried out between the hidden Markov model (HMM) consensus sequence for ecotin used by Pfam and the *Fnf* ecotin sequence, resulting in a 42.62 % identity. Putative conserved domains were then predicted by BLASTp search. The results of this were overlaid onto the sequence alignment (Figure 38). The four amino acids highlighted in green are those predicted to be the primary substrate binding site. Three of the four amino acids are conserved exactly, while the remaining position has been substituted for an amino acid with weakly similar properties. The two segments highlighted in yellow are predicted to make up the secondary substrate binding site. In the first of the two, four out of the five positions are conserved, and in the second, four out of the six are conserved. The L and M in the primary

substrate binding site are also predicted to form an inhibition loop and match the ecotin profile exactly. The single variation between the *Fnn* and *Fnf* sequences occurs in the amino acid immediately before part of the secondary substrate binding site. This may impact on substrate specificity or binding efficiency, but without laboratory analysis it is not known.

```
HMM     ------DLAPYPAPEEGQKRHVIKLPKLEDEADYKVELIIGKTLEVDCN-KQRLSGELEE
Fnf     MQHPDMNLEIYPKAKQGMKKVVYLLEKKEKEEDYKLEIKFGKDLVVDDNLHHFLGGKLEE
             :*    **   ::* *:  *   * * *.* ***:*: :** * ** * :: *.*:***

HMM     KTLEGWGYEYYEVEKASEAASTLMACPDDEKKKEKFVSLEEGEKLLLRYNSKLPVVVYLP
Fnf     KDVEGWGYPYYIFSGDSQMAQTLMAFPLGSEREKRVYY--PTATKILPYHSKLPLVLYVP
        *  :***** **  ..   *: *.**** *   .:::::.      . :* *.****:*:*:*

HMM     KDVELRYRVWK----------
Fnf     EDVKVEVSLWNRMQEIKEVSR
        :**::.   :*:
```

Figure 38: Clustal Omega alignment of the hidden Markov model consensus sequence of ecotin and the sequence of ecotin from the Fnf genomes. Dashes represent where there is no corresponding sequence, '*' represents an exact match, ':'represents amino acids with strongly similar properties and '.' represents those with weakly similar properties. The green highlighted sequence corresponds to the predicted primary substrate binding site and the yellow highlighted sequence corresponds to the predicted secondary binding site regions. The amino acid in red shows the position where the Fnn sequence differs.

6.2.2.4 Model of *F. necrophorum* ecotin

SWISS-MODEL and PyMol were used to create and visualise a protein prediction model of *F. necrophorum* ecotin (Figure 39) using an *E.coli* ecotin reference (Figure 40). The two regions of the protein sequence that make up the secondary binding site (highlighted in Figure 38) appear as one region in the 3 dimensional protein model prediction, which was as expected in order for it to form a single binding site. The primary (B1) and secondary (B2) binding sites were predicted to be exposed on opposite sides of the protein. A computer graphic model, based on a crystallised structure of *E. coli* ecotin shows a similar structure, with two ecotin monomers linked in such a way that when the ecotin dimer binds to two target proteases, both monomers bind to both proteases (Yang *et al.*, 1998). The prediction that an *F. necrophorum* ecotin monomer has a similar structure to an *E. coli* ecotin monomer suggests that there is a possibility that *F. necrophorum* ecotin is also

a functional inhibitor. Laboratory tests are needed to confirm this hypothesis (sections 6.2.3 – 6.2.5).
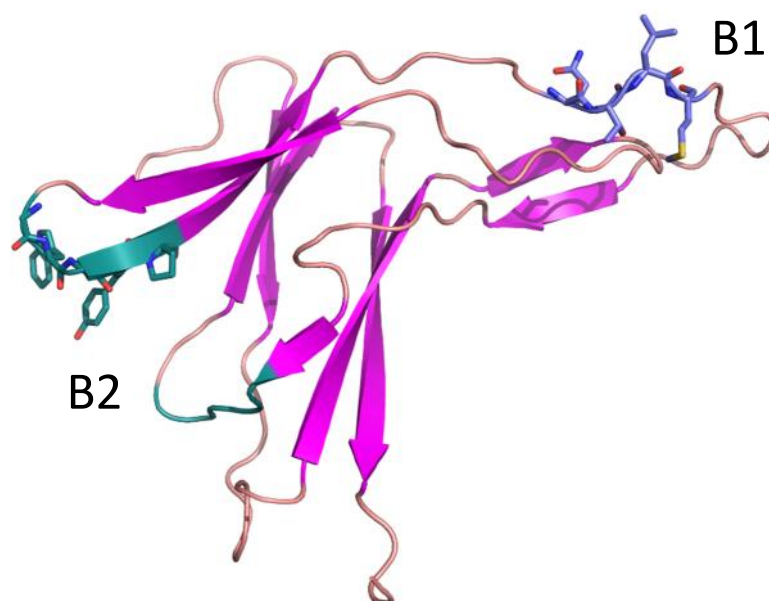
Figure 39: Predicted protein model structure of F. necrophorum ecotin highlighting the primary binding site (B1) and secondary binding site (B2).
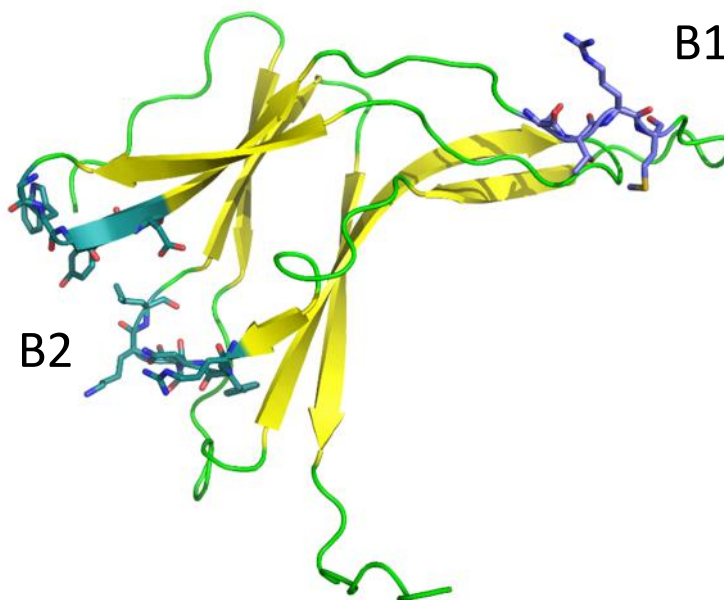
Figure 40: Protein model structure of E. coli ecotin highlighting the primary binding site (B1) and secondary binding site (B2). This model, visualised in PyMol, was the SWISS-MODEL reference used to predict the structure of F. necrophorum ecotin in Figure 39 and is based on crystallised structures. The structures of E. coli and F. necrophorum ecotin appear similar to each other and both contain B1 and B2 exposed on opposite sides of the protein.

6.2.3 Cloning, expression and purification of recombinant ecotin

6.2.3.1 Plasmid design

A recombinant *ecotin* gene was designed to be inserted into a plasmid vector to enable expression within *E. coli* cells and purification of the protein. A Novagen pET-16b plasmid was provided by Dr Anatoliy Markiv, University of Westminster. This plasmid was chosen for its histidine tag, T7 promoter and ampicillin resistance gene. A plasmid map and sequence of the multiple cloning site are shown in Appendix 22.

The cloning/expression region of the pET-16b plasmid includes a region cleaved by Factor Xa. Ecotin is thought to inhibit Factor Xa, therefore due to concerns with the Factor Xa cleavage site interfering with the ecotin assays this site was exchanged for an enterokinase site, making the pET-16b vector sequence match that of the pET-19b vector sequence (see Appendix 23).

The restriction sites chosen were NcoI and XhoI. A sequence was designed to be synthesised which included, in order, an NcoI restriction site, deca-histidine tag, enterokinase cleavage site, NdeI restriction site, the *F. necrophorum* ecotin sequence matching that of the two subspecies *funduliforme* strains, including the stop codon, and a XhoI restriction site. This allowed for the section between the NcoI and XhoI restriction sites to be removed from the pET-16b vector and be replaced by the equivalent sequence from the pET-19b sequence, with the addition of the ecotin sequence between the NdeI and XhoI restriction sites.

The *F. necrophorum* ecotin sequence contains some codons that are known to be rarely used in *E. coli* strains, therefore the codons were optimised to suit the codon bias of *E. coli* in order to facilitate high and stable expression rates while still producing the same translated protein sequence. This was carried out as part of the commercial synthesis package. An alignment of the original sequence designed for the plasmid and the final optimised sequence was produced to ensure the restriction sites had been conserved (Figure 41). These sites are where the DNA is cleaved during the cloning process and would be non-functional if altered. When translated into protein, the original and optimised sequences were 100 % congruent.

```
Optimised    CCATGGGCCATCATCATCACCATCACCACCATCATCATAGCAGCGGTCATGATGATGATG
Original     CCATGGGCCATCATCATCATCATCATCATCATCATCACAGCAGCGGCCATGACGACGACG
             ***************** ***** ** ******* ******* ***** ** ** *

Optimised    ATAAACATATGCAGCATCCGGATATGAACCTGGAAATTTATCCGAAAGCAAAACAGGGCA
Original     ACAAGCATATGCAGCATCCGGATATGAACTTGGAAATATATCCTAAGGCAAAACAAGGCA
             * ** .******************** ****** ****** **.******* .****

Optimised    TGAAAAAGTTGTTTATCTGCTGGAAAAAAAGAAAAAGAAGAAGATTACAAACTGGAAA
Original     TGAAGAAGGTTGTGTATCTTTTAGAGAAAAAAGAAAAAGAAGAAGACTATAAATTGGAAA
             ****.**.***** ***** *.**.****************** ** *** ******

Optimised    TCAAATTTGGCAAAGATCTGGTGGTGGATGATAACCTGCATCATTTTCTGGGTGGTAAAC
Original     TAAAATTTGGAAAAGATCTTGTTGTAGATGATAATCTTCATCACTTTTTAGGAGGAAAGC
             * .********.******** ** **.******** ** ***** *** *.**:**:**.*

Optimised    TGGAAGAAAAAGATGTTGAAGGTTGGGGCTATCCGTATTATATCTTTAGCGGTGATAGCC
Original     TGGAAGAGAAAGATGTAGAAGGTTGGGGCTATCCTTATTACATTTTTTCAGGAGATTCTC
             *******.*********:***************** ***** ** ***: .**:***: *

Optimised    AGATGGCACAGACCCTGATGGCATTTCCGCTGGGTAGCGAACGTGAAAAACGTGTTTATT
Original     AAATGGCACAAACTTTAATGGCGTTTCCTTTAGGAAGTGAACGAGAAAAAAGAGTATATT
             *.********.**   *.*****.***** * .**:** *****:******.*:**:****

Optimised    ATCCGACCGCAACCAAAATTCTGCCGTATCATAGCAAACTGCCGCTGGTTCTGTATGTTC
Original     ATCCCACAGCTACGAAAATATTGCCTTATCATTCAAAACTTCCTTTGGTTTTATATGTTC
             **** **.**:** *****: **** ******: .***** ** ***** *.*******

Optimised    CGGAAGATGTTAAAGTTGAAGTTAGCCTGTGGAATCGCATGCAAGAAATTAAAGAAGTGA
Original     CGGAAGATGTGAAGGTAGAAGTATCTCTTTGGAATCGAATGCAGGAAATCAAAGAAGTTT
             ********** **.**:*****::  ** ********.***** .***** ******** :

Optimised    GCCGTTAACTCGAG
Original     CTCGTTAACTCGAG
                 ************
```

Figure 41: Clustal Omega alignment showing the locations where DNA bases have been changed to suit the codon bias of E. coli. Sequence highlighted in yellow shows the locations of restriction sites, which have the DNA sequence preserved in order to retain the cleavage function. The section highlighted in green is the deca-histidine tag and the grey segment is the enterokinase cleavage site. The sequence coding for the ecotin gene is between the second and third restriction sites.

The plasmid containing the *ecotin* insert was commercially synthesised (methods section 2.14.2) and arrived as lyophilised DNA. After reconstitution (methods section 2.14.3), the plasmid was transformed into chemically competent TOP10 *E. coli* cells (methods section 2.14.4) and grown overnight at 37 °C on an LB agar plate with ampicillin. The next day, one colony was selected and inoculated into LB broth with ampicillin and grown for 16 hours at 37 °C and 330 rpm. A plasmid miniprep was carried out to extract plasmid DNA (methods section 2.14.5). Purified plasmid was quantified using a Nanodrop (methods section 2.8.6) and then digested using restriction enzymes to cleave the DNA at the NcoI and XhoI cloning sites (methods section 2.14.6). The pET-16b plasmid was cleaved simultaneously using the same enzymes and all digestions were run in a 0.8 % agarose gel. The resulting bands were excised and the DNA extracted using a DNA gel purification kit

(methods section 2.8.5). The nucleic acid was quantified using a Nanodrop. The cut insert was ligated into the pET-16b plasmid backbone (methods section 2.14.7) before being transformed into chemically competent TOP10 *E. coli* cells and grown overnight at 37 °C. One colony was selected and inoculated into LB broth with ampicillin and grown for 16 hours at 37 °C and 330 rpm. A plasmid miniprep was carried out to extract plasmid DNA. The digest was repeated and cut plasmids were run on a 0.8 % agarose gel to check for the presence of the insert. Successfully ligated plasmid DNA was transformed into chemically competent BL21(DE3) *E. coli* cells (methods section 2.14.8) for expression. A colony was selected from the LB agar with ampicillin, grown overnight in LB broth with ampicillin, and a plasmid miniprep was carried out to extract plasmid DNA. This was quantified using a Nanodrop before being sent for commercial Sanger sequencing using primers targeting the T7 promoter in order to verify that the final plasmid contained the correct sequence.

6.2.3.2 Protein expression induced with IPTG

E. coli BL21(DE3) expression strain transformed with pET-16b+ecotin was induced with IPTG to express the recombinant ecotin protein (methods section 2.15.1). BL21(DE3) containing no plasmid was grown in parallel to be used as a negative control. 1 ml samples were taken from each culture at IPTG induction (0 hours) and 1, 2 and 3 hours post-induction for analysis by SDS PAGE using the Tricine method and a denaturing buffer (Figure 42) (see methods 2.15.3).
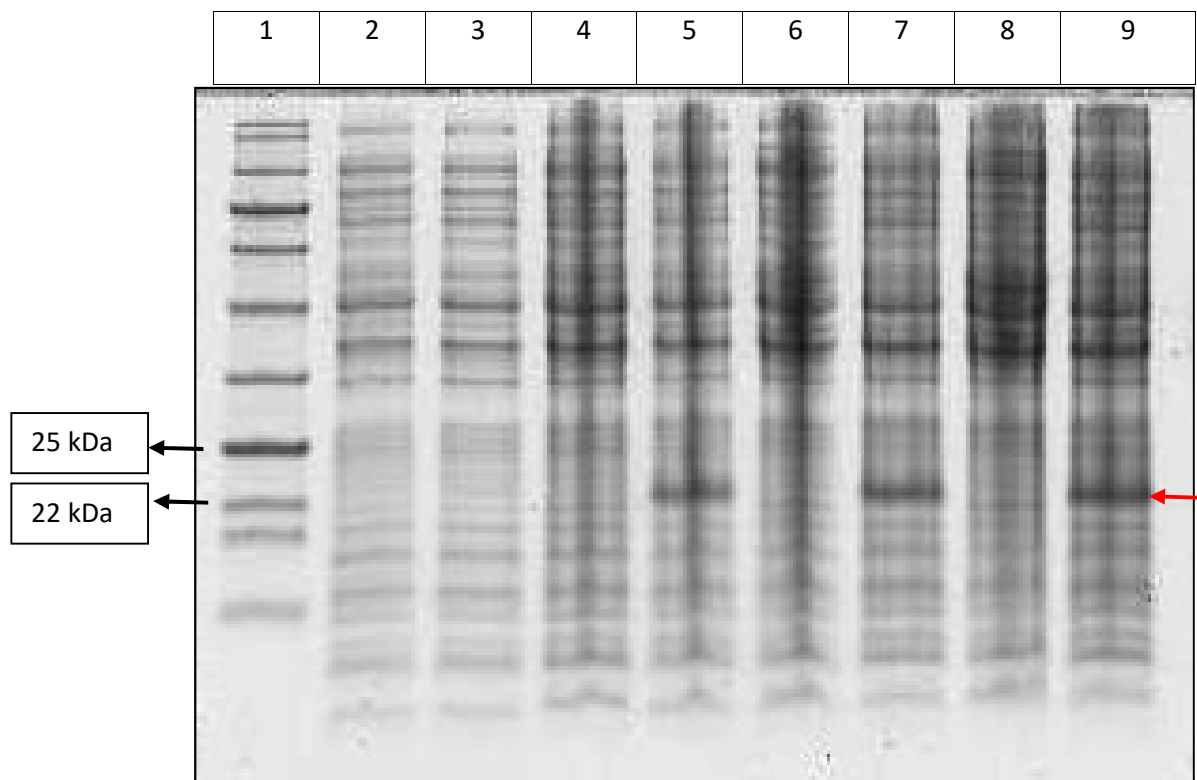
Figure 42: SDS PAGE analysis of expression strain with (BL21(DE3)+) and without (BL21(DE3)-) plasmid. Lane 1: NEB marker (11-190 kDa), 2: BL21(DE3)- 0 hours, 3: BL21(DE3)+ 0 hours, 4: BL21(DE3)- 1 hour, 5: BL21(DE3)+ 1 hour 6: BL21(DE3)- 2 hours, 7: BL21(DE3)+ 2 hours, 8: BL21(DE3)- 3 hours, 9: BL21(DE3)+ 3 hours. Presumed ecotin band highlighted with red arrow.

E. coli produces its own ecotin protein, so a band was expected in the 20 kDa region with or without the plasmid present. At zero hours, both the plasmid-containing strain and the negative control appeared the same. At one, two and three hours post-induction with IPTG, the plasmid-containing strain contained a more pronounced band of the size equivalent to the presumptive ecotin protein. This suggests that the IPTG has successfully induced production of the recombinant ecotin protein.

6.2.3.3 Purification of histidine-tagged protein by immobilised metal ion chromatography (IMAC) under native conditions

The histidine-tagged protein was purified under native conditions by immobilised metal affinity chromatography over a cross-linked agarose matrix charged with nickel ions (see methods 2.15.2). Samples were analysed using the Tricine-SDS PAGE method using denaturing loading buffer; Figure 43 shows samples of each step of the purification process. Lane 2 demonstrates that the protein of

interest is present in the total lysate of the cell pellet, and has therefore not been secreted. Lane 3 demonstrates that the protein is present in the soluble fraction, as opposed to being maintained in inclusion bodies, which would require purification under denaturing conditions. Lane 4 shows the column flow through, containing proteins that have not bound to the column. Some protein of interest did not bind to the column and can be seen in the flow through. Lanes 5 and 6 were samples from the first and final wash fractions, respectively. The absence of any bands in lane 6 showed that the column had been washed thoroughly enough to remove any protein not bound strongly to the resin. Lanes 7 to 10 were samples of elution aliquots, and contained protein at the approximate expected band size. There were also two high molecular weight bands faintly present, suggesting the elution had not yielded a protein that was 100 % pure.
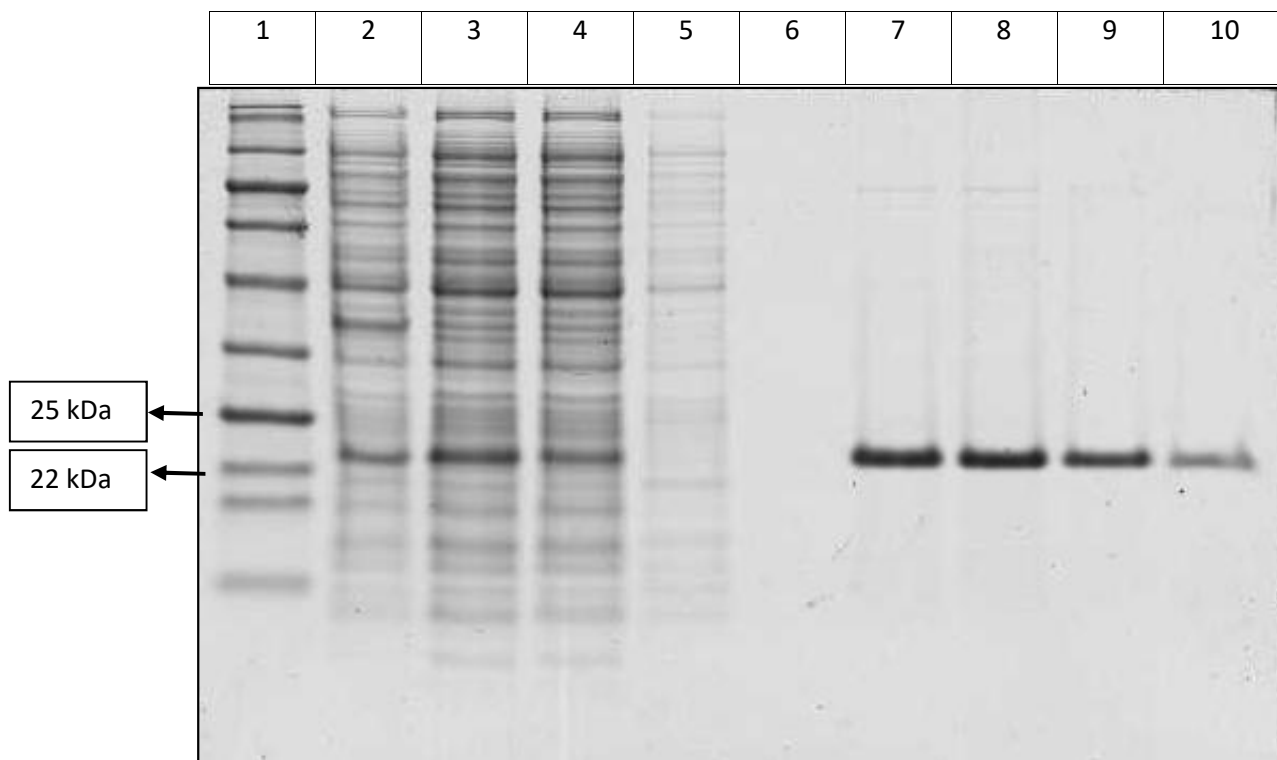
Figure 43: SDS PAGE analysis of protein purification fractions. Lane 1: NEB marker (11-190 kDa), 2: Total lysate, 3: Soluble fraction, 4: Flow through, 5: Wash 1, 6: Wash 5, 7: Elution 1, 8: Elution 2, 9: Elution 3, 10: Elution 4. Quantities loaded were: 5 µl of protein marker and 15 µl of each sample in 5 µl 4x Laemmli buffer.

6.2.3.4 Bradford assay to determine concentration of purified ecotin

A Bradford assay was carried out to measure the concentration of purified recombinant ecotin protein (methods section 2.15.4). The standard curve generated fitted the linear equation y=0.029x + 0.0513 and had an R squared value of 0.995.

Table 43: Protein concentrations of eluted recombinant ecotin products, as measured by Bradford assay.

Elution fraction	Concentration (mg/ml)
1	0.688
2	1.131
3	0.138
4	0.046

The above elution fractions were diluted as appropriate for use in enzyme inhibition and clotting assays.

6.2.4 Human plasma kallikrein and human neutrophil elastase inhibition assays

The inhibitory effect of *F. necrophorum* ecotin, which has not previously been investigated, may play a role in the virulence of the organism. The recombinant ecotin protein was therefore tested for inhibitory activity against human plasma kallikrein, part of the intrinsic clotting cascade and human neutrophil elastase, which functions as part of the immune response. These serine proteases have previously been shown to be inhibited by *E. coli* ecotin (Seymour *et al.*, 1994; Ulmer *et al.*, 1995).

6.2.4.1 Human plasma kallikrein inhibition assay

Human plasma kallikrein (HPK) was tested with a range of kallikrein substrate concentrations to establish normal, uninhibited enzyme kinetics (methods section 2.16). HPK concentration was kept constant at 10 nM throughout the experiment. Each concentration of substrate was added to the enzyme in triplicate and relative fluorescence units were measured every minute for 30 minutes (Figure 44). The kallikrein substrate contained a 7-Amino-4-(trifluoromethyl)coumarin (AFC) terminus, which is cleaved by HPK, resulting in fluorescence. In order to establish rate of product formation, a calibration curve was constructed with a concentration range of 0 – 0.5 mM AFC. The resulting line had an equation of y=244626x and an R squared value of 0.9978 (Appendix 24).

Standards
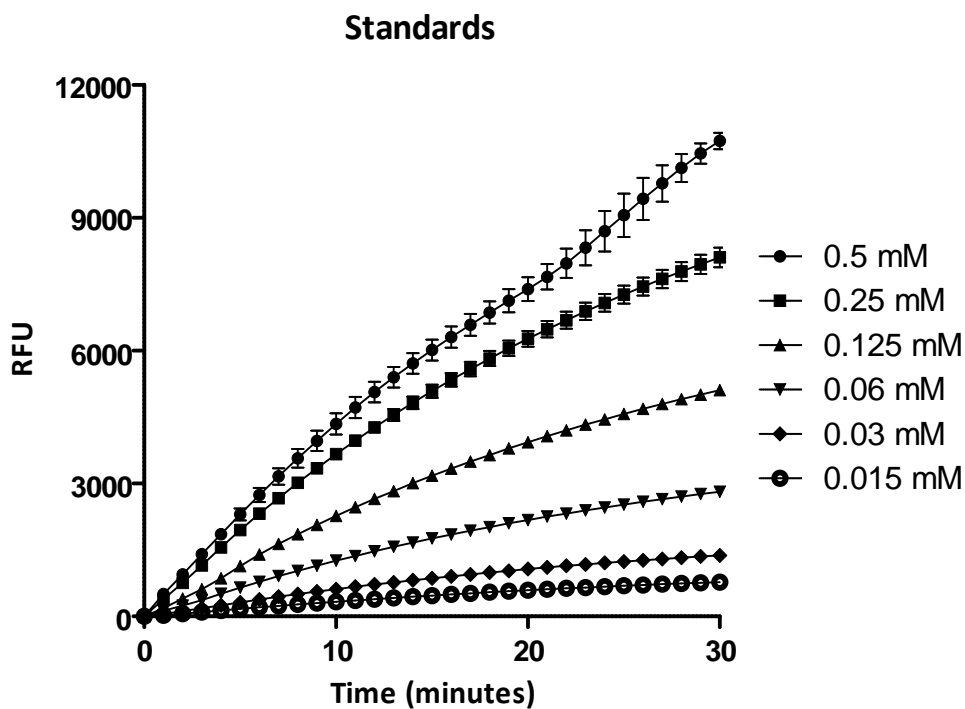
Figure 44: XY plot of relative fluorescence units (RFU) measured over 30 minutes from human plasma kallikrein. Substrate concentrations ranged from 0.015 mM to 0.5 mM and inhibitor was at 0 nM. The graph shows the mean of three replicate values and standard error of the mean is represented by bars.

The change in fluorescence at 5 minutes was used to calculate the rate of reaction (μmol min^{-1}) for each concentration of kallikrein substrate (see Table 44). The 5 minute time point was chosen for the initial rate as this was within the initial linear section of the graph. [P] (concentration of product) was calculated using the equation derived from the AFC calibration curve: y=244626x. Therefore [P]=y/244626.

Table 44: Calculations for the rate of reaction (μmol min⁻¹) of HPK with substrate.

[S] mM	Change in relative fluorescence	[P] after 5 minutes (mM)	P (μmol)	Rate of reaction (μmol min⁻¹)
0	0.00	0.0000	0.0000	0.0000
0.015	169.61	0.0007	0.0693	0.0139
0.03	317.51	0.0013	0.1298	0.0260
0.06	649.48	0.0027	0.2655	0.0531
0.125	1136.03	0.0046	0.4644	0.0929
0.25	1949.89	0.0080	0.7971	0.1594
0.5	2303.78	0.0094	0.9418	0.1884

The values in Table 44 were used to make an XY plot of substrate concentration against rate of reaction (Figure 45). Michaelis-Menten non-linear regression was applied to calculate the V_{max} and K_m values (see Table 45).
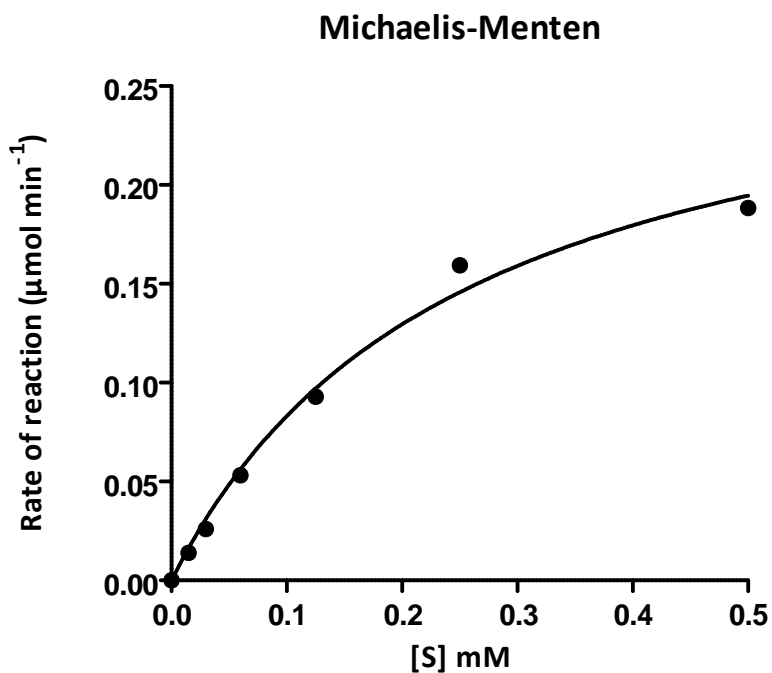
Michaelis-Menten

Figure 45: Michaelis-Menten HPK plot, showing the relationship between substrate concentration and rate of reaction. The curve was fitted by nonlinear regression in GraphPad Prism using the Michaelis-Menten settings. R squared value = 0.9912.

Table 45: Results of nonlinear regression analysis of data in Table 44 and Figure 45.

	Best fit values	Standard error	95% Confidence intervals
Vmax	0.293	0.026	0.226 to 0.360
Km	0.252	0.047	0.132 to 0.372

Using nonlinear regression analysis, the K_m for kallikrein substrate with human plasma kallikrein under the conditions specified (in methods section 2.16) was 252 µM. V_{max} was 0.293 µmol min^{-1}.

In order to confirm that autolysis of the substrate was not responsible for the increase in fluorescence, relative fluorescence for the substrate range 0.015 – 0.5 mM was monitored for 30 minutes with no enzyme present. The change in fluorescence for each substrate concentration was equal to less than 1 % of the fluorescence of the respective substrate standard.

HPK was then tested with the addition of ecotin (Figure 46), with fluorescence monitored as above.

See Appendix 25 for RFU readings.
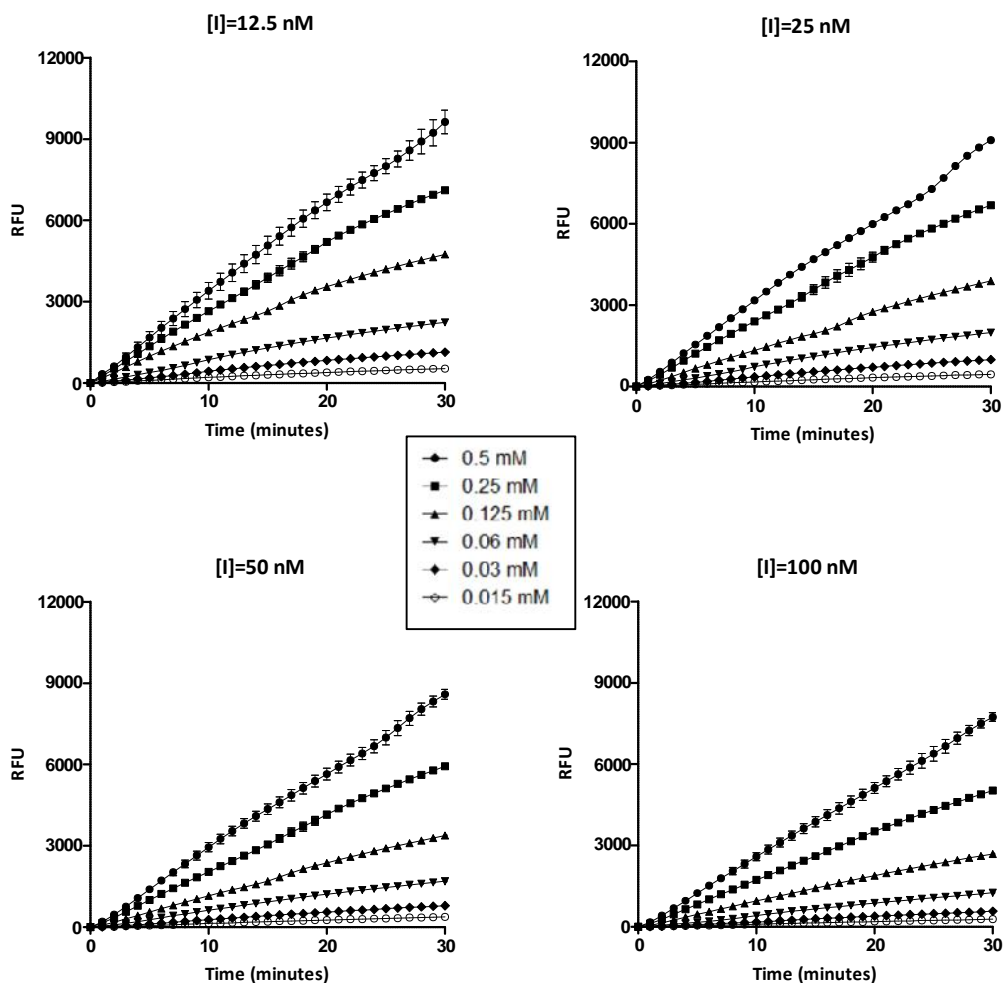
Figure 46: XY plots of RFUs monitored over 30 minutes in the presence of HPK (10 nM), substrate (0.015 – 0.5 mM) and inhibitor (12.5 – 100 nM). Enzyme and inhibitor were incubated for an hour at room temperature to equilibrate before addition of substrate.

In order to establish the type of inhibition caused by ecotin, a Lineweaver-Burk plot was produced (see Appendix 26). The plot showed characteristics of competitive inhibition, such as increasing slopes and K_m for increasing inhibitor concentrations while V_{max} remained stable.

Ecotin is thought to be a tight-binding inhibitor, so Morrison K_i nonlinear regression was applied. The ratio of inhibited to uninhibited rate of reaction was plotted against the inhibitor concentration for each substrate concentration (Figure 47). The Morrison equation was used as part of the

GraphPad Prism package to determine the estimated concentrations required to produce half maximal inhibition, known as the inhibition constant (K_i) (Table 46).
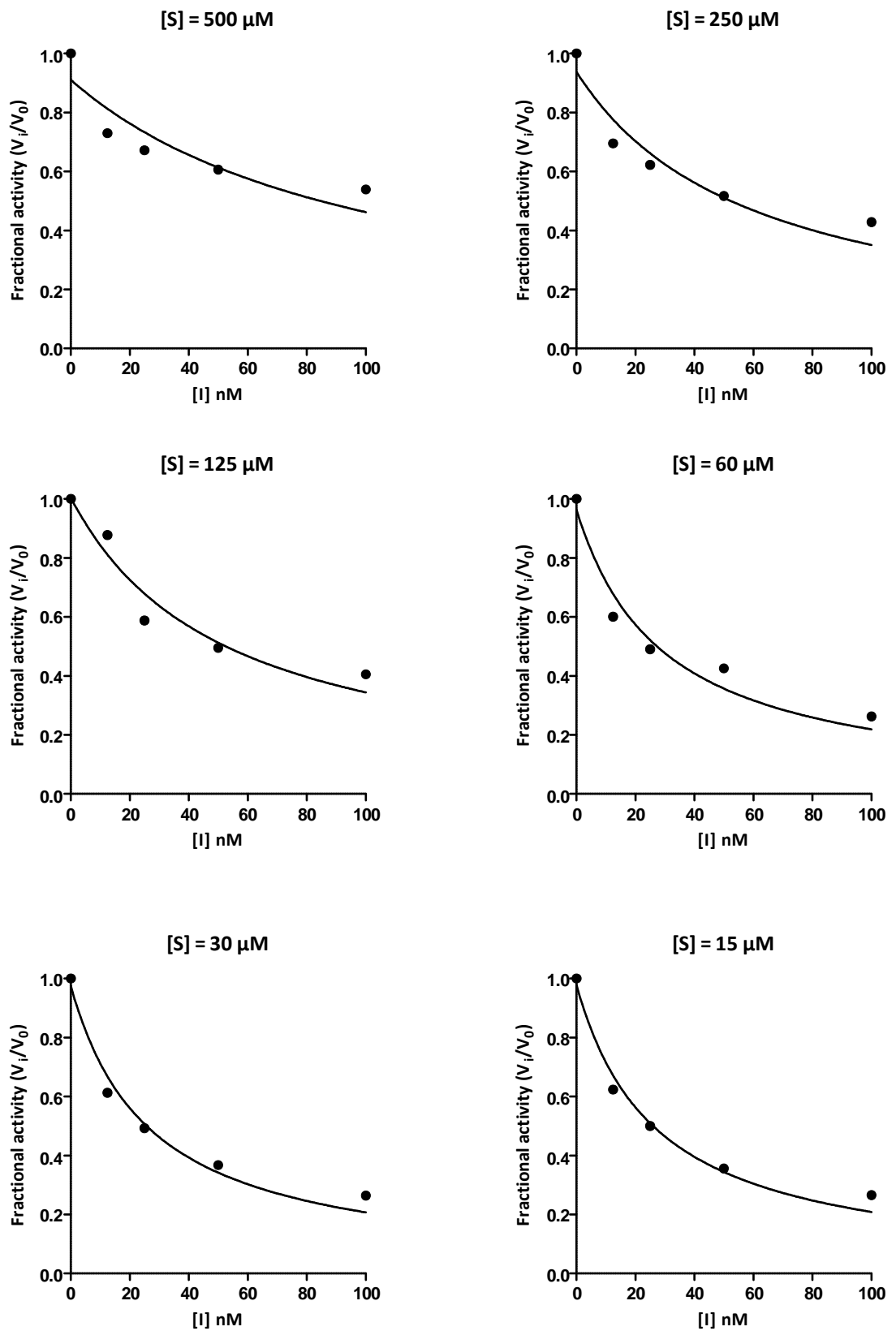
Figure 47: Fractional activity of HPK derived using Morrison K_i nonlinear regression.

Table 46: K_i values determined using Morrison tight binding kinetics within GraphPad Prism.

[S] μM	K_i	SE	95 % Confidence Intervals	R squared
15	**25.55**	**3.577**	**14.17 to 36.94**	**0.9815**
30	24.1	3.779	12.08 to 36.13	0.9766
60	23.71	5.239	7.043 to 40.38	0.9526
125	34.81	8.511	7.733 to 61.90	0.9348
250	29.99	8.247	3.754 to 56.24	0.9073
500	34.55	13	-6.815 to 75.91	0.8054

The Morrison K_i of *F. necrophorum* ecotin with human plasma kallikrein had the highest R squared value and the best visual fit at 15 μM, and was determined to be 26 nM.

6.2.4.2 Human neutrophil elastase inhibition assay

Human neutrophil elastase (HNE) was tested with a range of elastase substrate (Methoxysucinyl-Ala-Ala-Pro-Val-p-nitroanilide) concentrations to establish normal, uninhibited enzyme kinetics (methods section 2.16). Each concentration was added to the enzyme in triplicate and absorbance was measured every minute for 30 minutes (Figure 48). HNE concentration was kept constant at 17 nM throughout the experiment. The chromogenic elastase substrate contained a p-nitroanilide (pNA) terminus, which is cleaved by elastase, resulting in a shift in absorption spectrum. pNA has a molar extinction coefficient (ε) of 8800 M^{-1} cm^{-1} at 410 nm, as stated in the product information sheet provided by Merck, which when used with Beer-Lambert Law will establish concentration of product (Equation 1 in methods section 2.16).
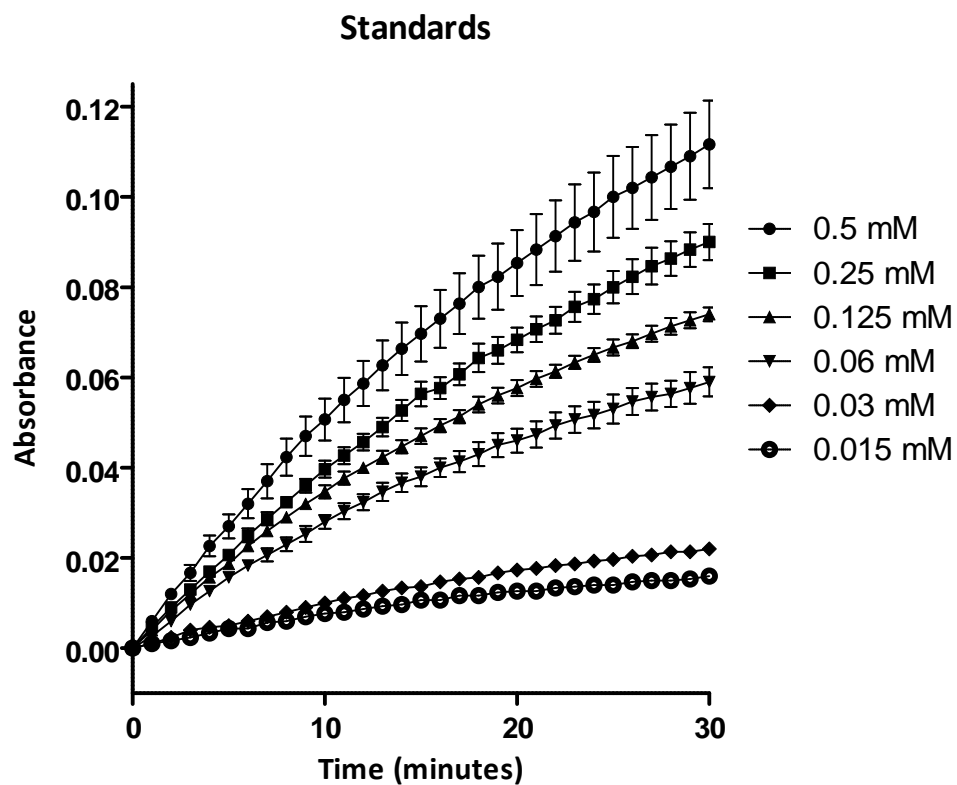
Figure 48: XY plot of absorbance measured over 30 minutes from human neutrophil elastase and a range of substrate concentrations (0.015 mM to 0.5 mM); no inhibitor was present. The graph shows the mean of three replicate values and standard error of the mean is represented by bars.

The change in absorbance at 5 minutes was used to calculate the rate of reaction (μmol min^{-1}) for each concentration of elastase substrate (see Table 47). Five minutes was chosen as the time point, as this was within the linear section of the graph. The concentration of product was calculated using Beer-Lambert Law, as discussed above.

Table 47: Calculations for the rate of reaction (μmol min^{-1}) of HNE with substrate.

[S] mM	Change in relative absorbance	[P] after 5 minutes (mM)	P (μmol)	Rate of reaction (μmol min^{-1})
0	0	0.000	0.000	0.000
0.015	0.004	0.000	0.064	0.013
0.03	0.005	0.000	0.080	0.016
0.06	0.016	0.001	0.255	0.051
0.125	0.019	0.002	0.302	0.060
0.25	0.021	0.002	0.334	0.067
0.5	0.027	0.002	0.430	0.086

The values in Table 47 were used to make an XY plot of substrate concentration against rate of reaction (Figure 49). Michaelis-Menten non-linear regression was applied to calculate the V_{max} and K_m values (see Table 48).
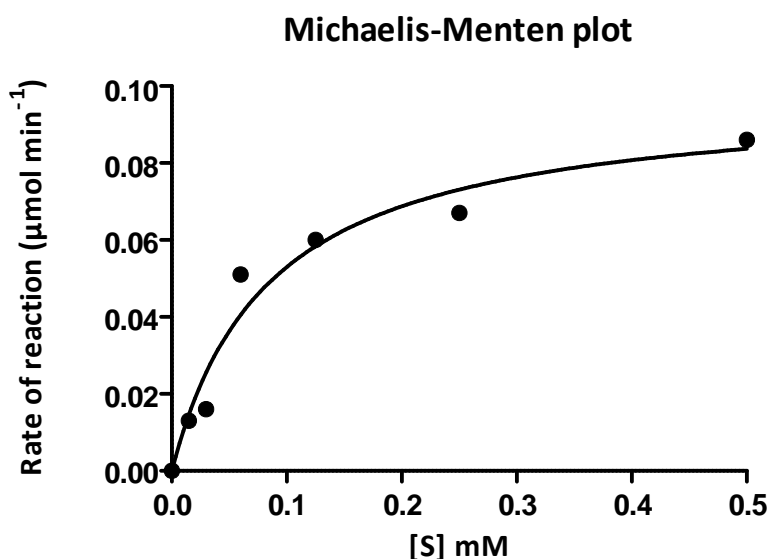
Figure 49: Michaelis-Menten HNE plot, showing the relationship between substrate concentration and rate of reaction. The curve was fitted by nonlinear regression in GraphPad Prism using the Michaelis-Menten settings. R squared value = 0.9627.

Table 48: Results of nonlinear regression analysis of data from Table 47 and Figure 49.

	Best fit values	Standard error	95% Confidence intervals
Vmax	0.098	0.010	0.072 to 0.124
Km	0.089	0.026	0.019 to 0.151

Using nonlinear regression analysis, the K_m for MeOSucc-AAPV-pNA with human neutrophil elastase under the conditions specified (in methods section 2.16) was 89 μM. V_{max} was 0.098 μmol min^{-1}.

In order to confirm that autolysis of the substrate was not responsible for the increase in absorbance, absorbance for the substrate range 0.015 – 0.5 mM was monitored for 30 minutes with no enzyme present. The change in absorbance for each substrate concentration was equal to less than 1 % of the absorbance of the respective substrate standard.

