Investigations into plasmid and chromosomally encoded outer membrane proteins of *Borrelia*

Abigail Helen Tindell BSc

A thesis submitted to the University of Huddersfield in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

September 2021

Copyright declaration

Abstract

Lyme disease is the most common vector-borne infection across temperate zones of the Northern Hemisphere and is caused by the spirochete, *Borrelia burgdorferi sensu lato* (s.l). The outer membrane (OM) of *Borrelia* possesses many unique features, including an abundance of lipoproteins and few integral β-barrel membrane proteins. The predicted small β-barrel OMPs from *Borrelia burgdorferi s.s*; BB_0027, BB_0405, BB_0406 and BB_0562, are all predicted to form an 8-stranded transmembrane β-barrel. Three truncations were made to each OMP to increase the chance of crystal formation and these, along with potentially soluble targets, were cloned and recombinantly expressed in *E. coli*. Following purification, a BB_0406 truncation was soluble in DDM and showed promise for crystal formation. Factor-H binding was investigated using a far-Western blot. No binding was detected, in agreement with the most recent literature.

The *Borrelia* genome consists of a linear chromosome and numerous linear and circular plasmids. While the chromosome of *Borrelia* has been subject to various searches for β-barrel OM proteins, the plasmids have so far remained unexplored. This research utilised a computational-framework approach to identify potential OM β-barrel proteins in the plasmid proteome of *B. burgdorferi s.l*. This approach identified two plasmid-encoded proteins: the annotated porin OMS28, and a previously uncharacterized protein BBJ25. Orthologs of BBJ25 are detected in two other major spirochaete families, *Brachyspira* and *Treponema*. Within the *Borrelia* species complex, BBJ25 is found within a predicted 7-gene operon at either the 3rd or 7th position, resulting in two allelic variants. Phylogenetic analysis of the allelic variants is inconsistent with vertical descent alongside the chromosome suggesting that horizontal transfer and/or recombination has occurred after the divergence of Relapsing Fever/Lyme Disease-type *Borrelia*. While the precise role of BBJ25 remains unknown, the identification of a novel conserved outer membrane protein is of interest as a diagnostic or therapeutic target.

Table of contents

List of figures

List of tables

Abbreviations

ALBI	Affinity ligand binding immunoblot
BbCRASP	*Borrelia burgdorferi* complement regulator-acquiring surface protein
BLAST	Basic Local Alignment Search Tool
C8E4	Tetraethylene Glycol Monooctyl Ether
CDC	Centres for Disease Control and Prevention
CFHR	Complement factor H–related
CheR	Chemotaxis methyltransferase R
DDAO	n-dodecyl-N,N-dimethylamine-N-oxide
DDM	n-dodecyl-β-D-maltoside
ECM	Extracellular matrix
EM	Erythema migrans
Erp	OspE-like protein
FH	Factor-H
FHL-1	Factor-H like-1
FPLC	Fast protein liquid chromatography
HMM	Hidden Markov Model
HPLC	High performance liquid chromatography
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LBRF	Louse borne relapsing fever
LD	Lyme disease
LIC	Ligation Independent Cloning
LPS	Lipopolysaccharide
MCP	Methyl-accepting chemotaxis protein
Ni-NTA	Nickel-Nitrilotriacetic acid
OM	Outer membrane
OMP	Outer membrane protein

Osp	Outer surface protein
PCNA	Proliferating Cell Nuclear Antigen
PD	Pentapeptide dependent
PI	Pentapeptide independent
PTLDS	Post treatment Lyme disease syndrome
RF	Relapsing fever
SAH	S-adenosyl-l-homocysteine
SAM	S-adenosylmethionine
SAXS	Small angle X-ray scattering
SDS-PAGE	Sodium docecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
sl	*sensu lato*
SOB	Super optimal broth
SodA	Superoxide dismutase A
ss	*sensu stricto*
TBRF	Tick borne relapsing fever
TEV	Tobacco Etch Virus
TrxT	Thioredoxin

Acknowledgements

Firstly, I would like to thank Dr Richard Bingham for his endless patience, support and advice throughout this PhD and the years prior. I extend this thanks to Dr Chris Cooper, for both his laboratory and pastoral support. Thank you both for not letting me give up when I wanted to most.

I thank the biological sciences technical staff; Sophie Sinclair, Elena Tillotson, Maggie Craven and Felix Owusu-Kwarteng for all of their help in the laboratory. Thank you to Dr Jarek Byrk for his assistance writing the R-code for this research.

A huge thank you to the friends I made during these years for being a much needed source of joy, emotional support and words of wisdom, which I miss dearly.

Thank you to David, for your patience and support throughout all these years of education and the many more to come. Finally, a heartfelt thank you to my family, especially my Mum and Grandma, for their unconditional love and support throughout everything I do. I could not have completed this without them and I am eternally grateful.

A special mention for my Grandad, who is not here to see this journey but shaped my life leading it up to it. This is for you.

Chapter 1 Introduction

1.1 Lyme disease

1.1.1 History of Lyme disease

Lyme disease, or Lyme borreliosis, is recognised as the most common vector-borne infection across temperate zones of the Northern Hemisphere (Lindgren & Jaenson, 2007), with reported cases increasing significantly over the past 20 years (Centers for Disease Control and Prevention, 2018a). While symptoms of the disease are recorded throughout history, modern emergence famously occurred in the small town of Lyme in Connecticut, USA when an epidemic of arthritis arose in children and young adults (Steere et al., 1977). Despite the initial diagnosis of juvenile arthritis, the presence of a typical bullseye rash, known as an erythema migrans (EM) rash was often seen, a skin condition first noted in 1909 by Swedish dermatologist Arvid Afzelius (Bergström & Normark, 2018; Marcus, 1912). The presence of EM, along with the development of later symptoms in many of these patients, suggested the initial diagnosis of juvenile arthritis was wrong and the symptoms were later linked with a tick bite (Steere et al., 1978).

Even though the vector of Lyme had been identified, the pathogen responsible remained unknown until its discovery in the Rocky Mountain Laboratories in Montana, USA. While trying to isolate Rickettsia from *Ixodes dammini* ticks, scientist Willy Burgdorfer identified and isolated a new *Treponema* like spirochete (Burgdorfer et al., 1982), which consequently became known as *Borrelia burgdorferi sensu lato*, the causative agent of Lyme disease. *Borrelia* species are generally divided into two major groups, those that cause Lyme disease, which are under the umbrella term *Borrelia burgdorferi sensu lato (s.l)*, and those causing relapsing fever. In addition, there are also several other species outside of these groups which are not yet characterised or do not cause human disease (Franke et al., 2013). There are three main genospecies associated with human infection; *Borrelia afzelii* and *Borrelia garinii* which are the most common in Europe and Asia and *Borrelia burgdorferi sensu stricto (s.s)* which is the main species in cases in the USA (Stanek et al., 2012).

1.1.2 Symptoms

Lyme disease can manifest in a variety of ways, with symptoms often depending on the species of *Borrelia* a person is infected with and the stage of infection. Infection occurs in three main stages; early localised infection, disseminated infection and persistent infection (Alao & Decker, 2012). During early infection, 3-30 days after the tick bite, one of the most common and well known symptoms is the presence of a round or oval shaped erythematous skin lesion (erythema migrans rash) (figure 1.1) that appears at the site of the tick bite and this manifests in around 80% of patients (Alao & Decker, 2012; Stanek et al., 2012). At this stage, non-specific symptoms can also occur such as fever, headache and fatigue (Wormser, 2006). Without appropriate treatment at this stage, the bacteria can disseminate from the tick bite area to other host tissues such as the musculoskeletal system, the heart, the nervous system and other areas of skin resulting in more EM lesions (Aguero-Rosenfeld et al., 2005). Manifestations in this early disseminated stage of infection appear to vary based on the species of *Borrelia* the person has been infected with; *B. burgdorferi s.s* is associated with Lyme arthritis, *B. afzelii* with skin disorders and *B. garinii* with Lyme neuroborreliosis (Rosa et al., 2005a).