HNE was then tested with the addition of ecotin (Figure 50), with absorbance monitored as before. See Appendix 27 for absorbance readings.
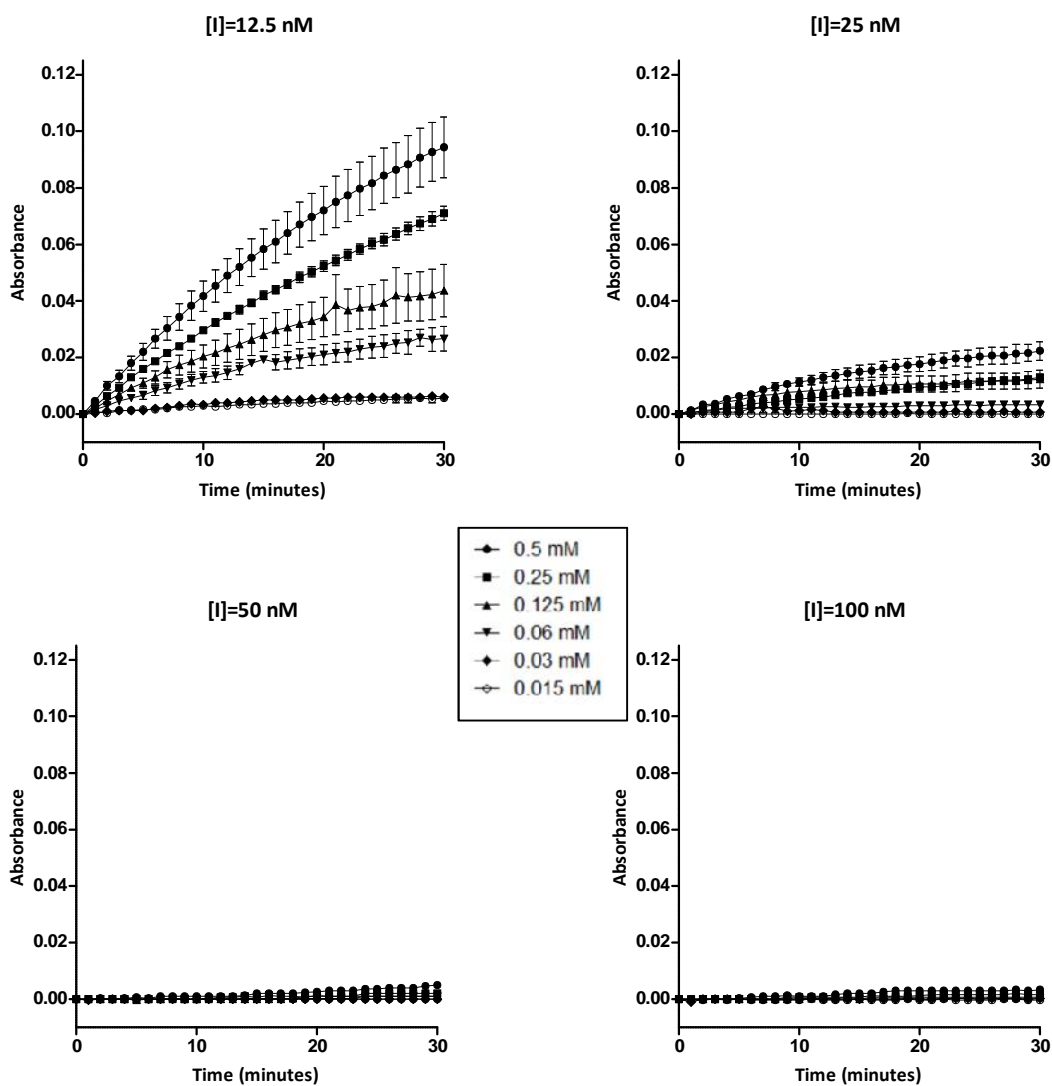
Figure 50: XY plots of absorbance monitored over 30 minutes in the presence of HNE (17 nM), substrate (0.015 − 0.5 mM) and inhibitor (12.5 − 100 nM). Enzyme and inhibitor were incubated for an hour at room temperature to equilibrate before addition of substrate.

Morrison K_i nonlinear regression was applied to the data. The ratio of inhibited to uninhibited rate of reaction was plotted against the inhibitor concentration for each substrate concentration (Figure 51). The Morrison equation was used as part of the GraphPad Prism package to determine the estimated inhibition constant (K_i) (Table 49).
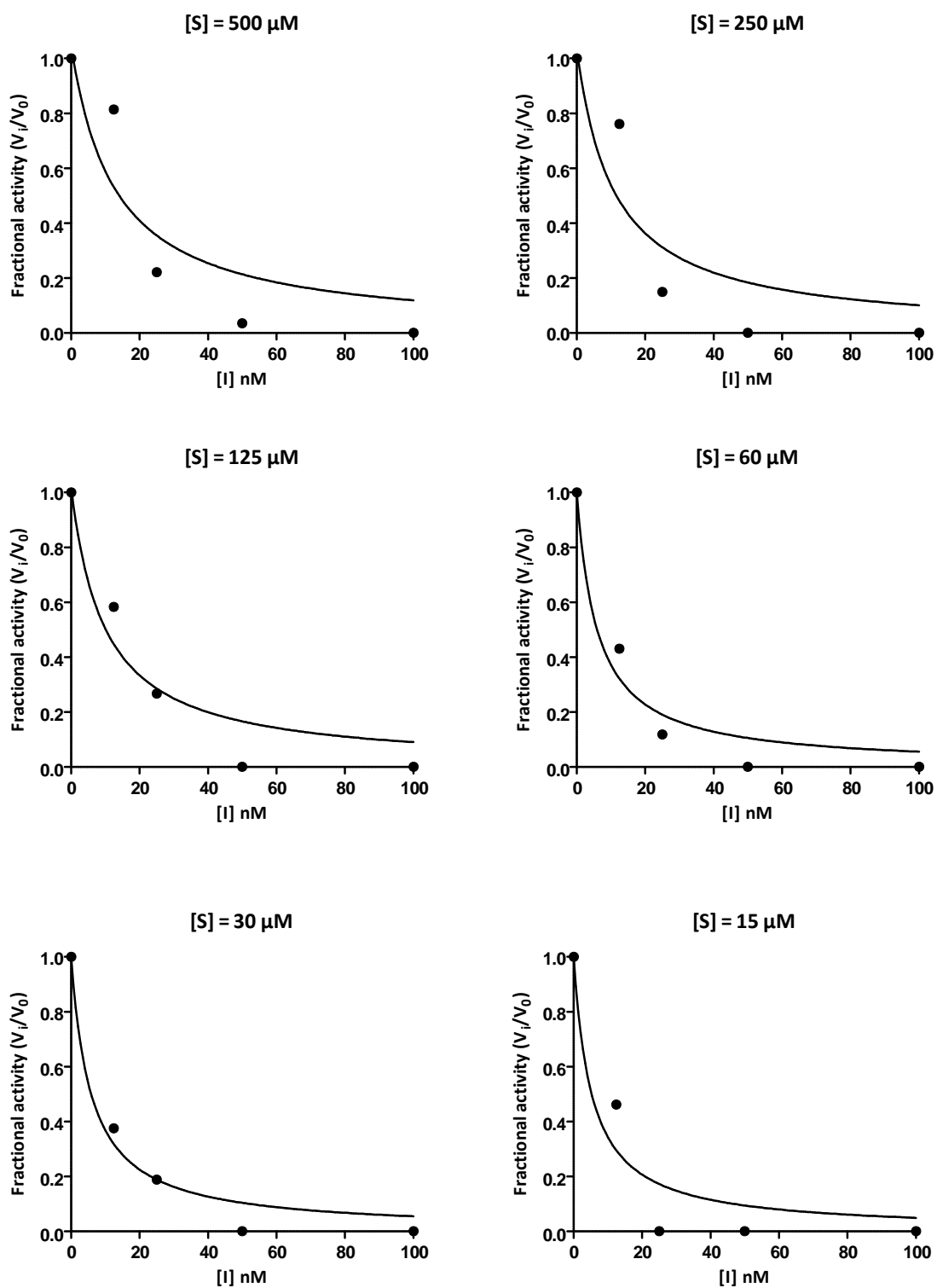
Figure 51: Fractional activity of HNE derived using Morrison K_i nonlinear regression.

Table 49: K_i values determined using Morrison tight binding kinetics equation

[S] µM	Ki	SE	95 % Confidence Intervals	R squared
15	4.404	2.619	-3.930 to 12.74	0.9120
30	**4.310**	**1.220**	**0.4290 to 8.192**	**0.9751**
60	3.486	1.330	-0.7465 to 7.718	0.9560
125	4.042	1.721	-1.434 to 9.518	0.9241
250	2.818	1.912	-3.267 to 8.903	0.8269
500	1.932	1.250	-2.047 to 5.911	0.8285

The Morrison K_i of *F. necrophorum* ecotin with human neutrophil elastase had the highest R squared value and the best visual fit at 30 µM, and was determined to be 4 nM. This demonstrates that the recombinant *F. necrophorum* ecotin inhibits HNE to a greater extent than it inhibits HPK (Morrison K_i of 4 nM compared to 26 nM, for HNE and HPK, respectively).

6.2.5 Plasma clotting assays

To test how the recombinant ecotin would affect the overall combination of clotting factors in human blood, three assays were used to measure clotting times; ecotin final concentrations ranged from 0 – 2 µM. Platelet-poor plasma was made from human donor blood (methods section 2.12.4). Thrombin time (TT) assay was used to monitor the conversion of fibrinogen to fibrin. Prothrombin time (PT) assay was used to monitor any abnormalities with the functioning of Factors II, V, VII, X and fibrinogen; it measures the extrinsic and common pathways. The activated partial thromboplastin time (APTT) assay measures the intrinsic and common pathways, and is sensitive to Factors II, V, VIII, IX, X, XI, XII and fibrinogen. Using a combination of these tests can be very suggestive to which factors are affected (Bain *et al.*, 2012).

Blood was used from three different donors, with each experiment being carried out on a different day. Platelet-poor plasma was separated and normal clotting times were established. The range of

inhibitor concentrations were tested, with a follow up test of uninhibited clotting, to confirm that the prolongation of clotting was not due to time bias. Each measurement was recorded in duplicate and the mean taken. The results of the clotting prolongation are shown in Figure 52; points represent the mean of the three sets of results. At 2 µM, ecotin caused a 1.1-fold increase in clotting time for TT, 7-fold increase for PT and 40-fold increase for APTT. A process of elimination can be used to determine the factors that may be inhibited, based on the factors measured in each test. See Appendix 28 for full clotting prolongation results.
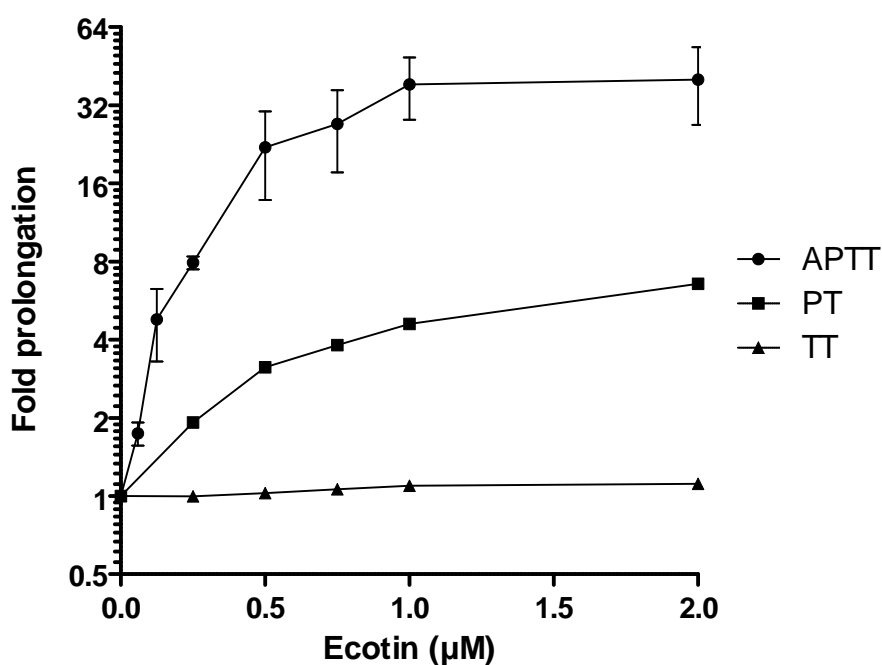
Figure 52: Fold prolongation of human donor plasma clotting tested with a range of ecotin concentrations. The y axis is shown as a log2 scale, however all data points have been unaltered.

6.3 Discussion

Ecotin has not previously been described in *F. necrophorum*. Strains JCM 3718, JCM 3724 and ARU 01 were shown to contain the *ecotin* gene by mining the sequence data and by PCR. It was hypothesised in section 6.1 that *F. necrophorum ecotin* would be universal within the strain collection. When tested using PCR methods, there was amplification of the *ecotin* gene in the 25 strains in the clinical collection. The protein sequences of strains JCM 3718, JCM 3724 and ARU 01 were searched in the Pfam database and when compared to the ecotin consensus sequence they had E values of 1.3e-32 and 2.3e-32 for *Fnf* and *Fnn* sequences, respectively, which indicate a very small chance that the sequence matched that of ecotin by chance.

When the DNA sequences were aligned, the two *Fnf* strains had identical sequences, whereas the *Fnn* sequence had two single point mutations. The predicted translated protein sequence resulted in a single amino acid difference between the two subspecies. This single amino acid difference occurs immediately before the region of the protein considered to be the secondary substrate binding site. The *Fnn* gene contains a lysine in place of the glutamic acid present in the *Fnf* genes. Lysine is a positively charged, basic amino acid, whereas glutamic acid is a negatively charged, acidic amino acid. This could have an impact on the binding abilities of the two ecotin proteins, although as demonstrated by Pál *et al.* (1994), the sequence can contain some variation and still be functional. The hidden Markov model ecotin sequence used for the Pfam alignments contains a glutamic acid in this position. As there was not sufficient funding to produce and test the activity of two recombinant ecotins, only the activity of the *Fnf* ecotin was assessed.

The presence of the signal peptide at the start of the protein sequence is consistent with the expectation that this gene is found in the periplasm of the bacteria, as opposed to the cytoplasm. The location of the protein in the periplasm, as found with the *E. coli* ecotin by Chung *et al.* (1983), is suggestive of the protein being a protective inhibitor. If the inhibitor was secreted then it would appear more likely that the aim is to target clotting factors. It is therefore possible that the inhibition of clotting is a coincidental by product of this serine protease inhibitor.

It had been hypothesised that protein model predictions would demonstrate a structure similar to that of *E. coli*, with both primary and secondary binding sites present. The Pfam domain results of the protein locates the presence of the primary substrate binding site and the two secondary substrate binding sites. While the substrate binding sites are not 100 % homologous, they are well conserved. As discussed previously, the sequence of the primary binding site can include mutations and still result in a functional inhibitor (Pál *et al.*, 1994), therefore it is not essential that the sequence is highly conserved and some sequence variation between species can be expected. The protein model prediction shows the primary and secondary binding sites are located in the expected positions, based on crystalised structures and *in silico* experiments. This therefore satisfies the hypothesis.

Expression of the recombinant ecotin protein by *E. coli* BL21(DE3) cells was indicated by the presence of a band at the expected molecular weight by Tricine-SDS PAGE analysis. At zero hours post-induction with IPTG, the BL21(DE3) strain both with and without the plasmid appear the same on the gel. At one, two and three hours post-induction there is a clear band for the strain containing the plasmid that is lacking for the strain not transformed with the plasmid. The negative control was chosen to be strain BL21(DE3) without the plasmid, as opposed to the strain with the plasmid but without IPTG, since there is anecdotal evidence of the T7 promoter being 'leaky', and thus inducing protein production in the absence of IPTG. As *E. coli* expresses an ecotin protein itself there is already likely to be a band in the region, so the plasmid was not used in the negative control to provide more clarity. The native *E. coli* ecotin does not contain a histidine tag, therefore was unlikely to contaminate the final recombinant ecotin product during purification.

A Bradford assay was used for protein quantification, as opposed to Nanodrop technology, as the Nanodrop is more reliable for proteins with a higher concentration. The Bradford assay works well for lower concentrations of protein and appeared more reproducible during initial optimisation of the assay. The Nanodrop was first used to provide an indication of the concentration before the protein was diluted for use in the Bradford assay.

The Tricine-SDS PAGE analysis of the purification steps (Figure 43) shows that native conditions were suitable as opposed to denaturing conditions. Sample fractions were run on the gel to demonstrate the location of the ecotin protein. If the protein had been present in the total fraction but not in the soluble fraction, this would have indicated that the protein was present in inclusion bodies within the cytoplasm, which would have required denaturing the protein and then refolding. This often occurs if the protein is expressed in too high a concentration, or if it is toxic to the cell. Under normal circumstances in *E. coli*, ecotin is expressed and the signal peptide causes the protein to be transported to the periplasm. For the recombinant ecotin, the signal peptide was removed and replaced with a histidine tag, therefore the protein would have remained in the cytoplasm. Fortunately this was not problematic as the inhibitor is not toxic to *E. coli*, however, the protein was not expressed in concentrations as high as had been anticipated.

A considerable amount of the recombinant protein appears to be present in the flow through (Figure 43, lane 4). This unfortunately diminishes the final concentration. The tag was a deca-histidine tag, as opposed to the frequently used hexa-histidine tag, and therefore should have bound strongly. The tag may have been partially obscured by the folding of the protein, reducing the strength of the interaction with the column, and this may be why some of the protein was lost prior to elution.

The washing steps during the purification were continued until protein concentration within the wash fraction was 0 mg/ml. The gel image supports this and the lane is clear for the final wash fraction. Despite this there are two faint bands of a high molecular weight present within the elution fractions, suggesting that the eluted protein isn't completely pure. It is unlikely that the small percentage of contaminating protein will inhibit serine proteases. It is likely, however, to have a minimal effect on the measurement of the protein concentration, causing the Bradford assay to measure a slightly higher concentration of inhibitor than is actually present.

The choice of methods used to assess the inhibitory properties of the *F. necrophorum* recombinant ecotin were designed to include a variety of tests on a limited budget. Ecotin from *E. coli* has already been shown to potently inhibit Factors Xa and XIIa, plasma kallikrein and human leukocyte elastase.

185

Factors X and XII and their respective substrates are very expensive and as kallikrein is part of the clotting cascade and was considered a likely target for ecotin, it was chosen due to the lower cost. Human neutrophil elastase, a protease that degrades bacteria, was chosen as it was anticipated that ecotin may act as a defence mechanism against this type of immune system protease released at inflammation sites. It is not part of the clotting cascade. The thrombin time, prothrombin time and activated partial thromboplastin time tests are standardised methods for testing clotting pathways.

For the enzyme assays, Michaelis-Menten kinetics were established to assess the activity of the enzymes under uninhibited conditions, to confirm appropriate concentrations of enzyme and substrate had been chosen and to allow for different types of data analysis in the event that the Morrison K_i for tight binding inhibitors had not been appropriate. Alternative options included non-linear regression for competitive inhibition, or reciprocal plots, which tend to amplify any errors in the data.

It had been hypothesised that the recombinant *F. necrophorum* ecotin protein would inhibit human neutrophil elastase and human plasma kallikrein. The data in section 6.2.4 demonstrates this to be the case. The set of four inhibition plots showing absorbance against time for HNE (Figure 50) shows stronger inhibition than for the equivalent HPK graphs plotting RFU against time (Figure 46). This is supported by the Morrison K_i values obtained for each enzyme, with 4 nM for HNE and 26 nM for kallikrein. For comparison, *E. coli* ecotin was reported as having Morrison K_i values of 54 pM for Factor Xa, 55 pM for human leukocyte elastase, 89 pM for Factor XIIa and 163 pM for plasma kallikrein (Ulmer *et al.*, 1995). This does appear to show *E. coli* ecotin as considerably more active against these serine proteases (73X more active against HNE and 160X more active against HPK), however without having performed the tests in the same laboratory under the same conditions it is difficult to speculate. Differences in the accuracy of measuring the concentration of ecotin as well as differences in the activity of the enzymes purchased could have an effect on the results. For example, the K_m values reported for kallikrein and human leukocyte elastase with their corresponding substrates were 165 µM and 150µM, respectively (Ulmer *et al.*, 1995), whereas

under the conditions used in this work, they were 252 µM and 89 µM for kallikrein and HNE, respectively. Variations in the results from this work compared to that of Ulmer *et al.* (1995) may also be due to the use of different substrates. Ideally, *E. coli* ecotin would have been tested alongside as a comparator, but due to the associated cost this was not possible.

The Morrison K_i data for the HPK assay (Figure 47) shows that the curve is flatter with more substrate present, which suggests that with a high enough concentration of substrate the inhibition can be overcome; this is typical of competitive inhibition. This supports the previous assumption that ecotin is a competitive inhibitor and was further supported by a Lineweaver Burk plot (Appendix 26) as discussed in section 6.2.4.1.

The Morrison K_i plots for the HNE data did not fit as closely as for the HPK data. This was reflected in the lower R squared values of the curves. The HNE fractional activity ratios rapidly dropped to zero, requiring a steep curve; a lower ecotin concentration range may have resolved this. The K_i value with the best fit (as measured by the R squared value) was reported, however the mean of all the K_i results were shown for comparison. The results did not vary greatly.

At the highest ecotin concentration tested (2 µM), the thrombin time clotting assay was prolonged only 1.1-fold, suggesting that ecotin had very little effect on the conversion of fibrinogen to fibrin. Prothrombin time was prolonged 7-fold, suggesting ecotin inhibited some of Factors II, V, VII and X. APTT was prolonged up to 40-fold, suggesting that ecotin potently inhibited some or all of Factors II, V, VIII, IX, X, XI and XII. Both PT and APTT are also measures of the conversion of fibrinogen to fibrin as part of the common pathway, but as the TT was not prolonged this can be ruled out as being a target for ecotin.

E. coli ecotin is known to be a potent inhibitor of Factors Xa, XIIa, kallikrein, human leukocyte elastase, and bovine trypsin and chymotrypsin, but not thrombin, Factors VIIa and XIa, activated protein C, plasmin, or tissue plasminogen activator (Seymour *et al.*, 1994; Ulmer *et al.*, 1995). *E. coli* ecotin and *F. necrophorum* ecotin are likely to have similar binding sites, as suggested by the protein model prediction in Figure 39, therefore information about *E. coli* ecotin inhibition can be used to

hypothesise which factors are inhibited by *F. necrophorum* ecotin. Firstly, as Factors II and V are present in both PT and APTT tests, these are not targets of potent inhibition by *F. necrophorum* ecotin, due to the variation in prolongation times between the tests. Factor VIIa was found not to be inhibited by *E. coli* ecotin, as discussed above, and no results were found regarding inhibition of Factor VIII or IX. Factors Xa and XIIa have been shown to be potently inhibited by *E. coli* ecotin, as also discussed above, while Factor XIa was not effectively inhibited. Consequently, it is likely that PT was prolonged by *F. necrophorum* ecotin due to the potent inhibition of Factor X, while APTT was prolonged to a much greater extent due to the combination of potent inhibition of Factor X, Factor XII and kallikrein, which is also part of the intrinsic pathway and was inhibited in enzyme assays. It is conceivable that *F. necrophorum* ecotin may also interfere with other factors in the intrinsic pathway that were not the focus of investigation.

The role, if any, that *F. necrophorum* ecotin plays in disseminated intravascular coagulation during cases of bacteraemia and Lemierre's syndrome remains unclear. *E. coli* is present in the gut of humans and is therefore highly likely to encounter many proteases against which it will need to protect itself. The need for *F. necrophorum* to inhibit serine proteases is less obvious. While the organism is present in the blood stream and at inflammatory sites it is likely to encounter neutrophils and the elastases they secrete to degrade pathogens. Having ecotin present in the periplasm may help protect the bacterial cell. It may also be released from dying *F. necrophorum* cells and subsequently inhibit factors of the intrinsic clotting system, thereby inhibiting chemotaxis of further neutrophils and blood clotting.

Chapter 7:

Discussion, conclusions and future work

7.1 Discussion

The aim of this project was to use next generation sequencing and subsequent bioinformatic analysis, and *in vitro* work to investigate the *F. necrophorum* genome and the virulence determinants that it contains. The type strains of the two subspecies, along with a clinical strain known to have caused Lemierre's syndrome in a patient, were collected and commercially sequenced and assembled. Draft genome sequences of JCM 3718 (*Fnn*), JCM 3724 (*Fnf*) and ARU 01 (*Fnf*) were found to have sizes of 2.3 Mb, 2.13 Mb, and 2.06 Mb, respectively. This is closely aligned with the findings of Calcutt *et al.* (2014), who published the draft genome of *Fnf* strain B35 with a reported size of 2.09 Mb. The genome of *F. nucleatum*, the most closely-related completed genome, was found to be 2.17 Mb (Kapatral *et al.*, 2002).

At the start of this project, in 2011, there was no available *F. necrophorum* genome sequence, therefore many of the genes the organism contained were unknown. With an increasing emphasis on the use of genomic data, it was clear that *F. necrophorum* needed to be sequenced. Since then, genome sequencing has dramatically increased in popularity and accessibility and the draft genomes of eleven *F. necrophorum* strains have been deposited in the NCBI database. The tools used for associated data analysis have also been rapidly changing over recent years. More extensive genome sequencing projects are being carried out, such as the Human Microbiome Project and the Genomics England 100,000 human genomes project. The increase in investment associated with these projects leads to more cost effective technologies and more advanced data analysis. The number of bacterial and archaeal genomes sequenced and submitted to the NCBI in 2011 was fewer than 2000. In 2014, more than 14,000 genomes were sequenced and submitted. Not only are the

number of genomes increasing, but also the diversity, due to projects such as the Genomic Encyclopedia of Bacteria and Archaea, which aims to sequence organisms from diverse branches of bacteria and archaea (Land *et al.*, 2015).

The genomic data contributed by this project represents the most complete assembly for strain JCM 3724/ATCC 51357 and the first genome sequence for strain JCM 3718/ATCC 25286. These strains are important as they are the type strains for *F. necrophorum* subsp. *funduliforme* and *F. necrophorum* subsp. *necrophorum*, respectively (Shinjo *et al.*, 1991). These type strains are used as references during various types of characterisation experiments and provide a method of external quality control. Strain ARU 01, also sequenced as part of this project is the first to be recognised as originating from human disease. This is important as there may be differences within genes that lead to increased virulence in some strains. Comparison tools were used to investigate this. If the genome is to be studied for virulence factors associated with disease, then isolates should therefore be used from disease sources. It would be beneficial for the study of *F. necrophorum* infections in humans if a type strain of *Fnf* isolated from human disease origin could be assigned as a type strain.

The *F. necrophorum* genome sequence data produced as part of this project provides a great deal of information for researchers to investigate *F. necrophorum* further and make comparisons to other organisms. With a greater availability of genomic data, more extensive phylogenetic analysis can be carried out.

The genomes of strains JCM 3718, JCM 3724 and ARU 01 were annotated using a semi-automatic approach. The use of the xBASE pipeline allowed for integration of various annotation packages, so that open reading frames, tRNAs and rRNAs were all annotated. BLASTp was used to manually investigate any ORFs that had not been assigned a function. This uncovered several interesting genes, such as the leukotoxin, which was expected, and ecotin, which was not previously known to be present in *F. necrophorum*. This appears to be the first *F. necrophorum* genome annotation to have been thoroughly manually curated.

The number of ORFs predicted in each draft genome in Table 28 ranges from 1,942 – 2,599. The average contig size was plotted against the estimated number of genes and non-linear regression was applied to demonstrate the trend of decreasing gene numbers with increasing average contig size (Figure 8). The curve was estimated to plateau at 1,950, suggesting this could be a more accurate prediction of the number of ORFs in the *F. necrophorum* genome.

The majority of the *F. necrophorum* genomes submitted to the NCBI database do not contain information regarding the subspecies identification. This therefore means that the data cannot be used for any subspecies comparisons. The use of a phylogenetic tree comprised of *gyraseB* sequences in Figure 9 provided an effective tool for subspeciation where *in vitro* work was not possible. The subspecies of the three strains sequenced during this project and two from the online database were known. Strain JCM 3724/ATCC 51357 was included only once. The subspecies identification of the remaining eight strains was deduced by the distinct grouping of the strains. When the subspecies identification result was paired with the origin of the strain, the results appeared to support the understanding that *Fnn* causes animal infections and that *Fnf* can be found in both animals and humans.

While 454 sequencing was an appropriate choice for *de novo* genome sequencing in 2012, now a more appropriate choice would be Pacific Biosciences for high quality *de novo* sequencing, or Illumina for resequencing. There is still not a complete genome for *F. necrophorum*, however, the sequence data for strain JCM 3724/ATCC 51357 contains only 35 contigs, which is a feasible number of gaps to bridge by PCR and Sanger sequencing.

To build on the genomics work carried out in this project, future work would involve additional bioinformatic analysis of the three genomes sequenced during this project as well as the sequence data from the *F. necrophorum* genomes in Table 28. A pangenomic analysis would be carried out to investigate the range of genes that are present within *F. necrophorum* genomes. The minimum core genome would also be analysed to determine which genes are shared by all the isolates. These two types of analyses would show which genes are shared and which are unique within *F. necrophorum* strains. The sequence output from the minimum core genome analysis would also be used to

produce a phylogenetic tree to reveal the diversity between the strains. This data set would also allow for the investigation of single nucleotide polymorphisms. A comparison of *F. necrophorum* core and accessory genes to those of related commensals and pathogens would be of great value, and may reveal pathogenic mechanisms utilised by *F. necrophorum*.

As discussed above, the number of contigs for strain JCM 3724/ATCC 51357 has been reduced to 35, following the merging of data sets in section 4.2.4. The additional genomes available are likely to contain information that may indicate the order of the contigs and the size of the gaps that need to be sequenced by Sanger method.

An additional use for the genomic data available from online databases is that these sequences can be used to design PCR primers for virulence determinants where the sequence wasn't available in the sequence data generated as part of this project. For example, the FadA gene, which is briefly mentioned below.

Genes found within the genomes of strains JCM 3718, JCM 3724 and ARU 01 that are thought to be associated with virulence potential include a leukotoxin, haemolysin, haemagglutinin, haemin receptors and a serine protease inhibitor known as ecotin. The leukotoxin was expected to be present due to the understanding that it is the organism's primary virulence factor (Tan *et al.*, 1994a). Ecotin has not previously been described in *F. necrophorum* and was therefore a novel and interesting find. It was unknown whether the FadA adhesin would be present, as this has been found to be present in several *Fusobacterium* species (Han *et al.*, 2005). It was not initially found within the *F. necrophorum* genome data, and was not amplified by PCR targeting the gene. The FadA adhesin gene was later found in *Fnn* genomes and would therefore be worth investigating as part of future work as this could contribute to the varying levels of pathogenicity between *Fnn* and *Fnf* strains.

The haemolysin and an adhesin that were annotated in the genomes contained motifs and domains associated with type V secretion systems. These systems are known to be associated with transporting large virulence proteins across the outer membrane of the bacterium, such as

haemagglutinins (Hodak and Jacob-Dubuisson, 2007). Strains JCM 3718 and JCM 3724 contain conserved motifs within the haemolysin gene that suggest the protein is secreted via a two-partner secretion system (Figure 13). These TPS motifs have not previously been described in *F. necrophorum*. Strain ARU 01 contains TPS motifs that are less well conserved and are in the reverse order. Not all TPS proteins need both motifs (Jacob-Dubuisson *et al.*, 2001), so secretion by this system may be unaffected. The YadA-like adhesin that was also found within the annotation data has head, stalk and anchor regions that strongly suggest it acts as an autotransporter. Future work on these secretion systems would include cloning the *YadA* adhesin gene and carrying out an adhesion assay. Also, cloning the *haemolysin* gene, both with and without the TPS domain, would be of interest, before carrying out a haemolysis assay with either human or animal whole blood.

There are limitations on what may be concluded by using bioinformatic tools alone, therefore to investigate the leukotoxin and ecotin further, *in vitro* techniques were used. For the leukotoxin chapter the aim was to investigate the leukotoxin sequence in strains JCM 3718, JCM 3724 and ARU 01 and also confirm cytotoxic activity of the strains. The HL-60 cell line was investigated to determine whether it would be an appropriate substitute for human white blood cells. The prevalence of the *leukotoxin* gene within the clinical strain collection was also tested using a range of PCR primers.

One copy of the leukotoxin operon, containing genes *lktB, lktA* and *lktC*, was found in each of strains JCM 3718, JCM 3724 and ARU 01. In order to ensure the sequence data was of a high quality, the next generation sequence data for the genes was resequenced by Sanger sequencing to remove the frameshift errors in the sequence that were causing the predicted protein sequence to terminate prematurely.

The lktB protein sequences were found to be highly conserved between the *Fnf* strains, which had a similarity of 99.8 %. The *Fnn* JCM 3718 lktB lacked this high level of similarity due to being shorter than the *Fnf* strains. The protein is 336 amino acids long compared to 447 due to a later start codon in the sequence. The lktB protein sequence contained several putative conserved domains, including a polypeptide transport associated (POTRA) domain and a haemolysin

activation/secretion protein. The POTRA domain was truncated at the N terminus in strains JCM 3724 and ARU 01 and was missing completely from strain JCM 3718. This domain has previously been described by Tadepalli *et al.* (2008b) in relation to leukotoxin secretion, however the haemolysin activation/secretion protein domain is present in all three strains and has a more significant E value. The secretion mechanism of the leukotoxin remains unknown, however with several domains within the lktB protein linked to secretion, it seems likely that this protein is involved, although potentially not via the POTRA domain, which had been expected.

The lktA protein sequences of the bovine JCM strains were shown to have a greater level of similarity than between the *Fnf* strains by using an Artemis Comparison Tool alignment (Figure 19). This was unexpected as it was thought that the two *Fnf* strains would share the most homology. It therefore seems that the leukotoxin sequence may be dependent on the source of the isolate more so than the subspecies identification. Each sequence may have evolved to be more effective against its host. The work by Narayanan *et al.* (2001b) discussed in section 5.1 implies that the active toxin region is within the first 3.5 kb of the *lktA* gene. The ACT alignment demonstrates that this region is well conserved between the two bovine strains, with similarities of 94 % between 1,000 bp and 4,000 bp. This not so well conserved between the two *Fnf* strains, which show several short regions of similarity within this window of the coding sequences with similarities of approximately 80 – 90 %.

A phylogenetic tree of the three *lktA* gene sequences with a further nine *lktA* gene sequences downloaded from the NCBI database (Figure 20) demonstrated how the sequences clustered with those from the same origin more than with those of the same subspecies. For example, strains JCM 3724 and B35, both *Fnf* strains of bovine origin, clustered more closely with the remaining bovine strains that were *Fnn* rather than with the human *Fnf* strains. The only animal strain to group with the human isolates is one isolated from a rumen microbiome, rather than a disease source. This suggests that this strain may be less virulent than the other animal strains, due to its grouping with the human isolates and it being present as a commensal rather than a pathogen.

The lktC protein sequences of the three sequenced strains were highly conserved. The top BLASTp result for the lktC of each strain was *Fusobacterium necrophorum* histidine kinase for JCM 3718 and JCM 3724, with E values of 2e-99 and 3e-94, respectively. For ARU 01 the top result was a sensory transduction regulator, with an E value of 4e-96. Given that histidine kinases are signal transduction enzymes (Wolanin *et al.*, 2002), and that the result of the Pfam search also resulted in a putative bacterial sensory transduction regulator domain, it seems that lktC has a signalling function.

The prevalence of the *lktA* gene within the clinical strain collection was shown to be 100 % by using five sets of PCR primers targeting the gene. Two sets of primers, lkt1 and lkt2, custom designed for this project, and Ludlam_LT1 (Ludlam *et al.*, 2009a) all had a 100 % amplification rate. Ludlam_LT2 (Ludlam *et al.*, 2009a) had amplification in 89 % of strains and lktA (Tadepalli *et al.*, 2008b) amplified product in 43 % of strains. The prevalence of the leukotoxin has been disputed in the literature based on the results Ludlam *et al.* (2009a) found using their primers. Bennett *et al.* (2010) claimed the leukotoxin was likely to be present in these cases but had been undetected. The reasons for the low amplification rate found by Ludlam *et al.* (2009a) are difficult to speculate on, given that the same primers amplified a product in all strains in this project. The sequences may have been more varied than those included in this project and the genes simply weren't detected, or the strain identification was incorrect and the strains were not *F. necrophorum*. It seems unlikely that the gene was absent in the strains. For future testing of leukotoxin prevalence, primer sets Ludlam_LT1, lkt1 and lkt2 would be considered appropriate for use.

High molecular weight culture supernatant fractions, expected to contain the leukotoxin, from strains JCM 3718, JCM 3724 and ARU 01 all demonstrated a cytotoxic effect against both the HL-60 cell line and human white blood cells (Figures 30 – 33), as measured by flow cytometry. For the HL-60 cytotoxicity assay, a dose-dependent response was demonstrated with high statistical significance between the negative control and the lowest concentration treatment group, as well as varying levels of significance with higher treatment concentrations.

It was anticipated that JCM 3724 would have the weakest cytotoxic effect, due to the understanding that the *funduliforme* subspecies is less virulent. However, this has been explained as being due to

Fnf having a less efficient *leukotoxin* promoter (Tadepalli *et al.*, 2008b). JCM 3718 and ARU 01 were expected to have a stronger cytotoxic effect, as JCM 3718 is a *Fnn* strain, and ARU 01 exhibited high virulence in a patient. The results showed that when the concentrations of the leukotoxin fraction were standardised, JCM 3724 had the greatest cytotoxic effect, followed by that of JCM 3718, the other bovine strain. Strain ARU 01, known to cause Lemierre's syndrome in a patient, had the weakest cytotoxic effect. However, the variations in promoter efficiency may change the order of cytotoxicity. For example, the concentration of high molecular weight protein measured by Bradford assay was higher for strain ARU 01 than for JCM 3724. The increase in production of leukotoxin may therefore result in a greater cytotoxic effect for ARU 01 than for JCM 3724, despite the differences in the standardised results. Given the lack of availability of an antibody for immunoaffinity purification, it cannot be confirmed that the protein fractions contained pure leukotoxin. Future work on the *F. necrophorum* leukotoxin would include carrying out qPCR on the promoter region of these three strains, in order to characterise them more fully. This has been carried out by Tadepalli *et al.* (2008b) on other strains in order to confirm that *Fnn* has a higher transcript level of *lktA* than *Fnf*, although this has not been carried out on the type strains.

It was concluded that the HL-60 cell line was an appropriate choice to use in place of human donor blood due the validation test using human donor blood as a comparison. For the concentration tested (150 µg/ml), the HL-60 cells had a slightly lower percentage viability compared to the human white blood cells, although there was no statistical significance ($P > 0.05$). Using this cell line instead of human WBCs provided more uniform cell samples, and would therefore be considered for use again.

Increasing the understanding of the leukotoxin is an important aspect of studying *F. necrophorum*. It is the primary virulence factor of the organism (Tan *et al.*, 1994a) and has been used in an inactive form as a vaccine for livestock (Saginala *et al.*, 1997). The mode of action and secretion of the leukotoxin are not yet understood. The POTRA and haemolysin activation/secretion protein domain within the lktB protein and the signal peptide found at the start of the lktA protein are strong indicators that this protein is secreted, rather than released upon cell death. The collection of a

leukotoxic supernatant fraction from fresh cultures is further evidence of this. The protein sequence was analysed for type V secretion systems, although none were found.

The leukotoxin work carried out on the JCM reference strains provides high quality sequence data of the leukotoxins of both *F. necrophorum* type strains along with cytotoxicity data. These strains have been well characterised in many other ways but have so far been lacking this data. This work also provides characterisation of the leukotoxin of a clinically important strain of human origin. It suggests that there is no mutation in the sequence that results in enhanced pathogenicity and that if the leukotoxin is responsible for greater virulence, then the promoter region is likely to be responsible, causing a greater rate of transcription.

The aim of the ecotin chapter was to determine the prevalence of the ecotin gene within the strain collection, analyse the sequence using bioinformatic tools, and assess the function of a recombinant ecotin against host serine proteases and the clotting cascade to assess the role of *F. necrophorum* ecotin in the host.

The *ecotin* gene, which has not previously been described in *F. necrophorum*, was found in 100 % of the strains. This was established by mining the BLASTp data of the three sequenced genomes and by PCR of these strains and the 25 isolates in the clinical collection. The sequences obtained from the genome data and by Sanger sequencing the first 12 strains of the clinical collection demonstrate that the gene is highly conserved, with percentage identities of 98.96 % or above between strains.

The sequence was searched for conserved domains by BLASTp and primary and secondary binding sites of ecotin were located. A protein model (Figure 39) of the *F. necrophorum* ecotin monomer predicted the primary and secondary binding sites were exposed on opposite sides of the protein. The two conserved domains of the secondary binding site in the protein sequence were separated by 36 amino acid residues, however the predicted folding of the protein results in those regions forming one binding pocket. *E. coli* ecotin has previously been shown to dimerise in such a way that when the ecotin dimer binds to two target proteases, both monomers bind to both proteases (Yang

et al., 1998). This was not tested for *F. necrophorum* ecotin, however, the 3-dimensional models of the *F. necrophorum* and *E. coli* ecotin monomers are similar, suggesting that the dimerised structure would be similar also, and that the binding sites would be arranged in such a way that each ecotin monomer would bind to each target protease.

The purified recombinant ecotin protein produced in section 6.2.3 acted as a potent inhibitor of human neutrophil elastase during enzyme inhibition assays. Non-linear regression was used to determine the inhibitor constants of ecotin, with a Morrison K_i of 4 nM for HNE. The inhibition was less potent against human plasma kallikrein, with a Morrison K_i of 26 nM. This is consistent with the findings of Ulmer *et al.* (1995), who reported that *E. coli* ecotin was a potent inhibitor of human leukocyte elastase and human plasma kallikrein, with human leukocyte elastase inhibited more strongly.

HNE is involved in bacterial degradation in phagosomes and at inflammatory sites (Korkmaz *et al.*, 2010). The ability for *F. necrophorum* to evade this enzyme would be advantageous for its survival. As this enzyme was inhibited the most potently in this study and in the work by Ulmer *et al.* (1995), it suggests that the primary function of ecotin may be as a defence mechanism. *E. coli* ecotin is known to be located in the periplasm which also supports this idea. If the primary target was the clotting cascade then the inhibitor would likely be secreted, however it is well placed in the periplasm to protect the organism from host proteases. It was established by Chung *et al.* (1983) that *E. coli* ecotin does not inhibit any of its own proteases.

HPK is part of the intrinsic pathway and therefore contributes to coagulation of blood plasma. Along with Factor XIIa, it also attracts neutrophils by chemotaxis (Bain *et al.*, 2012). While this serine protease was not affected as strongly as HNE, it still was susceptible to inhibition by the recombinant *F. necrophorum* ecotin. By inhibiting this enzyme, *F. necrophorum* may be able to reduce the number of neutrophils attracted to the infection site and also have an effect on blood clotting. It is unclear what the benefit of this may be as septic emboli are a mechanism the organism uses to transport around the body of the host (Riordan, 2007). Mild disseminated intravascular

coagulation has also been reported as a complication of Lemierre's syndrome (Hagelskjaer Kristensen and Prag, 2000), which again contradicts the inhibition of coagulation.

Human donor plasma treated with a range of ecotin concentrations was tested for clotting prolongation using tests: thrombin time, prothrombin time and activated partial thromboplastin time. At 2 µM, ecotin caused a 1.1-fold increase in clotting time for TT, 7-fold increase for PT and 40-fold increase for APTT. The thrombin time assay measures the effect of the conversion of fibrinogen to fibrin. The prolongation of 1.1-fold suggests that ecotin had very little effect on this reaction. Prothrombin time, which measures the effect on the extrinsic and common pathways was prolonged 7-fold, suggesting that ecotin inhibits some of Factors II, V, VII and X, but not all. APTT, which measures the effect on the intrinsic and common pathways, including kallikrein, high molecular weight kininogen, and Factors II, V, VIII, IX, X, XI and XII was prolonged 40-fold, suggesting that ecotin potently inhibits some or all of these factors. *E. coli* ecotin has been shown to be a potent inhibitor of Factors Xa, XIIa, kallikrein, human leukocyte elastase, and bovine trypsin and chymotrypsin, but not an effective inhibitor of thrombin, Factors VIIa and XIa, activated protein C, plasmin, or tissue plasminogen activator (Seymour *et al.*, 1994; Ulmer *et al.*, 1995). *E. coli* ecotin and *F. necrophorum* ecotin are likely to have similar binding sites, therefore information about *E. coli* ecotin inhibition can be applied to *F. necrophorum* ecotin to hypothesise which individual factors are inhibited in each test. Factors II and V are present in the both PT and APTT tests as part of the common pathway. These are unlikely to be targets of potent inhibition, due to the variation in prolongation times between the tests. Factor VIIa was found not to be inhibited by *E. coli* ecotin, as discussed above. No results were found regarding inhibition of Factor VIII or IX. Factor XIa was not effectively inhibited by *E. coli* ecotin, while Factors Xa and XIIa were potently inhibited, as also discussed above. It is therefore likely that PT was prolonged due to the inhibition of Factor X, while APTT was prolonged to a much greater extent due to inhibition of Factors X and XII and kallikrein.

While *F. necrophorum* ecotin is present in the host blood stream it is likely to encounter neutrophils and the elastases they secrete to degrade pathogens. Having ecotin present in the periplasm can

help protect the cell. It may possibly also be released from dying *F. necrophorum* cells and subsequently inhibit factors of the clotting cascade.

This work has resulted in the production of a functional recombinant ecotin protein that can be purified without the need to denature the protein. Future work on the *F. necrophorum* ecotin would involve testing additional clotting factors, and using *E.coli* ecotin as a comparator. It would also be of interest to determine the concentration of the *F. necrophorum* native inhibitor, to investigate the levels of ecotin that occur naturally and the effect it has on plasma clotting time.

7.2 Conclusions

Projects such as this aim to decrease the bias in studies of microorganisms, and focus on researching organisms that are not considered model strains. The leukotoxin and ecotin findings can be used to build up an understanding of how the organism behaves during infection. The organism has been reported to be increasing in incidence, thereby becoming more clinically relevant and important to study. The genome sequence data will be added to the *F. necrophorum* data available online, and enable further comparative study between strains. The data also provides genome sequence data for the type strain of both subspecies of *F. necrophorum*. A greater understanding of the genome of an organism can lead to more advanced diagnostic tests being introduced. Also, if human microbiome testing is to be incorporated into the field of personalised medicine in order to treat patients, then the genomes of organisms being tested for must be well understood. The major findings of this project are as follows:

1. The *F. necrophorum* genome was found to be approximately 2.1 – 2.3 Mb in size, with an estimated 1,950 ORFs and includes genes for a leukotoxin, ecotin, haemolysin, haemagglutinin, haemin receptor, adhesin and type Vb and Vc secretion systems.

2. The genomic data contributed by this project represents the most complete assembly for strain JCM 3724/ATCC 51357, the first genome sequence for strain JCM 3718/ATCC 25286 and the first *F. necrophorum* genome of a strain recognised as originating from human disease.

3. The subspeciation of the available *F. necrophorum* genomes online enables more meaningful comparisons and conclusions to be made with the data and supports the understanding that *Fnn* strains cause animal infections and that *Fnf* can be found in both animals and humans.

4. The *leukotoxin* operon was found to be universal within the *F. necrophorum* strain collection. Analysis of the sequence data found the lktB protein is highly conserved and contains domains associated with secretion. The lktA protein, containing a signal peptide, appears to be the most conserved among strains of the same host origin, and the lktC protein may have a signalling function.

5. Strains JCM 3718, JCM 3724 and ARU 01 were cytotoxic to HL-60 cells in a dose-dependent manner, and were shown to have a similar cytotoxic effect on human donor white blood cells. The results suggest that variations in virulence are likely to be due to bovine origin strains having a more efficient promoter for the leukotoxin. This hypothesis is based on the findings of a Bradford assay and the cytotoxicity results being inconsistent with the understanding that *Fnn* strains are more virulent.

6. The *ecotin* gene was found to be present in 100 % of the strain collection and had a highly conserved sequence. A protein prediction model of the *F. necrophorum* ecotin resulted in a structure very similar to that of *E. coli* ecotin.

7. A purified recombinant *F. necrophorum* ecotin protein inhibited human neutrophil elastase and human plasma kallikrein in enzyme inhibition studies and prolonged human plasma clotting times in both the intrinsic and extrinsic pathways.

References

Aagaard, K., Petrosino, J., Keitel, W., Watson, M., Katancik, J., Garcia, N., Patel, S., Cutting, M., Madden, T. & Hamilton, H. 2013. The Human Microbiome Project strategy for comprehensive sampling of the human microbiome and why it matters. *The FASEB Journal,* 27**,** 1012-1022.

Abbott, J., Aanensen, D. M., Rutherford, K., Butcher, S. & Spratt, B. G. 2005. WebACT—an online companion for the Artemis Comparison Tool. *Bioinformatics,* 21**,** 3665-3666.

Abrahamian, F. M. & Goldstein, E. J. 2011. Microbiology of animal bite wound infections. *Clinical Microbiology Reviews,* 24**,** 231-246.

Ali, M. T., Jain, P. & Narayanswami, G. 2003. Lemierre's Syndrome (LS): Sepsis complicating a dental procedure. *CHEST Journal,* 124**,** 300S-a-302S.

Aliyu, S. H., Marriott, R. K., Curran, M. D., Parmar, S., Bentley, N., Brown, N. M., Brazier, J. S. & Ludlam, H. 2004. Real-time PCR investigation into the importance of *Fusobacterium necrophorum* as a cause of acute pharyngitis in general practice. *Journal of Medical Microbiology,* 53**,** 1029–1035.

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

Amedei, A., Cappon, A., Codolo, G., Cabrelle, A., Polenghi, A., Benagiano, M., Tasca, E., Azzurri, A., D'elios, M. M. & Del Prete, G. 2006. The neutrophil-activating protein of Helicobacter pylori promotes Th1 immune responses. *The Journal of clinical investigation,* 116**,** 1092-1101.

Antiabong, J. F., Boardman, W., Smith, I., Brown, M. H., Ball, A. S. & Goodman, A. E. 2013. "Cycliplex PCR" confirmation of *Fusobacterium necrophorum* isolates from captive wallabies: a rapid and accurate approach. *Anaerobe,* 19**,** 44-49.

Appelbaum, P. C., Spangler, S. & Jacobs, M. 1990. Beta-lactamase production and susceptibilities to amoxicillin, amoxicillin-clavulanate, ticarcillin, ticarcillin-clavulanate, cefoxitin, imipenem, and metronidazole of 320 non-*Bacteroides fragilis Bacteroides* isolates and 129 fusobacteria from 28 US centers. *Antimicrobial Agents and Chemotherapy,* 34**,** 1546-1550.

Atcc: The Global Bioresource Center. 2015. *ATCC.org* [Online]. Available: http://www.atcc.org/ 2015].

Aziz, R. K., Bartels, D., Best, A. A., Dejongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M. & Kubal, M. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC genomics,* 9**,** 75.

Bagaitkar, J., Demuth, D. R. & Scott, D. A. 2008. Tobacco use increases susceptibility to bacterial infection. *Tobacco Induced Diseases,* 4**,** 12.

Bain, B. J., Bates, I., Laffan, M. A. & Lewis, S. M. 2012. *Dacie and Lewis Practical Haematology,* London, Churchill Livingstone.

Ball, S., Siou, G., Wilson, J., Howard, A., Hirst, B. & Hall, J. 2007. Expression and immunolocalisation of antimicrobial peptides within human palatine tonsils. *The Journal of Laryngology & Otology,* 121**,** 973-978.

Bang, B. 1890. Om Aarsagen til local Nekrose. *Maanedsskrift Dyrlæger,* 2**,** 235-259.

Barbieri, J. T. 2009. Exotoxins. Editor-in-Chief: Moselio, Schaechter. *Encyclopedia of Microbiology (Third Edition).* Oxford: Academic Press.

Barnes, L. 2000. *Surgical pathology of the head and neck. Volume 2.,* CRC Press, 404-405.

Barrow, G. & Feltham, R. K. A. 2004. *Cowan and Steel's manual for the identification of medical bacteria*, Cambridge university press, 94-98.

Basar, T., HavlíCek, V. R., Bezoušková, S., Hackett, M. & Sebo, P. 2001. Acylation of lysine 983 is sufficient for toxin activity of *Bordetella pertussis* adenylate cyclase: Substitutions of alanine 140 modulate acylation site selectivity of the toxin acyltransferase CyaC. *Journal of Biological Chemistry,* 276**,** 348-354.

Batch Entrez. 2015. *Ncbi.nlm.nih.gov* [Online]. Available: http://www.ncbi.nlm.nih.gov/sites/batchentrez 2015].

Batty, A. & Wren, M. D. W. 2005. Prevalence of *Fusobacterium necrophorum* and other upper respiratory tract pathogens isolated from throat swabs. *British Journal of Biomedical Science,* 62**,** 66-70.

Batty, A., Wren, M. W. D. & Gal, M. 2005. *Fusobacterium necrophorum* as the cause of recurrent sore throat: comparison of isolates from persistent sore throat syndrome and Lemierre's disease. *Journal of Infection,* 51**,** 299-306.

Belaaouaj, A. A., Kim, K. S. & Shapiro, S. D. 2000. Degradation of outer membrane protein A in *Escherichia coli* killing by neutrophil elastase. *Science,* 289**,** 1185-1187.

Bennett, G., Zhou, H. & Hickford, J. G. H. 2010. Undetected *lktA* genes within *Fusobacterium necrophorum*? *Journal of Medical Microbiology,* 59**,** 499-500.

Bennett, K. W. & Eley, A. 1993. Fusobacteria: new taxonomy and related diseases. *Journal of Medical Microbiology,* 39**,** 246-254.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L. & Bignell, H. R. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature,* 456**,** 53-59.

Biocyc. 2015. *BioCyc Pathway/Genome Database Collection* [Online]. Available: http://biocyc.org/ 2015].

Bird, J., Biggs, T. & King, E. 2014. Controversies in the management of acute tonsillitis: an evidence-based review. *Clinical Otolaryngology,* 39**,** 368-374.

Birnie, G. 1988. The HL60 cell line: A model system for studying human myeloid cell differentiation. *The British Journal of Cancer. Supplement 9***,** 41-45.

Björk, H., Bieber, L., Hedin, K. & Sundqvist, M. 2015. Tonsillar colonisation of *Fusobacterium necrophorum* in patients subjected to tonsillectomy. *BMC Infectious Diseases,* 15**,** 264.

Blast: Basic Local Alignment Search Tool. 2015. *Blast.ncbi.nlm.nih.gov* [Online]. Available: http://blast.ncbi.nlm.nih.gov/Blast.cgi 2015].

Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry,* 72**,** 248-254.

Brazier, J., Goldstein, E., Citron, D. & Ostovari, M. 1990. Fastidious anaerobe agar compared with Wilkins-Chalgren agar, brain heart infusion agar, and brucella agar for susceptibility testing of *Fusobacterium* species. *Antimicrobial Agents and Chemotherapy,* 34**,** 2280-2282.

Brazier, J. S. 2006. Human infections with *Fusobacterium necrophorum*. *Anaerobe,* 12**,** 165-172.

Brazier, J. S., Hall, V., Yusuf, E. & Duerden, B. I. 2002. *Fusobacterium necrophorum* infections in England and Wales 1990-2000. *Journal of Medical Microbiology,* 51**,** 269-272.

Broad Institute. 2009. *Prokaryotic Annotation Pipeline SOP* [Online]. http://hmpdacc.org/doc/sops/reference_genomes/annotation/Broad_SOP_DACC.pdf. [Accessed 07/02/13].

Brook, I. 2005. The role of anaerobic bacteria in tonsillitis. *International Journal of Pediatric Otorhinolaryngology,* 69**,** 9-19.

Brook, I. 2015. Fusobacterial head and neck infections in children. *International Journal of Pediatric Otorhinolaryngology,* 79**,** 953-958.

Brook, I. & Gillmore, J. D. 1996. Enhancement of growth of group A β-hemolytic streptococci in mixed infections with aerobic and anaerobic bacteria. *Clinical Microbiology and Infection,* 1**,** 179-182.

Brook, I. & Walker, R. I. 1986. The relationship between *Fusobacterium* species and other flora in mixed infection. *Journal of Medical Microbiology,* 21**,** 93-100.

Butler, J., Maccallum, I., Kleber, M., Shlyakhter, I. A., Belmonte, M. K., Lander, E. S., Nusbaum, C. & Jaffe, D. B. 2008. ALLPATHS: de novo assembly of whole-genome shotgun microreads. *Genome research,* 18**,** 810-820.

Calcutt, M. J., Foecking, M. F., Nagaraja, T. G. & Stewart, G. C. 2014. Draft genome sequence of *Fusobacterium necrophorum* subsp. *funduliforme* bovine liver abscess isolate B35. *Genome announcements,* 2**,** e00412-14.

Caroff, M. & Karibian, D. 2003. Structure of bacterial lipopolysaccharides. *Carbohydrate Research,* 338**,** 2431-2447.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M.-A., Barrell, B. G. & Parkhill, J. 2005. ACT: the Artemis comparison tool. *Bioinformatics,* 21**,** 3422-3423.

Castellarin, M., Warren, R. L., Freeman, J. D., Dreolini, L., Krzywinski, M., Strauss, J., Barnes, R., Watson, P., Allen-Vercoe, E. & Moore, R. A. 2012. *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Research,* 22**,** 299-306.

Castro, H. C., Monteiro, R. Q., Assafim, M., Loureiro, N. I., Craik, C. & Zingali, R. B. 2006. Ecotin modulates thrombin activity through exosite-2 interactions. *The International Journal of Biochemistry & Cell Biology,* 38**,** 1893-900.

Centor, R. M. 2009. Expand the pharyngitis paradigm for adolescents and young adults. *Annals of Internal Medicine,* 151**,** 812-815.

Centor, R. M., Atkinson, T. P., Ratliff, A. E., Xiao, L., Crabb, D. M., Estrada, C. A., Faircloth, M. B., Oestreich, L., Hatchett, J. & Khalife, W. 2015. The clinical presentation of *Fusobacterium*-positive and Streptococcal-positive pharyngitis in a university health clinic: A cross-sectional study. *Annals of Internal Medicine,* 162**,** 241-247.

Centor, R. M., Geiger, P. & Waites, K. B. 2010. *Fusobacterium necrophorum* bacteremic tonsillitis: 2 cases and a review of the literature. *Anaerobe,* 16**,** 626-628.

Chang, Y., Young, R., Post, D. & Struck, D. K. 1987. Identification and characterization of the *Pasteurella haemolytica* leukotoxin. *Infection and Immunity,* 55**,** 2348-2354.

Chaudhuri, R. R., Loman, N. J., Snyder, L. a. S., Bailey, C. M., Stekel, D. J. & Pallen, M. J. 2008. xBASE2: a comprehensive resource for comparative bacterial genomics. *Nucleic Acids Research,* 36**,** D543-D546.

Chikhi, R. & Medvedev, P. 2014. Informed and automated k-mer size selection for genome assembly. *Bioinformatics,* 30**,** 31-7.

Chirinos, J. A., Lichtstein, D. M., Garcia, J. & Tamariz, L. J. 2002. The evolution of Lemierre syndrome: report of 2 cases and review of the literature. *Medicine,* 81**,** 458-465.

Chow, A. W. 1992. Life-threatening infections of the head and neck. *Clinical Infectious Diseases,* 14**,** 991-1002.

Chung, C. H., Ives, H. E., Almeda, S. & Goldberg, A. L. 1983. Purification from *Escherichia coli* of a periplasmic protein that is a potent inhibitor of pancreatic proteases. *Journal of Biological Chemistry,* 258**,** 11032-8.

Clustal Omega. 2015. *Clustal Omega < Multiple Sequence Alignment < EMBL-EBI* [Online]. Available: http://www.ebi.ac.uk/Tools/msa/clustalo/ 2015].

Colman, R. W., Wachtfogel, Y. T., Kucich, U., Weinbaum, G., Hahn, S., Pixley, R. A., Scott, C. F., De Agostini, A., Burger, D. & Schapira, M. 1985. Effect of cleavage of the heavy chain of human plasma kallikrein on its functional properties. *Blood,* 65**,** 311-318.

Conrads, G., Citron, D. M., Mutters, R., Jang, S. & Goldstein, E. J. 2004. *Fusobacterium canifelinum* sp. nov., from the oral cavity of cats and dogs. *Systematic and Applied Microbiology,* 27**,** 407-413.

Courmont, P. & Cade, A. 1900. Sur une septico-pyohémie de l'homme simulant la peste et causée par un strepto-bacille anaérobie. *Arch. de Méd. Exp.,* 12**,** 393-418.

Culture Collections. 2015. *Phe-culturecollections.org.uk* [Online]. Available: https://www.phe-culturecollections.org.uk/collections/nctc.aspx 2015].

Darling, A. C., Mau, B., Blattner, F. R. & Perna, N. T. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research,* 14**,** 1394-1403.

Desvaux, M., Khan, A., Beatson, S. A., Scott-Tucker, A. & Henderson, I. R. 2005. Protein secretion systems in *Fusobacterium nucleatum*: genomic identification of Type 4 piliation and complete Type V pathways brings new insight into mechanisms of pathogenesis. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1713**,** 92-112.

Dorsch, M., Lovet, D. & Bailey, G. D. 2001. *Fusobacterium equinum* sp. nov., from the oral cavity of horses. *International Journal of Systematic and Evolutionary Microbiology,* 51**,** 1959-1963.

Duerden, B. I. 1990. *Topley & Wilson's Principles of Bacteriology, Virology and Immunity. Eighth Edition. Volume 2: Systematic Bacteriology,* Sevenoaks, Kent, Hodder and Stoughton Limited, 551-575.

Eaton, C. & Swindells, J. 2014. The significance and epidemiology of *Fusobacterium necrophorum* in sore throats. *Journal of Infection,* 69**,** 194-196.

Emery, D., Edwards, R. & Rothel, J. 1986. Studies of the purification of the leucocidin of *Fusobacterium necrophorum* and its neutralization by specific antisera. *Veterinary Microbiology,* 11**,** 357-372.

Enwonwu, C. O., Falkler, W., Idigbe, E. O., Afolabi, B., Ibrahim, M., Onwujekwe, D., Savage, O. & Meeks, V. I. 1999. Pathogenesis of cancrum oris (noma): confounding interactions of malnutrition with infection. *The American Journal of Tropical Medicine and Hygiene,* 60**,** 223-232.

European Nucleotide Archive. 2015. *European Nucleotide Archive < EMBL-EBI* [Online]. Available: http://www.ebi.ac.uk/ena 2015].

Expasy - Protscale. 2015. *ExPASy - ProtScale* [Online]. Available: http://web.expasy.org/protscale/ 2015].

Expasy - Translate Tool. 2015. *ExPASy - Translate tool* [Online]. Available: http://web.expasy.org/translate/ 2015].

Falkler, W., Enwonwu, C. O. & Idigbe, E. O. 1999. Isolation of *Fusobacterium necrophorum* from cancrum oris (noma). *The American journal of tropical medicine and hygiene,* 60**,** 150-156.

Farooq, S., Wani, S. A., Hassan, M. N., Nazir, N. & Nyrah, Q. J. 2015. The detection of Dichelobacter nodosus and *Fusobacterium necrophorum* from ovine footrot in Kashmir, India. *Anaerobe,* 35**,** 41-43.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J. & Punta, M. 2014. Pfam: the protein families database. *Nucleic Acids Research,* 42**,** D222-D230.

Fleischmann, R., Adams, M., White, O., Clayton, R., Kirkness, E., Kerlavage, A., Bult, C., Tomb, J., Dougherty, B., Merrick, J. & Al., E. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae Rd*. *Science,* 269**,** 496-512.

Flügge, C. 1886. Zweiter Abschnitt. II. Bacillen. B. Für Thiere pathogene Bacillen. Bacillus necrophorus (Löffler). *Die Mikroorganismen. Mit besonderer Berücksichtigung der Aetiologie der Infectionskrankheiten***,** 273.

Forrester, J. L., Campbell, B. J., Berg, J. N. & Barrett, J. T. 1985. Aggregation of platelets by *Fusobacterium necrophorum*. *Journal of Clinical Microbiology,* 22**,** 245-249.

Fr33.Net. 2015. *Sequence editor - Convert DNA and RNA sequences. Generate antiparallel, complement and inverse sequences.* [Online]. Available: http://www.fr33.net/seqedit.php 2015].

Gaboriaud, C., Gupta, R. K., Martin, L., Lacroix, M., Serre, L., Teillet, F., Arlaud, G. J., Rossi, V. & Thielens, N. M. 2013. The serine protease domain of MASP-3: enzymatic properties and crystal structure in complex with ecotin. *PLoS ONE,* 8**,** e67962.

Galioto, N. J. 2008. Peritonsillar abscess. *Steroids,* 8**,** 14.

Garcia, G., Amoako, K., Xu, D., Inoue, T., Goto, Y. & Shinjo, T. 1998. Chemical composition of endotoxins produced by *Fusobacterium necrophorum* subsp. *necrophorum* and *F. necrophorum* subsp. *funduliforme*. *Microbios,* 100**,** 175-179.

Gillmor, S. A., Takeuchi, T., Yang, S. Q., Craik, C. S. & Fletterick, R. J. 2000. Compromise and accommodation in ecotin, a dimeric macromolecular inhibitor of serine proteases. *Journal of Molecular Biology,* 299**,** 993-1003.

Golpe, R., Marín, B. & Alonso, M. 1999. Lemierre's syndrome (necrobacillosis). *Postgraduate Medical Journal,* 75**,** 141-144.

Gupta, R. S. & Sethi, M. 2014. Phylogeny and molecular signatures for the phylum Fusobacteria and its distinct subclades. *Anaerobe,* 28**,** 182-198.

Hagelskjær Kristensen, L. & Prag, J. 2008. Lemierre's syndrome and other disseminated *Fusobacterium necrophorum* infections in Denmark: a prospective epidemiological and clinical survey. *European Journal of Clinical Microbiology & Infectious Diseases,* 27**,** 779-789.

Hagelskjaer Kristensen, L. & Prag, J. 2000. Human necrobacillosis, with emphasis on Lemierre's syndrome. *Clinical Infectious Diseases,* 31**,** 524-532.

Hagelskjær, L. H., Prag, J., Malczynski, J. & Kristensen, J. H. 1998. Incidence and Clinical Epidemiology of Necrobacillosis, including Lemierre's Syndrome, in Denmark 1990-1995. *European Journal of Clinical Microbiology & Infectious Diseases,* 17**,** 561-565.

Hall, V., Duerden, B. I., Magee, J. T., Ryley, H. C. & Brazier, J. S. 1997. A comparative study of *Fusobacterium necrophorum* strains from human and animal sources by phenotypic reactions, pyrolysis mass spectrometry and SDS-PAGE. *Journal of Medical Microbiology,* 46**,** 865-871.

Hallé, J. 1898. Recherches sur la bactériologie du canal génital de la femme. *Faculté de Médicine de Paris, Paris, France*.

Han, Y. W. 2015. *Fusobacterium nucleatum*: a commensal-turned pathogen. *Current Opinion in Microbiology,* 23**,** 141-147.

Han, Y. W., Ikegami, A., Rajanna, C., Kawsar, H. I., Zhou, Y., Li, M., Sojar, H. T., Genco, R. J., Kuramitsu, H. K. & Deng, C. X. 2005. Identification and characterization of a novel adhesin unique to oral fusobacteria. *Journal of Bacteriology,* 187**,** 5330-5340.

Hedstrom, L. 2002. Serine protease mechanism and specificity. *Chemical Reviews,* 102**,** 4501-4524.

Highlander, S. K., Chidambaram, M., Engler, M. J. & Weinstock, G. M. 1989. DNA sequence of the *Pasteurella haemolytica* leukotoxin gene cluster. *DNA,* 8**,** 15-28.

Hodak, H. & Jacob-Dubuisson, F. 2007. Current challenges in autotransport and two-partner protein secretion pathways. *Research in Microbiology,* 158**,** 631-637.

Hodkinson, B. P. & Grice, E. A. 2015. Next-generation sequencing: a review of technologies and tools for wound microbiome research. *Advances in Wound Care,* 4**,** 50-58.

Hofstad, T. 1998. *Topley and Wilson's Microbiology and Microbial Infections. Ninth Edition. Volume 2: Systematic Bacteriology: Systematic Bacteriology,* New York, NY, Oxford University Press, 1355-1364.

Hoiczyk, E., Roggenkamp, A., Reichenbecher, M., Lupas, A. & Heesemann, J. 2000. Structure and sequence analysis of *Yersinia* YadA and *Moraxella* UspAs reveal a novel class of adhesins. *The EMBO journal,* 19**,** 5989-5999.

Hsiao, H.-J., Huang, Y.-C., Hsia, S.-H., Wu, C.-T. & Lin, J.-J. 2012. Clinical features of peritonsillar abscess in children. *Pediatrics & Neonatology,* 53**,** 366-370.

Jacob‑Dubuisson, F., Locht, C. & Antoine, R. 2001. Two-partner secretion in Gram-negative bacteria: a thrifty, specific pathway for large virulence proteins. *Molecular Microbiology,* 40**,** 306-313.

Jensen, A., Hagelskjær Kristensen, L. & Prag, J. 2007. Detection of *Fusobacterium necrophorum* subsp. *funduliforme* in tonsillitis in young adults by real-time PCR. *Clinical Microbiology and Infection,* 13**,** 695-701.

Jones, J., Riordan, T. & Morgan, M. 2001. Investigation of postanginal sepsis and Lemierre's syndrome in the South West Peninsula. *Communicable Disease and Public Health/PHLS,* 4**,** 278-281.

Jousimies-Somer, H., Savolainen, S., Makitie, A. & Ylikoski, J. 1993. Bacteriologic findings in peritonsillar abscesses in young adults. *Clinical Infectious Diseases,* 16**,** S292-S298.

Kachlany, S. C., Fine, D. H. & Figurski, D. H. 2002. Purification of secreted leukotoxin (LtxA) from *Actinobacillus actinomycetemcomitans*. *Protein Expression and Purification,* 25**,** 465-471.

Kanoe, M., Koyanagi, Y., Kondo, C., Mamba, K., Makita, T. & Kai, K. 1997. Location of haemagglutinin in bacterial cells of *Fusobacterium necrophorum* subsp. *necrophorum*. *Microbios,* 96**,** 33-38.

Kanoe, M. & Yamanaka, M. 1989. Bovine platelet aggregation by *Fusobacterium necrophorum*. *Journal of Medical Microbiology,* 29**,** 13-17.

Kapatral, V., Anderson, I., Ivanova, N., Reznik, G., Los, T., Lykidis, A., Bhattacharyya, A., Bartman, A., Gardner, W., Grechkin, G., Zhu, L., Vasieva, O., Chu, L., Kogan, Y., Chaga, O., Goltsman, E.,

Bernal, A., Larsen, N., D'souza, M., Walunas, T., Pusch, G., Haselkorn, R., Fonstein, M., Kyrpides, N. & Overbeek, R. 2002. Genome sequence and analysis of the oral bacterium *Fusobacterium nucleatum* strain ATCC 25586. *Journal of Bacteriology,* 184**,** 2005-2018.

Karakelian, D., Lear, J. D., Lally, E. T. & Tanaka, J. C. 1998. Characterization of *Actinobacillus actinomycetemcomitans* leukotoxin pore formation in HL60 cells. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease,* 1406**,** 175-187.

Karkos, P. D., Asrani, S., Karkos, C. D., Leong, S. C., Theochari, E. G., Alexopoulou, T. D. & Assimakopoulos, A. D. 2009. Lemierre's syndrome: a systematic review. *The Laryngoscope,* 119**,** 1552-1559.

Kirchner, M. & Schneider, S. 2015. CRISPR‐Cas: From the Bacterial Adaptive Immune System to a Versatile Tool for Genome Engineering. *Angewandte Chemie International Edition,* 54**,** 13508-13514.

Klug, T. E., Rusan, M., Fuursted, K. & Ovesen, T. 2009. *Fusobacterium necrophorum*: most prevalent pathogen in peritonsillar abscess in Denmark. *Clinical Infectious Diseases,* 49**,** 1467-1472.

Knorr, M. 1923. Ueber die fusospirilläre symbiose, die Gattung *Fusobacterium* (K. B. Lehmann) und *Spirillum sputigenum*. II Mitteilung. Die Gattung *Fusobacterium*. *Zentralbl. Bakteriol Parasitenkd Infektionskr Hyg. Abt,* 1**,** 4-22.

Korkmaz, B., Horwitz, M. S., Jenne, D. E. & Gauthier, F. 2010. Neutrophil elastase, proteinase 3, and cathepsin G as therapeutic targets in human diseases. *Pharmacological Reviews,* 62**,** 726-759.

Kumar, A., Anderson, D., Amachawadi, R. G., Nagaraja, T. G. & Narayanan, S. K. 2013. Characterization of *Fusobacterium necrophorum* isolated from llama and alpaca. *Journal of Veterinary Diagnostic Investigation,* 25**,** 502-507.

Kuppalli, K., Livorsi, D., Talati, N. J. & Osborn, M. 2012. Lemierre's syndrome due to *Fusobacterium necrophorum*. *The Lancet Infectious Diseases,* 12**,** 808-815.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* 227**,** 680-685.

Land, M., Hauser, L., Jun, S.-R., Nookaew, I., Leuze, M. R., Ahn, T.-H., Karpinets, T., Lund, O., Kora, G. & Wassenaar, T. 2015. Insights from 20 years of bacterial genome sequencing. *Functional & Integrative Genomics,* 15**,** 141-161.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M. & Fitzhugh, W. 2001. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

Langston, K. G., Worsham, L. M., Earls, L. & Ernst-Fonberg, M. L. 2004. Activation of hemolysin toxin: relationship between two internal protein sites of acylation. *Biochemistry,* 43**,** 4338-4346.

Langworth, B. F. 1977. *Fusobacterium necrophorum*: its characteristics and role as an animal pathogen. *Bacteriological Reviews,* 41**,** 373-90.

Lear, J. D., Karakelian, D., Furblur, U., Lally, E. T. & Tanaka, J. C. 2000. Conformational studies of *Actinobacillus actinomycetemcomitans* leukotoxin: partial denaturation enhances toxicity. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology,* 1476**,** 350-362.

Lemierre, A. 1936. On certain septicaemias due to anaerobic organisms. *The Lancet,* 227**,** 701-703.

Leo, J. C., Grin, I. & Linke, D. 2012. Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. *Philosophical transactions of the Royal Society of London. Series B,* 367**,** 1088-1101.

Li, Z., Chen, Y., Mu, D., Yuan, J., Shi, Y., Zhang, H., Gan, J., Li, N., Hu, X., Liu, B., Yang, B. & Fan, W. 2012. Comparison of the two major classes of assembly algorithms: overlap–layout–consensus and de-bruijn-graph. *Briefings in Functional Genomics,* 11**,** 25-37.

Linke, D., Riess, T., Autenrieth, I. B., Lupas, A. & Kempf, V. A. 2006. Trimeric autotransporter adhesins: variable structure, common function. *Trends in microbiology,* 14**,** 264-270.

Liu, L. & Fan, X.-D. 2014. CRISPR–Cas system: a powerful tool for genome engineering. *Plant molecular biology,* 85**,** 209-218.

Löffler, F. 1884. Die Diphtherie beim Kalbe. *Mitteilungen Kaiserlichen Gesundheitsamte,* 2**,** 489-499.

Lu, M. D., Vasavada, Z. & Tanner, C. 2009. Lemierre syndrome following oropharyngeal infection: a case series. *The Journal of the American Board of Family Medicine,* 22**,** 79-83.

Ludlam, H. A., Milner, N. J., Brazier, J. S., Davies, I. H., Perry, K., Marriott, R. K., Donachie, L. & Curran, M. D. 2009a. lktA-encoded leukotoxin is not a universal virulence factor in invasive *Fusobacterium necrophorum* infections in animals and man. *Journal of Medical Microbiology,* 58**,** 529-530.

Lyle, N., Rutherford, E. & Batty, V. 2011. A pain in the neck—Imaging in neck sepsis. *Clinical radiology,* 66**,** 876-885.

Madigan, M. T. & Martinko, J. M. 2006. *Brock Biology of Microorganisms, Eleventh edition,,* Upper Saddle River, NJ, USA, Pearson Prentice Hall.

Mardis, E. R. 2008. Next-Generation DNA Sequencing Methods. *Annual Review of Genomics and Human Genetics,* 9**,** 387-402.

Mardis, E. R. 2013. Next-Generation Sequencing Platforms. *Annual Review of Analytical Chemistry,* 6**,** 287-303.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J. & Chen, Z. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437**,** 376-380.

Markey, B., Leonard, F., Archambault, M., Cullinane, A. & Maguire, D. 2013. *Clinical veterinary microbiology*, Elsevier Health Sciences, China, 210-211.

Marshall, M., Bohach, G. & Boehm, D. 2000. Characterization of *Staphylococcus aureus* beta-toxin induced leukotoxicity. *Journal of Natural Toxins,* 9**,** 125-138.

Merck Manuals Professional Edition. 2015. *Tonsillopharyngitis - Ear Nose and Throat Disorders* [Online]. Available: http://www.merckmanuals.com/professional/ear-nose-and-throat-disorders/oral-and-pharyngeal-disorders/tonsillopharyngitis [Accessed 2nd November 2015].

Microbe Division (Jcm)(Riken Brc). 2015. *Jcm.brc.riken.jp* [Online]. Available: http://jcm.brc.riken.jp/en/ 2015].

Miller, J. R., Koren, S. & Sutton, G. 2010. Assembly algorithms for next-generation sequencing data. *Genomics,* 95**,** 315-327.

Miyazato, S., Shinjo, T., Yago, H. & Nakamura, N. 1978. Fimbriae (pili) detected in *Fusobacterium necrophorum*. *Japanese Journal of Veterinary Science,* 40**,** 619-21.

Nagai, S., Kanoe, M. & Toda, M. 1984. Purification and partial characterization of *Fusobacterium necrophorum* hemagglutinin. *Zentralblatt für Bakteriologie, Mikrobiologie und Hygiene. Series A: Medical Microbiology, Infectious Diseases, Virology, Parasitology,* 258**,** 232-241.

Nagaraja, T. G., Narayanan, S. K., Stewart, G. C. & Chengappa, M. M. 2005. *Fusobacterium necrophorum* infections in animals: pathogenesis and pathogenic mechanisms. *Anaerobe,* 11**,** 239-246.

Narayanan, S., Nagaraja, T. G., Okwumabua, O., Staats, J., Chengappa, M. M. & Oberst, R. D. 1997. Ribotyping To Compare *Fusobacterium necrophorum* Isolates from bovine liver abscesses, ruminal walls, and ruminal contents. *Applied and Environmental Microbiology,* 63**,** 4671–4678.

Narayanan, S., Stewart, G. C., Chengappa, M. M., Willard, L., Shuman, W., Wilkerson, M. & Nagaraja, T. G. 2002b. *Fusobacterium necrophorum* leukotoxin induces activation and apoptosis of bovine leukocytes. *Infection and Immunity,* 70**,** 4609-4620.

Narayanan, S. K., Nagaraja, T. G., Chengappa, M. M. & Stewart, G. C. 2001b. Cloning, sequencing, and expression of the leukotoxin gene from *Fusobacterium necrophorum*. *Infection and Immunity,* 69**,** 5447-5455.

Narayanan, S. K., Nagaraja, T. G., Chengappa, M. M. & Stewart, G. C. 2002a. Leukotoxins of gram-negative bacteria. *Veterinary Microbiology,* 84**,** 337-356.

Narongwanichgarn, W., Kawaguchi, E., Misawa, N., Goto, Y., Haga, T. & Shinjo, T. 2001. Differentiation of *Fusobacterium necrophorum* subspecies from bovine pathological lesions by RAPD-PCR. *Veterinary Microbiology,* 82**,** 383-388.

Narongwanichgarn, W., Misawa, N., Jin, J. H., Amoako, K. K., Kawaguchi, E., Shinjo, T., Haga, T. & Goto, Y. 2003. Specific detection and differentiation of two subspecies of *Fusobacterium necrophorum* by PCR. *Veterinary Microbiology,* 91**,** 183-195.

National Center for Biotechnology Information. 2015. *Ncbi.nlm.nih.gov* [Online]. Available: http://www.ncbi.nlm.nih.gov/ 2015].

Oelke, A. M., Nagaraja, T., Wilkerson, M. J. & Stewart, G. C. 2005. The leukotoxin operon of *Fusobacterium necrophorum* is not present in other species of *Fusobacterium*. *Anaerobe,* 11**,** 123-129.

Oikonomou, G., Machado, V. S., Santisteban, C., Schukken, Y. H. & Bicalho, R. C. 2012. Microbial diversity of bovine mastitic milk as described by pyrosequencing of metagenomic 16s rDNA. *PLoS ONE,* 7**,** e47671.

Pál, G., Sprengel, G., Patthy, A. & Gráf, L. 1994. Alteration of the specificity of ecotin, an *E. coli* serine proteinase inhibitor, by site directed mutagenesis. *FEBS Letters,* 342**,** 57-60.

Passy, V. 1994. Pathogenesis of peritonsillar abscess. *The Laryngoscope,* 104**,** 185-190.

Pennisi, E. 2013. The CRISPR Craze. *Science,* 341**,** 833-836.

Perry, M. & Whyte, A. 1998. Immunology of the tonsils. *Immunology Today,* 19**,** 414-421.

Petersen, T. N., Brunak, S., Von Heijne, G. & Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods,* 8**,** 785-6.

Pfam. 2015. *Pfam.xfam.org* [Online]. Available: http://pfam.xfam.org/ 2015].

Pinto, A., Scaglione, M., Scuderi, M. G., Tortora, G., Daniele, S. & Romano, L. 2008. Infections of the neck leading to descending necrotizing mediastinitis: Role of multi-detector row computed tomography. *European Journal of Radiology,* 65**,** 389-394.

Powell, E. L., Powell, J., Samuel, J. R. & Wilson, J. A. 2013. A review of the pathogenesis of adult peritonsillar abscess: time for a re-evaluation. *Journal of Antimicrobial Chemotherapy,* 68**,** 1941-1950.

Public Health England 2015. UK Standards for Microbiology Investigations: Investigation of Throat Related Specimens. *Bacteriology.*

Rath, D., Amlinger, L., Rath, A. & Lundgren, M. 2015. The CRISPR-Cas immune system: Biology, mechanisms and applications. *Biochimie,* 117**,** 119-128.

Ridgway, J. M., Parikh, D. A., Wright, R., Holden, P., Armstrong, W., Camilon, F. & Wong, B. J.-F. 2010. Lemierre syndrome: a pediatric case series and review of literature. *American journal of otolaryngology,* 31**,** 38-45.

Riordan, T. 2007. Human infection with *Fusobacterium necrophorum* (necrobacillosis), with a focus on Lemierre's syndrome. *Clinical Microbiology Reviews,* 20**,** 622–659.

Rubinstein, M. R., Wang, X., Liu, W., Hao, Y., Cai, G. & Han, Y. W. 2013. *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. *Cell Host & Microbe,* 14**,** 195-206.

Saginala, S., Nagaraja, T., Lechtenberg, K., Chengappa, M., Kemp, K. & Hine, P. 1997. Effect of *Fusobacterium necrophorum* leukotoxoid vaccine on susceptibility to experimentally induced liver abscesses in cattle. *Journal of Animal Science,* 75**,** 1160-1166.

Sangal, V., Nieminen, L., Tucker, N. P. & Hoskisson, P. A. 2014. Chapter 5 - Revolutionizing prokaryotic systematics through next-generation sequencing. *In:* MICHAEL GOODFELLOW, I. S. & JONGSIK, C. (eds.) *Methods in Microbiology.* Academic Press.

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., Hutchison, C. A., Slocombe, P. M. & Smith, M. 1977a. Nucleotide sequence of bacteriophage [phi]X174 DNA. *Nature,* 265**,** 687-695.

Sanger, F., Nicklen, S. & Coulson, A. R. 1977b. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences,* 74**,** 5463-5467.

Sanmillán, J. L., Pelegrín, I., Rodríguez, D., Ardanuy, C. & Cabellos, C. 2013. Primary lumbar epidural abscess without spondylodiscitis caused by *Fusobacterium necrophorum* diagnosed by 16S rRNA PCR. *Anaerobe,* 23**,** 45-47.

Schägger, H. & Von Jagow, G. 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analytical Biochemistry,* 166**,** 368-379.

Schmitt, C. K., Meysick, K. C. & O'brien, A. D. 1999. Bacterial toxins: friends or foes? *Emerging Infectious Diseases,* 5**,** 224.

Schmorl, G. 1891. Ueber ein pathogenes Fadenbacterium (*Streptothrix cuniculi*). *Deutsch Z. Thiermed,* 17**,** 375-408.

Seymour, J. L., Lindquist, R. N., Dennis, M. S., Moffat, B., Yansura, D., Reilly, D., Wessinger, M. E. & Lazarus, R. A. 1994. Ecotin is a potent anticoagulant and reversible tight-binding inhibitor of factor Xa. *Biochemistry,* 33**,** 3949-58.

Shinjo, T., Fujisawa, T. & Mitsuoka, T. 1991. Proposal of Two Subspecies of *Fusobacterium necrophorum* (Flügge) Moore and Holdeman: *Fusobacterium necrophorum* subsp. *necrophorum* subsp. nov., nom. rev. (ex Flügge 1886), and *Fusobacterium necrophorum* subsp. *funduliforme* subsp. nov., nom. rev. (ex Hallé 1898). *International Journal of Systematic and Evolutionary Microbiology,* 41**,** 395-397.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., Mcwilliam, H., Remmert, M. & Söding, J. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology,* 7**,** 539.

Silverman, G. A., Whisstock, J. C., Bottomley, S. P., Huntington, J. A., Kaiserman, D., Luke, C. J., Pak, S. C., Reichhart, J.-M. & Bird, P. I. 2010. Serpins Flex Their Muscle: I. Putting the clamps on proteolysis in diverse biological systems. *The Journal of Biological Chemistry,* 285**,** 24299-24305.

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. & Birol, I. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Research,* 19**,** 1117-1123.

Sims, D., Sudbery, I., Ilott, N. E., Heger, A. & Ponting, C. P. 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics,* 15**,** 121-132.

Smith, G. & Thornton, E. 1997. Classification of human and animal strains of *Fusobacterium necrophorum* by their pathogenic effects in mice. *Journal of Medical Microbiology,* 46**,** 879-882.

Smith, G. R., Wallace, L. M. & Noakes, D. E. 1990. Experimental observations on the pathogenesis of necrobacillosis, *Cambridge Journals Online*. 104**,** 73-78.

Stothard, P. & Wishart, D. S. 2005. Circular genome visualization and exploration using CGView. *Bioinformatics,* 21**,** 537-539.

Sugawara, S., Uehara, A., Tamai, R. & Takada, H. 2002. Innate immune responses in oral mucosa. *Journal of Endotoxin Research,* 8**,** 465-468.

Sun, D., Wu, R., Li, G., Zheng, J., Liu, X., Lin, Y. & Guo, D. 2009. Identification of three immunodominant regions on leukotoxin protein of *Fusobacterium necrophorum*. *Veterinary Research Communications,* 33**,** 749-755.

Syed, M. I., Baring, D., Addidle, M., Murray, C. & Adams, C. 2007. Lemierre syndrome: two cases and a review. *The Laryngoscope,* 117**,** 1605-1610.

Tadepalli, S., Stewart, G. C., Nagaraja, T., Jang, S. S. & Narayanan, S. K. 2008. *Fusobacterium equinum* possesses a leukotoxin gene and exhibits leukotoxin activity. *Veterinary Microbiology,* 127**,** 89-96.

Tadepalli, S., Stewart, G. C., Nagaraja, T. G. & Narayanan, S. K. 2008a. Human *Fusobacterium necrophorum* strains have a leukotoxin gene and exhibit leukotoxic activity. *Journal of Medical Microbiology,* 57**,** 225-231.

Tadepalli, S., Stewart, G. C., Nagaraja, T. G. & Narayanan, S. K. 2008b. Leukotoxin operon and differential expressions of the leukotoxin gene in bovine *Fusobacterium necrophorum* subspecies. *Anaerobe,* 14**,** 13-18.

Takayama, Y., Kanoe, M., Maeda, K., Okada, Y. & Kai, K. 2000. Adherence of *Fusobacterium necrophorum* subsp. *necrophorum* to ruminal cells derived from bovine rumenitis. *Letters in Applied Microbiology,* 30**,** 308-311.

Tan, T., Little, P. & Stokes, T. 2008. Antibiotic prescribing for self limiting respiratory tract infections in primary care: summary of NICE guidance. *BMJ,* 337**,** a437.

Tan, Z., Nagaraja, T., Chengappa, M. & Smith, J. 1994a. Biological and biochemical characterization of *Fusobacterium necrophorum* leukotoxin. *American Journal of Veterinary Research,* 55**,** 515-521.

Tan, Z., Nagaraja, T., Chengappa, M. & Staats, J. 1994b. Purification and quantification of *Fusobacterium necrophorum* leukotoxin by using monoclonal antibodies. *Veterinary Microbiology,* 42**,** 121-133.

Tan, Z. L., Nagaraja, T. G. & Chengappa, M. M. 1996. *Fusobacterium necrophorum* infections: Virulence factors, pathogenic mechanism and control measures. *Veterinary Research Communications,* 20**,** 113-140.

Thibodeau, G. A. & Patton, K. T. 1997. *The Human Body in Health & Disease, Second Edition,* Missouri, Mosby-Year Book.

Thusberg, J. & Vihinen, M. 2006. Bioinformatic analysis of protein structure–function relationships: case study of leukocyte elastase (ELA2) missense mutations. *Human Mutation,* 27**,** 1230-1243.

Tmpred. 2015. *TMpred Server* [Online]. Available: http://www.ch.embnet.org/software/TMPRED_form.html 2015].

Uitto, V.-J., Baillie, D., Wu, Q., Gendron, R., Grenier, D., Putnins, E. E., Kanervo, A. & Firth, J. D. 2005. *Fusobacterium nucleatum* increases collagenase 3 production and migration of epithelial cells. *Infection and Immunity,* 73**,** 1171-1179.

Ulmer, J. S., Lindquist, R. N., Dennis, M. S. & Lazarus, R. A. 1995. Ecotin is a potent inhibitor of the contact system proteases factor XIIa and plasma kallikrein. *FEBS Letters,* 365**,** 159-63.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M. & Rozen, S. G. 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Research,* 40**,** e115-e115.

Van Domselaar, G. H., Stothard, P., Shrivastava, S., Cruz, J. A., Guo, A., Dong, X., Lu, P., Szafron, D., Greiner, R. & Wishart, D. S. 2005. BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Research,* 33**,** W455-W459.

Webact. 2015. *WebACT* [Online]. Available: http://www.webact.org 2015].

Willimott, S., Barker, J., Jones, L. A. & Opara, E. I. 2007. Apoptotic effect of Oldenlandia diffusa on the leukaemic cell line HL60 and human lymphocytes. *Journal of Ethnopharmacology,* 114**,** 290-299.

Wilson, B. A., Salyers, A. A., Whitt, D. D. & Winkler, M. E. 2011. *Bacterial Pathogenesis A Molecular Approach,* Washington, ASM Press.

Wolanin, P. M., Thomason, P. A. & Stock, J. B. 2002. Histidine protein kinases: key signal transducers outside the animal kingdom. *Genome Biology,* 3**,** 3013.

Wright, W. F., Shiner, C. N. & Ribes, J. A. 2012. Lemierre syndrome. *Southern Medical Journal,* 105**,** 283-288.

Yang, S. Q., Wang, C.-I., Gillmor, S. A., Fletterick, R. J. & Craik, C. S. 1998. Ecotin: a serine protease inhibitor with two distinct and interacting binding sites. *Journal of Molecular Biology,* 279**,** 945-957.

Yarden-Bilavsky, H., Raveh, E., Livni, G., Scheuerman, O., Amir, J. & Bilavsky, E. 2013. *Fusobacterium necrophorum* mastoiditis in children–emerging pathogen in an old disease. *International Journal of Pediatric Otorhinolaryngology,* 77**,** 92-96.

Yoong, P. & Torres, V. J. 2013. The effects of *Staphylococcus aureus* leukotoxins on the host: cell lysis and beyond. *Current Opinion in Microbiology,* 16**,** 63-69.

Yusuf, E., Halewyck, S., Wybo, I., Piérard, D. & Gordts, F. 2015. *Fusobacterium necrophorum* and other *Fusobacterium* spp. isolated from head and neck infections: a 10-year epidemiology study in an academic hospital. *Anaerobe,* 34**,** 120-124.

Zerbino, D. R. & Birney, E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research,* 18**,** 821-829.

Zhang, F., Nagaraja, T. G., George, D. & Stewart, G. C. 2006. The two major subspecies of *Fusobacterium necrophorum* have distinct leukotoxin operon promoter regions. *Veterinary Microbiology,* 112**,** 73-78.

Zhou, H., Bennett, G., Buller, N. & Hickford, J. G. H. 2011b. Isolation of two novel *Fusobacterium necrophorum* variants from sheep in Australia. *Veterinary Microbiology,* 148**,** 448.

Zhou, H., Bennett, G. & Hickford, J. G. H. 2009a. Variation in *Fusobacterium necrophorum* strains present on the hooves of footrot infected sheep, goats and cattle. *Veterinary Microbiology,* 135**,** 363-367.

Zhou, H., Bennett, G., Kennan, R. M., Rood, J. I. & Hickford, J. G. 2009b. Identification of a leukotoxin sequence from *Fusobacterium equinum*. *Veterinary Microbiology,* 4**,** 394-395.

Zhou, H., Dobbinson, S. & Hickford, J. G. 2010. *Fusobacterium necrophorum* variants present on the hooves of lame pigs. *Veterinary Microbiology,* 3**,** 390.

Zhou, H., Meyer, K., Ganter, M. & Hickford, J. G. H. 2011a. Identification of a *Fusobacterium necrophorum* isolate that contains a new variant of the leukotoxin gene (lktA) from the hoof of a sheep with ovine footrot. *Veterinary Microbiology,* 149**,** 524-525.

Appendices

Appendix 1: Preparation of reagents and solutions

A1.1 Ethylenediaminetetraacetic acid (EDTA)

0.5 M ETDA was made by dissolving 73 g EDTA in 350 ml deionised water. A magnetic stir bar was used for stirring while the pH was adjusted to 8.0 with 10 M NaCl. The volume was then made up to 500 ml using deionised water.

A1.2 50X Tris base, acetic acid and EDTA (TAE) buffer

50X TAE buffer was made by adding 121 g Tris base in 250 ml deionised water while stirring. 28.6 ml, acetic acid was added and 50 ml 0.5 M EDTA solution, pH 8.0. Deionised water was added to bring the volume up to 500 ml. TAE buffer was diluted 50X in deionised water before use.

A1.3 6X Gel electrophoresis loading dye

6X Glycerol and bromophenol blue loading dye was made by adding 3 ml glycerol and 2.5 mg bromophenol blue to a 20 ml container and adding deionised water to bring the volume up to 10 ml. The loading dye was diluted to 1X by mixing in a ratio of 1 part loading dye to 5 parts DNA sample before samples were loaded into agarose gels.

A1.4 Laemmli buffer

6X Laemmli buffer was made as per Appendix Table for preparing and loading protein samples for SDS PAGE analysis. Non-reducing Laemmli buffer was also separately made by replacing the reducing agent β-mercaptoethanol with 1 ml of deionised water. Laemmli buffer was diluted to 1X by mixing in a ratio of 1 part buffer to 5 parts protein sample before samples were boiled for 5 minutes and loaded into the gel.

Component	Quantity
1 M Tris-HCl pH 6.8	2.4 ml
SDS	0.8 g
Glycerol	4 ml
Deionised water	2.8 ml
Bromophenol blue	Trace
β-mercaptoethanol	1 ml

A1.5 10X Tris-Glycine buffer

10X Tris-Glycine buffer was made by dissolving 30.29 g Tris base and 144.12 g glycine in 750 ml deionised water. 100 ml 10 % (w/v) Sodium dodecyl sulfate was added and the pH adjusted to 8.3. The volume was then brought up to 1 L with deionised water. The 10X buffer was diluted to 1X in deionised water before use.