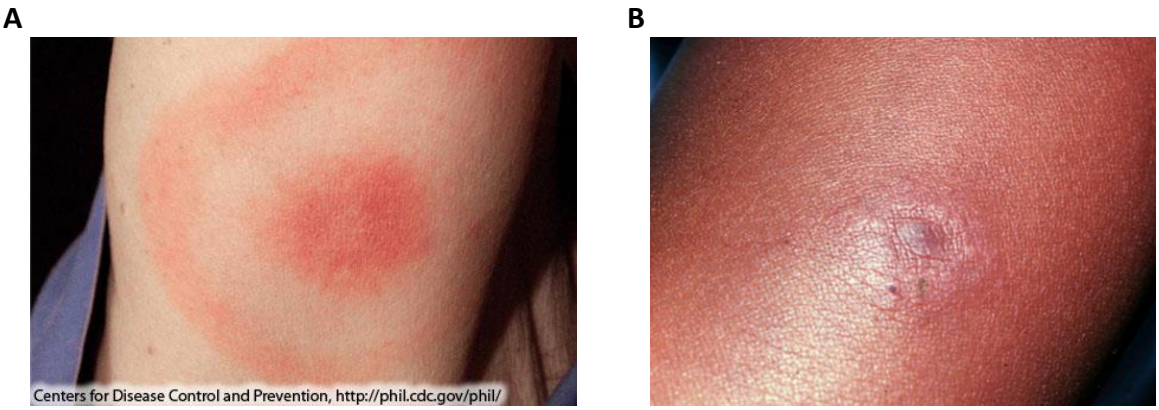


Figure 1.1. Erythema migrans (EM) rash. A. Example of a clear 'bullseye' shape EM rash (Centers for Disease Control and Prevention, 2018b). **B.** Example of an early EM nodule with no obvious rings around the tick bite mark in the centre (Bhate & Schwartz, 2011).

Approximately 60% of patients with EM will develop Lyme arthritis within ~6 months if left untreated (Cardenas de la Garza et al., 2019). Symptoms can be intermittent or persistent and patients usually present with monoarthritis (1 affected joint) or oligoarthritis (2-4 affected joints) of large and small joints, most commonly the knee joint (Arvikar & Steere, 2015).

Antibiotic treatment and the use of anti-inflammatory drugs is effective for most patients with Lyme arthritis, but some go on to develop chronic inflammatory arthritis and this is thought to be due to both host factors and the strain of *Borrelia* the person was infected with (Cardenas de la Garza et al., 2019; Sanchez, 2015).

Lyme carditis occurs in approximately 4-8% of patients within weeks of initial infection and most commonly results in arrhythmias and conduction delays, particularly varying degrees of atrioventricular (AV) block (Fish et al., 2008; Sanchez, 2015). While high-degree AV blocks in Lyme disease patients are usually resolved after antibiotic treatment (Yeung & Baranchuk, 2018), temporary cardiac pacing is required in one third of cases (Cadavid et al., 2004). Other less common serious cardiac complications include myocarditis, pericarditis, pericardial effusion, coronary artery aneurysm and myocardial infarction (Fish et al., 2008). While the outcome for patients with Lyme carditis is positive (~90% complete recovery), serious complications can result in fatality and Lyme carditis is considered the leading cause of mortality in Lyme disease patients (Rostoff et al., 2010).

Left untreated, around 15-25% of patients develop Lyme disease infection of the nervous system, also known as Lyme neuroborreliosis, which can cause acute meningitis, weakness and paralysis (Pearson, 2015). Early symptoms of Lyme neuroborreliosis include; cranial neuropathies with facial (Bell's) palsy being the most common, radicular pain caused by inflammation of spinal nerve roots (Bannwarth's syndrome), encephalitis and lymphocytic meningitis. Late Lyme neuroborreliosis is defined by disease persisting for 6 months or more and includes manifestations such as encephalitis, myelitis and chronic meningitis and in rare cases has the potential to cause stroke and intracerebral haemorrhage (Cardenas de la Garza et al., 2019).

Stage of infection	Manifestations	References
Stage 1 **Early localised infection**	Erythema migrans (EM) rash Flu-like illness Lymphocytoma	(Bhate & Schwartz, 2011; Steere et al., 2016)

Stage 2 **Disseminated infection**	Multiple EM lesions Arthritis Arrhythmias and conduction delays Myocarditis Pericarditis Cranial neuropathies Lymphocytic meningitis Severe fatigue	(Bhate & Schwartz, 2011; Cardenas de la Garza et al., 2019; Fish et al., 2008; Steere, 1989)
Stage 3 **Persistent infection**	Acrodermatitis chronica atrophicans Neuropathies with or without dementia Heart failure Refractory arthritis Fatigue	(Bhate & Schwartz, 2011; Steere, 1989)

Table 1.1. The stages of Lyme disease and associated clinical manifestations.

1.1.3 Diagnosis and treatment

Diagnosis of Lyme borreliosis is primarily based on the presentation of clinical manifestations, in particular the presence of an EM rash and also the likelihood of exposure to a tick. However, *Borrelia* infection can result in a wide range of nonspecific symptoms and a characteristic skin lesion is not always present making diagnosis difficult, so laboratory testing is often needed (Stanek et al., 2011). While there are various potential tests for diagnosing the disease, many of these lack sensitivity and specificity. Culturing spirochetes from patient samples is a possible direct test for diagnosing Lyme, but this method is expensive, time consuming and involves invasive sample collection such as a skin biopsy for tissue sample and lumbar puncture for collection of cerebrospinal fluid (Shapiro & Gerber, 2000). Another direct method is the use of PCR to target the genes encoding *Borrelia* antigens, but there is a risk of false positives as the presence of *Borrelia* DNA does not always mean the patient has an active *Borrelia* infection (Murray & Shapiro, 2010; Shapiro & Gerber, 2000). The current advised standard test procedure for patients with appropriate clinical manifestations is a two tiered approach, where an enzyme linked immunosorbent assay (ELISA) is to be carried out followed

by a Western blot and a positive result for Lyme is required from both tests in order to confirm a diagnosis of Lyme disease (Centers for Disease Control and Prevention, 1995).

Traditionally, an ELISA using whole cell *B. burgdorferi* lysate is followed by a confirmatory Western blot to detect IgM and IgG antibodies against *Borrelia* antigens such as outer surface protein C (OspC) and flagellin (Fla) (John & Taege, 2019). Newer ELISA testing detects antibodies against the invariable region 6 (IR6) section of the highly immunogenic variable surface protein (VlsE), using a synthetic peptide known as C6 which comprises the sequence of IR6 (Hu, 2016; Liang et al., 1999). Detection in the early stages of infection (< 1 month) requires 2 of 3 antigens tested to show positive for IgM antibodies on a Western blot and later stage diagnosis (> 1 month) requires 5 of 10 bands to be present for IgG antibodies (Centers for Disease Control and Prevention, 1995).

Lack of specificity is an issue in Lyme disease serological testing, particularly when using whole cell preparations for *B. burgdorferi* antigens, due to cross reactivity of *Borrelia* antigens with other bacterial proteins (Aguero-Rosenfeld et al., 2005). Specificity is improved in ELISAs using the synthetic C6 peptide instead of whole cell antigen preparations. A high frequency of false positives occur in IgM Western blot results, potentially due to misinterpretation of faint bands seen from low levels of reactivity to antigens such as OspC in patients with other infectious diseases (Aguero-Rosenfeld et al., 2005; Seriburi et al., 2012). This may lead to overdiagnosis and unnecessary treatment (Seriburi et al., 2012). Finally, one of the major issues with Lyme disease testing is that IgG antibodies can be present years after the onset of infection, due to the strong immunological response produced by *Borrelia*, so positive serological tests do not always indicate an active infection (Talagrand-Reboul et al., 2020).