A1.6 Bradford reagent

1X Bradford reagent was made by dissolving 100 mg of Coomassie Brilliant Blue in 47 ml 100 % methanol and stirring for an hour. 100 ml 85 % phosphoric acid was added and the volume was brought up to 1 L with deionised water. The solution was filtered twice with Whatman paper filters (Sigma, Gillingham) and stored at 4 °C in the dark.

A1.7 Phosphate buffered saline (PBS)

1X PBS was made by dissolving one PBS tablet (Sigma, Gillingham) in 200 ml deionised water and sterilising the solution by autoclaving. Final concentrations of 1X PBS are: 0.01 M phosphate buffer, 0.0027 M potassium chloride and 0.137 M sodium chloride, pH 7.4.

Appendix 2: Preparation of media and antibiotics

A2.1 LB agar and broth

LB agar and LB broth were obtained in powdered form from Fisher Scientific, Loughborough. LB agar was made by suspending 40 g in 1 L deionised water before autoclaving, and LB broth was made by suspending 25 g in 1 L deionised water before autoclaving, as per the manufacturer's instructions.

A2.2 Fastidious anaerobe agar (FAA) and fastidious anaerobe broth (FAB)

FAA and FAB were purchased from LabM, Heywood. FAA was made by suspending 46 g in 950 ml deionised water before autoclaving. Following sterilisation, the agar was cooled to 50 °C and 50 ml defibrinated horse blood (TCS Biosciences, Buckingham) was added aseptically. FAB was made by suspending 29.7 g in 1 L of deionised water and autoclaving, as per the manufacturer's instructions.

A2.3 Brain heart infusion (BHI) broth

BHI broth was purchased from Merck, Feltham and was made by suspending 37 g in 1 L deionised water before autoclaving, as per the manufacturer's instructions.

A2.4 Complete cell culture medium

IMDM with 2 mM L-glutamine and Foetal Bovine Serum (FBS) were purchased from Life Technologies, Paisley. IMDM medium was supplemented with 20 % FBS. Complete cell culture medium was stored at 4 °C for up to 4 weeks.

A2.5 SOC media

SOC media was made by adding the components from Appendix Table 2 to 90 ml deionised water. The volume was then made up to 100 ml. All solutions were first filter sterilised with 0.2 μm syringe filters (Millipore, Hertfordshire) and the final solution was autoclaved at 110 °C for 10 minutes. Media was stored at -20 °C in 1 ml aliquots.

Appendix Table 2: SOC media components.

Component	Quantity
Tryptone	2 g
Yeast extract	0.5 g
5 M NaCl	0.2 ml
1 M KCl	0.25 ml
1 M MgCl$_2$	1 ml
1 M MgSO$_4$	1 ml
1 M glucose	2 ml

A2.6 Ampicillin

Ampicillin was obtained from Sigma, Gillingham and a stock solution of 100 mg/ml was prepared by dissolving 1 g of ampicillin powder in 10 ml of deionised water. The solution was filter sterilised with 0.2 µm syringe filters (Millipore, Hertfordshire) and stored at -20 °C in 500 µl aliquots.

A2.7 100X Penicillin-Streptomycin solution

100X Penicillin-Streptomycin solution was purchased from Sigma, Gillingham and stored at -20 °C in 1 ml aliquots.

Appendix 3: Methods for strain identification using biochemical tests

A3.1 Gram stain

1 or 2 colonies were smeared into a drop of deionised water on a glass slide and left to dry, then heat fixed by passing through the flame upside down several times. The slide was covered with crystal violet for 30 seconds, followed by iodine for 30 seconds before a brief rinse with tap water. The slide was then quickly washed with acetone, followed immediately by a quick rinse with tap water. The slide was then covered with safranin for 30 seconds, washed with tap water and blotted dry. *F. necrophorum* should appear as a Gram negative pleomorphic rod, often with some long filaments present.

A3.2 Oxidase test

A 1 % (w/v) solution of oxidase reagent (Sigma, Gillingham) was made in sterile deionised water. This was poured onto filter paper in a petri dish until saturated. A cocktail stick was used to smear several colonies onto the filter paper. Colonies changed to purple if positive for cytochrome oxidase, or displayed no colour change if negative. *F. necrophorum* strains are oxidase negative.

A3.3 Catalase test

A drop of hydrogen peroxide (Sigma, Gillingham) was placed onto glass slide. A cocktail stick was used to mix several colonies into the drop. Bubbles were produced if positive for catalase or peroxidase. There was no bubble production if negative. *F. necrophorum* strains are catalase negative.

A3.4 Indole production

A vial of Bactidrop Spot Indole test (Remel, Dartford) was broken open and poured onto filter paper in a petri dish. A cocktail stick was used to smear several colonies onto the filter paper. Colonies changed to green if positive for the production of indole from tryptophan. There was no colour change if negative. *F. necrophorum* strains are indole positive.

Appendix 4: Method for gel extraction of nucleic acid product

The GenElute Gel extraction kit was used for purification of DNA fragments from agarose gels according to the manufacturer's instructions. The DNA fragment was excised from the gel with a clean, sharp scalpel under UV light and weighed in a microfuge tube. Three gel volumes of gel solubilisation solution were added to the tube and the mixture was incubated at 55 °C for 10 minutes with occasional vortexing. One gel volume of 100 % isopropanol was added and the solution was inverted several times. The binding column was then prepared by placing it into a 2 ml collection tube and adding 500 µl column preparation solution before centrifuging at 12,000 x g in an Eppendorf 5415 D benchtop centrifuge for 1 minute and discarding the flow through. The solubilised gel solution was added to the binding column in 700 µl portions and centrifuged at 12,000 x g for 1 minute. The column was then washed with 700 µl wash solution, and centrifuged again at 12,000 x g for 1 minute and the flow through discarded, followed by an additional spin at 12,000 x g for 1 minute. The binding column was transferred to a fresh collection tube and the DNA was eluted by adding 30 µl elution solution, incubating for 1 minute and then centrifuging at 12,000 x g for 1 minute.

Appendix 5: Methods for cell line culture, counting and storage

All sterile work involving the HL-60 cell line was carried out in a Bioair Safeflow microbiological safety cabinet.

A5.1 Propagation from frozen

A vial of HL-60 cells, purchased from Sigma, Gillingham, was thawed in a 37 °C incubator and decontaminated by spraying with ethanol. Thawed cells were added to a 15 ml conical based centrifuge tube containing 9 ml of pre-warmed complete culture medium (IMDM with 2 mM L-glutamine and 20 % foetal bovine serum). The cell suspension was centrifuged at 300 x g for 5 minutes. Culture medium was removed and the cells were resuspended in 5 ml of fresh medium. A 100 μl sample of the cell suspension was taken and the cells were counted and the concentration calculated (see section A5.2). The cell suspension was added to a 75 cm^2 cell culture flask and extra medium was added to make the cell density 3×10^5 cells/ml. The flask was incubated horizontally at 37 °C in 5 % CO_2.

A5.2 Cell counting

Under sterile conditions 100 μl of cell suspension was removed and an equal volume of 0.4 % trypan Blue was added and mixed by gentle pipetting. Approximately 5 – 10 μl of the mixture was added to a clean haemocytometer under a coverslip and viewed under an inverted phase contrast microscope using a x20 magnification lense. The number of viable cells, excluding the vital dye, in 0.1 mm^3 were counted and the concentration of viable cells was calculated using Appendix Equation 1.

Appendix Equation 1: Equation to calculate the concentration of viable HL-60 cells after addition of trypan blue in a 1:1 ratio.

$$C = N \times DF \times 10^4$$

Where C = Concentration of viable cells, N = Number of viable cells counted and DF = Dilution factor.

A5.3 Serial passage

Cell line cultures were maintained between $1 - 9 \times 10^5$ cells/ml at 5 % CO_2 in 37 °C by passaging every $2 - 3$ days in complete culture medium (IMDM with 2 mM L-glutamine and 20 % foetal bovine serum). Cultures were not passaged more than 20 times without thawing a new vial from the liquid nitrogen storage bank.

To passage, the cell suspension was decanted into a centrifuge tube and centrifuged at 300 x g for 5 minutes at room temperature to form a pellet. The media was removed and the cells gently resuspended in 5 ml warm fresh medium. Cells were counted, as described in section A5.2 and the concentration of viable cells calculated.

Cell suspension and pre-warmed medium were added into 75 ml cell culture flasks to provide a cell density of 1×10^5 cells/ml.

A5.4 Storage

After three weeks of propagating the cells, a portion of them were stored in a liquid nitrogen storage bank. Freezing medium was prepared by supplementing complete growth medium (IMDM with 2 mM L-glutamine and 20 % foetal bovine serum) with 5 % (v/v) DMSO. 50 ml of cell suspension was transferred to each of two 50 ml centrifuge tubes and were centrifuged at 300 x g for 5 minutes and the supernatant was discarded. 2.5 ml of freezing medium was added to each 50 ml centrifuge tube and the cell pellets were gently resuspended and pooled into one tube. Cells were added to cryovials in 1 ml aliquots in freezing medium before being put into a Mr. Frosty™ Freezing Container in a -80 °C freezer overnight. This ensured a cooling rate of approximately 1 °C per minute. The following day, when freezing was completed, the cryovials were transferred to a liquid nitrogen storage system.

PARTICIPATION INFORMATION SHEET

Cytotoxic effects of *Fusobacterium necrophorum* leukotoxin on human white blood cells

Researcher: Katie Wright

Staff Supervisor: Dr Patrick Kimmitt and Dr Pamela Greenwell

What is the study about?

You are being invited to take part in a research study on the effects of a leukotoxin on human white blood cells which involves taking a blood sample from you. *Fusobacterium necrophorum* is an anaerobic bacterium with the ability to cause infections in humans. We are specifically investigating the effects of the leukotoxin it produces on human white blood cells. There is evidence from other studies that it causes cell death to bovine white blood cells and minimal research showing that this happens to human white blood cells also. The aim of the research is to test a larger number of *Fusobacterium necrophorum* strains, isolated from previous human infections, than has been tested before to investigate the effect of this toxin on human white blood cells.

What is involved?

The study will involve you giving a small blood sample for us to extract your white blood cells from.

Blood samples will be taken by people experienced in the technique. Standard practice commonly used in medical environments with be followed. A needle will be inserted into your arm and up to 40ml of blood will be taken, or about 3 tablespoons per collection.

With your consent, we may wish to use photographs of your sample material in academic publications. This is most likely to be an image of a white blood cell before or after exposure to leukotoxin.

Will I get paid?

No, there is no cash payment for participation in this research project.

Please note:

- Participation is entirely voluntary.
- You have the right to withdraw at any time without giving a reason.
- You have the right to ask for your data to be withdrawn as long as this is practical, and for personal information to be destroyed.
- Your data will be confidential. No individuals will be identifiable from any collated data, written report of the research, or any publications arising from it.
- All personal data will be kept in a locked cupboard on University premises.
- Please notify us if any adverse symptoms arise during or after the research.
- If you wish you can receive information on the results of the research.
- The researcher can be contacted after participation by email (Katie.wright@my.westminster.ac.uk) or by telephone (0207 911 5000 ext 64404).

CONSENT FORM

Title of Study: Cytotoxic effects of *Fusobacterium necrophorum* leukotoxin on human white blood cells

Lead researcher: Katie Wright

I have read the information in the Participation Information Sheet, and I am willing to act as a participant in the above research study.

I also consent to the use of photographs of my sample material in academic publications.

Name: _____

Signature: _____ Date: _____

This consent form will be stored separately from any data you provide so that your responses remain anonymous.

I have provided an appropriate explanation of the study to the participant

Researcher Signature _____

UNIVERSITY OF WESTMINSTER⊞

PRIVATE   **AND**   **CONFIDENTIAL**

Katie Wright

14 December 2012

Dear Philip

App. No. 12_13_04

Katie Wright: School of Life Sciences

Mode: MPhil/PhD

Supervisor:

Cytotoxic effects of Fusobacterium necrophorum leukotoxin on human white blood cells

- I am writing to inform you that your application was considered by the Research Ethics Sub Committee (RESC) at its meeting of 22 November 2012, and the proposal was **approved.**

- However, you must provide confirmation that your Supervisor supports this application. Please ask your Supervisor to contact me by email and provide confirmation in electronic format (a signature is not necessary).
- Also you will need to provide me with the name of your Supervisor, this was missing from the Form.
- Please forward this letter to your Supervisor.

If your protocol changes significantly in the meantime, please contact me immediately, in case of further ethical requirements.

Yours sincerely

Huzma Kelly

Secretary, Research Ethics Sub Committee

cc.   Dr. John Colwell (Chair, Research Ethics sub Committee)

           Mike Fisher (Research Degrees Manager)

I am advised by the Committee to remind you of the following points:

1.   Your responsibility to notify the Research Ethics sub Committee immediately of any information received by you, or of which you become aware, which would cast doubt upon, or alter, any information contained in the original application, or a later amendment, submitted to the Research Ethics sub Committee and/or which would raise questions about the safety and/or continued conduct of the research.

2.   The need to comply with the Data Protection Act 1998

3.   The need to comply, throughout the conduct of the study, with good research practice standards

4. The need to refer proposed amendments to the protocol to the Research Ethics sub Committee for further review and to obtain Research Ethics sub Committee approval thereto prior to implementation (except only in cases of emergency when the welfare of the subject is paramount).

5. You are authorised to present this University of Westminster Ethics Committee letter of approval to outside bodies, e.g. NHS Research Ethics Committees, in support of any application for further research clearance.

6. The requirement to furnish the Research Ethics sub Committee with details of the conclusion and outcome of the project, and to inform the Research Ethics sub Committee should the research be discontinued. The Committee would prefer a concise summary of the conclusion and outcome of the project, which would fit no more than one side of A4 paper, please.

7. The desirability of including full details of the consent form in an appendix to your research, and of addressing specifically ethical issues in your methodological discussion.

Academic Services Department
101 New Cavendish Street
Cavendish House, University of Westminster
London, W1W 6XH
T: +44 (0) 20 7911 5051
E: H.Kelly01@westminster.ac.uk
westminster.ac.uk/research

Appendix 7: Ecotin ethics approval and associated participant information sheet and consent form

PARTICIPANT INFORMATION SHEET

The anticoagulant effect of ecotin, a *Fusobacterium necrophorum* serine protease inhibitor

Researcher: Katie Wright

Staff Supervisor: Dr Patrick Kimmitt and Dr Pamela Greenwell

What is the study about?

You are being invited to take part in a research study on the effects of a serine protease inhibitor, known as ecotin, on human blood plasma which involves taking a blood sample from you. *Fusobacterium necrophorum* is an anaerobic bacterium with the ability to cause infections in humans. We are specifically investigating the effects of the serine protease inhibitor it produces on human blood clotting. There is evidence from other studies that a similar protein in *E. coli* binds to and potently inhibits factors required for blood clotting. The aim of the research is to test the activity of ecotin from *Fusobacterium necrophorum* in order to establish whether it has the potential to cause an anticoagulant effect.

What is involved?

The study will involve you giving a small blood sample for us to extract your plasma from.

Blood samples will be taken by people experienced in the technique. Standard practice commonly used in medical environments will be followed. A needle will be inserted into your arm and about 20-40ml of blood will be taken.

Will I get paid?

No, there is no cash payment for participation in this research project.

Please note:

- Participation is entirely voluntary.
- You have the right to withdraw at any time without giving a reason.
- You have the right to ask for your data to be withdrawn as long as this is practical, and for personal information to be destroyed.
- Your data will be confidential. No individuals will be identifiable from any collated data, written report of the research, or any publications arising from it.
- All personal data will be kept in a locked cupboard on University premises.
- Please notify us if any adverse symptoms arise during or after the research.
- If you wish you can receive information on the results of the research.
- The researcher can be contacted after participation by email (Katie.wright@my.westminster.ac.uk) or by telephone (0207 911 5000 ext 64404).

CONSENT FORM

Title of Study: The anticoagulant effect of ecotin, a *Fusobacterium necrophorum* serine protease inhibitor

Lead researcher: Katie Wright

I have read the information in the Participation Information Sheet, and I am willing to act as a participant in the above research study.

Name: _____

Signature: _____ Date: _____

This consent form will be stored separately from any data you provide so that your responses remain anonymous.

I have provided an appropriate explanation of the study to the participant

Researcher Signature _____

UNIVERSITY OF WESTMINSTER▦

18 November 2014

Dear Katie,

Ethics Application: VRE1314-1070

Project title: Doctoral research project

Applicant: Miss Katie Wright

Thank you for providing the Committee with additional changes to protocol and a request for further approval.

The revision to the protocol was considered by correspondence and approved by Chair's Action on 18 November 2014.

Once again, if your protocol changes significantly in the meantime, please contact me immediately, in case of further ethical requirements.

Yours sincerely

Ms Mandy Walton

S&T Research Ethics Committee

I am advised by the Committee to remind you of the following points:

1. Your responsibility to notify the Research Ethics Committee immediately of any information received by you, or of which you become aware, which would cast doubt upon, or alter, any information contained in the original application, or a later amendment, submitted to the Research Ethics Committee and/or which would raise questions about the safety and/or continued conduct of the research.

2. The need to comply with the Data Protection Act 1998.

3. The need to comply, throughout the conduct of the study, with good research practice standards.

4. The need to refer proposed amendments to the protocol to the Research Ethics Committee for further review and to obtain Research Ethics Committee approval thereto prior to implementation (except only in cases of emergency when the welfare of the subject is paramount).

5. The requirement to furnish the Research Ethics Committee with details of the conclusion and outcome of the project, and to inform the Research Ethics Committee should the research be discontinued. The Committee would prefer a concise summary of the conclusion and outcome of the project, which would fit no more than one side of A4 paper, please.

6. The desirability of including full details of the consent form in an appendix to your research, and of addressing specifically ethical issues in your methodological discussion.

---

27 October 2014

Dear Katie,

Ethics Application: VRE1314-1070

Project title: Doctoral research project

Applicant: Miss Katie Wright

I am writing to inform you that your application was considered by the S&T Research Ethics Committee at its meeting of 21 October 2014. The proposal was approved subject to the following:

1. The application omits to cite any specific source of ecotin for use in the proposed project. This omission is further compounded by the attached (and unsigned) COSHH form entitled "Venous Blood Sampling" that lists human blood as a culture type. The committee seeks clarification of the source of ecotin to resolve any uncertainty regarding the potential of harm to the investigators through the culture of pathogens, namely Fusobacterium necrophorum.

2. • Consent forms cannot be destroyed at the end of the study. The length of time that they must be kept depends on the discipline, but the University directive is for at least two years after completion.

Please submit the above documentation or clarifications via the VRE no later than 27 November 2014 or at your earliest convenience.

Yours sincerely

Ms Mandy Walton

S&T Research Ethics Committee

I am advised by the Committee to remind you of the following points:

1. Your responsibility to notify the Research Ethics Committee immediately of any information received by you, or of which you become aware, which would cast doubt upon, or alter, any information contained in the original application, or a later amendment, submitted to the Research Ethics Committee and/or which would raise questions about the safety and/or continued conduct of the research.

2. The need to comply with the Data Protection Act 1998.

3. The need to comply, throughout the conduct of the study, with good research practice standards.

4. The need to refer proposed amendments to the protocol to the Research Ethics Committee for further review and to obtain Research Ethics Committee approval thereto prior to implementation (except only in cases of emergency when the welfare of the subject is paramount).

5. The requirement to furnish the Research Ethics Committee with details of the conclusion and outcome of the project, and to inform the Research Ethics Committee should the research be discontinued. The Committee would prefer a concise summary of the conclusion and outcome of the project, which would fit no more than one side of A4 paper, please.

6. The desirability of including full details of the consent form in an appendix to your research, and of addressing specifically ethical issues in your methodological discussion.

Appendix 8: Clustal alignments of *gyrB* sequences aligned to *Fnf* and *Fnn*-specific TaqMan probe sequences

```
JCM3724    AGGATTGCATGGAGTAGGAATTTCCGTTGTAAATGCCCTTTCAGAGTGGACCGAAGTTAA
ARU01      AGGATTGCATGGAGTAGGAATTTCCGTTGTAAATGCCCTTTCAGAGTGGACTGAAGTTAA
D12        AGGATTGCATGGAGTAGGAATTTCCGTTGTAAATGCCCTTTCAGAGTGGACTGAAGTTAA
Fnf1007    AGGATTGCATGGAGTAGGAATTTCCGTTGTAAATGCCCTTTCAGAGTGGACTGAAGTTAA
B35        AGGATTGCATGGAGTAGGAATTTCCGTTGTAAATGCCCTTTCAGAGTGGACCGAAGTTAA
HUN048     AGGATTGCATGGAGTAGGAATTTCCGTTGTAAATGCCCTTTCAGAGTGGACCGAAGTTAA
           **************************************************** ********

JCM3724    AGTAAAACGAGAAGGAAATGTATACTATCAAAAATATTTAAGAGGAAAACCGATAGAAGA
ARU01      AGTAAAACGAGAAGGAAATGTATACTATCAAAAATATTTAAGAGGAAAACCGATAGAAGA
D12        AGTAAAACGAGAAGGAAATGTATACTATCAAAAATATTTAAGAGGAAAACCGATAGAAGA
Fnf1007    AGTAAAACGAGAAGGAAATGTATACTATCAAAAATATTTAAGAGGAAAACCGATAGAAGA
B35        AGTAAAACGAGAAGGAAATGTATACTATCAAAAATATTTAAGAGGAAAACCGATAGAAGA
HUN048     AGTAAAACGAGAAGGAAATGTATACTATCAAAAATATTTAAGAGGAAAACCGATAGAAGA
           ************************************************************

JCM3724    TGTGAAAATAATTTCCGCTTTAGAGGCTGGAGAAACGACAGGAACCATTGTTACTTTTAA
ARU01      TGTGAAAATAATTTCCGCTTTAGAGGCTGGAGAAACGACAGGAACCATTGTTACTTTTAA
D12        TGTGAAAATAATTTCCGCTTTAGAGGCTGGAGAAACGACAGGAACCATTGTTACTTTTAA
Fnf1007    TGTGAAAATAATTTCCGCTTTAGAGGCTGGAGAAACGACAGGAACCATTGTTACTTTTAA
B35        TGTGAAAATAATTTCCGCTTTAGAGGCTGGAGAAACGACAGGAACCATTGTTACTTTTAA
HUN048     TGTGAAAATAATTTCCGCTTTAGAGGCTGGAGAAACGACAGGAACCATTGTTACTTTTAA
           ************************************************************

JCM3724    ACCGGATATAGAGATTTTTGAAACCGTTATTTTTGAATATGAGGTTTTACAACATCGTTT
ARU01      ACCGGATATAGAGATTTTTGAAACCGTTATTTTTGAATACGAGGTTTTACAACATCGTTT
D12        ACCGGATATAGAGATTTTTGAAACCGTTATTTTTGAATACGAGGTTTTACAACATCGTTT
Fnf1007    ACCGGATATAGAGATTTTTGAAACCGTTATTTTTGAATACGAGGTTTTACAACATCGTTT
B35        ACCGGATATAGAGATTTTTGAAACCGTTATTTTTGAATATGAGGTTTTACAACATCGTTT
HUN048     ACCGGATATAGAGATTTTTGAAACCATTATTTTTGAATACGAGGTTTTACAACATCGTTT
           ************************* ************ *******************

JCM3724    AAAAGAATTGGCATATTTAAATCGTGGACTGGAAATTAATTTATTGGATTGTCGAAATGA
ARU01      AAAAGAATTGGCATATTTAAATCGTGGACTGGAAATTAATTTATTGGATTGTCGAAATGA
D12        AAAAGAATTGGCATATTTAAATCGTGGACTGGAAATTAATTTATTGGATTGTCGAAATGA
Fnf1007    AAAAGAATTGGCATATTTAAATCGTGGACTGGAAATTAATTTATTGGATTGTCGAAATGA
B35        AAAAGAATTGGCATATTTAAATCGTGGACTGGAAATTAATTTATTGGATTGTCGAAATGA
HUN048     AAAAGAATTGGCATATTTAAATCGTGGACTGGAAATTAATTTATTGGATTGTCGAAATGA
           ************************************************************

JCM3724    AATAGG
ARU01      AATAGG
D12        AATAGG
Fnf1007    AATAGG
B35        AATAGG
HUN048     AATAGG
           ******
```

Appendix Figure 1: Fnf probe sequence alignment. All strains in the alignment were homologous to the Fnf probe sequence region, highlighted in yellow.

```
JCM3718    AGGATTGCATGGAGTAGGAATTTCCGTCGTGAATGCTCTTTCAGAGTGGACTGAAGTTAA
BL         AGGATTGCATGGAGTAGGAATTTCCGTCGTGAATGCTCTTTCAGAGTGGACTGAAGTTAA
DAB        AGGATTGCATGGAGTAGGAATTTCCGTCGTGAATGCTCTTTCAGAGTGGACTGAAGTTAA
BFTR-1     AGGATTGCATGGAGTAGGAATTTCCGTCGTGAATGCTCTTTCAGAGTGGACTGAAGTTAA
BFTR-2     AGGATTGCATGGAGTAGGAATTTCCGTCGTGAATGCTCTTTCAGAGTGGACTGAAGTTAA
DJ-1       AGGATTGCATGGAGTAGGAATTTCCGTCGTGAATGCTCTTTCAGAGTGGACTGAAGTTAA
DJ-2       AGGATTGCATGGAGTAGGAATTTCCGTCGTGAATGCTCTTTCAGAGTGGACTGAAGTTAA
           ************************************************************

JCM3718    AGTAAAACGAGAAGGAAATGTATACTATCAAAAATATTTAAGAGGAAAACCGGTAGAAGA
BL         AGTAAAACGAGAAGGAAATGTATACTATCAAAAATATTTAAGAGGAAAACCGGTAGAAGA
DAB        AGTAAAACGAGAAGGAAATGTATACTATCAAAAATATTTAAGAGGAAAACCGGTAGAAGA
BFTR-1     AGTAAAACGAGAAGGAAATGTATACTATCAAAAATATTTAAGAGGAAAACCGGTAGAAGA
BFTR-2     AGTAAAACGAGAAGGAAATGTATACTATCAAAAATATTTAAGAGGAAAACCGGTAGAAGA
DJ-1       AGTAAAACGAGAAGGAAATGTATACTATCAAAAATATTTAAGAGGAAAACCGGTAGAAGA
DJ-2       AGTAAAACGAGAAGGAAATGTATACTATCAAAAATATTTAAGAGGAAAACCGGTAGAAGA
           ************************************************************

JCM3718    TGTGAAAATAATTtCTACTTTGGAGGTTGGAGAAACAACAGGAACCATTGTTACTTTTAA
BL         TGTGAAAATAATTTCTACTTTGGAGGTTGGAGAAACAACAGGAACCATTGTTACTTTTAA
DAB        TGTGAAAATAATTTCTACTTTGGAGGTTGGAGAAACAACAGGAACCATTGTTACTTTTAA
BFTR-1     TGTGAAAATAATTTCTACTTTGGAGGTTGGAGAAACAACAGGAACCATTGTTACTTTTAA
BFTR-2     TGTGAAAATAATTTCTACTTTGGAGGTTGGAGAAACAACAGGAACCATTGTTACTTTTAA
DJ-1       TGTGAAAATAATTTCTACTTTGGAGGTTGGAGAAACAACAGGAACCATTGTTACTTTTAA
DJ-2       TGTGAAAATAATTTCTACTTTGGAGGTTGGAGAAACAACAGGAACCATTGTTACTTTTAA
           ************************************************************

JCM3718    ACCGGATATAGAGATTTTTGAAACTGTTATTTTTGAATATGAAGTTTTACAGCATCGTTT
BL         ACCGGATATAGAGATTTTTGAAACTGTTATTTTTGAATATGAAGTTTTACAGCATCGTTT
DAB        ACCGGATATAGAGATTTTTGAAACTGTTATTTTTGAATATGAAGTTTTACAGCATCGTTT
BFTR-1     ACCGGATATAGAGATTTTTGAAACTGTTATTTTTGAATATGAAGTTTTACAGCATCGTTT
BFTR-2     ACCGGATATAGAGATTTTTGAAACTGTTATTTTTGAATATGAAGTTTTACAGCATCGTTT
DJ-1       ACCGGATATAGAGATTTTTGAAACTGTTATTTTTGAATATGAAGTTTTACAGCATCGTTT
DJ-2       ACCGGATATAGAGATTTTTGAAACTGTTATTTTTGAATATGAAGTTTTACAGCATCGTTT
           ************************************************************

JCM3718    AAAAGAACTGGCATATTTAAATCGTGGACTCGAAATTAATTTATTGGATTGTCGAAATGA
BL         AAAAGAACTGGCATATTTAAATCGTGGACTCGAAATTAATTTATTGGATTGTCGAAATGA
DAB        AAAAGAACTGGCATATTTAAATCGTGGACTCGAAATTAATTTATTGGATTGTCGAAATGA
BFTR-1     AAAAGAACTGGCATATTTAAATCGTGGACTCGAAATTAATTTATTGGATTGTCGAAATGA
BFTR-2     AAAAGAACTGGCATATTTAAATCGTGGACTCGAAATTAATTTATTGGATTGTCGAAATGA
DJ-1       AAAAGAACTGGCATATTTAAATCGTGGACTCGAAATTAATTTATTGGATTGTCGAAATGA
DJ-2       AAAAGAACTGGCATATTTAAATCGTGGACTCGAAATTAATTTATTGGATTGTCGAAATGA
           ************************************************************

JCM3718    AATAGG
BL         AATAGG
DAB        AATAGG
BFTR-1     AATAGG
BFTR-2     AATAGG
DJ-1       AATAGG
DJ-2       AATAGG
           ******
```

Appendix Figure 2: Fnn probe sequence alignment. All strains in the alignment were homologous to the Fnn probe sequence region, highlighted in yellow.

242

Appendix 9: Additional BASys statistics

A

B
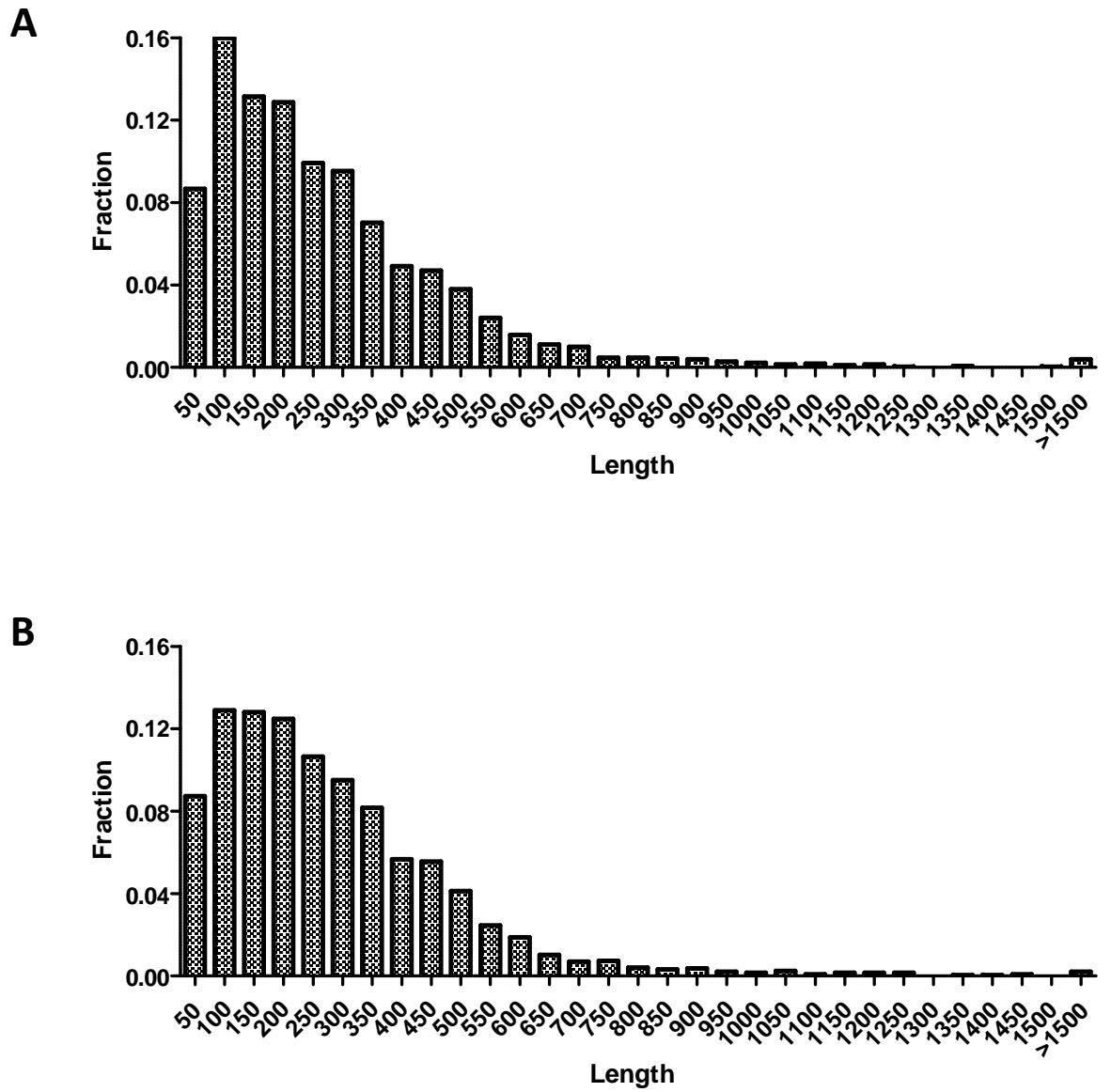
Appendix Figure 3: Gene lengths predicted by BASys in A) JCM 3718 genome data, and B) ARU 01 genome data.

Appendix 10: PCR primers for gap closing leukotoxin sequence

Appendix Table 3: Custom PCR primers for gap closing leukotoxin sequence.

Primer	5' – 3' primer sequence
JCM_3718_gap1_F	AGCGGGAATTGGAATAAGCC
JCM_3718_gap1_R	TCCGTGTTCAGCTCCCATAA
JCM_3718_gap2_F	GGGGAAAAGAATAGTACGGGG
JCM_3718_gap2_R	CCGATGCCTTCAACAGCATT
JCM_3724_gap1_F	TTTTTGTTGGAAGCGAGTATACAA
JCM_3724_gap1_R	AAATAACAGCCATTATCAAAATAACA
JCM_3724_gap2_F	ATTCATGCGGGATTAACTGG
JCM_3724_gap2_R	GCGAACCTGTAACAAAGCTG
JCM_3724_gap3_F	TTTCGGTTCTGGATTAGGAAA
JCM_3724_gap3_R	CCGCTTTGCTGCTTCTTTTA
JCM_3724_gap4_F	GGAACAGCTGTAGAAGATAGAAAAA
JCM_3724_gap4_R	CAATAATATCAGAAGTCTTCACAATGG
ARU_01_gap1_F	TGGGATTGGAATAGTTCATCCTG
ARU_01_gap1_R	TCCTCTTCCGTGTCCTCCTA

Appendix 11: PCR primers for Sanger sequencing the leukotoxin operons of JCM 3718, JCM 3724 and ARU 01

Appendix Table 4: Custom PCR primers for Sanger sequencing the leukotoxin operons.

Primer	5' – 3' primer sequence	Primer	5' – 3' primer sequence
op1_F [a, b]	ACAAGAAATGGAAAGGATTTGGG	op20_F [b]	GGGGATGCCAATGTGAAGAG
op1_R [a, b]	TCGGTAACAGTCCTTTGCCT	op20_R [b]	CAAGAAACAGCCCCTCCAAC
op2_F [a, b, c]	TGGGATTGGAATAGTTCATCCTG	op21_F [b]	GCTTGGCTTTTACAGGAGTAGG
op2_R [a, b, c]	TCCGATATCCTCTTCCGTGTC	op21_R [b]	ATCCCGTCTTCTGTTCCGTT
op3_F [a]	GGGTAAGACAAAATGAGTGGCA	op22_F [b]	GAAAGCGATGGAACGGGAAT
op3_R [a]	AGTTGAATTTTCTCCCACAGCT	op22_R [b]	GCTTCCACGTCATTGTCGAT
op4_F [a]	TCATGCGGGATTAACTGGAGA	op23_F [b]	CAACAACTCAGGTGACGGC
op4_R [a]	TCCTTCCGTCACCAATACTGA	op23_R [b]	TCCTCCGATTCCAACTTGCT
op5_F [a]	TGATGCTTCTGTTTCTGTTGGA	op24_F [b]	GCCGGTTCTCTTTCTACTGC
op5_R [a]	CCCTCTGAAGTCAACTCTGC	op24_R [b]	AACCGCTAACTCCAACTCCA
op6_F [a, c]	GGCTGCACGAAAGGAAGAAA	op25_F [b]	TAGGAGCCGGAGTAGCAGTT
op6_R [a, c]	TCCGGATGTACTGCTTCTCG	op25_R [b]	CGCCAACACCAAAGGAAGAA
op7_F [a, c]	AGGAGCTAAACTTGCTGCAA	op26_F [b]	CGGAGCATTAAACGCAGGAA
op7_R [a, c]	CCTTTATCGGGAGTGTGGGA	op26_R [b]	ATTCGCTCCAACAGAAACGC
op8_F [a, c]	CCGAAATCAGAAGAAAGGCCA	op27_F [b]	GCAAGTGGAGTGGTTTCTGT
op8_R [a, c]	CCTTTGATTGCCCTGCTTGT	op27_R [b]	CCCACAGAAGCTCCTTGAGA
op9_F [a]	TCTGAAGATAAGAGAGCGGATGT	op28_F [c]	CGTAAGGATTTGGGAGCCC
op9_R [a]	TTTTGACTGACAGCAAGCCC	op28_R [c]	GGCTTATTCCAATTCCCGCT
op10_F [a]	AGCGGAAGTAATTGGTGAGGA	op29_F [c]	GCTGTTGAAGGCATCGGTTT
op10_R [a]	CTTGCACAGAAGCCCCTTTT	op29_R [c]	GCAGAAATATATCCCGCTCCG
op11_F [a]	AGCAACAGTCGCTCATACAA	op30_F [c]	GCCAAAGTCGAAGCAACAGA
op11_R [a]	TCGGTTTCTTCTTTGGCTTCT	op30_R [c]	CCAATTCAGCATGATCGGCA
op12_F [a]	GAAGTCATGCAGGGGTAGGA	op31_F [c]	GAGCAGGTTTGGAAGCAGTT

op12_R [a]	ATTCCAGCTCCTTTTGCAGC	op31_R [c]	CCACCTTTGCAATCGTGTCA
op13_F [a, c]	TCAATTTGCAGGAAAGACGGA	op32_F [c]	GGAAGCAATAAGGAAGCCGG
op13_R [a, c]	TCCTTCTCCCTGAGTATCTCCT	op32_R [c]	CTCCTACTGCAGCTTGTCCA
op14_F [a, c]	TCTGAAGGAAAAGGAACGGAAG	op33_F [c]	ATAGCAGTCGGAGTTGGAGG
op14_R [a, c]	TACCGCCCCTACAGAAACTC	op33_R [c]	CCACCAACACCAACAGAACC
op15_F [a, b, c]	ATCGGAATTGACAGCGGAAG	op34_F [c]	TGGAGAAACAGAAGCTCTTGT
op15_R [a, b, c]	GCTCCTCCAGTTCCTTTTACA	op34_R [c]	CTCGTGTTTGATTCATCTTTGGT
op16_F [a, b, c]	TCAACTTCGTGCAAAAGCTTT	op35_F [c]	CGGATGCCAAAAGTCATGCT
op16_R [a, b, c]	TCCCTTCCAATCTTGCCAGA	op35_R [c]	TCTGCAATGGAGGTTACACAAC
op17_F [a]	ATGCCAAAAGTCATGCTGCA	op36_F [c]	TCGGATGCTGGAATGCTACT
op17_R [a]	TCTTCTGCAATTGAGGTTATGCA	op36_R [c]	GCACGACAAACCAAATAGCC
op18_F [b]	TCATGGAGAGAGTGCTTGGC	op37_F [b]	GAAGCAAAGGGTGTAGGAGC
op18_R [b]	TCCGCTTCAATACTTCCCGA	op37_R [b]	ACAAACATCTGCATCCTCTGTC
op19_F [b]	GCAAAGGTAGAGGAAGCAGG		
op19_R [b]	CCCTGCTCCGTTATCACTCA		

a=Used for JCM 3724, b=Used for ARU 01, c=Used for JCM 3718

Appendix 12: Clustal Omega DNA alignment of Sanger sequence data

from JCM 3718, JCM 3724 and ARU 01 leukotoxin operons

Contains genes of *lktB*, *lktA* and partial *lktC:*

```
3718_lkt    ATCGTAAGGatTTGGGAGCCCTATATAGAAAAAGAAGTTACCGTTGCAGATCTTTATACT     60
3724_lkt    ATGGAAAGGATTTGGGAATCCTATATAGAAAAAGAGATTACTATTACAGAGCTTTATACC     60
ARU01_lkt   ATGGAAAGGATTTGGGAATCCTATATAGAAAAAGAGATTACTATTACAGAGCTTTATACC     60
            ** * ************  ****************  ****  ** **** ********

3718_lkt    ATAGTTCAAAAAATCAATGAATTATATCAGGAAAAAGGCTATTTGGTTTGTCGTGCCGTA    120
3724_lkt    ATAGTTCAAAAAATCAATGAATTATATCAGGAAAAAGGCTATTTGGTTTGTCGTGCCGTA    120
ARU01_lkt   ATAGTTCAAAAAATCAATGAATTATATCAGGAAAAAGGCTATTTGGTTTGTCGTGCCGTA    120
            ************************************************************

3718_lkt    TTACCTGCACAAAAAATTCAAAATGGACTTGTGAATATTTTATTGATAGAAGGAAAAACA    180
3724_lkt    TTGCCTGCACAAAAAATTCAAACGGGATCGTAAATATTTTATTGATAGAAGGAAAAACA    180
ARU01_lkt   TTGCCTGCACAAAAAATTCAAACGGGATCGTAAATATTTTATTGATAGAAGGAAAAACA    180
            ** ******************* **   * ** **************************

3718_lkt    GGAGATATTATAATTCAAGGAAATCATTCCACTCGAGAAAAATATATCAGAGAAAGAATT    240
3724_lkt    GGAGATATTACAATTCAAGGAAATCATTCCACTCGAGAAAAATACATCAAAGAAAGAATT    240
ARU01_lkt   GGAGATATTACAATTCAAGGAAATCATTCCACTCGAGAAAAATACATCAAAGAAAGAATT    240
            **********  ********************************** ****  *********

3718_lkt    CCTTTAGAAAAAGACAGAGTTTCGAATTTTTAAAGAATTAGATCGAAGTTTAACACGTTTT    300
3724_lkt    CCCTTAGAAAAAGGTAAAATTTCAAATTTTTAAAGAATTGGATCGAAGTTTAACACGTTTT    300
ARU01_lkt   CCCTTAGAAAAAGGTAAAATTTCAAATTTTTAAAGAATTGGATCGAAGTTTAACACGTTTT    300
            ** **********  * * **** ************** **************** 

3718_lkt    AATCTTACAAATGACAGTCCTATACAGATCAATATGGCTTCCGGAAAAGTTCCGGGAACG    360
3724_lkt    AATCTTACAAATGACAGTCCTATACAGTTAATATGACTTCCGGAAAAGTTCTCGGAACG    360
ARU01_lkt   AATCTTACAAATGACAGTCCTATACAGGTTAATATGACTTCCGGAAAAGTTCTCGGAACG    360
            *************************** * ****** *************** ******

3718_lkt    ACCGATTATTTTGTGCAAATCTATGAACCAAAAAGACAGCAGTTTTTTGTTTTTGCAGAT    420
3724_lkt    ACGGATTATTTTCTGCAAATCTACGAGCCAAAAAGGCAGCAATTTTTTACTTTTGCAGAT    420
ARU01_lkt   ACGGATTATTTTCTGCAAATCTACGAGCCAAAAAGGCAGCAATTTTTTACTTTTGCAGAT    420
            ** ********* ********** ** ******** ***** *******  **********

3718_lkt    AATTTAGGACAAAAAAATACAGGAGAATTACGATGGGGGCTAAATTATATTAATAATAGT    480
3724_lkt    AATTTGGGACAAAAAAATACGGGAGAATTGCGATGGGGAATAAGTTATATCAATAATAGT    480
ARU01_lkt   AATTTGGGACAAAAAAATACGGGAGAATTGCGATGGGGAATAAGTTATATCAATAATAGT    480
            ***** ************** ******** *******  *** ****** *********

3718_lkt    GTTACAGGAAACAGAGATCAACTGTCTCTTACCTCTTTAGTAACAGAAGGAACGGCTTCT    540
3724_lkt    GTCACAGGAAATAGAGATCAGCTATCTCTGACTTCTTTATTAACGGAAGGAACGGCTTCT    540
ARU01_lkt   GTCACAGGAAATAGAGATCAGTTATCTCTGACTTCTTTAGTAACGGAAGGAACGGCTTCT    540
            ** ******** ********   * ***** ** ****** **** ***************

3718_lkt    CTATCTTCTTTTTTATACTTTTCCTGTTTCTAAAAAAGGAACCAAAATATCACTACAACAT    600
3724_lkt    CTATCTTCTCTATTTATACTTTTCCTGTTTCTAAAAAAGGAACCAAAGTGTCATTACAACAT    600
ARU01_lkt   CTATCTTCTCTATTTATACTTTTCCTGTTTCTAAAAAAGGAACCAAAGTGTCATTACAACAT    600
            ********* ****************************************** * *** ********

3718_lkt    TCTGTAGGAAAGTTGAAACATATACAAGGGGCTTTAAAGCATAAAATAACTGGAAACTCT    660
3724_lkt    TCTCTAGGAAAGTTGAAACATATACAAGGAGCTTTGAAACATAAAATAACCGGAAACTCT    660
ARU01_lkt   TCTCTAGGAAAGTTGAAACATATACAAGGAGCTTTGAAACATAAAATAACCGGAAACTCT    660
            *** ************************** ***** ** ********** *********

3718_lkt    TATAGTTATGGGGTTGGAATAGTTCATCCTATTCTGGTTCATGAAAAAAATAAAGTAGAA    720
3724_lkt    TATAGTTATGGGATTGGAATAGTTCATCCTGTTTTGGTTGATGAAAAAAACAAAATAGAA    720
ARU01_lkt   TATAGTTATGGGATTGGAATAGTTCATCCTGTTTTGGTTGATGAAAAAAACAAAATAGAA    720
            ************ ****************** ** ***** ********** *** *****

3718_lkt    CTTTCCTTGGATTGGGTAAAACAAAGGACTGTTACAGATCTATTGAAATTGAAATGGGTA    780
3724_lkt    CTTTCCCTGGATTGGGGAAGGCAAAGGACTGTTACCGATTTATTGAAATTGAATTGGGTA    780
ARU01_lkt   CTTTCCCTGGATTGGGGAAGGCAAAGGACTGTTACCGATTTATTGAAATTGAATTGGGTA    780
            ****** ******** ** ************** *** *** ************** ******

3718_lkt    AATAATAGACTTTCTAAGTATACAGCCGGGAATTGGAATAAGCCATTATGAGGAAGATAGT    840
3724_lkt    AATAATAGACTCTCTAAATATACAGCGGGGAATTGGAATGCATTATGAGGAAGACAGT    840
ARU01_lkt   AATAATAGACTCTCTAAATATACAGCGGGGAATTGGAATAAGTCATTATGAGGAAGACAGT    840
            ********** ***** ******* *********** * *** ****** *** ***

3718_lkt    GTTTTTCTATACAAAGCAAAATATTACAAAGGGAAAATTTATTCCAATTTCGGGAGATGCA    900
3724_lkt    ATTTTCTATACAAAACAAAATATTACAAAGGGAAAATTTATTCCTATCTCAGGAGATGAA    900
ARU01_lkt   ATTTTCTATACAAAACAAAATATTACAAAGGGAAAATTTATTCCTATCTCAGGAGATGAA    900
            ************** ***************************** ** ** ******* *
```

247

```
3718_lkt     AGAAATTATACAAAGTATGATATGTTTCTAATATATCAGAAAAACTTGAAATATAACACT      960
3724_lkt     AAAAAGTATACAAAGTATGATATGTTTCTAATGTATCAGAAAAACTTGAAATACCATACC      960
ARU01_lkt    AAAAAGTATACAAAGTATGATATGTTTCTAATGTATCAGAAAAACTTGAAATACCATACC      960
             * *** ************************ ****************** * **


3718_lkt     TTAGTAACACTAAAGATGGCAGGGCAATATTCTCTGAGTAAAAAATTACCCTCTGTCGAG     1020
3724_lkt     TTGGCAACACTAAGGATGACAGGGCAGTATTCTTTGAGTAAAAAATTGCCCTCTGTAGAA     1020
ARU01_lkt    TTGGCAACACTAAGGATGACAGGGCAGTATTCTTTGAGTAAAAAATTGCCCTCTGTAGAA     1020
             ** * ******** **** ******* ****** ************* ******** **


3718_lkt     CAAATTTATGCAGGAGGAGCCTATAATGTTCGTGGTTATCCGGAAAATTTTATGGGAGCT     1080
3724_lkt     CAAATTTATGCGGGAGGAGCTTATAATGTTCGCGGTTATCCGGAAAGTTTTATGGGAGCT     1080
ARU01_lkt    CAAATTTATGCGGGAGGAGCTTATAATGTTCGCGGTTATCCGGAAAGTTTTATGGGAGCT     1080
             ********** ******** ********** * ********* ************* *************


3718_lkt     GAACACGGAGTTTTTTTCAATGCTGAATTATCAAAATTAGTAGAGAATAAAGGAGAATTT     1140
3724_lkt     GAACATGGAATCTTTTTTAATATTGAATTATCAAAATTAGTAGAGAATAAAGGAGAATTT     1140
ARU01_lkt    GAACATGGAATCTTTTTTAATATTGAATTATCAAAATTAGTAGAGAATAAAGGAGAATTT     1140
             ***** *** * ***** ***  ************************************


3718_lkt     TTTGTTTTTTTAGATGGGGCTTCTCTTCATGGAGAGAGTGCTTGGCAGGAAAATAGAATT     1200
3724_lkt     TTTGTTTTTCTGGATGGAGCTTCTCTTCATGGAGAGAGTGCTTGGCAGGAAAATAGAATC     1200
ARU01_lkt    TTTGTTTTTCTGGATGGAGCTTCTCTTCATGGAGAGAGTGCTTGGCAGGAAAATAGAATC     1200
             ********* * ***** ************************************* ******


3718_lkt     TTTAGCTCAGGTTTTGGATATAAAATAAGGTTTTTAGAAAAAAATAATATTGCTGTTAGC     1260
3724_lkt     TTCAGTTCAGGTTTTGGATATAGACTGAGGTTTTTAGAAAAGAATAATATTGCTGTTAGC     1260
ARU01_lkt    TTCAGTTCAGGTTTTGGATATAGACTGAGGTTTTTAGAAAAGAATAATATTGCTGTTAGC     1260
             ** ** *************** * * ************* ******************


3718_lkt     ATGGCATTTCCATGGAAGAAAAAAATAAATAGTATTTCAGTAGATTCTAATCGAATCTAT     1320
3724_lkt     ATGGCATTTCCATGGAAGAAAACAATAAATAGTATTTCAGTAGATTCCAATCGAATCTAT     1320
ARU01_lkt    ATGGCATTTCCATGGAAGAAAACAATAAATAGTATTTCAGTAGATTCCAATCGAATCTAT     1320
             ********************** ******************** ************ ************


3718_lkt     ATTACAATAAATCATGAATTTTAAAGGGGGTAAGACAAAATGAGCGGCATCAAAAATAAC     1380
3724_lkt     ATTACAATAAATCATGAATTTTAAAGGGGGTAAGACAAAATGAGTGGCATCAAAAATAAT     1380
ARU01_lkt    ATTACAATAAATCATGAATTTTAAAGGGGGTAAGACAAAATGAGTGGCATCAAAAATAAT     1380
             *********************************************** **************


3718_lkt     GTTCAGAGGACAAGGAAGAGGATATCAGATTCTAAAAAAGTTTTAATGATTTTGGGATTG     1440
3724_lkt     GTTAGGAGGACACGGAAGAGGATATCGGATTCCAAAAAAGTGTTAATGATTCTGGGATTG     1440
ARU01_lkt    GTTAGGAGGACACGGAAGAGGATATCGGATTCCAAAAAAGTGTTAATGATTCTGGGATTG     1440
             ***  ******* ************* ***** ******** ********* ********


3718_lkt     TTGATTAACACTATGACGGTGAGGGCTAATGATACAATCACCGCGACTGAGAATTTTGGA     1500
3724_lkt     TTGATTAATACTATGACGGCAGTTGCCGACGATAGTATTAAAGAAACGCAAGGTTTTGGA     1500
ARU01_lkt    TTGATTAATACTATGACGGCAGTTGCCGACGATAGTATTAAAGAAACGCAAGGTTTTGGA     1500
             ******** ********** ** * **** ** * * ** * *******


3718_lkt     ACAAAAATAGAAAAAAAAGGATAATGTTTATGACATTACTACAAACAAGATTCAAGGGGAG     1560
3724_lkt     ACAAAAGTAGAAAAGCAGGGAAATGTTTATGATATTACTACAAATAAGATTAAAGATAAG     1560
ARU01_lkt    ACAAAAGTAGAAAAGCAGGGAAATGTTTATGATATTACTACAAATAAGATTAAAGATAAG     1560
             ****** ******* *** ********** *********** ****** ***    **


3718_lkt     AACGCTTTTAACAGTTTTAATAGATTTGCTTTAACAGAAAATAATATAGCAAATCTATAT     1620
3724_lkt     AATGCCTTTAATAGCTTTGATAAGTTTCATTTGGAACAAAATAATATAGCAAACATGCAC     1620
ARU01_lkt    AATGCCTTTAATAGCTTTGATAAGTTTCATTTGGAACAAAATAATATAGCAAACATGCAC     1620
             ** ** ***** ** *** ***   ***   *   * *************** *  *


3718_lkt     TTTGGGGAAAAGAATAGTACGGGGGTAAATAATCTTTTTAACTTTGTCAATGGAAAAATT     1680
3724_lkt     TTTGGAAGTAAAGGTGGGAAAGAGGCAGAAAATCTTTTTAACTTTGTTAAAGGGGAAAATT     1680
ARU01_lkt    TTTGGAAGTAAAGGTGGGAAAGAGGCAGAAAATCTTTTTAACTTTGTTAAAGGGGAAAATT     1680
             *****    **   * * *  * ** * * ***************** ** ** ******


3718_lkt     GAAGTAGATGGGATTATCAACGGAATTCGAGAAAATAAAATTGGAGGAAATTTATATTTC     1740
3724_lkt     GAAGTAAATGGAGTTATCAATGGAATCCGTGACAATAAAATAGGAGGAAATTTGTATTTC     1740
ARU01_lkt    GAAGTAAATGGAGTTATCAATGGAATCCGTGACAATAAAATAGGAGGAAATTTGTATTTC     1740
             ****** ****  ******* ***** ** ** ** ******** *********** ******


3718_lkt     TTAAGCTCGGAAGGGATGGCAGTAGGAAAAAATGGAGTTATCAATGCTGGTTCTTTTCAT     1800
3724_lkt     CTAAGTTCTGAAGGATTGCTTGTCGGAAAAACCGGGGTTATCAATGCCGGGACTTTCCAT     1800
ARU01_lkt    CTAAGTTCTGAAGGATTGCTTGTCGGAAAAACCGGGGTTATCAATGCCGGGACTTTCCGT     1800
              **** ** ***** **   ** ** ******* ** ********** ** **** * *


3718_lkt     TCTATTATTCCAAAACAAGATGATTTTAAGAAGGCTTTGGAAGAAGCCAAACATGGTAAA     1860
3724_lkt     GCTATTAGTCCGAAAAAGGAGGAATATGAAAAAGCTTTTAAAGACGCTCAAAATAGCAAA     1860
ARU01_lkt    GCTATTAGTCCGAAAAAGGAGGAATATGAAAAAGCTTTTAAAGACGCTCAAAATAGCAAA     1860
              ****** *** *** ** ** ** *  * ** ***** **** ** ** ** ** * ***


3718_lkt     GTTTTTAATGGAATCATTCCAGT---AGATGGAAAAGTAAAAATTCCATTGAATCCGAAT     1917
3724_lkt     GTTTTTGATGGAATTGTTCCACAACAGGATGGGAGTATCAAGATTCCGTTGAATCCAAAT     1920
ARU01_lkt    GTTTTTGATGGAATTGTTCCACAACAGGATGGGAGTATCAAGATTCCGTTGAATCCAAAT     1920
             ****** ******* *****        ***** *   * ** ***** ******** ***


3718_lkt     GGAAGCATTACGGTAGAAGGAAAAATCAATGCTGTTGAAGGCATCGGTTTATATGCGGCG     1977
3724_lkt     GGAAGTATTACGGTAGAAGGAAAAATCAATGCTGTTGAAGATATCGGTTTATATGCCGCA     1980
ARU01_lkt    GGAAGTATTACGGTAGAAGGAAAAATCAATGCGGTGGATGAAATTGGTTTGTATGCAGCA     1980
             ***** ************************** ** ** *  ** ***** ***** **
```

```
3718_lkt    GATATTAGATTGAAAGATACTGCAATACTAAAGACAGGAATTACAGATTTTAAAAATTTA    2037
3724_lkt    GATATTAGATTGAAAGATACTGCAAGACTAAAGACAGGAATTACAGAATTTAAAAATTTA    2040
ARU01_lkt   GATATTAAACTGCCTGAAACAGGAGCTTTGAAAACCGGGGTAACAGATTTTCATCAATTG    2040
            ******* * **    ** ** * *    * ** ** ** * ***** *** *  * **


3718_lkt    GTCAATATTAGTGATC---GAATAAATTCTGGTCTGACCGGAGATTTAAAAGCTACCAAG    2094
3724_lkt    GTAAATATTAATGAAA---CTATTCATGCGGGATTAACTGGAGATTTAACAGCTGTTAAA    2097
ARU01_lkt   GTAAATATAAAAGATTCTAATGTAAATGCAGGATTGTCTGGGGACCTGAAAGCTACTAAG    2100
            ** ***** *  **        *  ** * **  *  * ** ** * * ****   **


3718_lkt    ACAAAATCTGGAGATATTATTCTTTCAGCTCACATAGATTCTCCTCAAAAAGCTATGGGA    2154
3724_lkt    ACAAAGTCGGGAGATATTATTCTTTCAGCTCATGTAGATTCTCCTCAAAAAGCTGTGGGA    2157
ARU01_lkt   ACCGGAACGGGAGATATTATTCTTTCTGCAAAGGTAGAGGAAGCAGGACATGAATTGGAA    2160
            **      * **************** **  *  ****     *  * * *   *** *


3718_lkt    AAAAATTCAACTGTTGGAAAGAGAATAGAAGAATATGTAAAAGGAAATACCAAAGCAAAT    2214
3724_lkt    GAAAATTCAACTGTTGGAAAGAGAATAGAAGAATATGTAAAAGGAAATACCAAAGCAAAT    2217
ARU01_lkt   AGTTCTACCATTT---ATGAACAGATAGGGCGTAACTTCAAAGGAAAAATTAAGGCAAAC    2217
             * * * *    *         ****    * * ******** *  ** *****


3718_lkt    ATTGAATCTGATGCTGTATTGGAAGCAGATGGAAATATAAAAATTAGTGCGAAAGCTACA    2274
3724_lkt    ATTGAATCCGATGCCGTGTTGGAAGCAGATGGAAATATAAAAATTAGTGCGAGAGCTACA    2277
ARU01_lkt   ATTGAAACCTCGGGAAGTATTGAAGCGGAAGGGCATGCAAAAATTCATGCAGAAGCAAGC    2277
            ****** *     *       * ***** ** ** **  ******* ***    *** *


3718_lkt    AATGGGAGATTTATAAAGAAAGAAGGGGAAAAAGAAACTTATAACACTCCTTTAAGTTTA    2334
3724_lkt    AATGGGACATTTGTAAAGAAAGAAGGGGAAAAAGAAATTTACAATACTCCTTTAAGTTTA    2337
ARU01_lkt   AATGGAAAGTTGACTAAAAAAGACGGAGAAAAAGAAGTATACGCTCCGGAATTCAGTTTG    2337
            ***** * **    ** ***** ** ********* **    *    ** *****


3718_lkt    TCAGATGTGGAAGCTTCCGTAAGAGTAAATAAAGGAAAAGTCATAGGAAAGAATGTTGAC    2394
3724_lkt    TCAGATGTAGAAGCTTCCGTAAGAGTAAATAAAGGAAAAGTCATAGGAAAGAATGTTAGT    2397
ARU01_lkt   GCAGAGGTAGAGGCTAGCGTCAAAGTGAATAAAGGAAAGATCAAAGGAAAAAAGGTAGAT    2397
             **** ** ** ***  *** * * *** **********  *** ****** ** **


3718_lkt    ATTACAGCTGAAGCAAAGAATTTCTATGATGCAACTTTAGTTACTAAGCTTGCAAAGCAC    2454
3724_lkt    ATTACAGCTGAAGCAAAGAATTTCTATGATGCAACTTTAGTTACTAAGCTTGCAAAGCAC    2457
ARU01_lkt   ATCAGTGCAGAAGCAAAAAATTATTATGACACCCCCATCCTTACAAAAGTTGGAAAAATA    2457
            ** *  ** ******** **** ***** ** *   * **   *** ***  ** ** *


3718_lkt    TCTTTTAGCTTTGTTACAGGTTCTATTTCTCCTATCAATTTAAATGGATTTTTAGGTTTA    2514
3724_lkt    TCTTTCAGCTTTGTTACAGGTTCGCTTTCTCCTATCAATTTAAATGGATTTTTTAGGTTTA   2517
ARU01_lkt   GCTTTTTCTGTCGGGACAGGCTCTTTATCTCCTATCAATATGAACGGGGCCTTAGGTTTA    2517
             ****        * * ***** **     * ************ * ** **     ********


3718_lkt    TTGACAAGTAAGTCCAGTGTCGTTATTGGAAAAGATGCCAAAGTCGAAGCAACAGAAGGA    2574
3724_lkt    TTGACAAGTAAATCCAGTGTCGTTATTGGAAAAGATGCCAAAGTGGAGGCAAGGGAAGGA    2577
ARU01_lkt   TTAAAAAGTAAGGCGAGTGTTTTTTATTGGAAAAGATGCAACAGTAGAGTCAACGGAGGGA   2577
            ** * ******  * ***** *   **************** * ***  **  ** ***


3718_lkt    AAGGCAAATATTCATTCTTACAGTGGAGTAAGAGCAACTATGGGAGCAGCTACTTCTCCA    2634
3724_lkt    AGTGCAAATATTCATTCTTATAGTGGAGTAAGAGCGACTATGGGAGCAGCTACTTCTCCA    2637
ARU01_lkt   GAAGCGAATATTCGTTCTTATAGCGGAGTAAGAGCTTCAATGGGAGCCTCTACTTCTCCT    2637
             ** ******* ** ***** ** **********  *  ******** * *********


3718_lkt    TTAAAAATTACCAATTTATATTTGGAGAAAGCCAATGGAAAACTTCCTAGTATCGGAGCG    2694
3724_lkt    TTAAAAATTACTAATTTATATTTGGAGAATGCTGATGGAAAACTTCCCAGCATCGGAGCA    2697
ARU01_lkt   ATAAAACTTACTGACCTTTATTTGAAAACATTGGAGGGAAAATTACCTAGCGTTGGAGCT    2697
             ***** ****    * ** ***** *  *      * ****** * * ** ** * *****


3718_lkt    GGATATATTTCTGCAAAAAGTAATTCCAATGTAACTATTGAAGGAGAAGTAAAATCGAAG    2754
3724_lkt    GGGTATATTTCTGCAAAAAGTGACTCCAATGTAACCATTGAAGGAGAAGTAAAATCGAAG    2757
ARU01_lkt   GCATATATTTCTACTTCCAGTGAGTCTGATGTCACTGTGGAAGGAAAGGTGAGCTCAAAA    2757
            *  ********* *   * ***   ** * ** **  * ****** * **   ** *****


3718_lkt    GGAAGAGCAGATATTACTTCAAAATCTGAAAATACTATTGATGCTTCTGTTTCTGTTGGA    2814
3724_lkt    GAAAGAGCAGATATTACTTCAAAATCTGAAAATACTATTGATGCTTCTGTTTCTGTTGGA    2817
ARU01_lkt   GAAAAGGCAAATATTACTTCTAAATCAGAAAATACGATTGATGCTTCTGTTTCAGTCGGA    2817
            * ** *** ********** ***** * ******* ***************** ** ***


3718_lkt    ACGATGAGAGATTCCAATAAAGTAGCTCTTTCAGTATTGGTGACGGAAGGAGAAAATAAA    2874
3724_lkt    ACGATGAGAGATTCCAATAAAGTAGCTCTTTCAGTATTGGTGACGGAAGGAGAAAATAAA    2877
ARU01_lkt   ACTATAAGAGATTCAAATAAAGTAGGAATTTCTGTTTTGGTTACAGAGGGAAAAAATCAT    2877
            ** ** ******** ********* *  **** ** ***** ** ** *** ***** *


3718_lkt    TCTTCCGTCAAGATTGCTAAAGGAGCAAAAGTAGAATCAGAAACGGATGATGTAAATGTG    2934
3724_lkt    TCTTCCGTCAAGATTGCTAAAGGAGCAAAAGTAGAATCAGAAACGGATGATGTAAATGTG    2937
ARU01_lkt   TCTTCTGTCAAAATCGCTGAGGGAGCCAAAGTGGAATCAGACGGGG---ATGCCAATGTG    2934
            ***** ***** ** *** * ***** ***** ***** **  **    *** ******


3718_lkt    AGAAGTGAAGCGATTAATTCCATTCGAGCTGCTGTAAAAGGTGGATTGGGGGATAGTGGT    2994
3724_lkt    AGAAGTGAAGCGATTAACTCCGTTCGAGCTACTGTAAAAGGTGGATTGGGGGATAGCGGT    2997
ARU01_lkt   AAGAGTGAAGCCGTTAACTCTATTCGAGCAGCTGTTAAAAGCGGTTTGAGTGATAACGGA    2994
            * ******** **** ** *  ******* **** ***** ** ***  * **** **


3718_lkt    AATGGGGTTGTGGCTGCAAATATTTCTAACTATAATGCTTCCTCCCGTATAGATGTAGAT    3054
3724_lkt    AATGGGGTTGTGGCTGCAAATATTTCTAATTATAATGCTTCCTCCCGTGTAGATGTAGAT    3057
ARU01_lkt   GCAGGGGTTGTTGCAGCCAATATCTCCAATTATAACAGTTCTTCAAAGGTTGTAGTGGAT    3054
            *  ******** ** ***** ** ** ** *****   **** **    * * ** ***
```

```
3718_lkt    GGATATCTACATGCCAAGAAGCGACTAAATGTGGAGGCTCATAACATTACTAAAAATAGT    3114
3724_lkt    GGATATCTACATGCCAAGAAGCGACTAAATGTGGAGGCTCATAACATTACTAAAAATAGT    3117
ARU01_lkt   GGAGAAGTCCATGCAAAGAAGAGACTAAACACGGAAGCATATAATATTACGAAGCAAAAT    3114
            *** *   * ***** ****** *******   *** ** **** ***** **  * * *


3718_lkt    GTTCTGCAAACAGGATCTGATTTGGGAACTTCCAAGTTTATGAATGATCACGTTTATGAA    3174
3724_lkt    GTTCTACAAACAGGAACTAATTTGGGAACTTCCAAGTTTATGAATGATCACGTTTATGAA    3177
ARU01_lkt   GTTTTACAAACGGGAACGAAAACAGGGACTTCGAAACTGATGAATGAACTTGTTTTTGAA    3174
            *** * ***** *** * *   ** ***** **  * ******** *  **** ****


3718_lkt    TCAGGTCATCTAAAATCAATTTTAGATGCAATAAAACAGCGGTTTGGAGGAGACAGTGTC    3234
3724_lkt    TCAGGTCATCTAAAATCAATTTTAGATGCAATAAAACAGCGGTTTGGAGGGGACAGTGTC    3237
ARU01_lkt   TCCAGTCATACCAAAGCTTTGATAGATGCTTTACAATATCGATTTGGCGGAG--------    3226
            **  *****   *** *  * *******  ** ** * ** ***** ** *


3718_lkt    AATGAGGAAATAAAGAATAAGCTAACAGACTTATTTAGTGTCGGTGTGTCTGCAACCATA    3294
3724_lkt    AATGAGGAAATAAAGAATAAGCTAACAGACTTATTTAGTGTCGGTGTGTCTGCAACCATA    3297
ARU01_lkt   -ATGAAAAAATCAAATCTAAGTTGACTGATTTATTCAGTGTTGGAGCTTCTGCAACCGTA    3285
             ****  **** **   **** * ** ** ***** ***** **  *   ********* **


3718_lkt    GCAAATCATAATAATTCTGCTTCTGTGGCAATAGGAGAGAGTGGAAGACTTTCTTcaGGA    3354
3724_lkt    GCAAATCATAATAATTCTGCTTCTGTGGCAATAGGAGAGAGTGGAAGACTTTCTTCAGGA    3357
ARU01_lkt   GCAAATCACAATAATACTTCTTCTGTTAGCCTGGGGCAAAAATCCAAGTTAGTTTCCGGA    3345
            ******** ****** ** *******    * **  * *    *   *   *** ***


3718_lkt    GTGGAAGGGAGTAATGTAAGGGCATTAAATGAAGCTCAAAATCTTCGAGCGACtacGTca    3414
3724_lkt    GTGGAAGGGAGTAATGTAAGGGCATTAAATGAAGTTCAAAATCTTCGAGCGACTACGTCA    3417
ARU01_lkt   GTCGAAGGAACGAATATAAAAGCTTTGACAGAAACCCAACAACTTCGTTCTACCACAACA    3405
            ** ***** *  *** ***  ** ** *  ***   *** ****** * ** **  **


3718_lkt    aGTGGAAGTGTGGCTGTACGAAAGGAAGAAAAAAAGAAACTTATTGGAAATGCAGCAGTT    3474
3724_lkt    AGTGGAAGTGTGGCTGCACGAAAGGAAGAAAAAAAGAAATTTATTGGAAATGCAGCAGTT    3477
ARU01_lkt   AGTGGAAGCTTGGCAACAAGAGGAGAAGAAAAGAAAAAGTTGGTTGGAAATGCAGCTGTT    3465
            ********  ****    * **   ******** ** **    * ************* ***


3718_lkt    TTTTATGGAAACTATAAAAATAATGCTTCTGTGACAATTGCCGATCATGCTGAATTGGTA    3534
3724_lkt    TTTTATGGAAACTATAAAAATAACGCTTCTGTAACTATCGCAGATCAAGCAGAGTTGACT    3537
ARU01_lkt   TTCTATGGAAACTATAAAAATGATGCTTCCGTAACAATTGCGGACGATGCTCAAGTCGTA    3525
            ** ***************** * ***** ** ** ** ** ** ** **  * **  *  *


3718_lkt    TCGGAAGGAAAAATTGATATCAACAGTGAAAATAAAATTGAATATAAAAATCCTTCAAAA    3594
3724_lkt    TCAGAGGGTAAAATTGATACAATCAGTGAAAATAAAACTGAATATAAAAATCCTTCAAAA    3597
ARU01_lkt   TCGGAAGGAAAGATTGATACGATCAGTGAAAATAAGCTCGAGTACAAAAATCCGTCGAAA    3585
            ** ** ** ** ** *******  * ***********   ** **  ********* ** ***


3718_lkt    ATGGCAAAGTCTGTTATTGATAAATTAGAACTTTTAAAGAGAGCTTTTGGAAAAGAAACG    3654
3724_lkt    ATGGCAAAGTCTGTTATTGAAAAATTAGAACTTTTAAAGAGAGCTTTTGGAAAAGAAACG    3657
ARU01_lkt   ATGGCGGAAGAAGTTGTGGAAAAATTAGAAATTTTAAAAAGAGCCTTTGAAAAAGAGAAG    3645
            *****  *   *** * ** ********* ******* ***** **** ****** * *


3718_lkt    AAAAC---TCCAGAATATGATCCGAAAGATATTGAATCTATTGAAAAATTATTGAATGCA    3711
3724_lkt    AAAAC---TCCAGAATATGATCCGAAAGATATTGAATCTATTGAAAAATTATTGAATGCA    3714
ARU01_lkt   AAAGATGAAACGACATTTGATCCGAAAGACATTGATTCCATGAAAAAATTATTGACAGAA    3705
            ***        *   ** ************ ***** ** **  *********** * *


3718_lkt    TTTTCAGAAAAATTGGATGGAAAACCGGAGCTTTTACTAAATGGTGAAAGAATGACAATT    3771
3724_lkt    TTTTCAGAAAAATTGGATGGAAAACCGGAGCTTTTACTAAATGGTGAAAGAATGACAATT    3774
ARU01_lkt   TTTTCTAAAAAATTGGATGATAAGCCTGAAATTTTATTGAACGGAGAAAAAATAACGATT    3765
            *****  *********** ** ** ** **  ***** *** ** ** ** ** ** ***


3718_lkt    ATTCTTCCGGATGGAACTTCAAAAACAGGAACTGCTATAGAAATTGCAAACTATGTTCAG    3831
3724_lkt    ATTCTTCCGGATGGAACTTCAAAAACAGGAACTGCTATAGAAATTGCAAACTATGTTCAG    3834
ARU01_lkt   GTTCTTCCAGATGGAACTTCTAAAACAGGAACGGTTCCGGAACTTGCAGAATATGTGAAA    3825
             ******* *********** ********** *  *   *** * *** ***** * *


3718_lkt    GGAGAAATGAAAAAATTAGAGGAAAAATTACCGAAAGGATTTAAAGCTTTTTCAGAAGGA    3891
3724_lkt    GGAGAAATGAAAAAATTAGAGGAAAAATTACCGAAAGGATTTAAAGCTTTTTCAGAAGGA    3894
ARU01_lkt   GAGGAAATGAAAAAACTGGAAGCAAAATTACCGACAGGATTTAAAGCTTTTTCTGAAGGA    3885
            *  ************ * ** * ********** ***************** * ******


3718_lkt    TTGAGTGGACTGATTAAAGAAACTTTGAATTTTACAGGAGTAGGAAATTATGCAAATTTT    3951
3724_lkt    TTGAGTGGACTGATTAAAGAAACTTTGAATTTTACAGGAGTAGGAAATTATGCAAATTTT    3954
ARU01_lkt   TTAAGCGGATTGTTAAAAGAAAGCTTGGCTTTTACAGGAGTAGGAAATTATGCAAATTTC    3945
            ** ** *** ** * ******* ***   *****************************


3718_lkt    CACACTTTTACCTCTTCCGGAGCTAATGGAGAAAGAGATGTTTCTTCTGTGGGAGGAGCT    4011
3724_lkt    CACACTTTTACCTCTTCCGGAGCTAATGGAGAAAGAGATGTCTCTTCTGTGGGAGGAGCT    4014
ARU01_lkt   CATACATTTACCTCGGCAGGAACCAATGCAAAAAGAGATACTTCTTCTGTTGGAGGGGCT    4005
            ** ** ********  ** *** * **** * *********  ******** ***** ***


3718_lkt    GTTTCGTGGGTAGAACAGGAGAATTATAgcaaGGTATCCGTTGGAAAAGGAGCTAAACTT    4071
3724_lkt    GTTTCGTGGGTAGAACAGGAGAATTATATAGCAAGGTATCCATTGGAAAAGGAGCTAAACTT    4074
ARU01_lkt   GTTTCTTGGGTAGAAATGGATAATCACAGTAAGGTTTCTATTGGAAGAGGAAGCAAAATT    4065
            ***** *********   *** *** * * ** *** **  ****** ***** **   ** **


3718_lkt    GCTGCAAAAAAAGATTTAAATATAAAAGCTATCAATAAAGCAGAAACAGTGAATTTAGTT    4131
3724_lkt    GCTGCAAAAAAAGATTTAAATATAAAAGCTATCAATAAAGCAGAAACAGTGAATTTAGTT    4134
ARU01_lkt   ACTGCAAAGGGAGATTTGAATGTAAAGGCAATCAATAAAACGGAAACTGTAAATCTGGTT    4125
            *******    ****** *** **** ** * ******** *  ***** ** *** * ***
```

```
3718_lkt    GGAAATATTGGACTTGCGAGAAGCAGTACATCCGGAAGTGCAGTCGGAGGAAGATTAAAT    4191
3724_lkt    GGAAATATTGGACTTGCGAGAAGCAGTACATCCGGAAGTGCAGTCGGAGGAAGATTAAAT    4194
ARU01_lkt   GGAAAATTAGGTCTTTCAAAGAGTAGTGATTCGGCAACCGCAGTCGGTGGAGGATTGAAT    4185
            *****  * ** *** * *  ** ***   ** * **  ******* *** **** ***


3718_lkt    GTTCAAAGATCGAAAAATTCAGCTATCGTAGAAGCTAAAGAAAAAGCTGAATTATCAGGA    4251
3724_lkt    GTTCAAAGAACGAAAAATTCTGCTATTGTAGAAGCTAAAGAAAAAGCTGAATTATCAGGA    4254
ARU01_lkt   GTTCAAAAATCAAATACTTCCTCTATTGTAGTGACAAAAGAGAAAGCCGAGTTATCCGGG    4245
            *******  * * ** * ***  **** ****   * ***** ***** ** ***** **


3718_lkt    GAAAATATTAATGCAGATGCATTGAACAGACTTTTTCATGTAGCGGGATCTTTTAATGGT    4311
3724_lkt    GAAAATATTAATGCAGATGCATTGAACAGTCTTTTTCATGTAGCAGGATCTTTTAATGGT    4314
ARU01_lkt   GAAAATATTCATACAAATGCTTTCAATAATGTCTTCCATGTAGCGGCTTCTTTGAATGCC    4305
            ********* ** ** **** ** ** *   * ** ******** *  ***** ****


3718_lkt    GGCTCAGGTGGGAATGCAATCAATGGAATGGGAAGTTATAGTGGAGGTATCAGTAAGGCA    4371
3724_lkt    GGCTCAGGTGGGAATGCAATCAATGGAATGGGAAGTTATAGTGGAGGTATCAGTAAGGCA    4374
ARU01_lkt   GGAACGGGTGGAAACGTATCAATGGAATGGGAAGCTATAGCGGAGGAAGCAGTAAGTCG     4365
            **  * ***** ** *  ***************** ***** **** * ******* *


3718_lkt    AGAGTTTCCATTGATGACGAAGCATATTTGAAAGCTAATAAAAAAATTGCTTTAAACAGT    4431
3724_lkt    AGAGTTTCCATTGATGACGAAACATATTTGACAGCTAATAAAAAAATTGCTTTAAACAGT    4434
ARU01_lkt   AGAGTTTCCATTGATGATGAAACACACTTAAAGGCGAAGAAAAAAATATATTTAGACAGT    4425
            ***************** *** ** * ** *  ** ** ********   **** *****


3718_lkt    AAGAATGATACTTCTGTTTGGAATGCTGCCGGTTCAGCGGGAATCGGAACGAAAAATGCG    4491
3724_lkt    AAGAATGATACTTCTGTTTGGAATGTTGCCGGTTCAGCGGGAATCGGAACGAAAAATGCG    4494
ARU01_lkt   GAAAATAATACCTCTGTAACCAATGTTGCCGGAGCCGTAGCAATCGGAAGCAAAAATGCA    4485
             * *** **** *****   **** ****** *  *  * ******** ********


3718_lkt    GCGGTCGGGGTTGCTGTTGCGGTAAATGATTATGATATTTCAAACAAAGCTTCCATTGAA    4551
3724_lkt    GCAGTCGGGGTTGCTGTTGCGGTAAATGATTATGATATTTCAAACAAAGCTTCCATTGAA    4554
ARU01_lkt   GCAGTGGGGAGCTGCTGTGGCAATCAATGATTACGATATTTCAAACAAAGTTTCTATCGAA    4545
            ** ** ** * ****** **  * ******** **************** *** ** ***


3718_lkt    GATAATGACGAAG---------GACAAAGTAAATATGATAAGAATAAAGATGAAGTAACA    4602
3724_lkt    GATAATGACGAAG---------GACAAAGTAAATATGATAAGAATAAAGATGAAGTAACA    4605
ARU01_lkt   GACAATGATACCGAAGATGGAGGTAAAAGTAAGTACGATAAAAATAAAGAGGAAGTAACC    4605
            ** *****    *       *  ******* ** ***** ******** ********


3718_lkt    GTAACTGCGGAATCTTTAGAAGTAGATGCAAAAACGACCGGAACAATCAACAGTATTTCT    4662
3724_lkt    GTAACTGCGGAATCTTTAGAAGTAGATGCAAAAACGACCGGAACAATCAACAGTATTTCC    4665
ARU01_lkt   GTTACAGCAGAGGATTTACATGCAAATGCAAAAACGACAGGAACCATTAATGCTGTTTCC    4665
            ** ** ** **    **** * * * ************* ***** ** **    * ****


3718_lkt    GTTGCCGGAGGAATTAATAAGGTTGGAAGTAAACCGAGTGAAGAAAAACCGAAATCAGAA    4722
3724_lkt    GTTGCCGGAGGAATCAATAAGGTTGGAAGTAAACCAGTGAAGAAAAACCGAAATCAGAA    4725
ARU01_lkt   GTTGCCGGTGGAATTAGCAAGGTTGGAGAAGCGGAAAGCGA------------------    4706
            ******** ****** *  *********      ** **


3718_lkt    GAAAGACCAGAGGGATTTTTTGGCAAAATCGGAAACAAAGTGGACTCTGTAAAAAATAAA    4782
3724_lkt    GAAAGGCCAGAGGGATTTTTTGACAAAATCGGAAATAAAGTGGATTCTGTAAAAAATAAA    4785
ARU01_lkt   -----TGGAACGGGAATTTTTAGTAAAATTGGGAACAAGGTGGATTCTGTCAAAGGGAAA    4761
                 *  **** ***** ***** ** ** ** ** ** ** ***


3718_lkt    ATTACGGATAGTATGGATTCATTAACAGAAAAAATTACaAAtTAcATTTCTGAAGGAGTA    4842
3724_lkt    ATTACGGATAGTATGGATTCACTAACAGAAAAAATTACAAATTACATTTCTGAAGGAGTA    4845
ARU01_lkt   ATTACAGATGCTATCGGGTTTGTTACAGATCAAGTGACAAATTATGTCTATAGAGGAAGT    4821
            ***** ***  * **  * *  **  * ***** ** * ****** *  * *  ****


3718_lkt    AAAAAAGCGGGGAATCTTCCTTCGAACGTTTCTCATACTCCCGATAAAGGACCGTCTTTC    4902
3724_lkt    AAAAAAGCGGGAAATCTTCCTTCGAACGTTTCCCACACTCCCGATAAAGGACCGTCTTTC    4905
ARU01_lkt   AACGGAACAGAAGACGGGATTCCAAATATTTCGAAAGAATCTTCTAAATTGCCTTCTTTC    4881
            **    * * *    *      * ** **** *     *      ****   ** ******


3718_lkt    AGTTTGGGAGCTTCTGGAAGTGTTTCTTTCAATAATATTAAAAAGGAAACATCTGCTGTC    4962
3724_lkt    AGTTTAGGGGCTGCCGGCAGTGTTTCTGTTTAATAATAAAGAAACTTCCGCAGTG        4965
ARU01_lkt   AGTTTAGGAGCTTCCGGAAGTGTTTCCGTGAATAATATTAAAAAGGAAACTTCTGCTGTT    4941
            ***** ** *** ** ** ******  *  ****** ** ** ***** ** ** **


3718_lkt    GTAGATGGAGTAAAGATAAATTTGAAGGGAGCAAATAAAAAGGTAGAGGTGACTTCTTCT    5022
3724_lkt    GTTGACGGAGTGAAAGTAAACTTAAAAGGAGATAATAAAAAGGCGGAAGTTATTGCTTCA    5025
ARU01_lkt   ATGGATGGAGTGAAGATAAGTCTAAGCGGGGCAAATAAAGAGGTAGGAGTGACTTCTTCC    5001
             * ** ***** **  ***   *   ** ** ****** ***  *   ** * * ****


3718_lkt    GATTCTACTTTTGTTGGAGCATGGGGCGGATCTGCTGCACTTCAGTGGAATCATATTGGA    5082
3724_lkt    GATTCAGCTTTTATTGGTGCTTGGAGTGGAGCAGCAGCACTTCAATGGAATCATGTCGGA    5085
ARU01_lkt   GACTCCACTTTCGTTGGAGCATGGAGTGGATCTGCCGCACTTCAATGGAATCATATCGGA    5061
            ** ** ****   **** **  *** ** ** ** * *********  ** ********** * ***


3718_lkt    AGTGGAAATAGCAACATCAGTGCTGGTTTAGCTGGAGCGGCTGCTGTAAATAATATTCAA    5142
3724_lkt    AGTGGAGAGAGTAATACGACGGTTGGATTGGCTGGAGCGGCAGCTGTAAATAATGTTCAA    5145
ARU01_lkt   AGCGGAAATAGTAATTTCAGTGCTAGTTTAGCCGGAGCGGCTGCTGCAAATAACATTCAA    5121
            ** *** * ** **       * *  ** ** ******** ****** ***** *****


3718_lkt    AGTAAAACAAGTGCTTTGGTTAAAAATAGTGATATTCGAAATGCCAATAAATTTAAAGTA    5202
3724_lkt    AGTAAAACAAAGGCTTTTGTGAAAGACAGTAGTATTACCAATGCCAATAAGTTTAAGGTA    5205
ARU01_lkt   AGTAAAACAAATGCGGTGGTTAAAAATAGTAGTATTCAAAATGCCAATAAATTTAAAGTA    5181
            **********  **  * * * ** *** *** **** *   ********** ***** ***
```