Current treatment, after a diagnosis of Lyme disease, relies on the use of antibiotics. For early localised or disseminated infections, such as patients presenting with an EM rash, doxycycline or amoxicillin is administered orally over a period of 2 weeks. Antibiotics such as ceftriaxone are administered intravenously for 2-4 weeks and are only recommended for patients with late stage infection or those presenting with neurological symptoms, even in early stages (Wormser et al., 2006). Despite appropriate treatment, symptoms can persist in some

patients for many years, known as Post Treatment Lyme Disease Syndrome (PTLDS) and symptoms in these cases can be ongoing for years (Novak et al., 2017). PTLDS is an area surrounded by controversy and debate. The term PTLDS is used to refer to patients with ongoing symptoms of Lyme disease despite finishing the course of treatment, but the scientific basis of this phenomenon is still unknown. After antibiotic therapy, remaining symptoms are reported by 3-27% of patients with EM rash and by 10-40% of patients with Lyme neuroborreliosis (Borchers et al., 2014).

While there is currently no human vaccine for Lyme disease, there have been previous vaccines on the market. In the 1990's two vaccines were developed; LYMErix™ (GlaxoSmithKline) and ImuLyme™ (PasteurMérieuxConnaught), both targeting the immunogenic *Borrelia* outer surface protein A (OspA) (Nigrovic & Thompson, 2007; Sigal et al., 1998; Steere et al., 1998) despite the genetic heterogeneity among *B. burgdorferi* isolates (Mathiesen et al., 1997). LYMErix™ showed an efficacy of 76% following three injections in patients with confirmed Lyme (Steere et al., 1998), but was later withdrawn from the market, as sales were low due to reports of side effects such as arthritis and the requirement for additional booster injections to maintain high levels of antibodies (Federizon et al., 2019). ImuLyme™ showed an efficacy of 92% after three injections, but the manufacturers did not go ahead with vaccine licensing due to the events following the LYMErix™ vaccination (Poland, 2011). Recently, a potential new six-component vaccine has been developed from OspA-serotypes 1, 2, 3, 4, 5, and 7. The vaccine uses OspA fused to a bacterial ferritin to generate self-assembling nanoparticles which resulted in higher titres of antibodies and a more durable response in mice and non-human primates than previous marketed vaccines (Kamp et al., 2020).

1.1.4 Relapsing fever

Relapsing fever is a zoonotic disease transmitted when a person is bitten by an infected tick or a human body louse, *Pediculus humanus*. Cases of relapsing fever in the US are recorded much earlier in history than Lyme disease, with tick-borne relapsing fever (TBRF) first reported in 1905 and louse-borne relapsing fever (LBRF) reported in 1844 (Dworkin et al., 2008). Distributed globally, TBRF causing *Borrelia* species are mainly transmitted by *Ornithodoros*

(soft bodied) ticks and include; *B. hermsii, B. parkeri, B. turicatae, B. duttonii, B. hispanica, B. persica, B. caucasica* and *B. crocidurae* (Talagrand-Reboul et al., 2018). Interestingly, *Ixodes* (hard bodied) ticks, the vector of Lyme disease causing *Borrelia* species, have also been found to cause TBRF by the transmission of *B. miyamotoi* (Wagemakers et al., 2015). LBRF is exclusively transmitted by infection with *B. recurrentis* and although it has previously been prevalent globally, it is now only endemic in areas of Africa. In contrast to the tissue specific manifestations of Lyme disease, which affect the skin, joints, heart and nervous system, relapsing fever is characterised by high levels of spirochetes in the blood, causing flu-like symptoms and an intermittent recurring fever that ranges in severity (Bergström & Normark, 2018).

1.1.5 Distribution and transmission of *Borrelia*

Borrelia species known to cause human infection of Lyme disease are known collectively as *Borrelia burgdorferi sensu lato* (s.l) and as previously mentioned, there are three main genospecies responsible for the majority of infections; *Borrelia afzelii*, *Borrelia garinii* and *Borrelia burgdorferi sensu stricto. B. burgdorferi s.l* genospecies are mainly transmitted by four species of hard-bodied *Ixodes* ticks; *I. ricinus* in Europe, *Ixodes pacificus* and *I. scapularis* in the USA and *I. persulcatus* in Asia (Radolf et al., 2012). The global distribution of *Borrelia* species corresponds with the distribution of the tick vector each species is commonly associated with (summarised in table 1.2) and thus the availability of animal reservoirs.

***Borrelia* species**	**Vector**	**Geographical distribution**
B. afzelii	*I. ricinus, I. persulcatus, I. hexagonus*	Asia, Europe
B. americana	*I. pacificus, I. minor*	USA
B. andersonii	*I. dentatus*	USA
B. bavariensis	*I. ricinus*	Europe
B. bissettii	*I. ricinus, I. scapularis, I. pacificus, I. minor, I. affinis*	Europe, USA
B. burgdorferi sensu stricto	*I. ricinus, I. scapularis, I. pacificus, I. spinipalpis, I. hexagonis, I. affinis, I. minor, I. muris*	Europe, USA
B. californiensis	*I. pacificus, I. jellisonii, I. spinipalpis*	USA
B. carolinensis	*I. minor*	USA

B. garinii	*I. ricinus, I. persulcatus, I. nipponensis, I. uriae*	Asia, Europe
B. finlandensis	*I. ricinus*	Europe
B. japonica	*I. ovatus*	Japan
B. kurtenbachiia	*I. scapularis, (I. ricinus)*	Europe, USA
B. lusitaniae	*I. ricinus*	Europe, North Africa
B. sinica	*I. ovatus*	China
B. spielmanii	*I. ricinus*	Europe
B. tanukii	*I. tanuki*	Japan
B. turdi	*I. turdus*	Japan
B. valaisiana	*I. ricinus, I. granulatus, I. nipponensis, I. columnae*	Asia, Europe
B. yangtze	*I. granulatus, I. nipponensis, Haemaphysalis longicornis*	China

Table 1.2. Geographical distribution of species of *Borrelia burgdorferi sensu lato*. The species of *Borrelia* is listed in the first column, followed by its tick vector/s and finally the geographical location/s it is most commonly found in. Adapted from (Borchers et al., 2014).

Borrelia is maintained in an enzootic cycle by *Ixodes* ticks and a variety of vertebrate hosts such as small rodents, deer and birds, with humans being an incidental host and a dead end to the cycle (Cerar et al., 2015). Ixodes ticks go through a two year life cycle where they develop from larva to nymph and finally to the adult stage (figure 1.2). The female tick has a blood meal in each stage of its cycle and its generalist feeding behaviour allows transmission of *Borrelia* between numerous different animals, linking the spirochete to many different niches (Kurtenbach et al., 2006). Upon hatching from eggs, larval ticks are uninfected and acquire *B. burgdorferi* during their first blood meal from an infected animal, usually a small mammal or birds at this stage. After feeding, the larval tick moults to a nymph and the spirochete is able to persist in the tick's mid-gut despite the poor nutritional environment between feeds (Brisson et al., 2012).

The nymphal blood meal triggers replication of the spirochete and results in its migration from the tick mid-gut to the salivary glands, where it is able to enter the bloodstream of its host along with the saliva of the tick (Brisson et al., 2012). The nymph feeds on either a small mammal, maintaining the natural cycle, or a human which is a dead end host and nymphs are the main cause of human infection. When a nymph has fed on a small mammal, it moults to an adult and will feed usually on deer which are not competent hosts but are important for maintaining the tick population (Kurtenbach et al., 2006). The infected adult tick lays uninfected eggs and continues the cycle.

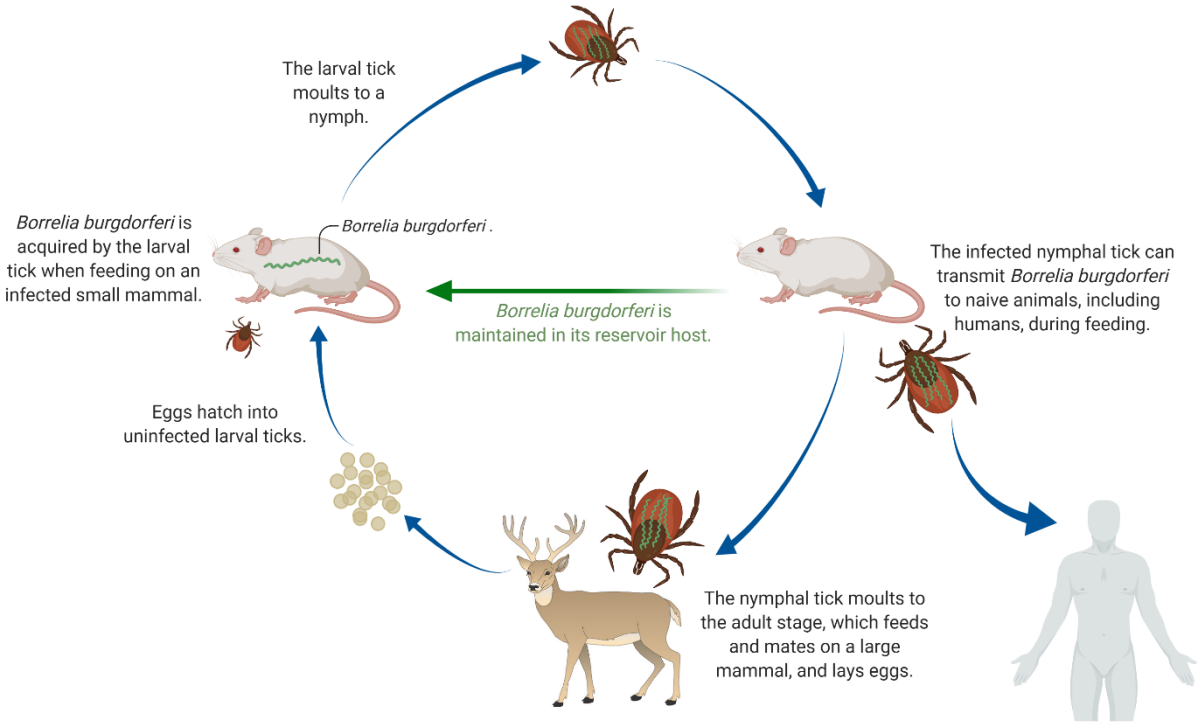


Figure 1.2. Enzootic cycle of *Borrelia burgdorferi*. *Borrelia burgdorferi* is transmitted and maintained through an enzootic cycle. Ticks acquire the pathogen upon feeding from an infected animal. The infected tick moults to a nymph and feeds on other hosts, including humans and transmits the bacteria. The tick moults to the adult stage and feeds on a larger animal, where it mates and lays eggs. The eggs hatch into uninfected larvae and the cycle continues. Adapted from (Brisson et al., 2012) and created in BioRender.

1.2 *Borrelia*

1.2.1 Introduction to spirochetes

Spirochetes are a diverse group of coiled, motile bacteria that make up the monophyletic phylum Spirocheates, many of which are responsible for serious diseases (Paster & Dewhirst, 2000). They are usually found in liquid environments (e.g. blood, plasma, mud) and can be either free living or host associated (Rosa et al., 2005a). The phylum comprises a single class, Spirochaetia, a single order, Spirochaetales, and four families; *Spirochaetaceae*, *Brachyspiraceae*, *Leptospiraceae*, and *Brevinemataceae*. The genus *Borrelia* come under the family *Spirochaetaceae*, along with *Treponema*, the causative agent of syphilis, yaws and periodontal disease (Fraser et al., 1998). Some *Leptospira* species in the family *Leptospiraceae*, such as *Leptospira interrogans*, are pathogenic and are responsible for Weil's disease (Jia et al., 2003). *Brachyspira* are members of the *Brachyspiraceae* family and are the causative agent of intestinal spirochaetosis (Gupta et al., 2013).

1.2.2 Known *Borrelia* species

The genus *Borrelia*, named after French biologist Amédée Borrel, encompasses around 50 species, many of which have been recognised over recent years (Kingry et al., 2018). While not all species cause disease, the genus is generally divided into two main clades; those that cause Lyme disease and those that cause relapsing fever. *B. afzelii, B. garinii* and *B. burgdorferi sensu stricto* are the three main species responsible for Lyme disease in humans and are known collectively as *B. burgdorferi sensu lato* (Stanek et al., 2012). As discussed in section 1.1.4, many species of *Borrelia* are responsible for relapsing fever, with *B. hermsii* and *B. duttonii* most common for TBRF from soft bodied ticks, *B. miyamotoi* for TBRF from hard bodies ticks and *B. recurrentis* responsible for LBRF. Recently, a new clade of *Borrelia* found in echidna (spiny anteater) and reptile hosts has been identified, consisting of two species; *B. turcica* and *B. tachyglossi* (Margos et al., 2018). This clade is genomically distinct from the Lyme disease and relapsing fever clades and shows early divergence from the relapsing fever clade (Gofton et al., 2018).

Over recent years, there has been much debate surrounding the genus *Borrelia* after a phylogenomic study by Adeolu and Gupta in 2014 proposed separating the genus into two

genera, with relapsing fever spirochetes coming under the genus *Borrelia* and Lyme disease spirochetes coming under the genus *Borreliella* (Adeolu & Gupta, 2014). This was disputed by Margos et al who argued that the evidence provided was insufficient for division of the genus and concerns about the impact on public health and the patient have also been raised (Margos et al., 2017; Stevenson et al., 2019).

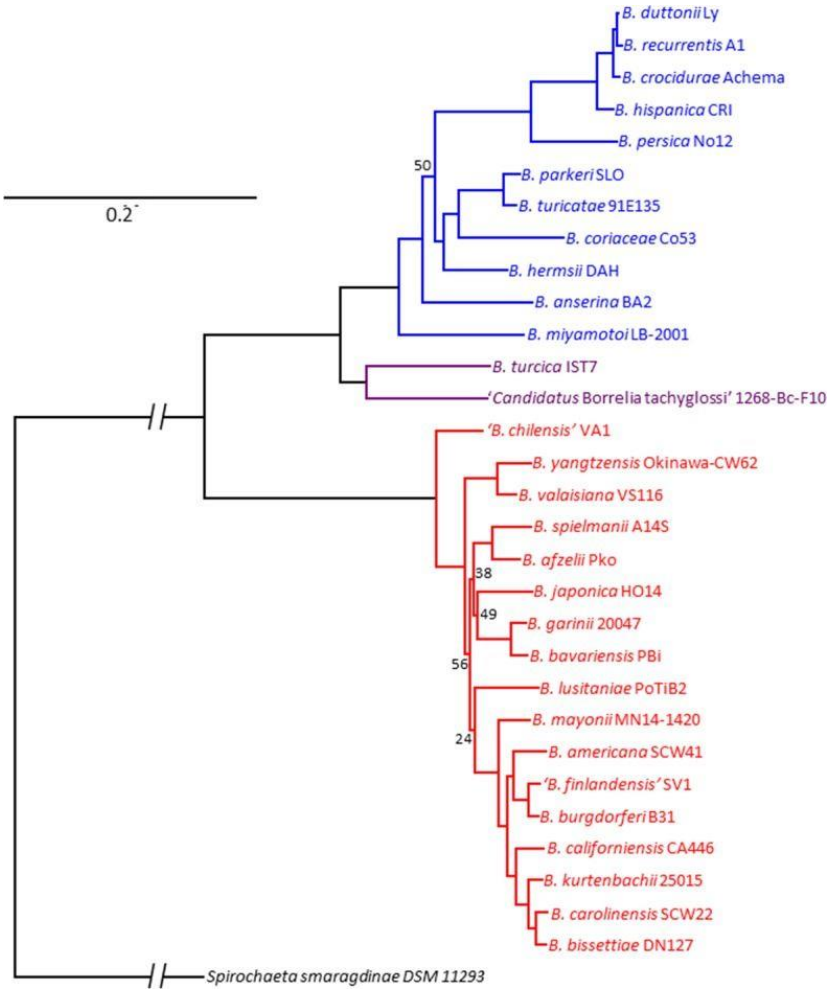


Figure 1.3. Phylogenetic tree of *Borrelia* species. Taken from (Margos et al., 2018). The phylogenetic reconstruction was based on 791 aligned protein homologs built with the PEPR pipeline and FastTree2 with 100 jackknifed resampling replicates. Lyme disease species are shown in red, relapsing fever species in blue and reptile associated species in purple.