251

```
3718_lkt     AATGCTTTGAGTGGAGGAACTCAAGTAGCAGCAGGAGCAGGTTTGGAAGCAGTTAAAGAA      5262
3724_lkt     AATGCAGTGAGTGGAGGAACACAGGTAGCAGCCGGTATGGGATTAGAAGCTGTTAAAGAA      5265
ARU01_lkt    AATGCTTTGAGTGGAGGAACTCAAGTAGCGGCGGGAGCAGGTTTGGAAGCAGCCAAAGAA      5241
             *****  ************* ** ***** ** **     ** ** ***** *  ******


3718_lkt     AGTGGAGGACAAGGAAAAAGTTATCTATTGGGAACTTCTGCTTCTATCAACTTAGTGAAC      5322
3724_lkt     AGTGGAGGACAAGGAAGAAGTTATTTATTAGGAACTTCAGCTTCTATGAACTTTATTGAT      5325
ARU01_lkt    AGTGGAGGTCAAGGAAAAAGTTATCTATTGGGAACATCCGCTTCTATCAACTTAGTGGAC      5301
             ******** ******* ******* ****  ***** ** ******** *****   *  *


3718_lkt     AATGAAGTTTCTGCAAAATCAGAAAATAATACAGTAGCAGGAGAATCTGAAAGCCAAAAA      5382
3724_lkt     AATGAAGTGTCTTCAGAATCTGAAAACAATATTATCACCGGAGAATCTGAAGATAAGAGA      5385
ARU01_lkt    AATGAAGTTTCTGCAAAATCAGAAAATAATATATAGTAGAAGGTGAATCTGAAAACAAAAA      5361
             ******** *** ** **** ***** ****     *      ** *********   * * *


3718_lkt     ATGGATGTTGATGTCACTGCTTATCAAGCGGACACCCAAGTGACAGGAGCTTTAAATTTA      5442
3724_lkt     GCGGATGTAGATGTTACAGCTTATGAAAGTGATACTCAAGTGACGGGAGGATTCAATTTA      5445
ARU01_lkt    ATGAATGTTGATGTCACTGCTTATGAAGCAGATACTCAAGTGACAGGAGCCTTGGATTTA      5421
             * **** ***** ** ****** **    ** ** ******** ****  **  *****


3718_lkt     CAAGCTGGAAAGTCAAATGGAACTGTAGGGGCTACTGTGACTGTTGCCAAATTAAACAAC      5502
3724_lkt     CAAGCAGGGCAATCAAAGGGAACAGTAGGAGCCGCCGTTACGGTTGCGAAATTAAATAAT      5505
ARU01_lkt    CAAGCCGGAAAAATCATCAGGAACTGCAGGAGCTGCTATCACTGTTGCTAAGTTAAACAAC      5481
             ***** **    * ***     ***** * *** **   * * ** ***** ** ***** **


3718_lkt     AAAGTAAATGCTTCTATTAGTGGTGGGAGATATACTAACGTTAATCGAGCGGACGCAAAA      5562
3724_lkt     AAGGTAAAAGCAGGTATTAAAGGAGGAAATTATACAAACATAAATCGGGCAGACACAAAA      5565
ARU01_lkt    AAGGTAAATGCTTCTATTAGTGGAGGAAAATATACCAAGGTCAATCGAGCGGATGCAAAA      5541
             ** ***** **  *****  ** ** *  ***** **  * ***** ** **    *****


3718_lkt     GCTCTTTTAGCAACCACTCAAGTGACTGCTGCAGTGACGACGGGAGGGACAATTAGTTCT      5622
3724_lkt     GCACTTTTGGCGACAACACAAGTGACTTCTGCTTTATCTGTGGGAGGGACTAATAGTTTC      5625
ARU01_lkt    GCTCTTTTAGCAACAACTCAGGTGACGGCTGCTGTGGCAACAGGAGGAACAATTAGTTCC      5601
             ** ***** ** ** ** ** ***** ***  ****  *       ***** ** * *****


3718_lkt     GGAGCGGGATTAGGAAAATTATCAAGGGGCTGTTTCTGTCAATAAGATTGACAATGACGTG      5682
3724_lkt     GGTTCTGGATTAGGAAAATTATCAGGGAGCTGTTTCTGTCAATAACATCAATAATGATGTA      5685
ARU01_lkt    GGAGCGGGATTAGGAAAATTATCAAGGAGTTGTTTCCGTCAATAAGATCGACAATGACGTG      5661
             **   * ***************** ** *  ****** ** ******** **  * ***** **


3718_lkt     GAAGCTAGCGTTGATAAATCTTCCATCGAAGGAGCTAATGAAATCAATGTCATTGCCAAA      5742
3724_lkt     TCAGCCACTGTAGATAAGTCTTCGATTAAAAATGCAAAGAATTGAATGTAATTGCGAAA      5745
ARU01_lkt    GAAGCCGGCATTGATAAGTCTTCCATCGAAGGAGCTGATGAAGTTAATGTCATTGCCAAA      5721
             ***       * ***** ***** **  **     ** * ** ***** ***** ***


3718_lkt     GATGTCAAAGGAAGTTCTGATCTAGCAAAAGAATATCAGGCTTTACTAAATGGAAAAGAT      5802
3724_lkt     GATATCAAAGGGAGTTCTGAATTAGCAAAAGAATATCAATCCTTATTAAACGGGAAAGAT      5805
ARU01_lkt    GATGTCAAAGGAAGTTCTAATCTGGCAAAAGAGTATCAAGCTTTACTGAACGGAAAAGAT      5781
             *** ******* ****** *  * * ******** *****  * *** * ** ** ******


3718_lkt     AAAAAATATTTAGAAGATCGTGGTATTAATACGACTGGAAATGGTTATTATACGAAGGAA      5862
3724_lkt     AGAGAATACTTGGAAGCACATGGTATTGATACTACAGGAAAAAAATTACTATACTGAAGAA      5865
ARU01_lkt    GAAAAATATTTAAAAGATCGTGGAATTGATACGACCGGAAATGGTTATTATACAAAAGAA      5841
             * **** **   *** * * **  *** *** ** ** *****     *** ***** * ***


3718_lkt     CAACTAGAAAAAGCAAAGAAAAAAGAAGGAGCGGTCATTGTAAATGCTGCTTTATCGGTT      5922
3724_lkt     CAATTAAAAGAAGCAGCAAAGCGGGATGGAGCTGTAATTGTCAATGCGGCAGCATCTATA      5925
ARU01_lkt    CAATTGGAAGAAGCAAAGAAAAAAGAAGGAGCGGTCATTGTAAATGCTGCTTTATCGGTT      5901
             *** *  ** *****  **    **  ** ***** ** ***** ***** **    *** *


3718_lkt     GCTGGAACGGATAAATCCGCTGGAGGAGTAGCTATTGCAGTCAATACTGTTAAAAATAAA      5982
3724_lkt     GCGGGAAGTGATAAATCTTCCGGAGGAGCCGGAGTTGCGGTGAATCTTGTAAAAAATAAA      5985
ARU01_lkt    GCCGGAACGGATAAATCCGCCGGAGGAGTAGCTATTGCAGTCAATACTGTTAAAAATAAA      5961
             **  **** ********  *  ******* ***   **** ** ***    *** *********


3718_lkt     TTTAAAGCAGAATTGAGTGGAAGCAATAAGGAAGCCGGAGAGGATAAAATTCATGCGAAA      6042
3724_lkt     TTTAAAGCGGAAGTAATTGGTGAGGATAAGGTTGATAAAACAAAGGATATTGCTGCAGAA      6045
ARU01_lkt    TTTAAAGCGGAATTGAATGGAAAAGATAAGGAAGTTGAAGAGGGTAAAATTCGTGCGAAG      6021
             ******** *** * * ***     * ****** *  *      *  * *** *** *


3718_lkt     CATGTAAATGTGGAGGCAAAATCATCTACTGTTGTTGTGAATGCGGCTTCTGGACTTGCT      6102
3724_lkt     AAAATAAATGTGGATGCAAAATCATCTACAGTTATTGTAAACACAGCTGCAGGGCTTGCT      6105
ARU01_lkt    CATGTAAATGTCGAAGCTAAATCGTCCACTGTCGTTGTAAATGCGGCTTCCGGACTTGCT      6081
              * ******* ** ** ***** ** ** ** ** ** ** **** ** * ** ******


3718_lkt     ATCAGCAAAGATGCTTTTTCAGGAATGGGATCTGGAGCATGGCAAGACTTATCAAATGAC      6162
3724_lkt     GTCAGTCAAAATAGTTTTGCAGGATTGGGTTCTTTCGCATGGCAAGAGTTGAACAATGAA      6165
ARU01_lkt    GTCAGCAAAGACGCTTTTTCAGGAATGGGATCCGGAGCATGGCAAGATTTATCGAATGAA      6141
              ****  ** *  **** **** ***** **  * ***********  **    *****


3718_lkt     ACGATTGCAAAGGTGGATAAAGGAAGAATTTCTGCTGATTCCTTAAATGTGAACGCAAAT      6222
3724_lkt     GTAAGTACAAAGGTAGAGAATATAAAAGCGGAAACCGATCGGTTAAATGTCGTTGCTAGA      6225
ARU01_lkt    ACCGTTGCAAAGGTCAATAAAGGAAGAATTTCTACCGATTCCTTGAATGTGACTGCAAAT      6201
              * *******  * **    ** *     * ***    ** *** ** ***** ** *


3718_lkt     AATTCCATTCTTGGGGTGAATGTTGCGGGAACCATTGCCGGTTCTCTTTCTACGGCGGTA      6282
3724_lkt     AACAATACTTTAGGAGTTAATATAGCAGGAAATATTGCCGGAACAGGTTCTGCTGCAGTG      6285
ARU01_lkt    AATTCCGTTCTTGGAGTGAATGTTGCTGGAACCATTGCCGGTTCTCTTTCTACTGCAGTA      6261
             **       * * ** ** *** * ** **** ********  *    **** * ** **
```

```
3718_lkt    GGAGCTGCTTTTGCGAATAATACTCTTCATAATAAAACCTCTGCTTTGATTACAGGAACG    6342
3724_lkt    GGAGCTGCTTTAGCTCATAATTCTCTGAAAAATAAAGCATCATCTTCTGTGACCGCTTCT    6345
ARU01_lkt   GGAGCTGCTTTTGCGAATAATACTCTTCATAATAAAACCTCTGCTTTGATTACGGAAACA    6321
            *********** **  ***** ****  * ****** * **  ***   * ** *    *


3718_lkt    AAGGTAAATCCTTTTAGTGGAAAG------AATACAAAAGTCAATGTACAAGCTTTGAAT    6396
3724_lkt    AATTTTAAGCCTTTTTCAAAAACAAATGCTAAGATGGGTATCAATGTACAGGCATTGAAT    6405
ARU01_lkt   AAGGTAAATTCTTTTAGTGGAAAG------GATACAAAATCAATGCACAAGCTTTGAAT    6375
            **  * **  *****      **       * *    ****** *** ** ******


3718_lkt    GATTCTCATATTACAAACGTTTCTGCTGGAGGCGCTGCAAGTATTAAGCAGGCTGGAATC    6456
3724_lkt    GATTCTCATATCACGAATGTGTCTGCGGGAGTTGTTGCCAGTGCAAAAGCTTCCGGAATT    6465
ARU01_lkt   GATTCTCATATTACAAATGTTTCTGCCGGAGTCGCTGCAAGTGTTAAGCAAGTTGGAATC    6435
            *********** ** ** ** ***** ****  * *** ***    **      *****


3718_lkt    GGAGGAATGGTATCTGTCAATCGTGGTTCTGATGAAACGGAAGCTTTAGTTAGTGATTCT    6516
3724_lkt    TCCGGAATGGCATCTATTAATCGAGGTTCTGATGAAACGGAAGCTTTAGTCAGTGATTCT    6525
ARU01_lkt   GGAGGAATGGTATCTGTCAATCGCGGTTCTGATGAAACAGAAGCTTTGGTCAGTGATTCT    6495
            ******* **** * ***** ************** ** ******** ** *********


3718_lkt    GAGTTTGAAGGAGTAAGTTCTTTCAATGTAGATGCAAAAGATCAAAAAACAATAAATACA    6576
3724_lkt    GAATTTGAAGGAGTAAATTCTTTCAATGTAGATGCAAAAGATCAAAAAATAGTAAATACA    6585
ARU01_lkt   GAGTTTGATGGAGTAAATTCTGTCGATGTAAGAGCTGAAGATAAAAAAGTATTGAATACC    6555
            ** ***** ******* **** *** ***** **  ***** *****    * * *****


3718_lkt    ATTGCCGGAAATGCAAATGGAGGAAAAGCGGCTGGAGTTGGAGCAACAGTTGCTCATACA    6636
3724_lkt    ATTGCCGGGAACATAAATGGAGGAAAAGCAGCCGGAGTTGGAGCAACAGTCGCTCATACA    6645
ARU01_lkt   ATTGCCGGAAATATCAATGGAGGGCAAGCAGCCGGAGTTGGAGCGACAGTTGCTCATACC    6615
            ******** **    ********  **** ** ***********  ***** ********


3718_lkt    AATATTGGAAAACAATCAGTTATAGCTATTGTAAAAAAACAGTAAAATTACAACGGCGAAT    6696
3724_lkt    AATATTGGAAAACAATCTGTTACAGCTATTGTAAAAAATAGTAAAATTAAAACAGCAAAT    6705
ARU01_lkt   AATATCGGAAAGCAATCTGTTACAGCTGCTATAAAAAATAGTAAAATTACAACGGCTAAG    6675
            ***** ***** ***** **** **** *  * ******* ********** *** ** **


3718_lkt    GATCAAGATAGAAAAAAATATCAATGTGACTGCAAAAGATTATACTATGACTAATACTATA    6756
3724_lkt    AATCAAGATAAAAAAAAaTATCCATGTTACTGCGAAAGATAGTGTTATTATGAATACAGTG    6765
ARU01_lkt   AATCAAGACAAGAAAAATATAAATGTTACTGCAAAAGACAGTGTTGTTATGAATACAGTG    6735
             ******* *  ******** * ***** ***** ***** *  * * * *  ***** *


3718_lkt    GCAGTCGGAGTTGGAGGAGCAAAAGGAGCCTCTGTGCAAGGAGCTTCTGCAAGTACTACC    6816
3724_lkt    GCAGCCGGAGTTGGAGGAGCAAAAGGGGCTTCTGTGCAAGGAGCTTCTGCAAGTACCGCT    6825
ARU01_lkt   GCAGCCGGAGTCGGAGGGGCAAAAGGAGCTTCTGTGCAAGGATCTTCTGCAAGTACCACT    6795
            **** ****** ***** ******** **  ************  ** *********** *


3718_lkt    TTGAATAAGACAGTTTCTTCTCATGTTGATCAAACTGATATTGaCAAAGATTTAGAGGAA    6876
3724_lkt    TTCAATAAAACAGTTGCTTCTCATGTCGAGACTACTGATATTGACAAAGATTTAGAAAAA    6885
ARU01_lkt   TTGAATAAAACGGTTTCCGCTCATGTTGAGAAAACTGATATTGATAA-----------A    6843
            ** ***** ** *** *  ******* **  * *********** **             *


3718_lkt    GAAAATAATGGAAATAAGGAAAAGGCAAATGTTAATGTTCTAGCTGAAAATACGAGTCAA    6936
3724_lkt    GAAAATGATGAAAATAAAGAAAAAGCAAATTTCAATATTCTGGCTGAGAACACAAGTCAA    6945
ARU01_lkt   GAAAATCATGAAAGTAAAGAAAAAGCAGATTTCAACGTTCTGGCTGAAAATACAAGTCAA    6903
            ****** *** ** *** *** ** ** *** **  **** ***** ** ** ** ******


3718_lkt    GTGGTCACAAATGCGACAGTGCTTTCCGGAGCAAGTGGACAAGCTGCAGTAGGAGCTGGA    6996
3724_lkt    GTAGTCACAAATGCGACAGTGCTTTCCGGAGCGAGTGGACAAGCCGCAGTAGGAGCCGGA    7005
ARU01_lkt   GTAGTAACAAATGCGACAGTACTTGCCGGAGCGAGCGGACAAGCTGCGGTAGGAGCCGGA    6963
            ** ** ************** ***  ******* ** ******** ** ********* ***


3718_lkt    GTAGCAGTTAATAAAATTACACAAAATACTTCTGCACATATAAAAAAATAGTACTCAAAAT    7056
3724_lkt    GTAGCAGTTAATAAAATTACACAAAATACTTCTGCACATATAAAAAACAGTACTCAAAAT    7065
ARU01_lkt   GTAGCAGTTAATAAAATTACACAAAATACTTCTGCACATATAAAAAATAGTACCCAAAAT    7023
            *********************************************** ***** *******


3718_lkt    GTACGAAATGCTTTGGTAAAAAGCAAATCTCATTCATCTATTAAAACAATTGGAATTGGA    7116
3724_lkt    GTACGAAATGCTTTAGTAAAAAGTAAAGCGAATTCCTCTATTAAAACGATCGGAATTGGA    7125
ARU01_lkt   GTACGAAATGCTTTGCTAAAAAGTAAAGCTAATTCTTCCATTAAAACAATTGGAATTGGA    7083
            *************   ******* *** *  **** ** ** ******* ** ********


3718_lkt    GCTGGAGTTGGAGCTGGAGGAGCTGGAGTGACAGGTTCTGTAGCAGTGAATAAGATTGTA    7176
3724_lkt    GCGGGAATTGGAGCTGGAGGAGTTGGATCCAGCGGTTCTGTAGCTGTGAATAAAATTACA    7185
ARU01_lkt   GCAGGAGTCGGAGTAGGTGGAGTTGGAGTTAGCGGTTCTGTAGCTGTAAATAAAATCACA    7143
            ** *** * **** ** *** ** *****  * ********** ** ***** **    *


3718_lkt    AATAATACGATAGCAGAATTAAATCATGCAAAAATCACTGCGAAGGGAAATGTCGGAGTT    7236
3724_lkt    AATAATAATACAGCTTCTGTTGAACATTCTAATATTTTTGCAAGAGGAAATGTAGGAGTT    7245
ARU01_lkt   AATAATAACACAGCTGTTGTTGAACATTCTGATATTTTTGCAAAAGGAAATATAGGAGTA    7203
            *******  * ***     * * ** *** ** * *  ***** ****** * *****


3718_lkt    ATTACAGAGTCTGATGCGGTAATTGCTAATTATGCAGGAACAGTGTCTGGAGGGGGCCCGT    7296
3724_lkt    ATTACAGAATCTGATGCGGTCATTGCTAATTATGCCGGAACAGTTTCCGGAGGAAGTCAT    7305
ARU01_lkt   ATTGCAGAATCCGATGCAGTCATTGCTAATTATGCCGGAACAGTTTCCGGAGGAGGTCAT    7263
            *** **** ** ***** ** ************** ******** ** ***** ** ** **


3718_lkt    GCAGCAATAGGAGCCTCAACCAGTGTGAATGAAATTACAGGATCTACAAAAGCATATGTA    7356
3724_lkt    GCAGGGGTAGGAGGATCCACCAGTGTCAATGAAATCCAAGGAGATACGATTGCTTCTGTT    7365
ARU01_lkt   GTAGGGGTAGGAGGATCTACCGGTGTCAATGAAATCGAAGGAGATACGATTGCTTCCGTT    7323
            * **  ****** ** *** ** **** ********   ****  *** * ** * * **
```

```
3718_lkt    AAAGATTCTACAGTGATTGCTAAAGAAGAAACAGATGATTATATTACTACTCAAGGGCAA          7416
3724_lkt    AAACATTCAAAAATAGAAGCCAAAGAAGAAACCGAAGATGTCATTGAAACACAGGGAAA          7425
ARU01_lkt   AAAAAATCAAAACTGGAGGCAAAAGAAGAAACGGAAGATGTTATTGAAACGCAAGGGAAA          7383
            *** * ** * * *      ** ********** ** ***      ***    ** ** *** **


3718_lkt    GTAGATAAAGTGGTAGATAAAGTATTCAAAAATCTTAATATTAACGAAGACTTATCACAA          7476
3724_lkt    GTAAAAGAAGTTGTAGAAAAGGTATTTCAAGATATAGGGAATTAATGAGGATTTGAGCAAG         7485
ARU01_lkt   GTAAAAGAAGTTGTAGAAAACGTACTTCAAGATATAGGGAATCAATGATGATTTGAGTAAG         7443
            *** * **** ***** ** *** *  ** ** *      ** ** ** ** **      *


3718_lkt    AAAAGAAAaatAagtAATAAAAAAAGGATTTGTTaccAATAGTTCAGCTACTCATACTTTA         7536
3724_lkt    AAAAGAGAAAAATCAAAGAAAAGAGGGATTTGTTGTAAACAGCTCCTCTACTCATACTTTA         7545
ARU01_lkt   AAAAGAGAAAAATCAAATAAAAAAAGGGATTTGTTGTAAACAGCTCTTCCACTCACACGATA         7503
            ****** *** *   ** **** **********    ** ** **  * ***** **   **


3718_lkt    AAATCTTTATTGGCAAATGCCGCTGGTTCaggAcaaGCCGGAGTGGCAGGAACTGttaAT         7596
3724_lkt    AAATCCTTATTGGCAAATGCAGCCGGTTCAGGACAAGCCGGAGTGGCAGGAACTGTCAAT         7605
ARU01_lkt   AAATCATTGTTGGCAAATGCTGCAGGAGCAGGTAAAGCTGCAGTGGCAGGAACTGTAAAT         7563
            ***** ** ********** ** ** ****  **** * *************** ***


3718_lkt    ATCAacaaGGTTTatgGAGAAacAGAAGCTCTTGTAGAAAATTCTATATTAAATGCAAAA         7656
3724_lkt    ATCAACAAGGTTTATGGAGCAACAGAAGCTCTTGTGTAGAAGATTCCATATTGAATGCAAAA        7665
ARU01_lkt   GTAAATAAAATTACAGGAGAAACAAAAGCCCTTGTAGAAAGTTCCGTATTAAATGTAAAA         7623
             * ** ** **    **** **** **** ********* ***  **** **** ****


3718_lkt    CATTATTCTGTAAAATCAGGAGATTACACGAATTCAATCGGAGTAGTAGGTTCTGTTGGT         7716
3724_lkt    CATTACTCTGTGAAATCGGGAGATTACACAAACTCGATCGGAGCAGTAGGCTCTGTTGGT         7725
ARU01_lkt   AATTCTTCGATACATGCAGGAGATTATACTAACAGTATTGGAGCGGTAGGATCCGCTTCC         7683
             ***  **  *  *  * ********* ** **      ** **** ***** ** *  *


3718_lkt    GTTGGTGGAAATGTAGGAGTAGGAGCTTCTTCTGATACCAATATTATAAAAAGAAATACC         7776
3724_lkt    GTTGGTGGAAATGTAGGAGCAGGAGCTTCTTCTGATACCAATATTATAAAAAGAAATACC         7785
ARU01_lkt   GGAGCATTAAACGCAGGAATAGGAGCTTCTTCCAATACAAATCTTTTCAGTAGAAATACA         7743
            *  *   *** * ****  ** ************ **** *** ** * *  ********


3718_lkt    AAGACAAGAGTTGGAAAAACTACAATGTCTGATGAAGGTTTCGGAGAAGAAGCTGAAATT         7836
3724_lkt    AAGACAAGAGTTGGAAAAACTACAATGTCTGATGAAGGTTTCGGAGAAGAAGCTGAAATT         7845
ARU01_lkt   AAAACAATAGTGCACGATACGAAAATATCTACAGAAAAAGAAGGACAATCTTCAGAAATT         7803
            ** **** ***      * ** * *** ***      *** **      * ******


3718_lkt    ACAGCAGATTCTAAGCAAGGAATTTCCTCTTTTGGAGTCGGAGTCGCAGCAGCCGGGGTA         7896
3724_lkt    ACAGCAGATTCTAAGCAAGGAATTTCCTCTTTTGGAGTCGGAGTCGCAGCAGCCGGGGTA         7905
ARU01_lkt   ACAGCGGATTCTAAACAAGGAATTTCTTCCTTTGGTGTTGGCGTTGGAGCTGGAGTAGTC         7863
            ***** ******** ********** **  ** *** ** ** ** * * *** *   **


3718_lkt    GGAGCCGGAGTGGCAGGAACCGTTTCCGTAAATCAATTTGCAGGAAAGACGGAAGTAGAT         7956
3724_lkt    GGAGCCGGAGTGGCAGGAACCGTTTCCGTAAATCAATTTGCAGGAAAGACGGAAGTAGAT         7965
ARU01_lkt   GGAGCCGGAGTGGCAGGAACGGTTTCTGTCAATAAATTATCTGGAAAAACCGAAGTAGAT         7923
            ******************** ***** ** *** ****    * ***** ** *********


3718_lkt    GTGGAAGAAGCAAAGATTTTGGTAAAAAAAGCTGAGATTACAGCAAAACGTTATAGTTCT         8016
3724_lkt    GTGGAAGAAGCAAAGATTTTTGTAAAAAAAGCTGAGATTACAGCAAAACGTTATAGTTCG         8025
ARU01_lkt   ATTCAAAAATCTGATATTTCTACAAAATCAGCAGATATTTCCGCAAAACATTATGGAGTG         7983
             *  ** ** *  * ****    **** *** ** *** * *** ****** **** *


3718_lkt    GTTGCAATtGGAAATGCCGCAGTCGGAGTGGCTGCAAAAGGAGCTGGAATTGGAGCAGCA         8076
3724_lkt    GTTGCAATTGGAAATGCCGCAGTCGGAGTGGCTGCAAAAGGAGCTGGAATTGGAGCAGCA         8085
ARU01_lkt   CTTTCTACAGGAAACGGATCTGTTGGGGCGGCAGTAAAGGGAGCCGGAGTCGGAGCAGCG         8043
             ** * *  ***** *     * ** ** * *** * *** ***** *** * ********


3718_lkt    GTGGCAGTTACCAAAGATGAATCAAACACGAGAGCAAGAGTGAAAAATTCTAAAATTATG         8136
3724_lkt    GTGGCAGTTACCAAAGATGAATCAAACACGAGAGCAAGAGTGAAAAATTCTAAAATTATG         8145
ARU01_lkt   GTTGCCGTGACGAAAGATTTGACAAATACAAATGTAAGAATTAAAGATTCTAAAATTGTC         8103
            ** ** ** ******      **** ** *     * **** * *** *********** *


3718_lkt    ACTCGAAACAAGTTAGATGTAATAGCAGAAAATGAGATAAAATCAGGTACTGGAATCGGT         8196
3724_lkt    ACTCGAAACAAGTTAGATGTAATAGCAGAAAATGAGATAAAATCAGGTACTGGAATCGGT         8205
ARU01_lkt   ACAAAAACAAAATTAGATGTCATTGCACAAAATCATACTAAAGTAAATGGAGGAATGGTA         8163
            **    ** **  ** ******** ** *** **  * *** * *    ** ***** *


3718_lkt    TCAGCCGGAGCTGGAATTCTTGCAGCCGGAGTATCTGGAGTTGTTTCTGTCAATAATATT         8256
3724_lkt    TCAGCCGGAGCTGGAATTCTTGCAGCCGGAGTATCTGGAGTTGTTTCTGTCAATAATATT         8265
ARU01_lkt   GGAATTGGAGCAGCCGGAGTTGGAGCCGGATTAGCAGGAACCGTTTCTGTCAATAATATC         8223
             *   ***** *          *** ******* ** * ***     ****************


3718_lkt    GCAAATAAGGTAGAAACAGATATCGATCATAGTACTTTACACTCTTCTACTGATGTAAAT         8316
3724_lkt    GCAAATAAGGTAGAAACAGATATCGATCATAGTACTTTACACTCTTCTACTGATGTAAAT         8325
ARU01_lkt   ACCAGCAAAGTAGGAACAGAAATAGATCATAGTGAATTGCTGTCTGAAGAAGATGTAAAT         8283
             * *  ** **** ****** ** ********* **  *  ***      * ********


3718_lkt    GTAAAAGCTCTTAATAAAATTTCGAATTCCTTGACAGCCGGTGGAGGAGCCGCAGGTCTT         8376
3724_lkt    GTAAAAGCTCTTAATAAAATTTCAAATTCCTTGACAGCCGGTGGAGGAGCCGCAGGTCTT         8385
ARU01_lkt   GTGAAAGCTTTGAATAAAGTTGATAGTTCTATGATGGCTGCCGGAGGAGCTGCAGGTCTT         8343
            ** ****** * ****** **  * ** ** *     ** *  * ******** *********


3718_lkt    GCAGCAGTTACCGGAGTGGTTTCTGTTAACACTATAAATAGTTCTGTGATAGCTCGAGTT         8436
3724_lkt    GCGGCAGTTACCGGAGTGGTTTCTGTTAACACTATAAATAGTTCTGTGATAGCTCGAGTT         8445
ARU01_lkt   GGTGCAGCAAGTGGAGTGGTTTCTGTGAATACTGTCAATACTTCCGTAGTAACCAGAGTT         8403
            *  **** *  ************** ** *** * ** * *** ** ** **  *  *****
```

254

```
3718_lkt      CACAATAACTCTGATTTGACTTCCGTACGAGAAAAAGTAAATGTAACGGCAAAAGAGGAA      8496
3724_lkt      CACAATAACTCTGATTTGACTTCCGTACGAGAAAAAGTAAATGTAACGGCAAAAGAGGAA      8505
ARU01_lkt     CATAATAATTCAAAAATAACTTCATCCAAAGGAAGCACCAATGTAAAAGCAGAAGAAGAA      8463
              ** ***** **   *  * *****      ** **      *******  *** **** ***


3718_lkt      AAAAATATTAAGCAAACAGCAGCAAATGCAGGAATCGGAGGAGCAGCAATCGGAGCCAAT      8556
3724_lkt      AAAAATATTAAGCAAACGGCAGCAAATGCAGGAATCGGAGGAGCAGCAATCGGAGCCAAT      8565
ARU01_lkt     AAAAATATAAAACAAGTAGTTGCAAATGCAGTGGCAGGGGGCGTTTCTGTTGGAGCGAAT      8523
              ******** ** ***    *  *********    ** ** *   *  * ***** ***


3718_lkt      GTCTTGGTAAATAATTTTGGAACAGCTGTAGAAGATAGAAAAAATTCTGAAGGAAAAGGA      8616
3724_lkt      GTCTTGGTAAATAATTTTGGAACAGCTGTAGAAGATAGAAAAAATTCTGAAGGAAAAGGA      8625
ARU01_lkt     GTCTTGGTTAATAATTTTGGAACTGCTGTAGAAGATAAGGCAGCTGAAACAGGAAAAGGA      8583
              ******** ************** ************   *  *        *********


3718_lkt      ACAGAAGTTTTAAAAACTTTAGACGAAGTTAACAAAGAACAAGATAAAAAAGTAAATGAT      8676
3724_lkt      ACGGAAGTTTTGAAAACTTTAGACGAAGTTAACAAAGAACAAGATAAAAAAGTAAATGAT      8685
ARU01_lkt     ACTGAAATCTTAAAAACTTTAGAAGAAATCAATAAAGATCAAGATCAAAAAATCAGTGAA      8643
              ** ***  * ** ********** *** * ** ***** ****** *  ***** * * ***


3718_lkt      GCTACGAAAAAAATCTTACAATCAGCAGGTATTTCTACAGAAGATACTTCTGTAAAAGCG      8736
3724_lkt      GCTACGAAAAAATCTTACAATCAGCAGGTATTTCTACAGAAGATACTTCTGTAAAAGCG      8745
ARU01_lkt     GATGCCGCTAAAATATTAAAGTCAAATGGAATTGATACAAAAGACTATGCTGTCAAGGCG      8703
              *  *  *    ***** *** * ***    ** ***  **** ****   * **** ** ***


3718_lkt      GATAGAGGAGATACTCAGGGAGAAGGAATTAAAGCCATTGTGAAGACTTCTGATATTATT      8796
3724_lkt      GATAGAGGAGATACTCAGGGAGAAGGAATTAAAGCCATTGTGAAGACTTCTGATATTATT      8805
ARU01_lkt     GATCGTGGAGAAACACAAGCAGAAGGAATTAAAGTATTTGTAGAAAATTCTTCGATAAAA      8763
              *** * ***** ** ** * **************    ****   * * ****    ** *


3718_lkt      GGAAAAAATGTAGATATTACAACAGAGGACAAGAATAATATCACTTCTACTGGTGGTTTG      8856
3724_lkt      GGAAAAAATGTAGATATTACAACAGAGGACAAGAATAATATCACTTCTACTGGTGGTTTG      8865
ARU01_lkt     GGAAAAAATACGAATATTGCAGCAAATCAAACGGATAATGTGACTTCCACCGGTGGAGCA      8823
              *********    ***** ** ** *  * * * ***** * ***** ** *****


3718_lkt      GGAACTGCAGGTCTTGCTTCCGCATCAGGAACAGTGGCAGTTACAAATATTAAAAGAAAT      8916
3724_lkt      GGAACTGCAGGTCTTGCTTCCGCATCAGGAACAGTGGCAGTTACAAATATTAAAAGAAAT      8925
ARU01_lkt     GGAAGTGCAGGAGTGGCGTCCGCATTGGGAACTGTAGGAGTAACAAATATCAAAAGAAAC      8883
              **** ******   * ** *******  ***** ** * *** ******** ********


3718_lkt      TCCGGAGTTACTGTTGAAAATTCTTTTGTGAAAGCAGCTGAAAAAGTAAATGTTAGATCG      8976
3724_lkt      TCCGGAGTTACTGTTGAAAATTCTTTTGTGAAAGCGATTGAAAAACTAAATGTTAGATCG      8985
ARU01_lkt     ATAGGAGTTTTCATAAATAATGCGGATGTGAAAGCGGATGAGAAACTAAGTGTACGATCC      8943
              ******    *  * *** *  *********    ***  *** *** *** ***   ****


3718_lkt      GATATTACAGGAAATGTTGCTTTAACAGCATATCAAGGTTCTGTAGGAGCATTGGGAATA      9036
3724_lkt      GATATTACAGGAAATGTCGCTTTAACAGCATACCAAGGTTCCGTAGGAGCCATTGGGAATA      9045
ARU01_lkt     GATATTTCAGGAAACGTATCCCTAACTTCTTATCAAGGAACTGTTGGAGCTTTGGGATTA      9003
              ****** ******* **   *   **** *  * ***** * *** ***** ****** **


3718_lkt      GGAGCTGCCTATGCAGAATTAAATTCTAATGGAAGATCAAATATCAGTATTAAAAATTCT      9096
3724_lkt      GGAGCTGCCTATACAGAATTAAATTCTAATGGAAGATCAAATATCAACATTAGAAATTCT      9105
ARU01_lkt     GGAGCGGCTTATTCACAAATAAATGCTAACGGACAATCGAATATTACTATTAAAAACTCT      9063
              ***** ** *** ** ** ***** **** *** ** ** **** *  **** *** ***


3718_lkt      AAGCTATTAGGAAAAAATATTGATGTTATTGTAAAAGATAAATCGGAATTGAGAGCGGAA      9156
3724_lkt      AAGCTATTAGGAAAAAATATTGATGTTATTGTAAAAGATAAATCGGAATTGACAGCGGAA      9165
ARU01_lkt     AGATTTGCAGCAAAACATATTGATGTTATTGCAAAGATAAATCTCAATTGAAGGCGGAA      9123
              *    *     ** **** ************** *** ******** ******  ******


3718_lkt      GCAAAAGGATTAACCGTAGGAGCGGTAGCTGCCGGAGCCATTATCTCAAAAGCAAAGAAT      9216
3724_lkt      GCAAAAGGATTAACCGCCGGAGCGGTAGCCGCCGGAGCTATTATCTCAAAAGCAAAGAAT      9225
ARU01_lkt     GCAAAAGGATTAACCGTAGGAGCGGTAGCTGCCGGAGCTATTATCTCAAAAGCAAACAAT      9183
              ****************  ********** ******** ********************  ***


3718_lkt      GAAATGAATTCAGAGGTTGAAATTGAGAAGAGTATTTTCAATGAAGAAAATAGAGTAACT      9276
3724_lkt      GAAATGAATTCAGAGATTGAAATCGAGAAGAGTATTTTCAATGAAGAAAATAGAGTAAAC      9285
ARU01_lkt     GAAATGAATTCAGAAATTGAAATTGAGAAGAGCATTTTCAATGAAGAAAATAGAGTAAGC      9243
              **************   ****** ******** *********************


3718_lkt      AGCCCTTCTAAAGGAATTGGAAGAGAAATCAATGTCAAAGTGGAAAAAGAAACAGAGTG      9336
3724_lkt      AGCCCTTCTAAAGGAATTGGAAGAGAAATCAATGTCAAAGTGGAAAAAGAAAATAGAGTG      9345
ARU01_lkt     AGCCCTTCTAAGGGAATTGGAAGAGAAATCAATGTTAAAGTGGAAAAAGAAACAGAGTG      9303
              ********** ******************* ***** *** ***************** ******


3718_lkt      ACTGCTGAATCTCAAGGAGCTTCTGTAGGAGCAGTAGCAGGGGCAGGAATTATTTCCGAA      9396
3724_lkt      ACTGCTGAATCTCAGGGAGTTTCTGTAGGGGCGGTAGCAGGGGCAGGAATCATTTCCGAT      9405
ARU01_lkt     ACTGCTGAATCTCAAGGAGCTTCTGTGGGAGCAGTAGCGGGGGCAGGAATTATTTCCGAT      9363
              ************** **** ** ****** ** ** ***** ********** ********


3718_lkt      GCAAAAGATGCCGGAAGCTCTTATTTGAAAGTTAGTACAAAATCCGGAAGAAGTATTTTT      9456
3724_lkt      GCAAAAGATGCGGGAAGTTCTTATTTAAAAGTTAGTACAAAATCCGGAAGAAGTATTTTC      9465
ARU01_lkt     GCAAAAGATGCGGGAAATTCTTATTTAAAAGTTAGTACAAAATCCGGAAGAAGTATTTTC      9423
              *********** ***  ******* ******************************


3718_lkt      CATGCAGATAATGTGAATATGGAAGCAACACATAAAATGAAAGTAACAGCAGTTTCTAAA      9516
3724_lkt      CATGCAGATAATGTGAATATGGAAGCAACACATAAAATGAAAGTAACTGCAGCTTCTAAA      9525
ARU01_lkt     CATGCAGATAATGTGAATATGGAAGCAACACATAAAATGAAAGTAACTGCAGGTTCTAAA      9483
              ********************************************** **** ** *******
```

```
3718_lkt    GCAGTAACAGGTTCTGTATTGGGAGGAGTTGGAGTCACCAAGGCAGAAGCTACTGCTGCA            9576
3724_lkt    GCGGTAACGGGTTCTGTATTGGGAGGAGTTGGAGTCACCAAAGCGGAAGCTGTTGCTGCC            9585
ARU01_lkt   GCGGTAACAGGTTCTGTATTGGGAGGAGTTGGAGTTACAAAGGCGGAAGCAACTTCTGCA            9543
            ** ***** ************************ ** ** ** ***** * ****


3718_lkt    GGTAAAACTATGGTAGAAGTTGAGGAAGGAAATTTGTTCAGAACAAATCGATTGAATGCA            9636
3724_lkt    GGTAAAACTATGATAGAAGTTGAAGAAGGAAATTTGTTCAGAACAAATCGATTGAATACA            9645
ARU01_lkt   GGAAAAGTAATGGTAGAGATTGAGGAAGGAAATTTGTTCAGAACAAATCGATTGAATGCA            9603
            ** ***   *** ****  **** *********************************** **


3718_lkt    ATTTCTAAAGTAGAAGGTTTGGATGAAGATAAAGTAACTGCTAAATCTTCTGTAGTATCA            9696
3724_lkt    ATTTCCAAAGTAGAAGGTTTGGATGAAGATAAAGTAACTGCTAAATCTTCTGCAATATCA            9705
ARU01_lkt   ATTTCTAAAGTAGAAGGATTGGATGAAGATAAAGTAACTGCGAAATCTTTTGTAGTATCA            9663
            ***** *********** ******************** ******* ** * *****


3718_lkt    GGAAATGGAGGAGGAATTGCCGGAGCAGGAGTGAATACTTCTACAGCACAAAGTAATACT            9756
3724_lkt    GGAAATGGAGGAGGAATTTCTGGAGCGGGAGTGAATACTTCTACAGCACAAAGTAATACC            9765
ARU01_lkt   GGAAATGGAGGAGGAATTGCCGGAGCGGGAGTGAATACTTCTACAGCACAAAGTAATACG            9723
            ****************** * ***** **********************************


3718_lkt    GAATCCGTAGTTCGTTTACGAAAGCAAGATTATGAAAATAATGATTACACAAAAAAATAT            9816
3724_lkt    GAATCCGTAGTTCATTTACAAAGACAAGACTATGAAAATAATGATTATACAAAAAAATAT            9825
ARU01_lkt   GAATCCGTCGTTCGTTTACGAAAGCAAGATTACGAAAATAACGACAGTACGAAAAAATAT            9783
            ******** **** **** ** **  ***** ** ******** **   ** ** ********


3718_lkt    ATTTCAGAAGTCAATGCTCTTGCTTTAAATGATACAAAGAATGAAGCGAATATAGAATCT            9876
3724_lkt    ATTTCAGAAGTCAATGCTCTTGCTTTAAATGATACAAAGAATGAAGCGAATATAGAATCT            9885
ARU01_lkt   ATTTCAGAAGTGAATGCTCTTGCTGTAAATGACACAAAGAATGAAGCAAATATAGAATCT            9843
            *********** ************ ******* ************** ************


3718_lkt    TTAGCGGTAGCCGGTGTGCATGCACAAGGAACAAACAAAGCATTTacGAGATCAAACAAG            9936
3724_lkt    TTAGCGGTAGCCGGTGTGCATACACAAGGAACAAACAAAGCATTTACGAGATCAAACAAG            9945
ARU01_lkt   TTAGAGGTAGCGGGAGTATACGCACAAGGAACAAACAAAGCATTTACGAAATCGGATAAG            9903
            **** ****** ** **  *  ***************************** * * ***


3718_lkt    TTAACTTCTACAACTGTAAATGGAGGaAACGTAtCtCAACTTCGTGCAAAAgCTTTGGCT            9996
3724_lkt    TTAACTTCTACAACTGTAAATGGAGGAAACGTATATCAACTTCGTGCAAAAGCTTTGGCT            10005
ARU01_lkt   TTAACTTCTACAACTGTAAATGGAGGAAAAGTATATCAACTTCGTGCAAAAGCTTTGGCT            9963
            ************************** **** *********************************


3718_lkt    AAAAATGAAAATTATggaAaTGTAAAAggAACTGGAGGGAGCcttAGTCGGAGCGGAAACA            10056
3724_lkt    AAAAATGAAAATTATGGAAATGTAAAAGGAACTGGAGGGAGCCTTAGTCGGAGCGGAAACA            10065
ARU01_lkt   AAAAATGAAAATTACGGAAATGTAAAAGGAAGCGGAGGGAGCCTTAGTCGGAGCGGAAACA            10023
            ************** **************** ***********  **************************


3718_lkt    GCAGCCGTTGAAAATTATACAAAGAGTACTACAGGAGCATTGGTTGCAGGAAATTGGGAA            10116
3724_lkt    GCAGCCGTTGAAAATTATACAAAGAGTACTACAGGAGCATTGGTTGCAGGAAATTGGGAA            10125
ARU01_lkt   GCAGCTGTTGAGAATTATGTAAAGAGTACTACAGGAGCATCTGTTGCAGGAGAATGGGAA            10083
            ***** ***** ****** ***************** ****  ********* * ******


3718_lkt    ATTGGAGATAAATTAGAAACGATTGCAAGAGATAATACGATTGTAAGAGTCAACGGAGAC            10176
3724_lkt    ATTGGAGATAAATTAGAAACGATTGCAAGAGATAATACGATTGTAAGAGTCAACGGAGAC            10185
ARU01_lkt   ATTGGAGATAAATTAGAAACGATTGCAAAGATAATACGATTGTAAAAGTTAACGGAGAC            10143
            ***************************** ***************** *** *********


3718_lkt    GGAACCAAAGGAGGTCTTGTCGGAAAGAATGGTATTTCTGTGAAAAATACAATTTCAGGG            10236
3724_lkt    GGAACCAAAGGAGGTCTTGTCGGAAAGAATGGTATTTCTGTGAAAAATACAATTTCAGGG            10245
ARU01_lkt   GGAACAAAAGGAGGTCTTGTCGGAAAGAATGGTATTTCTGTGAAAAATACAATTTTAGGA            10203
            ***** **************************************************** ***


3718_lkt    GAAACAAAATCATCCATTGAAGATAAAGCCAGAATTGTTGGAACCGGAAGTGTAAATGTA            10296
3724_lkt    GAAACAAAATCATCCATTGAAGATAAAGCCAGAATTGTTGGAACCGGAAGTGTAAATGTA            10305
ARU01_lkt   GAAACAAAATCATTTATTGACGATAAGGCTAAAATCGCTCGCGGAACCGGAAGCGTAAATGTA            10263
            ************* ***** ***** ** * *** *  *********** * *********


3718_lkt    GATGCTTTGAATGAACTTGATGTAGATCTACAAGGAAAAAGTGGTGGCTATGGTGGAATT            10356
3724_lkt    GATGCTTTGAATGAACTTGATGTAGATCTACAAGGAAAAAGTGGTGGTTATGGTGGAATT            10365
ARU01_lkt   GAGGCTCTGAATGAACTTGATGTAGATTTACAAGGAAAAAGTGGTGGATATGGTGGAATT            10323
            ** *** ******************** ***************** * ************


3718_lkt    GGTATTGGAAATGTTGATGTAAATAATGTGATTAAGAAAAATGTAGAAGCCAAAATCGGA            10416
3724_lkt    GGTATTGGAAATGTTGATGTAAATAATGTGATTAAGAAAAATGTAGAAGCCAAAATCGGA            10425
ARU01_lkt   GGTGTTGGAAATGTCGATATAAGCAATGTTATTAAAAAAGATACAGAAGCAAAAATAGGA            10383
            *** ********* *** ** *** ***** ***** *  *** ****** ***** ***


3718_lkt    AGACATGCTATTGTAGAAACTACTGGAAAACAAGAATATCAAGCATTTACAAGAGCAAAA            10476
3724_lkt    AGACATGCTATTGTAGAAACTACTGGAAAACAAGAATATCAAGCATTTACAAGAGCAAAA            10485
ARU01_lkt   CAACATGCTGTTGTAGAAACCACCGGAAAACAAGAATATCAAGCATTTACAAAAGCGAAA            10443
            ******* ********** ** ************************** *** ***


3718_lkt    GTAAATATTCTTGGAAAAGGAGACGCTGCAGCTGCAGCTGCAATATCGAATGTACACATT            10536
3724_lkt    GTAAATATTCTTGGAAAAGGAGACGCTGCAGCTGCAGCTGCAATATCGAATGTACACATT            10545
ARU01_lkt   GTAAACATTCTTGGAAAAGGAGATGCTGCAGCTGCAGCTGCGATATCCAATGTACATGTT            10503
            ***** ***************** ***************** ***** ******** **


3718_lkt    TCCAATGAGATGGATATTAAAAATTTGGCAAAGCAGTATGCATCTTCTCAATTAATAACC            10596
3724_lkt    TCCAATGAGATGGATATTAAAAATTTGGCAAAGCAGTATGCATCTTCTCAATTAATAACC            10605
ARU01_lkt   TCTAATGAGATGGATATTAAAAATTTAGCAAAGCAGTATGCATATTCTCAATTAACAACA            10563
            ** *********************** ***************** ********** ***
```

256

```
3718_lkt      AAAAATTCAAAAAATAATATTACTTTAGCATCAAGTAGTGAATCGAATGTGAATGTTCAT                10656
3724_lkt      AAAAATTCAAAAAATAATATTACTTTAGCATCAAGTAGTGAATCGAATGTGAATGTTCAT                10665
ARU01_lkt     AAAAATTCAAAAAATGATATTACTTTAGCTTCAAGTAGTGAGTTGGATGTGAATGTTCAC                10623
              *************** ************* ********** ***** * * ************

3718_lkt      GGGGTGGCTGAAGCAAGAGGTGCAGGAGCCAAAGCGACAGTTAGTGTAAAGAATCAAATA                10716
3724_lkt      GGGGTGGCTGAAGCAAGAGGTGCAGGAGCCAAAGCGACAGTTAGTGTAAAGAATCAAATA                10725
ARU01_lkt     GGAGTGGCTGAAGCAAAGGGTGTAGGAGCCAAAGCGACGTCTAGTGTAAAAAATCAAATA                10683
              ** ************* **** ***************   ********* *********

3718_lkt      AATAGAACTAATAATGTTGATTTAGCAGGAAAAATTAAAAcaGAGGGAAACATCAATGTA                10776
3724_lkt      AATAGAACTAATAATGTTGATTTAGCAGGAAAAATTAAAACAGAGGGAAACATCAATGTA                10785
ARU01_lkt     AATAGAGCGAATAATGTTGATTTGGCAGGAAAAATTAAGACAGAAGGAAACATTAATGTA                10743
              ****** * ************** ************** ***** ******** ******

3718_lkt      TATGCCGGATATGATAAAAATTATAATATAAGTAAGACAAATTCTAAGGCTATTGCGGAT                10836
3724_lkt      TATGCCGGATATGATAAAAATTATAATATAAGTAAGACAAATTCTAAGGCTATTGCGGAT                10845
ARU01_lkt     TATGCCGGATATGATAAAAATTATAATATAAGTAAGACAAATTCTAAGGCCATTGCGGAT                10803
              ************************************************** *********

3718_lkt      GCCAAAAGTCATGCTGCAGCTGCTTCGGCAACTGCCACTGTTGAAAAAAATGAAGTAAAA                10896
3724_lkt      GCCAAAAGTCATGCTGCAGCTGCTTCAGCAACTGCCACTATTGAAAAAAATGAAGTAAAA                10905
ARU01_lkt     GCTAAAAGTCATGCTGCAGCTGCTTCTGCAACTGCTACCATTGAAAAAAATGAAGTGAAA                10863
              ** ********************** ******** **  *************** ***

3718_lkt      TTTAATAATGCGATCCGAGAATTTAAAAATAATCTGGCAAGATTGGAAGGGAAAGCTAAT                10956
3724_lkt      TTTAATAATGCGATCCGAGAATTTAAAAATAATCTGGCAAGATTGGAAGGGAAAGCTAAT                10965
ARU01_lkt     TTCAATAACGCGATCCGAGAGTTTAAAAATAATCTGGCAAAATTGGAAGGTAAAGTTAAT                10923
              ** ***** *********** ***************** ********* **** ****

3718_lkt      AAAAAAACGTCGGTAGGATCTAATCAGGTAGACTGGTATACGGATAAATATACATGGCAT                11016
3724_lkt      AAAAAAACGTCGGTAGGATCTAATCAGGTAGACTGGTATACGGATAAATATACATGGCAT                11025
ARU01_lkt     AAAGAAGTGTCAACTGGATTGAATCAGGTAGATTGGTATACAGATAAATATACATGGCAT                10983
              *** ** ***    **** ********** ******** ******************

3718_lkt      TCTTCTGAAAAAGCATACAAAAAATTGACATATCAATCAAAGAGAGGAGAAAAAGGGAAA                11076
3724_lkt      TCTTCTGAAAAAGCATACAAAAAATTGACATATCAATCAAAGAGAGGAGAAAAAGGGAAA                11085
ARU01_lkt     TCTTCTGAAAAAGCATATAAAAAATTGACATATCAATCAAAGAAAGGGGGAAAAAAATAAA                11043
              ***************** ********************** *** ******   ***

3718_lkt      AAATGAATTTAAGAGAGAGTAAATTTAGTGAGTTTTTAAAAAAATTCAAACATAACTTGTT                11136
3724_lkt      AAATGAATTTAAGAGAGAGTAAATTTAGTGAGTTTTTAAAAAAATTCAAACATAACTTGTT                11145
ARU01_lkt     AAATGAATTTAAGAGAGAGTAAATTTAATGATTTTTTAAAAAAATTCAAACATAACTTGTT                11103
              ************************** *** **************************

3718_lkt      TTGAAAGAGAAGAAGTGAAAGATGAGTTAGAAACAGTTGTATATCGAAGTTTTATGGAAG               11196
3724_lkt      TTGAAAGAGAAGAAGTGAAAGATGAGTTGGAAACGGTTGTATATCGAAGTTTTATGGAAG               11205
ARU01_lkt     TTGAAAGAGAAGAAGTGAAAGATGAATTGGAAACAGTGGTATATCGAAGTTTTATGGAAG               11163
              ************************* ** ***** ** *********************

3718_lkt      TAGAGGGACAAAATTTACCTATGGTAATTGTGATGGATAACAGTATTTATACGAATATCC               11256
3724_lkt      TAGAGGGACAAAATTTACCTATGGTAATTGTGATGGATAATAGTATCTATACGAATATTC               11265
ARU01_lkt     TAGAGGGACAAAATTTACCTATGGTAATTGTGCTGGATAACAGTATCTATACAAATATTC               11223
              ******************************** ******* ***** ***** ***** *

3718_lkt      GAGTGCAAATTGCTCCAAAAGTCATAAAAGATACTAATAAAGAAGCGGTACTTTCCTATA               11316
3724_lkt      GAGTGCAAATTGCCCCAAAAGTTATAAAAGATAGTAACAGGGAAGCAGTACTTTCCTATA               11325
ARU01_lkt     GAGTACAAATTGCCTCAAAAGTCATAAAAGATAGCAACAAAGAAGCAGTACTTTCCTATA               11283
              **** ********  ******* ********** ** *  ***** ************

3718_lkt      TCAATGAATTGAACCGAGAATACAAAGTATTTAAATATTATGTGACAGAGGATG      11370
3724_lkt      TCAATGAATTGAACCGAGAATACAAAGTATTTAAATATTATGTAACAGAGGATG      11379
ARU01_lkt     TCAATGAATTGAACCGAGAATACAAAGTATTTAAATACTATGTGACAGAGGATG      11337
              ************************************* ***** **********
```

Appendix Figure 4: Clustal Omega DNA alignment of Sanger sequence data from JCM 3718, JCM 3724 and ARU 01

leukotoxin operons. Contains genes of lktB, lktA and partial lktC.