1.2.3 *Borrelia* genomes

The *Borrelia* genome is perhaps the most unusual bacterial genome, segmented into a linear chromosome, circular plasmids and linear plasmids. The first whole *Borrelia* genome to be

sequenced was that of *B. burgdorferi s.s B31*, consisting of a linear chromosome nearly 911,000 base pairs in size, 9 circular and 12 linear plasmids ranging in size from ~5-56 kilo base pairs (figure 1.4) (Beaurepaire & Chaconas, 2005; Fraser et al., 1997). Many other *B. burgdorferi s.s* genospecies have been sequenced since, along with other related *Borrelia* species including *B. afzelii, B. garinii* and *B. bavariensis*. The size and gene content of the linear chromosomes do not appear to differ greatly between different *Borrelia* species, suggesting the chromosome is evolutionary stable (Casjens et al., 2012). The linear chromosome of *B. burgdorferi s.s B31* contains 815 predicted genes, which encode housekeeping proteins such as those required for replication of DNA, transcription and translation (Casjens et al., 2012; Fraser et al., 1997). Highly conserved β-barrel integral membrane proteins are also coded for on the chromosome.

The plasmids make up ~47% of the Borrelial genome and encode many outer surface lipoproteins, some of which are essential for virulence (Norris et al., 2011), but carry very few genes that share homology with other bacterial proteins. There is a high variability in plasmids between not only different species of *Borrelia*, but also between strains such as those of *B. burgdorferi s.s*. The linear plasmids of *B. burgdorferi* contain an unusually high amount of non-protein coding DNA and also contain many pseudogenes which are described as being in a state of mutational decay (Casjens, 2000). Another interesting feature of the *B. burgdorferi* genome is its high level of redundancy, particularly among the cp32 circular plasmids which contain long stretches of identical DNA interrupted by variable lipoprotein genes (Radolf et al., 2012).

While the plasmids of *Borrelia* are variable between species and many are lost during *in vitro* culture, loss of the circular plasmid cp-26 has not been observed and it appears to be an essential plasmid. An essential gene coded on cp-26 is the telomere resolvase, ResT, which is required for the replication of linear plasmids due to their covalently closed hairpin ends (Brisson et al., 2012; Kobryn & Chaconas, 2002). It also carries the gene for outer surface protein C (OspC), a lipoprotein involved in the transmission between the tick and host, further emphasising the importance of cp-26 for infectivity and survival (Byram et al., 2004). The

plasmids also contain a large number of pseudogenes and DNA sequence rearrangements between plasmids are common (Casjens et al., 2000).

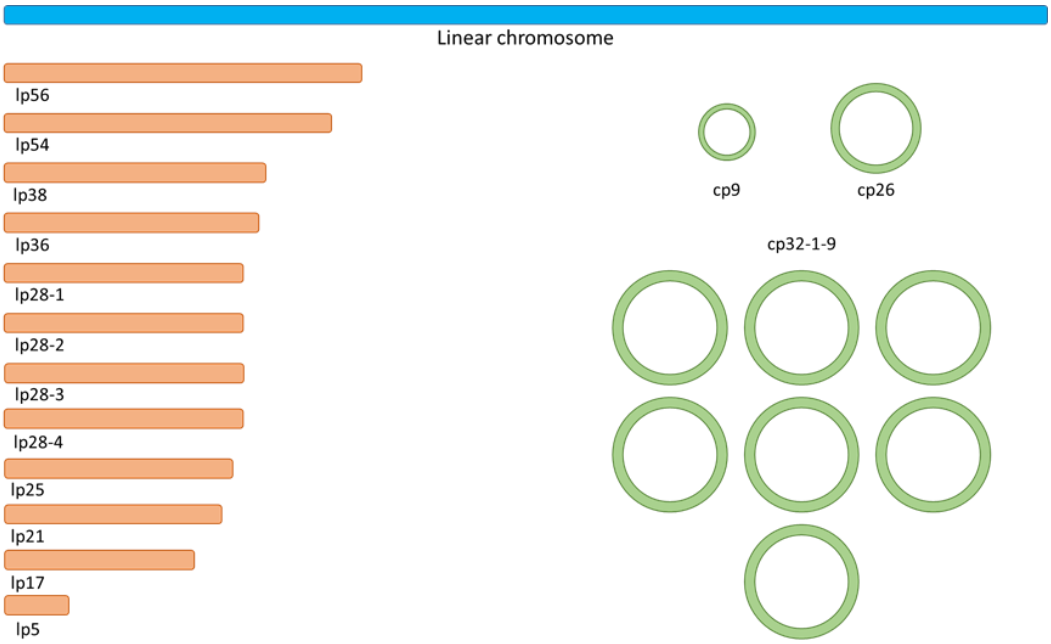


Figure 1.4. The arrangement of the *Borrelia burgdorferi* B31 genome. The genome of the B31 strain of *B. burgdorferi* includes a linear chromosome, 9 circular plasmids and 12 linear plasmids.

A further unusual feature of the *Borrelia* genome is the lack of genes for metabolic functions, such as those for synthesising amino acids, fatty acids, nucleotides and enzyme cofactors. It instead has genes encoding 16 distinct membrane transporters in order to obtain these factors (Fraser et al., 1997). This is possibly an evolutionary result of the spirochete acquiring a lifestyle where it parasitizes required nutrients from the tick vector and its hosts. *B. burgdorferi* also lacks genes encoding enzymes for oxidative phosphorylation and the tricarboxylic acid cycle and instead relies on energy from the fermentation of sugar into lactic acid (Radolf et al., 2012). Many studies have found that *B. burgdorferi* does not have a requirement for iron, an unusual feature for a bacterium, and its genome does not code for any iron-dependent metalloproteins (Posey & Gherardini, 2000). The spirochetes' metalloproteins may instead rely on other metal co-factors, which is seen in the case of superoxide dismutase, which is suggested to be manganese dependent (Brown et al., 2019).

1.2.4 Cellular architecture

Borrelia are motile spirochetes of 10-30 µM in length and are described as having a flat wave structure formed by periplasmic flagella wrapping round a rod shaped protoplasmic cylinder (Krupka et al., 2007; Motaleb et al., 2000). The ultrastructure of the spirochete is shown in figure 1.5 and the main structural features are the outer membrane, the periplasmic space containing the flagella, the protoplasmic cylinder and the cytoplasmic membrane (Penza et al., 2016). *B. burgdorferi s.l* have 2 bundles of 7–11 periplasmic flagella attached at each end of the cytoplasmic membrane and consist of minor filament protein FlaA and major filament protein FlaB. The flagella of *Borrelia* species are important for both motility and morphology, with mutants lacking the major flagellar filament protein gene FlaB becoming rod-shaped and non-motile (Motaleb et al., 2000).

Borrelia burgdorferi is often considered a Gram negative bacteria as it is composed of an inner cytoplasmic membrane surrounded by peptidoglycan and a loosely associated outer membrane. However, the outer membrane of *Borrelia* has unique characteristics that differentiate it from other Gram negative bacteria (section 1.3.1.), such as its lack of lipopolysaccharide (LPS) (Takayama et al., 1987) and its abundance of surface lipoproteins.