Appendix 13: Clustal Omega protein sequence alignment of JCM 3718, JCM 3724 and ARU 01 leukotoxinA

```
3718_lktA    MSGIKNNVQRTRKRISDSKKVLMILGLLINTMTVRANDTITATENFGTKIEKKDNVYDIT    60
3724_lktA    MSGIKNNVRRTRKRISDSKKVLMILGLLINTMTAVADDSIKETQGFGTKVEKQGNVYDIT    60
ARU01_lktA   MSGIKNNVRRTRKRISDSKKVLMILGLLINTMTAVADDSIKETQGFGTKVEKQGNVYDIT    60
             ********:**********************. *.*:*. *: ****:**: ******

3718_lktA    TNKIQGENAFNSFNRFALTENNIANLYFGEKNSTGVNNLFNFVNGKIEVDGIINGIRENK    120
3724_lktA    TNKIKDKNAFNSFDKFHLEQNNIANMHFGSKGGKEAENLFNFVKGKIEVNGVINGIRDNK    120
ARU01_lktA   TNKIKDKNAFNSFDKFHLEQNNIANMHFGSKGGKEAENLFNFVKGKIEVNGVINGIRDNK    120
             ****: :******::* * :*****::**.* .. .:*****:*****:*:*****:**

3718_lktA    IGGNLYFLSSEGMAVGKNGVINAGSFHSIIPKQDDFKKALEEAKHGKVFNGIIPV-DGKV    179
3724_lktA    IGGNLYFLSSEGLLVGKTGVINAGTFHAISPKKEEYEKAFKDAQNSKVFDGIVPQQDGSI    180
ARU01_lktA   IGGNLYFLSSEGLLVGKTGVINAGTFRAISPKKEEYEKAFKDAQNSKVFDGIVPQQDGSI    180
             ************: ***.******:*:* **::::*:*.**:::*:..***:**:*  **.:

3718_lktA    KIPLNPNGSITVEGKINAVEGIGLYAADIRLKDTAILKTGITDFKNLVNISD-RINSGLT    238
3724_lktA    KIPLNPNGSITVEGKINAVEDIGLYAADIRLKDTARLKTGITEFKNLVNINE-TIHAGLT    239
ARU01_lktA   KIPLNPNGSITVEGKINAVDEIGLYAADIKLPETGALKTGVTDFHQLVNIKDSNVNAGLS    240
             *******************: ********:* :*. ****:*:*::****.:  :.:**:

3718_lktA    GDLKATKTKSGDIILSAHIDSPQKAMGKNSTVGKRIEEYVKGNTKANIESDAVLEADGNI    298
3724_lktA    GDLTAVKTKSGDIILSAHVDSPQKAVGENSTVGKRIEEYVKGNTKANIESDAVLEADGNI    299
ARU01_lktA   GDLKATKTGTGDIILSAKVEEAGHELE-SSTIYEQIGRNFKGKIKANIETSGSIEAEGHA    299
             ***.*.** :*******::. : : .**: ::* . .**: *****:.. :**:*.

3718_lktA    KISAKATNGRFIKKEGEKETYNTPLSLSDVEASVRVNKGKVIGKNVDITAEAKNFYDATL    358
3724_lktA    KISARATNGTFVKKEGEKEIYNTPLSLSDVEASVRVNKGKVIGKNVSITAEAKNFYDATL    359
ARU01_lktA   KIHAEASNGKLTKKDGEKEVYAPEFSLAEVEASVKVNKGKIKGKKVDISAEAKNYYDTPI    359
             ** *.*:** : **:**** *   :**::*****:***** **:*.*:*****:**. :

3718_lktA    VTKLAKHSFSFVTGSISPINLNGFLGLLTSKSSVVIGKDAKVEATEGKANIHSYSGVRAT    418
3724_lktA    VTKLAKHSFSFVTGSLSPINLNGFLGLLTSKSSVVIGKDAKVEAREGSANIHSYSGVRAT    419
ARU01_lktA   LTKVGKIAFSVGTGSLSPINMNGALGLLKSKASVFIGKDATVESTEGEANIRSYSGVRAS    419
             :**:.* :**. ***:*****:** ****.**:** ***** **: **.***:********:

3718_lktA    MGAATSPLKITNLYLEKANGKLPSIGAGYISAKSNSNVTIEGEVKSKGRADITSKSENTI    478
3724_lktA    MGAATSPLKITNLYLENADGKLPSIGAGYISAKSDSNVTIEGEVKSKERADITSKSENTI    479
ARU01_lktA   MGASTSPIKLTDLYLKTLEGKLPSVGAAYISTSSESDVTVEGKVSSKEKANITSKSENTI    479
             ***:***:*:*:****:  :******:** .***:.*:*:*:**:**:* .** :*:*********

3718_lktA    DASVSVGTMRDSNKVALSVLVTEGENKSSVKIAKGAKVESETDDVNVRSEAINSIRAAVK    538
3724_lktA    DASVSVGTMRDSNKVALSVLVTEGENKSSVKIAKGAKVESETDDVNVRSEAINSVRATVK    539
ARU01_lktA   DASVSVGTIRDSNKVGISVLVTEGKNHSSVKIAEGAKVESDG-DANVKSEAVNSIRAAVK    538
             ********:******.:*******:**:******:******:   *.**:***:**:**:**

3718_lktA    GGLGDSGNGVVAANISNYNASSRIDVDGYLHAKKRLNVEAHNITKNSVLQTGSDLGTSKF    598
3724_lktA    GGLGDSGNGVVAANISNYNASSRVDVDGYLHAKKRLNVEAHNITKNSVLQTGTNLGTSKF    599
ARU01_lktA   SGLSDNGAGVVAANISNYNSSSKVVVDGEVHAKKRLNTEAYNITKQNVLQTGTKTGTSKL    598
             .**.*.* ************:**:: *** :******.**:****:.*****:. ****:

3718_lktA    MNDHVYESGHLKSILDAIKQRFGGDSVNEEIKNKLTDLFSVGVSATIANHNNSASVAIGE    658
3724_lktA    MNDHVYESGHLKSILDAIKQRFGGDSVNEEIKNKLTDLFSVGVSATIANHNNSASVAIGE    659
ARU01_lktA   MNELVFESSHTKALIDALQYRFGGD---EKIKSKLTDLFSVGASATVANHNNTSSVSLGQ    655
             **: *:**.* *::**:: *****    *:**.*********.***:*****:.**:..::

3718_lktA    SGRLSSGVEGSNVRALNEAQNLRATTSSGSVAVRKEEKKKLIGNAAVFYGNYKNNASVTI    718
3724_lktA    SGRLSSGVEGSNVRALNEVQNLRATTSSGSVAARKEEKKKFIGNAAVFYGNYKNNASVTI    719
ARU01_lktA   KSKLVSGVEGTNIKALTETQQLRSTTTSGSLATRGEEKKKLVGNAAVFYGNYKNDASVTI    715
             ..:* *****:*::**.*.*:*:**:**:***.* *  ******:*************:*****

3718_lktA    ADHAELVSEGKIDINSENKIEYKNPSKMAKSVIDKLELLKRAFGKETK-TPEYDPKDIES    777
3724_lktA    ADQAELTSEGKIDTISENKTEYKNPSKMAKSVIEKLELLKRAFGKETK-TPEYDPKDIES    778
ARU01_lktA   ADDAQVVSEGKIDTISENKLEYKNPSKMAEEVVEKLEILKRAFEKEKKDETTFDPKDIDS    775
             **.*::******  **** *********:.*::**:***** **.*   :*****:*

3718_lktA    IEKLLNAFSEKLDGKPELLLNGERMTIILPDGTSKTGTAIEIANYVQGEMKKLEEKLPKG    837
3724_lktA    IEKLLNAFSEKLDGKPELLLNGERMTIILPDGTSKTGTAIEIANYVQGEMKKLEEKLPKG    838
ARU01_lktA   MKKLLTEFSKKLDDKPEILLNGEKITIVLPDGTSKTGTVPELAEYVKEEMKKLEAKLPTG    835
             :::***. **:*** ***:*****::**:**********. *:*:** ****** ***.*

3718_lktA    FKAFSEGLSGLIKETLNFTGVGNYANFHTFTSSGANGERDVSSVGGAVSWVEQENYSKVS    897
3724_lktA    FKAFSEGLSGLIKETLNFTGVGNYANFHTFTSSGANGERDVSSVGGAVSWVEQENYSKVS    898
ARU01_lktA   FKAFSEGLSGLLKESLAFTGVGNYANFHTFTSAGTNAKRDTSSVGGAVSWVEMDNHSKVS    895
             ***********:**:* *****************:**.*.:**.*********** :*:*****

3718_lktA    VGKGAKLAAKKDLNIKAINKAETVNLVGNIGLARSSTSGSAVGGRLNVQRSKNSAIVEAK    957
3724_lktA    IGKGAKLAAKKDLNIKAINKAETVNLVGNIGLARSSTSGSAVGGRLNVQRTKNSAIVEAK    958
ARU01_lktA   IGRGSKITAKGDLNVKAINKTETVNLVGKLGLSKSSDSATAVGGGLNVQKSNTSSIVVTK    955
             :*:*:*::* ***.*:*****:.:***.:*** *.:*** ****::.*:** :*
```

```
3718_lktA    EKAELSGENINADALNRLFHVAGSFNGGSGGNAINGMGSYSGGISKARVSIDDEAYLKAN      1017
3724_lktA    EKAELSGENINADALNSLFHVAGSFNGGSGGNAINGMGSYSGGISKARVSIDDETYLTAN      1018
ARU01_lktA   EKAELSGENIHTNAFNNVFHVAASLNAGTGGNGINGMGSYSGGSSKSRVSIDDETHLKAK      1015
             **********.::*:* :****.*:*.*:***.*********  **:*******::*.*:


3718_lktA    KKIALNSKNDTSVWNAAGSAGIGTKNAAVGVAVAVNDYDISNKASIEDND---EGQSKYD      1074
3724_lktA    KKIALNSKNDTSVWNVAGSAGIGTKNAAVGVAVAVNDYDISNKASIEDND---EGQSKYD      1075
ARU01_lktA   KKIYLDSENNTSVTNVAGAVAIGSKNAAVGAAVAINDYDISNKVSIEDNDTEDGGKSKYD      1075
             *** *:*:*:*** *.**:..**:******.***:******** ******    *:****


3718_lktA    KNKDEVTVTAESLEVDAKTTGTINSISVAGGINKVGSKPSEEKPKSEERPEGFFGKIGNK      1134
3724_lktA    KNKDEVTVTAESLEVDAKTTGTINSISVAGGINKVGSKPSEEKPKSEERPEGFFDKIGNK      1135
ARU01_lktA   KNKEEVTVTAEDLHANAKTTGTINAVSVAGGISKVGEAE--------SDGTGIFSKIGNK      1127
             ***:*******.*..:*********::******.***.           .  *:* *****


3718_lktA    VDSVKNKITDSMDSLTEKITNYISEGVKKAGNLPSNVSHTPDKGPSFSLGASGSVSFNNI      1194
3724_lktA    VDSVKNKITDSMDSLTEKITNYISEGVKKAGNLPSNVSHTPDKGPSFSLGAAGSVSVNNI      1195
ARU01_lktA   VDSVKGKITDAIGFVTDQVTNYVYRGSNGTEDGIPNISKESSKLPSFSLGASGSVSVNNI      1187
             ***** ****::  :*::***: .* : : :   *:*: .* *******:****.***


3718_lktA    KKETSAVVDGVKINLKGANKKVEVTSSDSTFVGAWGGSAALQWNHIGSGNSNISAGLAGA      1254
3724_lktA    KKETSAVVDGVKVNLKGDNKKAEVIASDSAFIGAWSGAAALQWNHVGSGESNTTVGLAGA      1255
ARU01_lktA   KKETSAVMDGVKISLSGANKEVGVTSSDSTFVGAWSGSAALQWNHIGSGNSNFSASLAGA      1247
             *******:****:.*.* **:. * :***:*:***.*:*******:***:**  :..****


3718_lktA    AAVNNIQSKTSALVKNSDIRNANKFKVNALSGGTQVAAGAGLEAVKESGGQGKSYLLGTS      1314
3724_lktA    AAVNNVQSKTKAFVKDSSITNANKFKVNAVSGGTQVAAGMGLEAVKESGGQGRSYLLGTS      1315
ARU01_lktA   AAANNIQSKTNAVVKNSSIQNANKFKVNALSGGTQVAAGAGLEAAKESGGQGKSYLLGTS      1307
             **.**:****.*.**:*.* *********:********* ****.*******:*******


3718_lktA    ASINLVNNEVSAKSENNTVAGESESQKMDVDVTAYQADTQVTGALNLQAGKSNGTVGATV      1374
3724_lktA    ASMNFIDNEVSSESENNIITGESEDKRADVDVTAYESDTQVTGGFNLQAGQSKGTVGAAV      1375
ARU01_lktA   ASINLVDNEVSAKSENNIVEGESENKKMNVDVTAYEADTQVTGALDLQAGKSSGTAGAAI      1367
             **:*::::****:.**** : ****.::  :*******::******.::****:*.**.**::


3718_lktA    TVAKLNNKVNASISGGRYTNVNRADAKALLATTQVTAAVTTGGTISSGAGLGNYQGAVSV      1434
3724_lktA    TVAKLNNKVKAGIKGGNYTNINRADTKALLATTQVTSALSVGGTNSFGSGLGNYQGAVSV      1435
ARU01_lktA   TVAKLNNKVNASISGGKYTKVNRADAKALLATTQVTAAVATGGTISSGAGLGNYQGVVSV      1427
             *********:*.*.**:**:*****:*********:::.*** * *:*******.***


3718_lktA    NKIDNDVEASVDKSSIEGANEINVIAKDVKGSSDLAKEYQALLNGKDKKYLEDRGINTTG      1494
3724_lktA    NNINNDVSATVDKSSIKNAKELNVIAKDIKGSSELAKEYQSLLNGKDREYLEAHGIDTTG      1495
ARU01_lktA   NKIDNDVEAGIDKSSIEGADEVNVIAKDVKGSSNLAKEYQALLNGKDEKYLKDRGIDTTG      1487
             *:*:***.* :******: *.*:******:****:******:******.:**: :**:***


3718_lktA    NGYYTKEQLEKAKKKEGAVIVNAALSVAGTDKSAGGVAIAVNTVKNKFKAELSGSNKEAG      1554
3724_lktA    KNYYTEEQLKEAAKRDGAVIVNAAASIAGSDKSSGGGAGVAVNLVKNKFKAEVIGEDKVDK      1555
ARU01_lktA   NGYYTKEQLEEAKKKEGAVIVNAALSVAGTDKSAGGVAIAVNTVKNKFKAELNGKDKEVE      1547
             : ***:***:* *::********* *:**:***:*** .:*** ********: *.:*


3718_lktA    EDKIHAKHVNVEAKSSTVVVNAASGLAISKDAFSGMGSGAWQDLSNDTIAKVDKGRISAD      1614
3724_lktA    TKDIAAEKINVDAKSSTVIVNTAAGLAVSQNSFAGLGSFAWQELNNEVSTKVENIKAETD      1615
ARU01_lktA   EGKIRAKHVNVEAKSSTVVVNAASGLAVSKDAFSGMGSGAWQDLSNETVAKVNKGRISTD      1607
              .* *:::**:*******:**:*:***:*:::*:*:*:** ***:*.*:. :**:: : .:*


3718_lktA    SLNVNANNSILGVNVAGTIAGSLSTAVGAAFANNTLHNKTSALITGTKVNPFSGKNTK--      1672
3724_lktA    RLNVVARNNTLGVNIAGNIAGTGSAAVGAALAHNSLKNKASSVTASNFKPFSKTNAKMG      1675
ARU01_lktA   SLNVTANNSVLGVNVAGTIAGSLSTAVGAAFANNTLHNKTSALITETKVNSFSGKDTK--      1665
              *** *.*. ****:**.***: *:*****:*.*:*:**:*  :* ::.: ** .::*


3718_lktA    VNVQALNDSHITNVSAGGAASIKQAGIGGMVSVNRGSDETEALVSDSEFEGVSSFNVDAK      1732
3724_lktA    INVQALNDSHITNVSAGVVASAKASGISGMASINRGSDETEALVSDSEFEGVNSFNVDAK      1735
ARU01_lktA   INAQALNDSHITNVSAGVAASVKQVGIGGMVSVNRGSDETEALVSDSEFDGVNSVDVRAE      1725
             :*.************** .** *  **.**.*:*****************:**.*:.:* *:


3718_lktA    DQKTINTIAGNANGGKAAGVGATVAHTNIGKQSVIAIVKNSKITTANDQDRKNINVTAKD      1792
3724_lktA    DQKIVNTIAGNKAAGVGATVAHTNIGKQSVTAIVKNSKIKTANNQDKKNIHVTAKD      1795
ARU01_lktA   DKKVLNTIAGNINGGQAAGVGATVAHTNIGKQSVTAAIKNSKITTAKNQDKKNINVTAKD      1785
             *:* :****** ***:***************** * :*****.**::**:*** *****


3718_lktA    YTMTNTIAVGVGGAKGASVQGASASTTLNKTVSSHVDQTDIDKDLEEENNGNKEKANVNV      1852
3724_lktA    SVIMNTVAAGVGGAKGASVQGASASTAFNKTVASHVEHTDIDKDLEKENDENKEKANFNI      1855
ARU01_lktA   SVVMNTVAAGVGGAKGASVQGSSASTTLNKTVSAHVEKTDIDK----ENHESKEKADFNV      1841
              .: **:*.**************:****::*****::**     **. .****:.*:


3718_lktA    LAENTSQVVTNATVLSGASGQAAVGAGVAVNKITQNTSAHIKNSTQNVRNALVKSKSHSS      1912
3724_lktA    LAENTSQVVTNATVLSGASGQAAVGAGVAVNKITQNTSAHIKNSTQNVRNALVKSKANSS      1915
ARU01_lktA   LAENTSQVVTNATVLAGASGQAAVGAGVAVNKITQNTSAHIKNSTQNVRNALLKSKANSS      1901
             ***************:*****************************************:***.**


3718_lktA    IKTIGIGAGVGAGGAGVTGSVAVNKIVNNTIAELNHAKITAKGNVGVITESDAVIANYAG      1972
3724_lktA    IKTIGIGAGIGAGGVGSSGSVAVNKITNNNTASVEHSNIFARGNVGVITESDAVIANYAG      1975
ARU01_lktA   IKTIGIGAGVGVGGVGVSGSVAVNKITNNNTAVVEHSDIFAKGNIGVIAESDAVIANYAG      1961
             *********:*.**.* :*********.**. * ::*:.* *:**:***:***********


3718_lktA    TVSGGARAAIGASTSVNEITGSTKAYVKDSTVIAKEETDDYITTQGQVDKVVDKVFKNLN      2032
3724_lktA    TVSGGSHAGVGGSTSVNEIQGDTIASVKHSKIEAKEETEDVIETQGKVKEVVEKVFQDIG      2035
ARU01_lktA   TVSGGGHVGVGGSTGVNEIEGDTIASVKKSKLEAKEETEDVIETQGKVKEVVENVLQDIG      2021
             *****.:..:*.**.**** *.* .*  * **.*.: *****:* * ***:*.:**::*::::
```

259

```
3718_lktA     INEDLSQKRKISNKKGFVTNSSATHTLKSLLANAAGSGQAGVAGTVNINKVYGETEALVE        2092
3724_lktA     INEDLSKKREKSKKRGFVVNSSSTHTLKSLLANAAGSGQAGVAGTVNINKVYGATEALVE        2095
ARU01_lktA    INDDLSKKREKSNKKGFVVNSSSTHTIKSLLANAAGAGKAAVAGTVNVNKITGETKALVE        2081
              **:***:**: *:*:*** .***:***:*********:*:*.******:**: * *:****


3718_lktA     NSILNAKHYSVKSGDYTNSIGVVGSVGVGGNVGVGASSDTNIIKRNTKTRVGKTTMSDEG        2152
3724_lktA     DSILNAKHYSVKSGDYTNSIGAVGSVGVGGNVGAGASSDTNIIKRNTKTRVGKTTMSDEG        2155
ARU01_lktA    SSVLNVKNSSIHAGDYTNSIGAVGSASGALNAGIGASSNTNLFSRNTKTIVHDTKISTEK        2141
              .*:**.*. *:::********.***.. . *.* ****:**:: ***** * .*.:* *


3718_lktA     FGEEAEITADSKQGISSFGVGVAAAGVGAGVAGTVSVNQFAGKTEVDVEEAKILVKKAEI        2212
3724_lktA     FGEEAEITADSKQGISSFGVGVAAAGVGAGVAGTVSVNQFAGKTEVDVEEAKIFVKKAEI        2215
ARU01_lktA    EGQSSEITADSKQGISSFGVGVGAGVVGAGVAGTVSVNKLSGKTEVDIQKSDISTKSADI        2201
               *:.:*************** .*. ************:::******:::: .* .*.*:*


3718_lktA     TAKRYSSVAIGNAAVGVAAKGAGIGAAVAVTKDESNTRARVKNSKIMTRNKLDVIAENEI        2272
3724_lktA     TAKRYSSVAIGNAAVGVAAKGAGIGAAVAVTKDESNTRARVKNSKIMTRNKLDVIAENEI        2275
ARU01_lktA    SAKHYGVLSTGNGSVGAAVKGAGVGAAVAVTKDLTNTNVRIKDSKIVTKTKLDVIAQNHT        2261
              :**:*. :: **.:**.*.****:********* :**..*:*:***:*:.******:*.


3718_lktA     KSGTGIGSAGAGILAAGVSGVVSVNNIANKVETDIDHSTLHSSTDVNVKALNKISNSLTA        2332
3724_lktA     KSGTGIGSAGAGILAAGVSGVVSVNNIANKVETDIDHSTLHSSTDVNVKALNKISNSLTA        2335
ARU01_lktA    KVNGGMVGIGAAGVGAGLAGTVSVNNITSKVGTEIDHSELLSEEDVNVKALNKVDSSMMA        2321
              *    *: . **. :.**::*.******:.** *:**** * *. *********:..*: *


3718_lktA     GGGAAGLAAVTGVVSVNTINSSVIARVHNNSDLTSVREKVNVTAKEEKNIKQTAANAGIG        2392
3724_lktA     GGGAAGLAAVTGVVSVNTINSSVIARVHNNSDLTSVREKVNVTAKEEKNIKQTAANAGIG        2395
ARU01_lktA    AGGAAGLGAASGVVSVNTVNTSVVTRVHNNSKITSSKGSTNVKAEEEKNIKQVVANAVAG        2381
              .******.*.:*******:*:**:*****.:** : ..**.*:*******.*** *


3718_lktA     GAAIGANVLVNNFGTAVEDRKNSEGKGTEVLKTLDEVNKEQDKKVNDATKKILQSAGIST        2452
3724_lktA     GAAIGANVLVNNFGTAVEDRKNSEGKGTEVLKTLDEVNKEQDKKVNDATKKILQSAGIST        2455
ARU01_lktA    GVSVGANVLVNNFGTAVEDKAAETGKGTEILKTLEEINKDQDQKISEDAAKILKSNGIDT        2441
              *.::***************:   . *****:****:*:**:**:*:. : ***:* **.*


3718_lktA     EDTSVKADRGDTQGEGIKAIVKTSDIIGKNVDITTEDKNNITSTGGLGTAGLASASGTVA        2512
3724_lktA     EDTSVKADRGDTQGEGIKAIVKTSDIIGKNVDITTEDKNNITSTGGLGTAGLASASGTVA        2515
ARU01_lktA    KDYAVKADRGETQAEGIKVFVENSSIKGKNTNIAANQTDNVTSTGGAGSAGVASALGTVG        2501
              :* :*******:**.****:.*:.* *** .*:*::::.:*:***** *:**:*** ***.


3718_lktA     VTNIKRNSGVTVENSFVKAAEKVNVRSDITGNVALTAYQGSVGALGIGAAYAELNSNGRS        2572
3724_lktA     VTNIKRNSGVTVENSFVKAIEKLNVRSDITGNVALTAYQGSVGALGIGAAYTELNSNGRS        2575
ARU01_lktA    VTNIKRNIGVFINNADVKADEKLSVRSDISGNVSLTSYQGTVGALGLGAAYSQINANGQS        2561
              ******* ** ::*: *** **:.*****:*** **:***:*****:****:::*:**:*


3718_lktA     NISIKNSKLLGKNIDVIVKDKSELRAEAKGLTVGAVAAGAIISKAKNEMNSEVEIEKSIF        2632
3724_lktA     NININRNSKLLGKNIDVIVKDKSELTAEAKGLTAGAVAAGAIISKAKNEMNSEIEIEKSIF        2635
ARU01_lktA    NITIKNSRFAAKHIDVIAKDKSQLKAEAKGLTVGAVAAGAIISKANNEMNSEIEIEKSIF        2621
              **.*:**:: .*.****.****:* ******* ************:******:*******


3718_lktA     NEENRVTSPSKGIGREINVKVEKENRVTAESQGASVGAVAGAGIISEAKDAGSSYLKVST        2692
3724_lktA     NEENRVNSPSKGIGREINVKVEKENRVTAESQGVSVGAVAGAGIISDAKDAGSSYLKVST        2695
ARU01_lktA    NEENRVSSPSKGIGREINVKVEKENRVTAESQGASVGAVAGAGIISDAKDAGNSYLKVST        2681
              ******.********************** .*************:*****.*******


3718_lktA     KSGRSIFHADNVNMEATHKMKVTAVSKAVTGSVLGGVGVTKAEATAAGKTMVEVEEGNLF        2752
3724_lktA     KSGRSIFHADNVNMEATHKMKVTAASKAVTGSVLGGVGVTKAEAVAAGKTMIEVEEGNLF        2755
ARU01_lktA    KSGRSIFHADNVNMEATHKMKVTAGSKAVTGSVLGGVGVTKAEATSAGKVMVEIEEGNLF        2741
              ************************ .***************** :*** .*:*:******


3718_lktA     RTNRLNAISKVEGLDEDKVTAKSSVVSGNGGGIAGAGVNTSTAQSNTESVVRLRKQDYEN        2812
3724_lktA     RTNRLNTISKVEGLDEDKVTAKSSAISGNGGGISGAGVNTSTAQSNTESVVHLQRQDYEN        2815
ARU01_lktA    RTNRLNAISKVEGLDEDKVTAKSFVVSGNGGGIAGAGVNTSTAQSNTESVVRLRKQDYEN        2801
              ******.********************  .:*******:***************:*::*****


3718_lktA     NDYTKKYISEVNALALNDTKNEANIESLAVAGVHAQGTNKAFTRSNKLTSTTVNGGNVSQ        2872
3724_lktA     NDYTKKYISEVNALALNDTKNEANIESLAVAGVHTQGTNKAFTRSNKLTSTTVNGGNVYQ        2875
ARU01_lktA    NDSTKKYISEVNALAVNDTKNEANIESLEVAGVYAQGTNKAFTKSDKLTSTTVNGGKVYQ        2861
              ** *************:*********** ****:.*********:*:**********:* *


3718_lktA     LRAKALAKNENYGNVKGTGGALVGAETAAVENYTKSTTGALVAGNWEIGDKLETIARDNT        2932
3724_lktA     LRAKALAKNENYGNVKGTGGALVGAETAAVENYTKSTTGALVAGNWEIGDKLETIARDNT        2935
ARU01_lktA    LRAKALAKNENYGNVKGSGGALVGAETAAVENYVKSTTGASVAGEWEIGDKLETIAKDNT        2921
              *****************:*************** .****** ***.**********:***


3718_lktA     IVRVNGDGTKGGLVGKNGISVKNTISGETKSSIEDKARIVGTGSVNVDALNELDVDLQGK        2992
3724_lktA     IVRVNGDGTKGGLVGKNGISVKNTISGETKSSIEDKARIVGTGSVNVDALNELDVDLQGK        2995
ARU01_lktA    IVKVNGDGTKGGLVGKNGISVKNTILGETKSFIDDKAKIAGTGSVNVEALNELDVDLQGK        2981
              **:********************** ***** *:***:* .*******:************


3718_lktA     SGGYGGIGIGNVDVNNVIKKNVEAKIGRHAIVETTGKQEYQAFTRAKVNILGKGDAAAAA        3052
3724_lktA     SGGYGGIGIGNVDVNNVIKKNVEAKIGRHAIVETTGKQEYQAFTRAKVNILGKGDAAAAA        3055
ARU01_lktA    SGGYGGIGVGNVDISNVIKKDTEAKIGHAVVETTGKQEYQAFTKAKVNILGKGDAAAAA         3041
              ********:****:.*****:.*****:**.************:******.***********


3718_lktA     AISNVHISNEMDIKNLAKQYASSQLITKNSKNNITLASSSESNVNVHGVAEARGAGAKAT        3112
3724_lktA     AISNVHISNEMDIKNLAKQYASSQLITKNSKNNITLASSSESNVNVHGVAEARGAGAKAT        3115
ARU01_lktA    AISNVHVSNEMDIKNLAKQYAYSQLTTKNSKNDITLASSSELDVNVHGVAEAKGVGAKAT        3101
              ******:************** *** ******:*:*******  :**********:*.*****
```

260

```
3718_lktA      VSVKNQINRTNNVDLAGKIKTEGNINVYAGYDKNYNISKTNSKAIADAKSHAAAASATAT      3172
3724_lktA      VSVKNQINRTNNVDLAGKIKTEGNINVYAGYDKNYNISKTNSKAIADAKSHAAAASATAT      3175
ARU01_lktA     SSVKNQINRANNVDLAGKIKTEGNINVYAGYDKNYNISKTNSKAIADAKSHAAAASATAT      3161
               ********:***************************************************


3718_lktA      VEKNEVKFNNAIREFKNNLARLEGKANKKTSVGSNQVDWYTDKYTWHSSEKAYKKLTYQS      3232
3724_lktA      IEKNEVKFNNAIREFKNNLARLEGKANKKTSVGSNQVDWYTDKYTWHSSEKAYKKLTYQS      3235
ARU01_lktA     IEKNEVKFNNAIREFKNNLAKLEGKVNKEVSTGLNQVDWYTDKYTWHSSEKAYKKLTYQS      3221
               :******************:****.**:.*.* ************************


3718_lktA      KRGEKGKK      3240
3724_lktA      KRGEKGKK      3243
ARU01_lktA     KKGEKNKK      3229
               *:*** **
```

Appendix Figure 5: Clustal Omega protein sequence alignment of JCM 3718, JCM 3724 and ARU 01 leukotoxinA

.

Appendix 14: PCR primers for *lktA* detection within the clinical strain collection

Appendix Table 5: PCR primers for lktA detection within the clinical strain collection.

Primer	5' – 3' primer sequence
Ludlam_LT1_F	CGAAAACTCCAGAATATGATCCGAAAGA
Ludlam_LT1_R	CTACCCACGAAACAGCTCCTCCCACAG
Ludlam_LT2_F	ATCGGAGTAGTAGGTTCTGTTGGTGTTG
Ludlam_LT2_R	GGCTGCTGCGACTCCGACTC
lktA_F	AAATGGTGAAAGAATGACAA
lktA_R	TGCATAATTTCCTACTCCTG
lkt1_F	CTTAGTCGGAGCGGAAACAG
lkt1_R	TCTTTCCGACAAGACCTCCT
lkt2_F	TGTATTGGGAGGAGTTGGAGT
lkt2_R	TCGTAAATGCTTTGTTTGTTCCT

Appendix 15: Full leukotoxin PCR results

Appendix Table 6: Full leukotoxin PCR results, showing amplification (+) or non-amplification (-) for each isolate and primer

pair combination.

Isolate number	lktA	Ludlam_LT1	Ludlam_LT2	lkt1	lkt2
1	+	+	+	+	+
5	-	+	+	+	+
11	-	+	-	+	+
21	+	+	+	+	+
24	-	+	+	+	+
30	-	+	+	+	+
39	+	+	+	+	+
40	+	+	+	+	+
41	-	+	-	+	+
42	-	+	+	+	+
52	-	+	-	+	+
59	-	+	+	+	+
62	+	+	+	+	+
70	+	+	+	+	+
80	-	+	+	+	+
82	-	+	+	+	+
86	-	+	+	+	+
87	-	+	+	+	+
89	-	+	+	+	+
90	+	+	+	+	+
91	+	+	+	+	+
92	+	+	+	+	+
93	-	+	+	+	+
94	-	+	+	+	+
95	-	+	+	+	+
JCM 3718	+	+	+	+	+
JCM 3724	+	+	+	+	+
ARU 01	+	+	+	+	+

Appendix 16: Alignment of Ludlam_LT1 amplicon Sanger sequences

```
LT1_1    AAAATTGGATGGAAAACCGGAGCTTTTACTAAATGGTGAAAGAATGACAATTATTCTTCCG    60
LT1_5    AAAATTGGATGATAAGCCTGAAATTTTATTGAACGGAGAAAAAATAACGATTGTTCTTCCA    60
LT1_11   AAAATTGGATGATAAGCCTGAAATTTTATTGAACGGAGAAAAAATAACGATTGTTCTTCCA    60
LT1_21   AAAATTGGATGGAAAACCGGAGCTTTTACTAAATGGTGAAAGAATGACAATTATTCTTCCG    60
LT1_24   AAAATTGGATGATAAGCCTGAAATTTTATTGAACGGAGAAAAAATAACGATTGTTCTTCCA    60
LT1_30   AAAATTGGATGATAAGCCTGAAATTTTATTGAACGGAGAAAAAATAACGATTGTTCTTCCA    60
LT1_39   AAAATTGGATGGAAAACCGGAGCTTTTACTAAATGGTGAAAGAATGACAATTATTCTTCCG    60
LT1_40   AAAATTGGATGGAAAACCGGAGCTTTTACTAAATGGTGAAAGAATGACAATTATTCTTCCG    60
LT1_41   AAAATTGGATGATAAGCCTGAAATTTTATTGAACGGAGAAAAAATAACGATTGTTCTTCCA    60
LT1_42   AAAATTGGATGATAAGCCTGAAATTTTATTGAACGGAGAAAAAATAACGATTGTTCTTCCA    60
LT1_52   AAAATTGGATGATAAGCCTGAAATTTTATTGAACGGAGAAAAAATAACGATTGTTCTTCCA    60
LT1_59   AAAATTGGATGATAAGCCTGAAATTTTATTGAACGGAGAAAAAATAACGATTGTTCTTCCA    60
         **********  **  **  **   ***** * **  ** **** *** ** *** *******

LT1_1    GATGGAACTTCAAAAACAGGAACTGCTATAGAAATTGCAAACTATGTTCAGGGAGAAATG    120
LT1_5    GATGGAACTTCTAAAACAGGAACGGTTCCGGAACTTGCAGAATATGTGAAAGAGGAAATG    120
LT1_11   GATGGAACTTCTAAAACAGGAACGGTTCCGGAACTTGCAGAATATGTGAAAGAGGAAATG    120
LT1_21   GATGGAACTTCAAAAACAGGAACTGCTATAGAAATTGCAAACTATGTTCAGGGAGAAATG    120
LT1_24   GATGGAACTTCTAAAACAGGAACGGTTCCGGAACTTGCAGAATATGTGAAAGAGGAAATG    120
LT1_30   GATGGAACTTCTAAAACAGGAACGGTTCCGGAACTTGCAGAATATGTGAAAGAGGAAATG    120
LT1_39   GATGGAACTTCAAAAACAGGAACTGCTATAGAAATTGCAAACTATGTTCAGGGAGAAATG    120
LT1_40   GATGGAACTTCAAAAACAGGAACTGCTATAGAAATTGCAAACTATGTTCAGGGAGAAATG    120
LT1_41   GATGGAACTTCTAAAACAGGAACGGTTCCGGAACTTGCAGAATATGTGAAAGAGGAAATG    120
LT1_42   GATGGAACTTCTAAAACAGGAACGGTTCCGGAACTTGCAGAATATGTGAAAGAGGAAATG    120
LT1_52   GATGGAACTTCTAAAACAGGAACGGTTCCGGAACTTGCAGAATATGTGAAAGAGGAAATG    120
LT1_59   GATGGAACTTCTAAAACAGGAACGGTTCCGGAACTTGCAGAATATGTGAAAGAGGAAATG    120
         *********** ***********  * *    *** ***** * *****    *  ******

LT1_1    AAAAAATTAGAGGAAAAATTACCGAAAGGATTTAAAGCTTTTTCAGAAGGATTGAGTGGA    180
LT1_5    AAAAAACTGGAAGCAAAATTACCGACAGGATTTAAAGCTTTTTCTGAAGGATTAAGCGGA    180
LT1_11   AAAAAACTGGAAGCAAAATTACCGACAGGATTTAAAGCTTTTTCTGAAGGATTAAGCGGA    180
LT1_21   AAAAAATTAGAGGAAAAATTACCGAAAGGATTTAAAGCTTTTTCAGAAGGATTGAGTGGA    180
LT1_24   AAAAAACTGGAAGCAAAATTACCGACAGGATTTAAAGCTTTTTCTGAAGGATTAAGCGGA    180
LT1_30   AAAAAACTGGAAGCAAAATTACCGACAGGATTTAAAGCTTTTTCTGAAGGATTAAGCGGA    180
LT1_39   AAAAAATTAGAGGAAAAATTACCGAAAGGATTTAAAGCTTTTTCAGAAGGATTGAGTGGA    180
LT1_40   AAAAAATTAGAGGAAAAATTACCGAAAGGATTTAAAGCTTTTTCAGAAGGATTGAGTGGA    180
LT1_41   AAAAAACTGGAAGCAAAATTACCGACAGGATTTAAAGCTTTTTCTGAAGGATTAAGCGGA    180
LT1_42   AAAAAACTGGAAGCAAAATTACCGACAGGATTTAAAGCTTTTTCTGAAGGATTAAGCGGA    180
LT1_52   AAAAAACTGGAAGCAAAATTACCGACAGGATTTAAAGCTTTTTCTGAAGGATTAAGCGGA    180
LT1_59   AAAAAACTGGAAGCAAAATTACCGACAGGATTTAAAGCTTTTTCTGAAGGATTAAGCGGA    180
         ****** * **  * ***********  ***************** ******** ** ***

LT1_1    CTGATTAAAGAAACTTTGAATTTTACAGGAGTAGGAAATTATGCAAATTTTCACACTTTT    240
LT1_5    TTGTTAAAAGAAAGCTTGGCTTTTACAGGAGTAGGAAATTATGCAAATTTCCATACATTT    240
LT1_11   TTGTTAAAAGAAAGCTTGGCTTTTACAGGAGTAGGAAATTATGCAAATTTCCATACATTT    240
LT1_21   CTGATTAAAGAAACTTTGAATTTTACAGGAGTAGGAAATTATGCAAATTTTCACACTTTT    240
LT1_24   TTGTTAAAAGAAAGCTTGGCTTTTACAGGAGTAGGAAATTATGCAAATTTCCATACATTT    240
LT1_30   TTGTTAAAAGAAAGCTTGGCTTTTACAGGAGTAGGAAATTATGCAAATTTCCATACATTT    240
LT1_39   CTGATTAAAGAAACTTTGAATTTTACAGGAGTAGGAAATTATGCAAATTTTCACACTTTT    240
LT1_40   CTGATTAAAGAAACTTTGAATTTTACAGGAGTAGGAAATTATGCAAATTTTCACACTTTT    240
LT1_41   TTGTTAAAAGAAAGCTTGGCTTTTACAGGAGTAGGAAATTATGCAAATTTCCATACATTT    240
LT1_42   TTGTTAAAAGAAAGCTTGGCTTTTACAGGAGTAGGAAATTATGCAAATTTCCATACATTT    240
LT1_52   TTGTTAAAAGAAAGCTTGGCTTTTACAGGAGTAGGAAATTATGCAAATTTCCATACATTT    240
LT1_59   TTGTTAAAAGAAAGCTTGGCTTTTACAGGAGTAGGAAATTATGCAAATTTCCATACATTT    240
         ** * *******  ***   ************************** ** ** ***

LT1_1    ACCTCTTCCGGAGCTAATGGAGAAAGAGATGTCTCTTCTGTGGGAGGAGCT    291
LT1_5    ACCTCGGCAGGAACCAATGCAAAAAGAGATACTTCTTCTGTGGGAGGAGCT    291
LT1_11   ACCTCGGCAGGAACCAATGCAAAAAGAGATACTTCTTCTGTGGGAGGAGCT    291
LT1_21   ACCTCTTCCGGAGCTAATGGAGAAAGAGATGTCTCTTCTGTGGGAGGAGCT    291
LT1_24   ACCTCGGCAGGAACCAATGCAAAAAGAGATACTTCTTCTGTGGGAGGAGCT    291
LT1_30   ACCTCGGCAGGAACCAATGCAAAAAGAGATACTTCTTCTGTGGGAGGAGCT    291
LT1_39   ACCTCTTCCGGAGCTAATGGAGAAAGAGATGTCTCTTCTGTGGGAGGAGCT    291
LT1_40   ACCTCTTCCGGAGCTAATGGAGAAAGAGATGTCTCTTCTGTGGGAGGAGCT    291
LT1_41   ACCTCGGCAGGAACCAATGCAAAAAGAGATACTTCTTCTGTGGGAGGAGCT    291
LT1_42   ACCTCGGCAGGAACCAATGCAAAAAGAGATACTTCTTCTGTGGGAGGAGCT    291
LT1_52   ACCTCGGCAGGAACCAATGCAAAAAGAGATACTTCTTCTGTGGGAGGAGCT    291
LT1_59   ACCTCGGCAGGAACCAATGCAAAAAGAGATACTTCTTCTGTGGGAGGAGCT    291
         *****  * *** * **** * ********    *****************
```

Appendix Figure 6: Clustal Omega alignment of Ludlam_LT1 amplicon Sanger sequence data. lktA forward and reverse

PCR primer binding positions are highlighted in yellow. Those not highlighted in yellow lack homology and did not amplify

lktA product.

Appendix 17: Full results of the HL-60 cell line cytotoxicity assay

Appendix Table 7: Full results of the HL-60 cell line cytotoxicity assay. All viabilities are expressed as a percentage of the negative control.

	Percentage viabilities per treatment group						
	Negative Control	125 µg/ml	150 µg/ml	175 µg/ml	250 µg/ml	Positive Control	Flow Through
JCM 3718	100.00	86.79	81.77	76.39	64.82	86.49	98.28
	100.00	86.72	84.96	81.02	61.71	85.13	97.86
	100.00	87.82	84.83	83.07	63.33	86.82	99.14
JCM 3724	100.00	79.60	66.12	55.74	NT	90.34	96.25
	100.00	80.87	74.03	60.30	NT	80.27	94.15
	100.00	77.27	73.56	62.49	NT	82.46	95.19
ARU 01	100.00	91.61	88.76	86.06	74.09	90.49	98.64
	100.00	92.46	91.22	87.75	74.26	84.49	97.71
	100.00	92.18	87.32	85.82	70.55	80.05	99.07

NT= Not Tested

Appendix 18: Full statistical analysis of the HL-60 cell line cytotoxicity assay, using Fisher's least significant differences test

Appendix Table 8: Full statistical analysis of the HL-60 cell line cytotoxicity assay, using Fisher's least significant differences test.

	Treatment groups		Significance
JCM 3718	Negative control	125 µg/ml	0.000
		Positive control	0.000
		Flow through	0.261
	125 µg/ml	150 µg/ml	0.029
	150 µg/ml	175 µg/ml	0.016
	175 µg/ml	250 µg/ml	0.000
	Positive control	Flow through	0.000
JCM 3724	Negative control	125 µg/ml	0.000
		Positive control	0.000
		Flow through	0.097
	125 µg/ml	150 µg/ml	0.011
	150 µg/ml	175 µg/ml	0.001
	Positive control	Flow through	0.002
ARU 01	Negative control	125 µg/ml	0.001
		Positive control	0.000
		Flow through	0.432
	125 µg/ml	150 µg/ml	0.137
	150 µg/ml	175 µg/ml	0.197
	175 µg/ml	250 µg/ml	0.000
	Positive control	Flow through	0.000

Appendix 19: Full results of the cytotoxicity assay with human donor white blood cells

Appendix Table 9: Full results of the cytotoxicity assay with human donor white blood cells. All viabilities are expressed as a percentage of the negative control.

Negative Control	JCM 3718 HL-60	JCM 3718 WBCs	JCM 3724 HL-60	JCM 3724 WBCs	ARU 01 HL-60	ARU 01 WBCs	Positive Control
100.00	81.77	85.87	66.12	73.40	88.76	91.22	91.77
100.00	84.96	89.83	74.03	77.46	91.22	92.21	77.82
100.00	84.83	92.84	73.56	74.56	87.32	92.26	80.33

Appendix 20: Statistical analysis of the cytotoxicity assay with human donor white blood cells, using Fisher's least significant differences test

Appendix Table 10: Statistical analysis of the cytotoxicity assay with human donor white blood cells, using Fisher's least significant differences test.