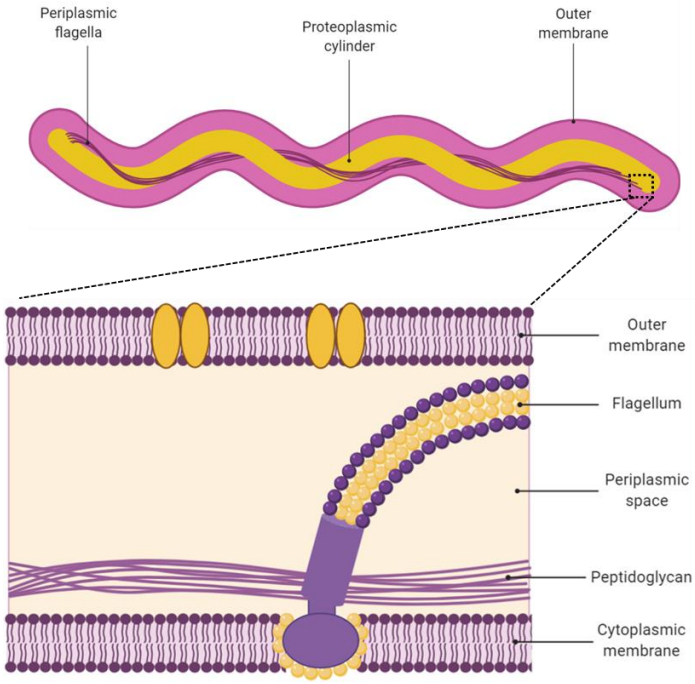


Figure 1.5. The structure of *Borrelia burgdorferi*. Adapted from (Rosa et al., 2005b) and made in BioRender.

1.2.5 Motility and chemotaxis

Borrelia have complex mechanisms of motility that allow the spirochete to move through gel-like viscous solutions that would usually inhibit bacterial motility. As previously mentioned, *B. burgdorferi* has two bundles of 7-11 endoflagella attached at each end of the cytoplasmic membrane which wrap around the protoplasmic cylinder and overlap in the middle (Motaleb et al., 2000). It has three different modes of movement; run, flex and reverse and this is controlled by the movement of its periplasmic flagella. During a run, the two bundles of flagella rotate in opposite directions (asymmetrically), with one group rotating clockwise and the other rotating counter-clockwise, generating propagating waves that travel backward along the cell and move it forward (Sal et al., 2008). In well studied bacteria such as *Escherichia coli* and *Salmonella enterica*, counter clockwise rotation of the flagella results in tumble movement, which is thought to be equivalent to flex in *Borrelia,* where both groups of flagella rotate in the same direction rather than asymmetrically. *B. burgdorferi* is also able to simultaneously alter the movement of both groups of periplasmic flagella and reverse directions in under 0.3 seconds, achieving rapid communication between opposite poles of the cell (Samuels & Radolf, 2010). The motors of the flagella are located at the poles of cell and are next to methyl-accepting chemotaxis proteins (MCPs) which are involved in bacterial chemotaxis transduction pathway and modulate flagella rotation (Steere et al., 2016).

Chemotaxis is important and advantageous in motile bacteria as it allows them to move towards favourable chemicals, such as sources of energy and oxygen, and avoid unfavourable conditions (Adler, 1966). *Borrelia* is anaerobic and does not require oxygen, but does rely on nutrient acquisition from the host. It is thought that *B. burgdorferi* depends on a reliable chemotaxis response for virulence and to migrate in both the tick and host. In the bacterial chemotactic response, MCPs are located in the cytoplasmic membrane and bind to ligands in the periplasmic space to regulate the autophosphorylation of histidine kinase CheA. The phosphate group is transferred by CheA to the response regulator CheY forming CheY-P, which interacts with the flagella switch complex and regulates flagella motor rotation (Samuels & Radolf, 2010). In *E. coli*, chemotaxis protein methyltransferase CheR catalyses the transfer of methyl groups from S-adenosylmethionine (SAM) to the membrane bound MCPs, counterbalancing ligand binding effects and allows continuous detection of chemicals

(Djordjevic & Stock, 1997; Schmidt et al., 2011). CheB is a methylesterase and catalyses demethylation of MCPs upon dephosphorylation of CheY-P, allowing a return to pre-stimulus state. However, both CheR and CheB remain uncharacterised in *Borrelia*. While the chemotaxis genes of *B. burgdorferi* are similar to those found in rod shaped bacteria such as *E. coli*, it contains multiple copies of these genes including two CheA (CheA1 and CheA2), three CheY (CheY1, CheY2, and CheY3), three CheW (CheW1, CheW2, and CheW3), two CheB (CheB1 and CheB2) and two CheR (CheR1 and CheR2) (Fraser et al., 1997). *B. burgdorferi* also encodes CheX, which is required for CheY-P dephosphorylation (Motaleb et al., 2005).

Chemotaxis protein methyltransferases (CheRs) are important enzymes involved in the chemotactic response in Gram negative bacteria. CheR catalyses the methylation of glutamate residues in the cytosolic signalling domain of chemoreceptors, using S-adenosyl-L-methionine (SAM) as a substrate and methyl donor (García-Fontana et al., 2013). The reaction produces S-adenosyl-L-homocysteine (SAH), which is able to bind to CheR at the same site as SAM and inhibit methylation. Some CheRs bind to a pentapeptide sequence present on the C-terminal of chemoreceptors, which is thought to aid tethering of the enzyme to the receptor and these are referred to as pentapeptide dependent CheRs (PDCheRs). Those which do not bind the pentapeptide sequence are known as pentapeptide independent CheRs (PICheRs).

The chemotactic response in *Salmonella typhimurium* and *E. coli* is well studied and these bacteria contain single copies of their chemotaxis genes, in contrast to *Borrelia* which contains multiple copies. The x-ray structure of *S. typhimurium* PDCheR in complex with SAH has been solved (figure 1.6) and shows consistency with other methyltransferase structures, with an overall Rossmann fold topology (Djordjevic & Stock, 1997). It is comprised of an α helical N-terminal domain thought to be important for receptor recognition and binding, an α/β domain involved in SAM/SAH binding and a β subdomain containing a pentapeptide binding site (Djordjevic & Stock, 1997). This overall topology is also seen in the x-ray structure of a PICheR in complex with SAH from *Bacillus subtilis* (figure 1.6), but sequence identity of the N-terminal region of StCheR and BsCheR was low at ~18%, suggesting differing mechanisms for recognition of chemoreceptors between different classes of CheRs (Batra et al., 2016).

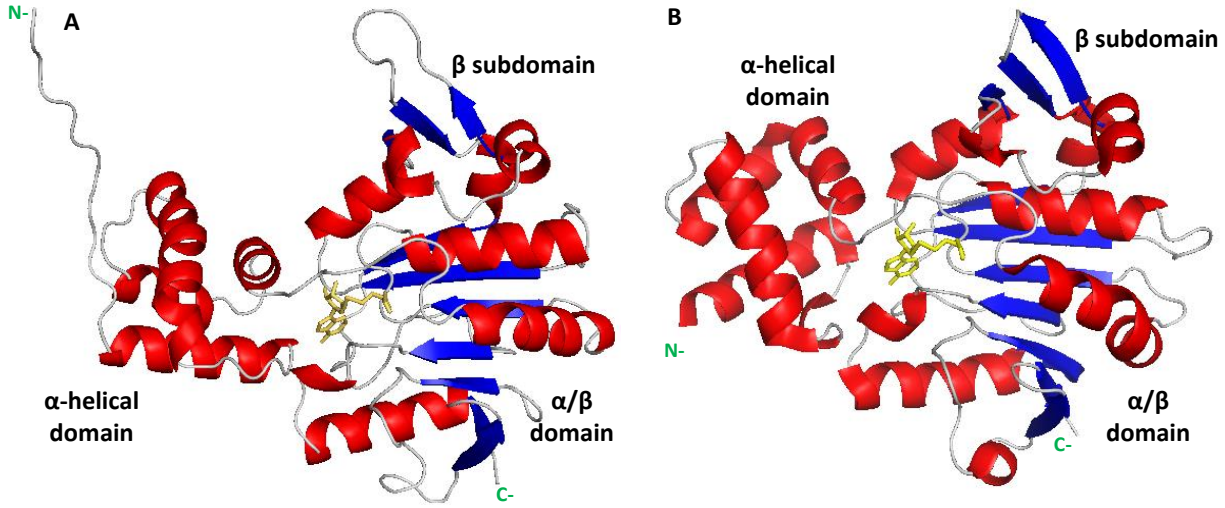


Figure 1.6. The crystal structure of StCheR and BsCheR in complex with SAH. A. The crystal structure of *Salmonella typhimurium* CheR in complex with S-adenosyl-L-homocysteine (PDB: 1AF7). **B.** The crystal structure of *Bacillus subtilis* CheR in complex with S-adenosyl-L-homocysteine (PDB: 5FTW). The α helical N-terminal domain, the α/β C-terminal domain and the β subdomain are indicated. Bound S-adenosyl-L-homocysteine is shown in yellow. Images produced using PyMol (Schrödinger, 2010).

1.3 Borrelial outer membrane

1.3.1 Introduction to the bacterial outer membrane

The Gram negative class of bacteria have a diderm structure, with an inner lipid bilayer and an outer lipid bilayer. The bacterial outer membrane (OM) is essential, allowing selective transport and serving as a protective barrier against potential toxins, including antibiotics (Silhavy et al., 2010). The outer membrane proteins (OMPs) of Gram negative bacteria have a vital role in virulence by various mechanisms, including facilitating host-pathogen interactions required for adhesion, dissemination and immune evasion (Koebnik et al., 2000). Furthermore, the surface exposed nature of these proteins make them important diagnostic and therapeutic targets in the treatment of bacterial infections.

While *Borrelia* is considered Gram negative, its outer membrane is quite distinct compared to the proteobacteria. A notable difference is the lack of highly immunogenic glycolipid lipopolysaccharide (LPS), which are frequently found in Gram negative bacteria such as *E. coli*, but are also absent in related spirochete *Treponema* (Takayama et al., 1987). The spirochetes *Leptospira* and *Brachyspira* do synthesise LPS, but these genera are more distantly related to

Borrelia than *Treponema* (Saier, 2000). While lacking in LPS, *Borrelia* has three glycolipids; cholesteryl 6-O-acyl-β-D-galactopyranoside (ACGal) and cholesteryl-β-D-galacto-pyranoside (CGal), which both contain cholesterol and mono-α-galactosyl-diacylglycerol (MGalD) which does not (LaRocca et al., 2010). The cholesterol glycolipids make up a significant percentage of *Borrelia*'s glycolipid content, with ACGal constituting ~45% of the glycolipid content and ~22% of the lipid content (Stübs et al., 2009).

As well as cholesterol glycolipids, *B. burgdorferi* has also been found to contain free cholesterol, which is rarely found in bacteria (LaRocca et al., 2010). Interestingly, *Borrelia* is unable to synthesise cholesterol or any unsaturated or saturated long chain fatty acids and instead has been shown to acquire cholesterol through a two-way lipid exchange with eukaryotic cells which can occur by direct contact between cells or through outer membrane vesicles (Crowley et al., 2013). During this exchange, the spirochete acquires cholesterol from the plasma membrane of epithelial cells and transfers cholesterol glycolipids to the host cells. The cholesterol glycolipids in *Borrelia* are able to form lipid rafts, which were long thought to be exclusive to eukaryotic cells since the presence of sterols in prokaryotes is uncommon. Lipid rafts are ordered microdomains in the OM that are rich in cholesterol and some lipid anchored proteins and are important for lateral protein sorting, maintaining heterogeneity in membranes, elasticity and vesicle formation (LaRocca et al., 2010).

1.3.2 Cell surface lipoproteins

Bacterial lipoproteins are anchored to the membrane by fatty acids linked to an N-terminal cysteine residue (Wilson & Bernstein, 2015) and around two thirds of the expressed lipoproteins in *B. burgdorferi* localise to the cell surface (Zückert, 2014). In contrast to most Gram negative bacteria, *Borrelia* encode many surface lipoproteins and few transmembrane proteins. One of the major survival and immune evasion strategies of *B. burgdorferi* is the differential expression of its cell surface lipoproteins during transmission and infection. Borrelial cell surface lipoproteins facilitate in cell adhesion, host-pathogen interactions and immune evasion. At different stages of *Borrelia* infection, different groups of lipoproteins are upregulated and downregulated in response to environmental changes such as temperature and this variable expression allows the spirochete to adapt to host conditions (Kenedy et al.,

2012; Pulzova & Bhide, 2014). A summary of some major outer surface proteins can be seen in table 1.3.

The outer surface proteins OspA and OspB are both encoded on *B. burgdorferi* linear plasmid lp54 and are important lipoproteins for survival and migration in the tick vector. They are 31 kDa and 34 kDa respectively and share 53% sequence identity. OspA is formed of 21 antiparallel β-strands with a single α-helix at the C-terminus (figure 1.7) (Li et al., 1997). While the complete structure of OspA has been solved by x-ray crystallography, only the structure of the 16.6 kDa C-terminus structure of OspB is available (figure 1.8) and was found to be homologous to the C-terminus of OspA (Becker et al., 2005). OspA is upregulated in the mid-gut of an unfed tick and is required for spirochete adhesion to the tick mid-gut, as mutants lacking OspA are unable to colonise and survive (Yang et al., 2004). The protein mediates this adhesion by specific binding to a tick receptor found in the mid-gut known as the TROSPA (Tick Receptor for Outer Surface Protein A) receptor (Pal et al., 2004). When the tick feeds and has a blood meal, the increase in temperature in the mid-gut results in the downregulation of OspA on the spirochete surface and TROSPA in the tick to allow migration to the salivary glands of the tick and the spirochete instead upregulates the highly immunogenic outer surface protein C (Pal et al., 2000; Schwan et al., 1995).

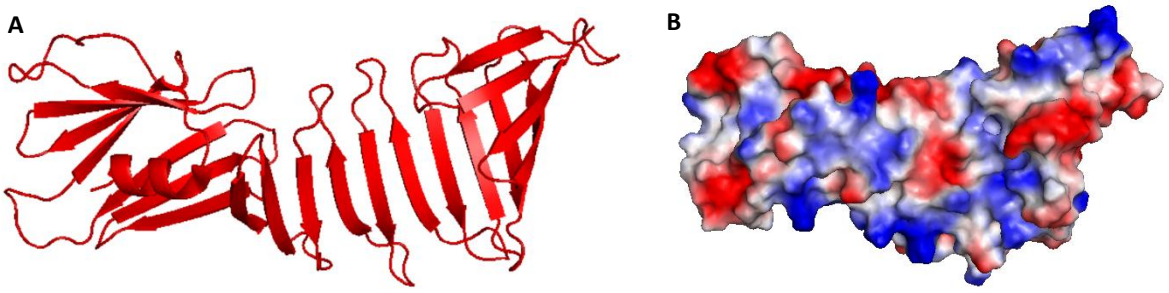


Figure 1.7. Crystal structure of *B. burgdorferi* OspA (PDB: 2G8C). A. Cartoon representation of OspA . **B.** Surface electrostatic potential of OspA. Images produced using PyMol (Schrödinger, 2010).

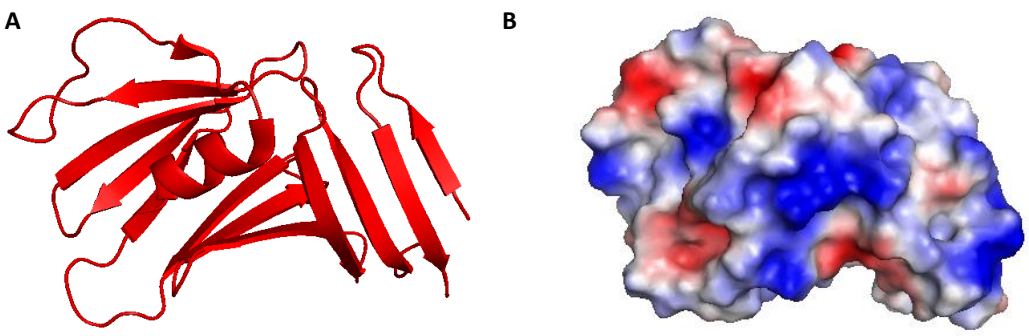


Figure 1.8. Crystal structure of the C-terminus of *B. burgdorferi* OspB (PDB: 1P4P). A. Cartoon representation of the C-terminus of OspB . **B.** Surface electrostatic potential of the C-terminus of OspB. Images produced using PyMol (Schrödinger, 2010).