Treatment groups		Significance
JCM 3718 HL-60	JCM 3718 WBCs	0.067
JCM 3724 HL-60	JCM 3724 WBCs	0.193
ARU 01 HL-60	ARU 01 WBCs	0.345

Appendix 21. Clustal Omega alignment of ecotin DNA sequences

```
Eco_42    ATGAAAAAATGTATTTATGCTATCGGCTTACTGTTTTCTTTTTCCGTCAGTGTTTTTGCA    60
Eco_11    ATGAAAAAATGTATTTATGCTATCGGCTTACTGTTTTCTTTTTCCGTCAGTGTTTTTGCA    60
Eco_41    ATGAAAAAATGTATTTATGCTATCGGCTTACTGTTTTCTTTTTCCGTCAGTGTTTTTGCA    60
Eco_52    ATGAAAAAATGTATTTATGCTATCGGCTTACTGTTTTCTTTTTCCGTCAGTGTTTTTGCA    60
JCM3718   ATGAAAAAATGTATTTATGCTATCGGCTTACTGTTTTCTTTTTCCGTCAGTGTTTTTGCA    60
Eco_01    ATGAAAAAATGTATTTATGCTATCGGCTTACTGTTTTCTTTTTCCGTCAGTGTTTTTGCA    60
Eco_05    ATGAAAAAATGTATTTATGCTATCGGCTTACTGTTTTCTTTTTCCGTCAGTGTTTTTGCA    60
Eco_21    ATGAAAAAATGTATTTATGCTATCGGCTTACTGTTTTCTTTTTCCGTCAGTGTTTTTGCA    60
Eco_24    ATGAAAAAATGTATTTATGCTATCGGCTTACTGTTTTCTTTTTCCGTCAGTGTTTTTGCA    60
Eco_30    ATGAAAAAATGTATTTATGCTATCGGCTTACTGTTTTCTTTTTCCGTCAGTGTTTTTGCA    60
Eco_39    ATGAAAAAATGTATTTATGCTATCGGCTTACTGTTTTCTTTTTCCGTCAGTGTTTTTGCA    60
Eco_40    ATGAAAAAATGTATTTATGCTATCGGCTTACTGTTTTCTTTTTCCGTCAGTGTTTTTGCA    60
JCM3724   ATGAAAAAATGTATTTATGCTATCGGCTTACTGTTTTCTTTTTCCGTCAGTGTTTTTGCA    60
ARU01     ATGAAAAAATGTATTTATGCTATCGGCTTACTGTTTTCTTTTTCCGTCAGTGTTTTTGCA    60
Eco_59    ATGAAAAAATGTATTTATGCTATCGGCTTACTGTTTTCTTTTTCCGTCAGTGTTTTTGCA    60
          ************************************************************

Eco_42    ATGCAGCATCCGGATATGAACTTGGAAATATATCCTAAGGCAAAACAAGGCATGAAGANG    120
Eco_11    ATGCAGCATCCGGATATGAACTTGGAAATATATCCTAAGGCAAAACAAGGCATGAAGAAG    120
Eco_41    ATGCAGCATCCGGATATGAACTTGGAAATATATCCTAAGGCAAAACAAGGCATGAAGAAG    120
Eco_52    ATGCAGCATCCGGATATGAACTTGGAAATATATCCTAAGGCAAAACAAGGCATGAAGAAG    120
JCM3718   ATGCAGCATCCGGATATGAACTTGGAAATATATCCTAAGGCAAAACAAGGCATGAAGAAG    120
Eco_01    ATGCAGCATCCGGATATGAACTTGGAAATATATCCTAAGGCAAAACAAGGCATGAAGAAG    120
Eco_05    ATGCAGCATCCGGATATGAACTTGGAAATATATCCTAAGGCAAAACAAGGCATGAAGAAG    120
Eco_21    ATGCAGCATCCGGATATGAACTTGGAAATATATCCTAAGGCAAAACAAGGCATGAAGAAG    120
Eco_24    ATGCAGCATCCGGATATGAACTTGGAAATATATCCTAAGGCAAAACAAGGCATGAAGAAG    120
Eco_30    ATGCAGCATCCGGATATGAACTTGGAAATATATCCTAAGGCAAAACAAGGCATGAAGAAG    120
Eco_39    ATGCAGCATCCGGATATGAACTTGGAAATATATCCTAAGGCAAAACAAGGCATGAAGAAG    120
Eco_40    ATGCAGCATCCGGATATGAACTTGGAAATATATCCTAAGGCAAAACAAGGCATGAAGAAG    120
JCM3724   ATGCAGCATCCGGATATGAACTTGGAAATATATCCTAAGGCAAAACAAGGCATGAAGAAG    120
ARU01     ATGCAGCATCCGGATATGAACTTGGAAATATATCCTAAGGCAAAACAAGGCATGAAGAAG    120
Eco_59    ATGCAGCATCCGGATATGAACTTGGAAATATATCCTAAGGCAAAACAAGGCATGAAGAAG    120
          ********************************************************** *

Eco_42    GTTGTGTATCTTTTAGAGAAAAAAGAAAAAGAAGAAGACTATAAATTGGAAATAAAATTT    180
Eco_11    GTTGTGTATCTTTTAGAGAAAAAAGAAAAAGAAGAAGACTATAAATTGGAAATAAAATTT    180
Eco_41    GTTGTGTATCTTTTAGAGAAAAAAGAAAAAGAAGAAGACTATAAATTGGAAATAAAATTT    180
Eco_52    GTTGTGTATCTTTTAGAGAAAAAAGAAAAAGAAGAAGACTATAAATTGGAAATAAAATTT    180
JCM3718   GTTGTGTATCTTTTAGAGAAAAAAGAAAAAGAAGAAGACTATAAATTGGAAATAAAATTT    180
Eco_01    GTTGTGTATCTTTTAGAGAAAAAAGAAAAAGAAGAAGACTATAAATTGGAAATAAAATTT    180
Eco_05    GTTGTGTATCTTTTAGAGAAAAAAGAAAAAGAAGAAGACTATAAATTGGAAATAAAATTT    180
Eco_21    GTTGTGTATCTTTTAGAGAAAAAAGAAAAAGAAGAAGACTATAAATTGGAAATAAAATTT    180
Eco_24    GTTGTGTATCTTTTAGAGAAAAAAGAAAAAGAAGAAGACTATAAATTGGAAATAAAATTT    180
Eco_30    GTTGTGTATCTTTTAGAGAAAAAAGAAAAAGAAGAAGACTATAAATTGGAAATAAAATTT    180
Eco_39    GTTGTGTATCTTTTAGAGAAAAAAGAAAAAGAAGAAGACTATAAATTGGAAATAAAATTT    180
Eco_40    GTTGTGTATCTTTTAGAGAAAAAAGAAAAAGAAGAAGACTATAAATTGGAAATAAAATTT    180
JCM3724   GTTGTGTATCTTTTAGAGAAAAAAGAAAAAGAAGAAGACTATAAATTGGAAATAAAATTT    180
ARU01     GTTGTGTATCTTTTAGAGAAAAAAGAAAAAGAAGAAGACTATAAATTGGAAATAAAATTT    180
Eco_59    GTTGTGTATCTTTTAGAGAAAAAAGAAAAAGAAGAAGACTATAAATTGGAAATAAAATTT    180
          ************************************************************

Eco_42    GGAAAAGATCTTGTTGTAGATGATAATCTTCNTCACTTTTTAGGAGGAAAGCTGGAAGAG    240
Eco_11    GGAAAAGACCTTGTTGTAGATGATAATCTTCATCACTTTTTAGGAGGAAAGCTGGAAGAG    240
Eco_41    GGAAAAGACCTTGTTGTAGATGATAATCTTCATCACTTTTTAGGAGGAAAGCTGGAAGAG    240
Eco_52    GGAAAAGACCTTGTTGTAGATGATAATCTTCATCACTTTTTAGGAGGAAAGCTGGAAGAG    240
JCM3718   GGAAAAGACCTTGTTGTAGATGATAATCTTCATCACTTTTTAGGAGGAAAGCTGGAAGAG    240
Eco_01    GGAAAAGATCTTGTTGTAGATGATAATCTTCATCACTTTTTAGGAGGAAAGCTGGAAGAG    240
Eco_05    GGAAAAGATCTTGTTGTAGATGATAATCTTCATCACTTTTTAGGAGGAAAGCTGGAAGAG    240
Eco_21    GGAAAAGATCTTGTTGTAGATGATAATCTTCATCACTTTTTAGGAGGAAAGCTGGAAGAG    240
Eco_24    GGAAAAGATCTTGTTGTAGATGATAATCTTCATCACTTTTTAGGAGGAAAGCTGGAAGAG    240
Eco_30    GGAAAAGATCTTGTTGTAGATGATAATCTTCATCACTTTTTAGGAGGAAAGCTGGAAGAG    240
Eco_39    GGAAAAGATCTTGTTGTAGATGATAATCTTCATCACTTTTTAGGAGGAAAGCTGGAAGAG    240
Eco_40    GGAAAAGATCTTGTTGTAGATGATAATCTTCATCACTTTTTAGGAGGAAAGCTGGAAGAG    240
JCM3724   GGAAAAGATCTTGTTGTAGATGATAATCTTCATCACTTTTTAGGAGGAAAGCTGGAAGAG    240
ARU01     GGAAAAGATCTTGTTGTAGATGATAATCTTCATCACTTTTTAGGAGGAAAGCTGGAAGAG    240
Eco_59    GGAAAAGATCTTGTTGTAGATGATAATCTTCATCACTTTTTAGGAGGAAAGCTGGAAGAG    240
          ********* ******************** ****************************
```

```
Eco_42    AAAGATGTAGAAGGTTGGGGCTATCCTTATTACATTTTTTCAGGAGATTCTCAAATGGCA    300
Eco_11    AAAGATGTAAAAGGTTGGGGCTATCCTTATTACATTTTTTCAGGAGATTCTCAAATGGCA    300
Eco_41    AAAGATGTAAAAGGTTGGGGCTATCCTTATTACATTTTTTCAGGAGATTCTCAAATGGCA    300
Eco_52    AAAGATGTAAAAGGTTGGGGCTATCCTTATTACATTTTTTCAGGAGATTCTCAAATGGCA    300
JCM3718   AAAGATGTAAAAGGTTGGGGCTATCCTTATTACATTTTTTCAGGAGATTCTCAAATGGCA    300
Eco_01    AAAGATGTAGAAGGTTGGGGCTATCCTTATTACATTTTTTCAGGAGATTCTCAAATGGCA    300
Eco_05    AAAGATGTAGAAGGTTGGGGCTATCCTTATTACATTTTTTCAGGAGATTCTCAAATGGCA    300
Eco_21    AAAGATGTAGAAGGTTGGGGCTATCCTTATTACATTTTTTCAGGAGATTCTCAAATGGCA    300
Eco_24    AAAGATGTAGAAGGTTGGGGCTATCCTTATTACATTTTTTCAGGAGATTCTCAAATGGCA    300
Eco_30    AAAGATGTAGAAGGTTGGGGCTATCCTTATTACATTTTTTCAGGAGATTCTCAAATGGCA    300
Eco_39    AAAGATGTAGAAGGTTGGGGCTATCCTTATTACATTTTTTCAGGAGATTCTCAAATGGCA    300
Eco_40    AAAGATGTAGAAGGTTGGGGCTATCCTTATTACATTTTTTCAGGAGATTCTCAAATGGCA    300
JCM3724   AAAGATGTAGAAGGTTGGGGCTATCCTTATTACATTTTTTCAGGAGATTCTCAAATGGCA    300
ARU01     AAAGATGTAGAAGGTTGGGGCTATCCTTATTACATTTTTTCAGGAGATTCTCAAATGGCA    300
Eco_59    AAAGATGTAGAAGGTTGGGGCTATCCTTATTACATTTTTTCAGGAGATTCTCAAATGGCA    300
          ********* ***************************************************

Eco_42    CAAACTTTAANGGCGTTTCCTTTAGGAAGTGAACGAGAAAAAGAGTATATTATCCCACA    360
Eco_11    CAAACTTTAATGGCGTTTCCTTTAGGAAGTGAACGAGAAAAAGAGTATATTATCCCACA    360
Eco_41    CAAACTTTAATGGCGTTTCCTTTAGGAAGTGAACGAGAAAAAGAGTATATTATCCCACA    360
Eco_52    CAAACTTTAATGGCGTTTCCTTTAGGAAGTGAACGAGAAAAAGAGTATATTATCCCACA    360
JCM3718   CAAACTTTAATGGCGTTTCCTTTAGGAAGTGAACGAGAAAAAGAGTATATTATCCCACA    360
Eco_01    CAAACTTTAATGGCGTTTCCTTTAGGAAGTGAACGAGAAAAAGAGTATATTATCCCACA    360
Eco_05    CAAACTTTAATGGCGTTTCCTTTAGGAAGTGAACGAGAAAAAGAGTATATTATCCCACA    360
Eco_21    CAAACTTTAATGGCGTTTCCTTTAGGAAGTGAACGAGAAAAAGAGTATATTATCCCACA    360
Eco_24    CAAACTTTAATGGCGTTTCCTTTAGGAAGTGAACGAGAAAAAGAGTATATTATCCCACA    360
Eco_30    CAAACTTTAATGGCGTTTCCTTTAGGAAGTGAACGAGAAAAAGAGTATATTATCCCACA    360
Eco_39    CAAACTTTAATGGCGTTTCCTTTAGGAAGTGAACGAGAAAAAGAGTATATTATCCCACA    360
Eco_40    CAAACTTTAATGGCGTTTCCTTTAGGAAGTGAACGAGAAAAAGAGTATATTATCCCACA    360
JCM3724   CAAACTTTAATGGCGTTTCCTTTAGGAAGTGAACGAGAAAAAGAGTATATTATCCCACA    360
ARU01     CAAACTTTAATGGCGTTTCCTTTAGGAAGTGAACGAGAAAAAGAGTATATTATCCCACA    360
Eco_59    CAAACTTTAATGGCGTTTCCTTTAGGAAGTGAACGAGAAAAAGAGTATATTATCCCACA    360
          ********* **************************************************

Eco_42    GCTACGAAAATATTGCCTTATCATTCAAAACTTCCTTTGGTTTTATATGTTCCGGAAGAT    420
Eco_11    GCTACGAAAATATTGCCTTATCATTCAAAACTTCCTTTGGTTTTATATGTTCCGGAAGAT    420
Eco_41    GCTACGAAAATATTGCCTTATCATTCAAAACTTCCTTTGGTTTTATATGTTCCGGAAGAT    420
Eco_52    GCTACGAAAATATTGCCTTATCATTCAAAACTTCCTTTGGTTTTATATGTTCCGGAAGAT    420
JCM3718   GCTACGAAAATATTGCCTTATCATTCAAAACTTCCTTTGGTTTTATATGTTCCGGAAGAT    420
Eco_01    GCTACGAAAATATTGCCTTATCATTCAAAACTTCCTTTGGTTTTATATGTTCCGGAAGAT    420
Eco_05    GCTACGAAAATATTGCCTTATCATTCAAAACTTCCTTTGGTTTTATATGTTCCGGAAGAT    420
Eco_21    GCTACGAAAATATTGCCTTATCATTCAAAACTTCCTTTGGTTTTATATGTTCCGGAAGAT    420
Eco_24    GCTACGAAAATATTGCCTTATCATTCAAAACTTCCTTTGGTTTTATATGTTCCGGAAGAT    420
Eco_30    GCTACGAAAATATTGCCTTATCATTCAAAACTTCCTTTGGTTTTATATGTTCCGGAAGAT    420
Eco_39    GCTACGAAAATATTGCCTTATCATTCAAAACTTCCTTTGGTTTTATATGTTCCGGAAGAT    420
Eco_40    GCTACGAAAATATTGCCTTATCATTCAAAACTTCCTTTGGTTTTATATGTTCCGGAAGAT    420
JCM3724   GCTACGAAAATATTGCCTTATCATTCAAAACTTCCTTTGGTTTTATATGTTCCGGAAGAT    420
ARU01     GCTACGAAAATATTGCCTTATCATTCAAAACTTCCTTTGGTTTTATATGTTCCGGAAGAT    420
Eco_59    GCTACGAAAATATTGCCTTATCATTCAAAACTTCCTTTGGTTTTATATGTTCCGGAAGAT    420
          ************************************************************

Eco_42    GTGAAGGTAGAAGTATCTCTTTGGAATCGAATGCAGGAAATCAAAGAAGTTTCTCGTTAA    480
Eco_11    GTGAAGGTAGAAGTATCTCTTTGGAATCGAATGCAGGAAATCAAAGAAGTTTCTCGTTAA    480
Eco_41    GTGAAGGTAGAAGTATCTCTTTGGAATCGAATGCAGGAAATCAAAGAAGTTTCTCGTTAA    480
Eco_52    GTGAAGGTAGAAGTATCTCTTTGGAATCGAATGCAGGAAATCAAAGAAGTTTCTCGTTAA    480
JCM3718   GTGAAGGTAGAAGTATCTCTTTGGAATCGAATGCAGGAAATCAAAGAAGTTTCTCGTTAA    480
Eco_01    GTGAAGGTAGAAGTATCTCTTTGGAATCGAATGCAGGAAATCAAAGAAGTTTCTCGTTAA    480
Eco_05    GTGAAGGTAGAAGTATCTCTTTGGAATCGAATGCAGGAAATCAAAGAAGTTTCTCGTTAA    480
Eco_21    GTGAAGGTAGAAGTATCTCTTTGGAATCGAATGCAGGAAATCAAAGAAGTTTCTCGTTAA    480
Eco_24    GTGAAGGTAGAAGTATCTCTTTGGAATCGAATGCAGGAAATCAAAGAAGTTTCTCGTTAA    480
Eco_30    GTGAAGGTAGAAGTATCTCTTTGGAATCGAATGCAGGAAATCAAAGAAGTTTCTCGTTAA    480
Eco_39    GTGAAGGTAGAAGTATCTCTTTGGAATCGAATGCAGGAAATCAAAGAAGTTTCTCGTTAA    480
Eco_40    GTGAAGGTAGAAGTATCTCTTTGGAATCGAATGCAGGAAATCAAAGAAGTTTCTCGTTAA    480
JCM3724   GTGAAGGTAGAAGTATCTCTTTGGAATCGAATGCAGGAAATCAAAGAAGTTTCTCGTTAA    480
ARU01     GTGAAGGTAGAAGTATCTCTTTGGAATCGAATGCAGGAAATCAAAGAAGTTTCTCGTTAA    480
Eco_59    GTGAAGGTAGAAGTATCTCTTTGGAATCGAATGTAGGAAATCAAAGAAGTTTCTCGTTAA    480
          ********************************* ***************************
```

Appendix Figure 7: Clustal Omega alignment of ecotin DNA sequences of the first 12 strains from the clinical collection,

and strains JCM 3718, JCM 3724 and ARU 01.

Appendix 22. Plasmid map and sequence of pET-16b

Appendix Figure 8: Plasmid map of pET-16b vector.
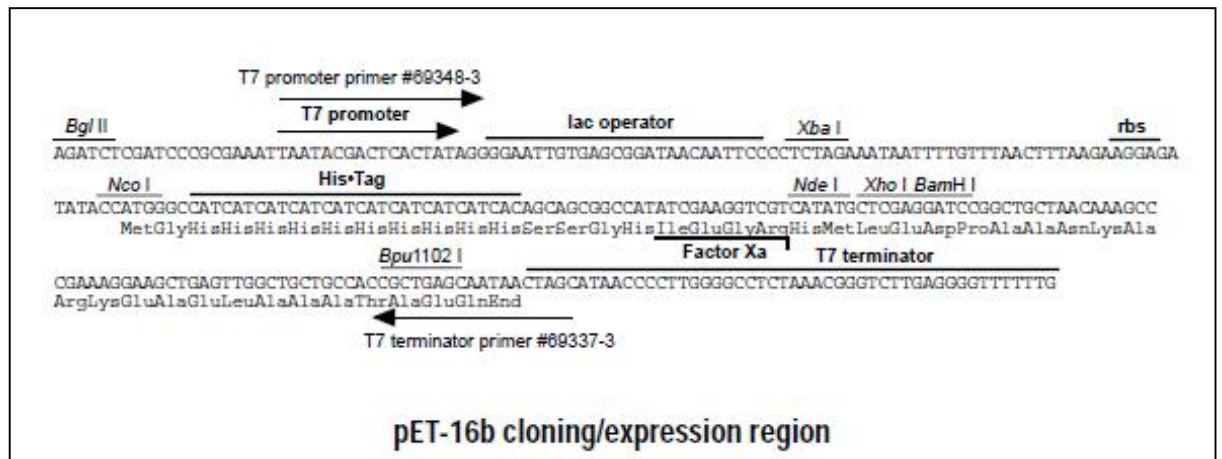
Appendix Figure 9: DNA sequence of the cloning/expression region of vector pET-16b.

Appendix 23. Plasmid sequence of pET-19b
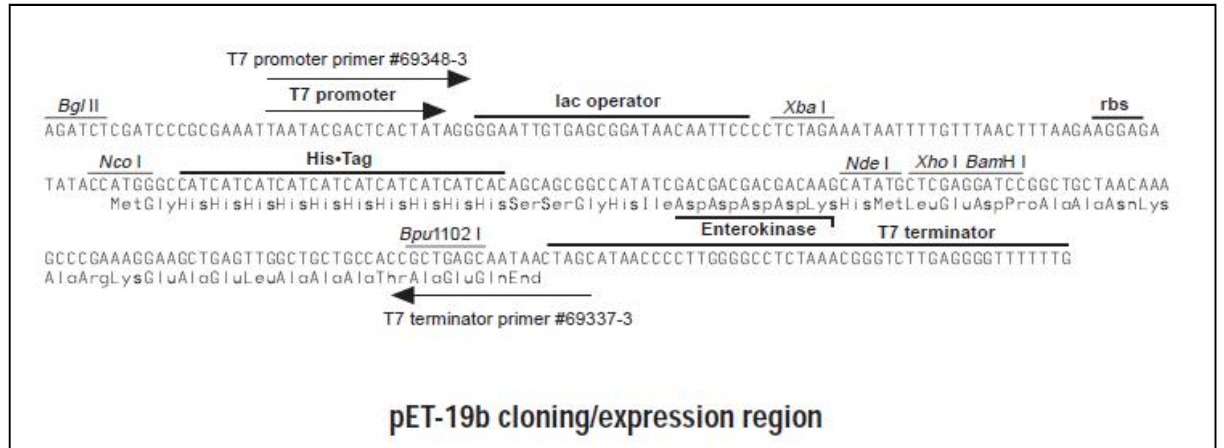
Appendix Figure 10: DNA sequence of the cloning/expression region of vector pET-19b, including an enterokinase cleavage

site in place of the Factor Xa cleavage site present in pET-16b.

AFC calibration curve
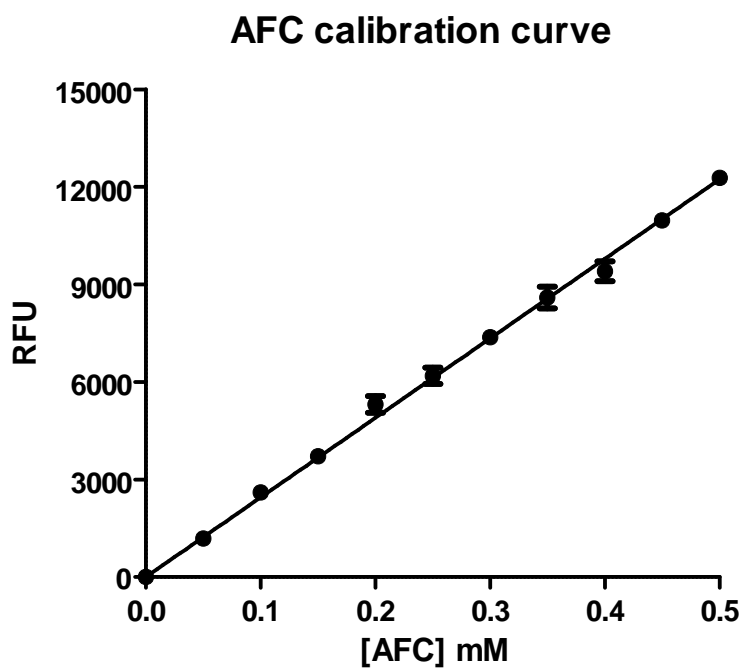
Appendix Figure 11: Calibration curve for 7-Amino-4-(trifluoromethyl)coumarin (AFC). Ten concentrations were tested, covering the full range of fluorescence generated from the substrate standard curves. The AFC calibration curve has an equation of y=244626x and an R squared value of 0.9978.

Appendix 25. Human plasma kallikrein (HPK) RFU readings

Appendix Table 11: Mean relative fluorescent units for HPK standards, with no ecotin present.

Time (minutes)	Substrate concentration					
	0.5 mM	0.25 mM	0.125 mM	0.06 mM	0.03 mM	0.015 mM
0	0.00	0.00	0.00	0.00	0.00	0.00
1	502.23	387.95	187.82	124.54	53.08	28.88
2	944.47	758.78	400.45	239.76	116.66	63.83
3	1402.98	1152.92	615.32	369.83	183.12	100.69
4	1857.98	1559.57	861.20	511.80	251.62	135.66
5	2303.78	1949.89	1136.03	649.48	317.51	169.61
6	2738.23	2319.86	1399.72	782.38	382.08	203.93
7	3159.66	2671.34	1634.09	908.09	443.43	237.48
8	3565.63	3013.27	1853.90	1029.18	502.89	270.04
9	3961.50	3343.21	2064.39	1144.81	559.98	301.45
10	4344.97	3661.04	2267.55	1257.16	615.23	332.13
11	4713.54	3968.56	2462.09	1364.66	668.57	361.83
12	5062.98	4263.88	2650.69	1468.97	720.07	390.53
13	5396.19	4547.22	2830.15	1569.18	769.01	418.32
14	5711.41	4820.72	3005.64	1665.34	816.78	445.27
15	6012.46	5084.33	3173.68	1758.52	862.95	471.71
16	6303.00	5339.29	3336.28	1847.57	907.04	496.82
17	6584.13	5581.94	3492.45	1934.72	949.70	521.39
18	6859.10	5820.54	3643.40	2017.69	990.75	545.04
19	7125.29	6048.48	3790.46	2098.67	1030.10	567.83
20	7388.89	6266.99	3931.15	2175.66	1068.32	590.09
21	7664.26	6482.32	4068.33	2250.93	1104.45	611.45
22	7970.73	6686.01	4199.11	2322.75	1140.45	632.27
23	8320.06	6885.34	4327.26	2392.58	1174.28	652.38
24	8694.00	7075.54	4449.47	2459.20	1206.63	671.70
25	9052.66	7264.36	4569.30	2523.85	1238.86	690.39
26	9423.72	7444.28	4683.51	2586.06	1268.61	708.58
27	9772.61	7615.70	4794.83	2646.39	1297.71	726.28
28	10120.45	7783.29	4901.07	2702.78	1325.62	743.17
29	10449.32	7945.26	5004.14	2758.69	1352.98	759.48
30	10729.96	8102.58	5104.08	2813.27	1378.90	775.74

Time	Substrate concentration					
(minutes)	0.5 mM	0.25 mM	0.125 mM	0.06 mM	0.03 mM	0.015 mM
0	0.00	0.00	0.00	0.00	0.00	0.00
1	331.83	265.46	192.19	71.63	35.26	17.87
2	622.00	523.49	384.72	146.64	73.15	35.91
3	968.13	799.20	591.10	228.48	113.75	57.10
4	1324.57	1078.84	797.32	310.38	154.29	81.86
5	1681.06	1355.95	997.55	390.52	194.53	105.86
6	2037.81	1629.45	1184.39	470.93	235.95	128.27
7	2384.17	1895.97	1361.22	556.42	280.42	149.58
8	2727.80	2158.91	1535.18	652.31	329.82	170.36
9	3064.03	2409.28	1706.46	758.25	384.62	190.86
10	3403.49	2652.79	1875.42	865.02	436.59	211.14
11	3735.09	2894.18	2040.24	962.95	483.37	231.10
12	4071.56	3132.03	2193.77	1051.75	527.47	250.77
13	4404.29	3372.88	2342.74	1135.76	569.78	269.79
14	4740.37	3624.79	2490.91	1217.90	611.27	288.55
15	5079.19	3886.08	2655.59	1297.90	651.63	306.85
16	5416.13	4145.69	2856.14	1374.52	691.25	324.87
17	5741.61	4405.22	3072.92	1448.87	729.58	342.40
18	6060.54	4660.25	3258.29	1520.83	767.06	359.62
19	6369.85	4928.92	3413.98	1591.04	803.51	376.29
20	6667.65	5200.94	3556.98	1659.41	838.81	392.63
21	6956.42	5438.80	3694.17	1726.09	873.29	408.52
22	7227.91	5650.54	3827.47	1789.88	906.48	424.22
23	7489.04	5851.10	3954.29	1852.42	938.86	439.40
24	7739.52	6047.85	4078.88	1913.06	970.74	454.37
25	7994.61	6236.61	4200.27	1971.42	1001.40	468.89
26	8269.29	6418.94	4316.19	2028.71	1030.85	483.20
27	8574.05	6600.43	4427.28	2084.30	1059.86	497.09
28	8904.51	6775.53	4536.72	2138.10	1088.17	510.60
29	9228.53	6944.67	4639.93	2190.65	1115.22	523.64
30	9623.73	7109.77	4742.09	2241.79	1141.75	536.82

Appendix Table 13: Mean relative fluorescent units for HPK, with an ecotin concentration of 25 nM.

Time (minutes)	Substrate concentration					
	0.5 mM	0.25 mM	0.125 mM	0.06 mM	0.03 mM	0.015 mM
0	0.00	0.00	0.00	0.00	0.00	0.00
1	247.78	210.31	105.65	48.58	23.77	14.33
2	545.81	440.03	227.82	111.80	53.04	30.09
3	882.66	696.43	371.43	180.44	87.34	47.33
4	1220.36	957.55	520.62	249.34	121.45	66.35
5	1547.42	1213.92	668.07	318.27	156.52	84.85
6	1870.10	1465.82	810.34	388.60	195.64	103.12
7	2194.28	1711.14	946.16	464.30	236.09	120.93
8	2520.36	1947.73	1079.44	547.87	278.60	138.60
9	2849.26	2178.70	1210.68	637.96	320.08	155.84
10	3178.82	2403.95	1340.97	728.63	360.07	173.05
11	3502.90	2625.16	1473.19	813.21	398.87	189.88
12	3820.65	2842.65	1601.64	891.34	436.97	206.55
13	4125.77	3066.18	1726.52	966.27	474.40	222.74
14	4416.60	3309.53	1845.80	1039.60	511.09	238.69
15	4691.26	3575.51	1952.26	1111.18	547.06	254.58
16	4956.32	3837.39	2068.64	1180.91	582.35	269.99
17	5216.88	4081.89	2217.86	1249.38	616.86	285.11
18	5476.87	4303.45	2412.92	1315.61	650.47	299.90
19	5733.93	4531.40	2604.15	1380.04	683.55	314.52
20	5992.10	4775.17	2755.68	1443.07	715.91	328.86
21	6245.98	5025.01	2886.14	1504.89	747.28	342.85
22	6495.13	5248.30	3009.67	1564.61	777.89	356.51
23	6719.52	5453.04	3128.95	1623.09	808.19	369.91
24	6980.46	5647.13	3246.33	1680.15	837.33	383.11
25	7288.28	5826.39	3362.78	1735.16	866.39	396.13
26	7687.97	6007.56	3474.29	1789.62	894.18	408.85
27	8127.55	6186.91	3583.64	1842.26	921.19	421.20
28	8511.51	6358.33	3689.45	1893.45	947.90	433.59
29	8817.79	6527.27	3792.43	1944.27	973.91	445.41
30	9083.76	6689.05	3893.47	1993.14	999.16	457.12

Appendix Table 14: Mean relative fluorescent units for HPK, with an ecotin concentration of 50 nM.

| Time (minutes)	Substrate concentration					
	0.5 mM	0.25 mM	0.125 mM	0.06 mM	0.03 mM	0.015 mM
0	0.00	0.00	0.00	0.00	0.00	0.00
1	202.43	154.67	79.30	36.82	17.18	9.64
2	463.75	348.54	183.29	86.32	35.83	20.60
3	760.39	568.23	305.41	146.57	58.83	32.62
4	1074.00	790.59	434.21	210.32	85.63	45.35
5	1397.19	1008.16	562.80	276.55	116.73	60.37
6	1715.18	1222.78	687.34	349.35	150.04	77.21
7	2026.16	1434.39	807.67	423.34	183.43	94.12
8	2333.56	1638.34	925.47	493.59	215.93	109.47
9	2642.48	1842.56	1042.20	560.26	247.62	123.77
10	2950.82	2044.10	1157.87	625.53	278.47	137.98
11	3254.70	2244.31	1271.79	689.46	309.02	151.86
12	3547.80	2443.81	1379.69	752.90	339.07	165.48
13	3828.88	2640.12	1479.47	814.62	368.43	178.93
14	4097.39	2840.22	1579.40	875.56	397.42	192.24
15	4352.25	3055.42	1696.40	934.95	425.63	205.31
16	4607.17	3278.92	1848.94	993.60	453.55	218.43
17	4865.00	3501.04	2008.86	1050.86	480.94	231.12
18	5125.84	3711.56	2142.51	1107.10	507.78	243.56
19	5390.52	3924.93	2260.82	1161.99	534.49	256.09
20	5645.85	4152.75	2373.65	1215.71	560.61	268.23
21	5902.94	4372.95	2484.00	1268.38	586.03	280.22
22	6159.61	4570.75	2592.72	1320.10	611.08	292.10
23	6404.48	4756.17	2698.37	1370.62	635.75	303.66
24	6674.53	4936.11	2802.78	1420.01	659.94	315.20
25	6986.83	5112.00	2904.60	1468.64	683.69	326.56
26	7341.09	5282.84	3004.18	1515.73	707.12	337.53
27	7702.85	5451.83	3101.15	1562.32	729.94	348.48
28	8036.73	5613.82	3197.01	1607.81	752.38	359.22
29	8322.44	5774.86	3289.94	1652.13	774.37	369.80
30	8581.79	5932.68	3380.86	1696.00	796.22	380.22

Appendix Table 15: Mean relative fluorescent units for HPK, with an ecotin concentration of 100 nM.

Time (minutes)	Substrate concentration					
	0.5 mM	0.25 mM	0.125 mM	0.06 mM	0.03 mM	0.015 mM
0	0.00	0.00	0.00	0.00	0.00	0.00
1	167.02	119.55	57.49	22.69	11.68	6.31
2	417.13	282.96	145.48	58.20	27.81	15.41
3	688.72	463.79	249.90	95.71	46.03	25.52
4	963.02	649.11	356.08	134.09	64.62	35.43
5	1242.05	834.77	460.35	170.57	83.85	45.18
6	1520.16	1018.79	561.46	207.16	102.74	54.83
7	1791.33	1199.73	661.19	248.47	121.51	64.96
8	2056.43	1378.24	760.45	297.65	140.45	75.45
9	2321.45	1555.13	858.76	356.14	159.94	86.74
10	2588.51	1732.40	953.58	418.31	179.56	98.19
11	2851.47	1908.75	1047.21	476.89	201.65	109.79
12	3113.07	2083.12	1142.79	529.61	225.31	120.42
13	3371.30	2258.64	1239.89	576.43	250.67	130.41
14	3623.59	2432.75	1337.31	621.77	276.20	140.22
15	3874.34	2611.00	1433.08	666.21	299.33	150.00
16	4122.95	2789.16	1527.63	709.97	320.15	159.67
17	4372.27	2969.23	1619.98	753.81	340.19	169.30
18	4618.95	3155.34	1711.88	796.48	359.71	178.70
19	4867.30	3341.62	1801.22	838.36	379.00	188.14
20	5119.24	3515.02	1889.27	880.24	397.89	197.45
21	5370.87	3681.81	1976.50	921.06	416.61	206.70
22	5623.93	3844.16	2061.75	961.45	435.14	215.75
23	5873.34	4003.88	2144.98	1000.95	453.48	224.79
24	6119.88	4160.65	2227.53	1039.80	471.54	233.65
25	6381.72	4311.81	2309.21	1078.29	489.84	242.52
26	6658.18	4460.37	2388.24	1116.21	507.25	251.17
27	6956.29	4607.23	2465.93	1153.46	524.64	259.81
28	7244.23	4749.82	2542.92	1190.19	541.88	268.20
29	7497.64	4889.30	2618.25	1226.57	558.90	276.42
30	7733.40	5028.88	2692.92	1262.11	575.64	284.61

Appendix 26. Lineweaver-Burk plot for kallikrein data
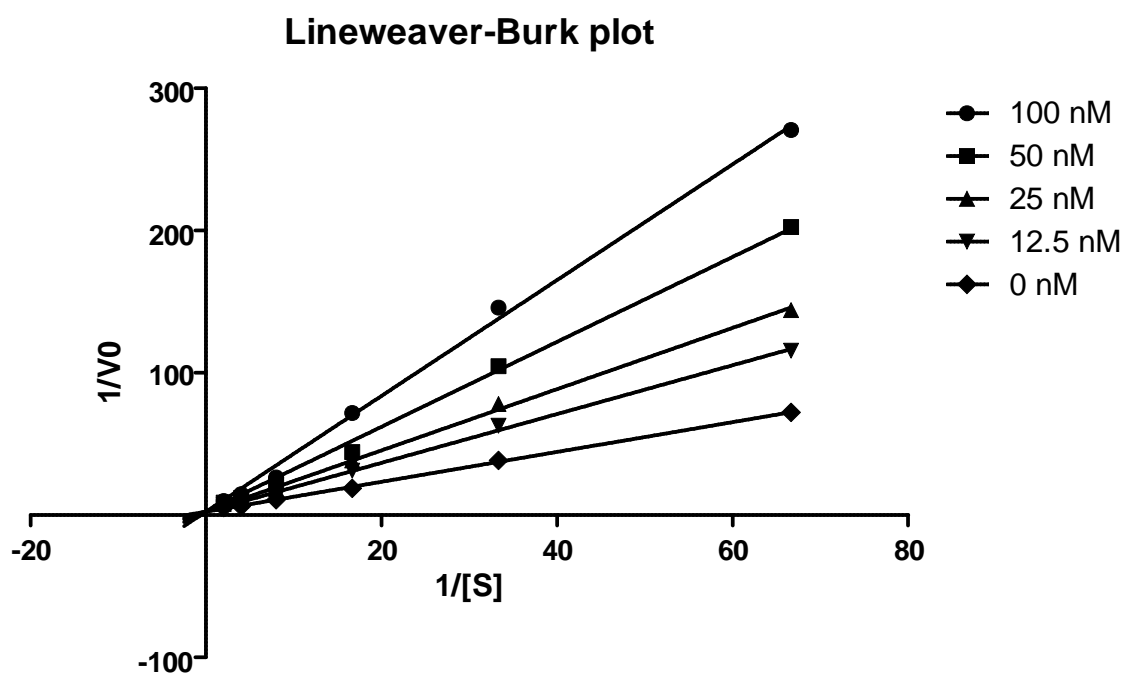
Appendix Figure 12: A Lineweaver-Burk plot showing characteristics of competitive inhibition. Y intercept was set at 0.466 resulting in a shared Vmax of 0.406 μmol min-1. The x intercept is equal to -1/Km and the y intercept is equal to 1/Vmax.

Appendix 27. Human neutrophil elastase (HNE) absorbance readings

Appendix Table 16: Mean absorbance units for HNE standards, with no ecotin present.

Time (minutes)	Substrate concentration					
	0.5 mM	0.25 mM	0.125 mM	0.06 mM	0.03 mM	0.015 mM
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0060	0.0047	0.0040	0.0027	0.0010	0.0010
2	0.0120	0.0090	0.0083	0.0060	0.0023	0.0017
3	0.0167	0.0130	0.0120	0.0097	0.0040	0.0023
4	0.0227	0.0170	0.0157	0.0127	0.0047	0.0033
5	0.0270	0.0207	0.0187	0.0157	0.0050	0.0043
6	0.0320	0.0250	0.0227	0.0183	0.0060	0.0043
7	0.0370	0.0287	0.0260	0.0207	0.0070	0.0057
8	0.0423	0.0323	0.0290	0.0230	0.0080	0.0060
9	0.0470	0.0360	0.0320	0.0253	0.0090	0.0070
10	0.0507	0.0397	0.0347	0.0280	0.0100	0.0077
11	0.0550	0.0427	0.0377	0.0303	0.0110	0.0080
12	0.0587	0.0457	0.0400	0.0323	0.0117	0.0087
13	0.0627	0.0490	0.0423	0.0347	0.0127	0.0093
14	0.0663	0.0527	0.0447	0.0367	0.0133	0.0097
15	0.0697	0.0563	0.0470	0.0380	0.0137	0.0107
16	0.0730	0.0577	0.0493	0.0400	0.0147	0.0107
17	0.0763	0.0607	0.0513	0.0413	0.0153	0.0117
18	0.0800	0.0643	0.0540	0.0430	0.0157	0.0117
19	0.0823	0.0660	0.0560	0.0450	0.0167	0.0123
20	0.0853	0.0683	0.0577	0.0460	0.0173	0.0127
21	0.0883	0.0707	0.0597	0.0473	0.0177	0.0127
22	0.0913	0.0727	0.0613	0.0493	0.0183	0.0133
23	0.0943	0.0757	0.0633	0.0507	0.0187	0.0137
24	0.0967	0.0773	0.0650	0.0517	0.0193	0.0140
25	0.1000	0.0800	0.0667	0.0530	0.0197	0.0140
26	0.1020	0.0823	0.0680	0.0547	0.0203	0.0147
27	0.1043	0.0847	0.0697	0.0557	0.0207	0.0150
28	0.1067	0.0863	0.0713	0.0563	0.0213	0.0150
29	0.1090	0.0883	0.0727	0.0577	0.0213	0.0153
30	0.1117	0.0900	0.0740	0.0590	0.0220	0.0160

Appendix Table 17: Mean absorbance units for HNE, with an ecotin concentration of 12.5 nM.

	Substrate concentration					
Time (minutes)	0.5 mM	0.25 mM	0.125 mM	0.06 mM	0.03 mM	0.015 mM
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0047	0.0030	0.0020	0.0017	0.0003	0.0010
2	0.0100	0.0063	0.0047	0.0030	0.0010	0.0003
3	0.0133	0.0093	0.0070	0.0047	0.0013	0.0013
4	0.0180	0.0130	0.0093	0.0057	0.0013	0.0013
5	0.0220	0.0160	0.0110	0.0067	0.0013	0.0017
6	0.0267	0.0187	0.0130	0.0083	0.0020	0.0017
7	0.0303	0.0217	0.0157	0.0093	0.0023	0.0020
8	0.0343	0.0240	0.0173	0.0107	0.0030	0.0027
9	0.0383	0.0267	0.0187	0.0120	0.0037	0.0027
10	0.0417	0.0297	0.0203	0.0130	0.0033	0.0030
11	0.0453	0.0323	0.0217	0.0137	0.0040	0.0027
12	0.0490	0.0347	0.0233	0.0147	0.0040	0.0030
13	0.0520	0.0370	0.0247	0.0160	0.0043	0.0033
14	0.0553	0.0393	0.0263	0.0180	0.0047	0.0033
15	0.0583	0.0420	0.0280	0.0193	0.0050	0.0037
16	0.0610	0.0440	0.0297	0.0183	0.0050	0.0037
17	0.0640	0.0460	0.0307	0.0190	0.0050	0.0040
18	0.0670	0.0483	0.0320	0.0197	0.0050	0.0040
19	0.0697	0.0503	0.0330	0.0203	0.0053	0.0040
20	0.0720	0.0523	0.0343	0.0210	0.0057	0.0047
21	0.0750	0.0543	0.0387	0.0217	0.0057	0.0043
22	0.0773	0.0563	0.0367	0.0220	0.0060	0.0047
23	0.0797	0.0583	0.0377	0.0230	0.0060	0.0047
24	0.0817	0.0603	0.0380	0.0237	0.0060	0.0050
25	0.0843	0.0617	0.0393	0.0240	0.0060	0.0050
26	0.0863	0.0637	0.0420	0.0250	0.0060	0.0053
27	0.0883	0.0657	0.0413	0.0253	0.0060	0.0053
28	0.0907	0.0673	0.0417	0.0270	0.0060	0.0057
29	0.0927	0.0690	0.0423	0.0263	0.0063	0.0053
30	0.0943	0.0710	0.0437	0.0267	0.0060	0.0057

Appendix Table 18: Mean absorbance units for HNE, with an ecotin concentration of 25 nM.

Time (minutes)	Substrate concentration					
	0.5 mM	0.25 mM	0.125 mM	0.06 mM	0.03 mM	0.015 mM
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0013	0.0003	0.0013	0.0007	0.0000	0.0000
2	0.0033	0.0013	0.0027	0.0010	0.0007	0.0000
3	0.0037	0.0017	0.0033	0.0013	0.0007	0.0000
4	0.0053	0.0023	0.0043	0.0020	0.0007	0.0000
5	0.0063	0.0027	0.0050	0.0020	0.0007	0.0000
6	0.0070	0.0037	0.0060	0.0020	0.0020	0.0000
7	0.0087	0.0040	0.0067	0.0020	0.0027	0.0000
8	0.0097	0.0043	0.0070	0.0023	0.0020	0.0000
9	0.0107	0.0050	0.0077	0.0023	0.0010	0.0000
10	0.0113	0.0053	0.0080	0.0023	0.0013	0.0000
11	0.0120	0.0057	0.0083	0.0023	0.0017	0.0000
12	0.0130	0.0060	0.0087	0.0027	0.0013	0.0000
13	0.0137	0.0067	0.0090	0.0027	0.0007	0.0000
14	0.0143	0.0073	0.0093	0.0023	0.0007	0.0000
15	0.0150	0.0077	0.0097	0.0023	0.0007	0.0000
16	0.0153	0.0077	0.0097	0.0027	0.0007	0.0000
17	0.0163	0.0083	0.0103	0.0027	0.0007	0.0000
18	0.0167	0.0087	0.0103	0.0027	0.0007	0.0000
19	0.0173	0.0090	0.0107	0.0030	0.0007	0.0003
20	0.0177	0.0097	0.0107	0.0030	0.0007	0.0000
21	0.0183	0.0097	0.0107	0.0030	0.0007	0.0000
22	0.0190	0.0100	0.0107	0.0030	0.0007	0.0000
23	0.0197	0.0107	0.0110	0.0033	0.0010	0.0000
24	0.0197	0.0107	0.0117	0.0033	0.0007	0.0000
25	0.0203	0.0113	0.0117	0.0030	0.0007	0.0000
26	0.0207	0.0117	0.0117	0.0033	0.0007	0.0000
27	0.0207	0.0120	0.0117	0.0033	0.0010	0.0000
28	0.0213	0.0120	0.0117	0.0033	0.0007	0.0000
29	0.0217	0.0123	0.0120	0.0033	0.0007	0.0000
30	0.0223	0.0130	0.0123	0.0033	0.0007	0.0000

Appendix Table 19: Mean absorbance units for HNE, with an ecotin concentration of 50 nM.

Time (minutes)	Substrate concentration					
	0.5 mM	0.25 mM	0.125 mM	0.06 mM	0.03 mM	0.015 mM
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.0000	0.0000	0.0000	-0.0003	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003
3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0007	0.0000	-0.0003	0.0000	0.0000	0.0000
6	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.0010	0.0003	0.0000	0.0000	0.0000	0.0000
9	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000
10	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000
11	0.0010	0.0003	0.0000	0.0000	0.0000	0.0000
12	0.0010	0.0007	0.0000	0.0003	0.0000	0.0000
13	0.0010	0.0007	0.0000	0.0000	0.0000	0.0000
14	0.0013	0.0010	0.0000	0.0003	0.0000	0.0000
15	0.0020	0.0007	0.0000	0.0003	0.0000	0.0000
16	0.0020	0.0013	0.0000	0.0003	0.0000	0.0000
17	0.0020	0.0010	0.0000	0.0003	0.0000	0.0000
18	0.0020	0.0010	0.0003	0.0003	0.0000	0.0000
19	0.0023	0.0007	0.0010	0.0003	0.0000	0.0000
20	0.0027	0.0013	0.0010	0.0003	0.0000	0.0000
21	0.0030	0.0010	0.0010	0.0003	0.0000	0.0000
22	0.0030	0.0010	0.0010	0.0003	0.0000	0.0000
23	0.0030	0.0010	0.0010	0.0003	0.0000	0.0000
24	0.0037	0.0020	0.0010	0.0010	0.0000	0.0000
25	0.0037	0.0020	0.0010	0.0010	0.0000	0.0000
26	0.0040	0.0020	0.0010	0.0010	0.0000	0.0000
27	0.0040	0.0020	0.0010	0.0010	0.0000	0.0000
28	0.0040	0.0020	0.0010	0.0010	0.0000	0.0000
29	0.0047	0.0020	0.0010	0.0010	0.0000	0.0000
30	0.0050	0.0020	0.0010	0.0010	0.0000	0.0000

Appendix Table 20: Mean absorbance units for HNE, with an ecotin concentration of 100 nM.

Time (minutes)	Substrate concentration					
	0.5 mM	0.25 mM	0.125 mM	0.06 mM	0.03 mM	0.015 mM
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.0000	0.0000	-0.0010	-0.0010	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0003
3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0003
5	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0007	0.0000	0.0000	0.0000	0.0000	-0.0003
7	0.0010	0.0000	0.0000	0.0000	0.0000	-0.0003
8	0.0010	0.0000	0.0000	0.0000	0.0000	-0.0003
9	0.0013	0.0007	0.0000	0.0000	0.0000	-0.0003
10	0.0010	0.0007	0.0000	0.0000	0.0000	-0.0003
11	0.0010	0.0007	0.0003	0.0003	0.0003	0.0000
12	0.0010	0.0007	0.0003	0.0000	0.0000	-0.0003
13	0.0017	0.0007	0.0003	0.0000	0.0000	-0.0003
14	0.0020	0.0007	0.0007	0.0000	0.0000	-0.0003
15	0.0020	0.0007	0.0000	0.0000	0.0000	0.0000
16	0.0020	0.0013	0.0007	0.0000	0.0000	-0.0003
17	0.0027	0.0010	0.0007	0.0000	0.0000	-0.0003
18	0.0030	0.0010	0.0007	0.0000	0.0000	-0.0003
19	0.0030	0.0010	0.0007	0.0003	0.0003	0.0000
20	0.0030	0.0013	0.0007	0.0000	0.0000	-0.0003
21	0.0030	0.0017	0.0007	0.0000	0.0000	-0.0003
22	0.0030	0.0017	0.0007	0.0003	0.0003	0.0000
23	0.0030	0.0017	0.0007	0.0003	0.0003	-0.0003
24	0.0030	0.0017	0.0007	0.0003	0.0003	-0.0003
25	0.0030	0.0017	0.0007	0.0000	0.0000	-0.0003
26	0.0030	0.0017	0.0007	0.0003	0.0003	-0.0003
27	0.0030	0.0017	0.0007	0.0003	0.0003	0.0000
28	0.0033	0.0020	0.0007	0.0003	0.0003	0.0000
29	0.0030	0.0017	0.0007	0.0000	0.0000	-0.0003
30	0.0033	0.0017	0.0007	0.0003	0.0003	-0.0003

Appendix 28. Thrombin time, prothrombin time and activated partial thromboplastin time results

Appendix Table 21: Fold prolongation values of thrombin time, prothrombin time and activated partial thromboplastin time when exposed to a range of ecotin concentrations.

Ecotin (µM)	Thrombin time			Prothrombin time			Activated partial thromboplastin time		
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.06	NT	NT	NT	NT	NT	NT	2.06	1.71	1.45
0.125	NT	NT	NT	NT	NT	NT	4.72	7.39	2.24
0.25	1.02	1.00	0.98	1.94	1.84	1.98	7.93	8.69	7.14
0.5	1.08	0.98	1.02	2.89	3.34	3.17	38.48	14.48	13.18
0.75	1.09	1.02	1.07	3.77	4.00	3.66	45.93	15.84	19.61
1	1.11	1.02	1.16	4.51	4.72	4.57	59.17	28.20	28.14
2	1.13	1.04	1.18	6.92	6.28	6.49	66.21	22.19	32.28