Outer surface protein C (OspC) is a dimeric 22 kDa lipoprotein encoded for on the circular plasmid cp26 and is essential for infection of the host (Grimm et al., 2004; Schwan et al., 1995). It is thought that OspC is important for invasion of the tick salivary glands and OspC deficient mutants were unable to colonise the salivary glands (Pal et al., 2004). While OspC is highly immunogenic, it has been shown to bind to the tick salivary protein Salp15, which protects the spirochete from antibody-mediated killing during initial mammalian infection (Ramamoorthi et al., 2005). OspC is also down-regulated once initial infection is established, again protecting the spirochete from the host humoral response (Seemanapalli et al., 2010). The structure of monomeric OspC consists of four long α-helices and a fifth short α-helix and upon dimerization forms a hydrophobic core (figure 1.9) (Kumaran et al., 2001). The protein also has a highly negative electrostatic patch on the top surface of the protein (not seen in figure 1.9), which is potentially involved in interactions with a positively charged ligand and suggests a role for OspC in host interactions.

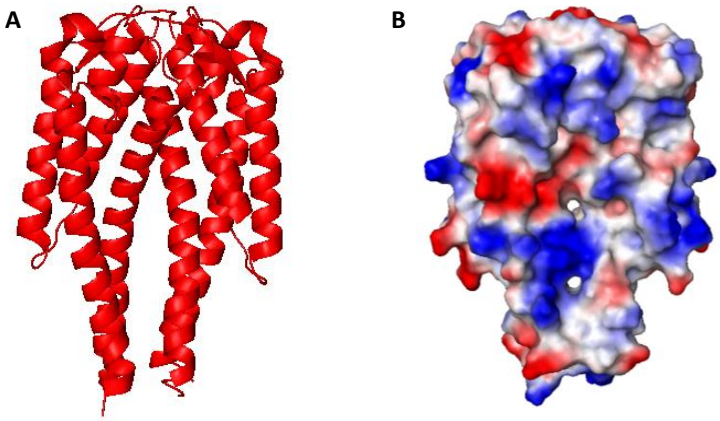


Figure 1.9. Crystal structure of dimeric *B. burgdorferi* OspC (PDB: 1GGQ). A. Cartoon representation of OspC . **B.** Surface electrostatic potential of OspC. Images produced using PyMol (Schrödinger, 2010).

Borrelia is transferred to the human host along with the tick saliva through the bite wound and, upon transmission, other surface proteins are upregulated, including decorin binding proteins DbpA and DbpB, BBK32 and complement regulator-acquiring surface proteins (CRASPs). The ability of *Borrelia* to adhere to the extracellular matrix is essential for colonisation of tissues and this is facilitated by proteins such as DbpA/B and BBK32 (Fikrig et al., 2000; Fischer et al., 2003). BBK32 is a 47 kDa lipoprotein which has been found to bind to fibronectin, a large (~440 kDa) glycoprotein found as a component of the extracellular matrix (ECM) and also in body fluids (Probert & Johnson, 1998). As well as binding to fibronectin, BBK32 has also been found to bind glycosaminoglycans (GAGs) in the ECM, further facilitating bacterial adhesion and colonisation of host tissues (Fischer et al., 2006). Decorin binding proteins A and B bind to decorin, a proteoglycan that interacts with collagen and can be found in connective tissue. Studies have found that DbpA and DbpB are not required for mammalian infection but may be important for dissemination and chronic infection of certain tissues (Kenedy et al., 2012).

Complement regulator-acquiring surface proteins (CRASPs) are an important group of surface expressed proteins in *Borrelia* that contribute to immune response evasion as they bind with a high affinity to the human blood components factor H (FH), factor H like-1 (FHL1) and complement factor H-related proteins (CFHR) giving resistance to complement-mediated killing by inhibiting the formation of the terminal membrane attack complex (further discussed in section 1.4.2.) (Kenedy et al., 2009; Lackum et al., 2005). *B. burgdorferi s.s* encodes five CRASPS; BbCRASP-1 (CspA), BbCRASP-2 (CspZ), BbCRASP-3 (ErpP), BbCRASP-4 (ErpC) and BbCRASP-5 (ErpA). Differing from BbCRASPs 1 and 2, BbCRASPs 3-5 have similar sequences and functions to one another and are collectively known as OspE related proteins (Erps) (Kraiczy & Stevenson, 2013). CRASP-1 is a 27.5 kDa protein consisting of five crossing α-helices and with two monomers making up its dimeric structure, creating a cleft thought to be the FH binding site (figure 1.10) (Cordes et al., 2005; Cordes et al., 2006).

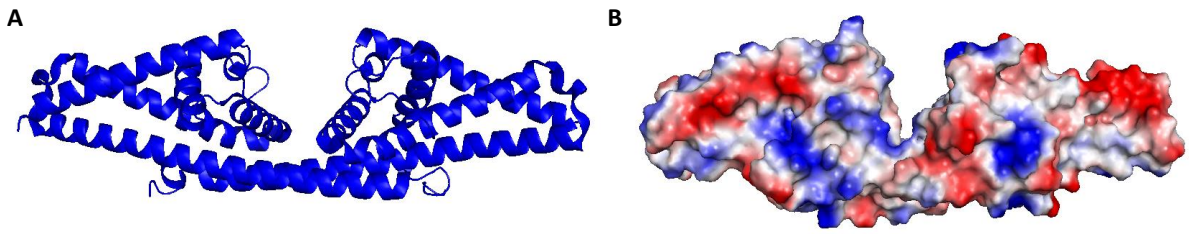


Figure 1.10. Crystal structure of BbCRASP-1 from *B. burgdorferi* (PDB: 1W33). A. Cartoon representation of BbCRASP-1. **B.** Surface electrostatic potential of BbCRASP-1. Images produced using PyMol (Schrödinger, 2010).

Protein	Alternative Names	Function	Expressed in	References
OspA	BBA15	Adhesion to tick gut, spirochete-spirochete aggregation, degradation of ECM, adhesion and activation of host cells.	Vector and host	(Fuchs et al., 1994; Pal et al., 2000; Pal et al., 2004)
OspB		Adhesion to tick gut.	Vector	(Fikrig et al., 2004)
OspC		Adhesion to tick salivary glands, establishment of early host infection, ECM degradation.	Vector/host	(Lagal et al., 2006; Norgard et al., 2005; Pal et al., 2004)
BBK32		Adhesion to ECM.	Vector/host	(Fikrig et al., 2000)
CRASP-1	BbCRASP-1 CspA BBA68	Complement system evasion, ECM degradation, adhesion to host cells, adhesion to ECM.	Vector/Host	(Hallström et al., 2010; Kraiczy & Stevenson, 2013)
CRASP-2	BbCRASP-2 CspZ BBH06	Complement system evasion.	Host	(Brissette et al., 2009; Kraiczy et al., 2001)
CRASP-3	BbCRASP-3 ErpP BBN38	Complement system evasion, ECM degradation.	Vector/Host	(Brissette et al., 2009; Haupt et al., 2007; Schwab et al., 2013; Siegel et al., 2010)

CRASP-4	BbCRASP-4 ErpC	Complement system evasion, ECM degradation.	Vector/Host	(Brissette et al., 2009; Haupt et al., 2007; Schwab et al., 2013; Siegel et al., 2010)
CRASP-5	BbCRASP-5 ErpA BBP38 BBL39	Complement system evasion, ECM degradation.	Vector/Host	(Brissette et al., 2009; Haupt et al., 2007; Schwab et al., 2013; Siegel et al., 2010)
DbpA	BBA24	Adhesion to ECM.	Host	(Fischer et al., 2003)
DbpB	BBA25	Adhesion to ECM.	Host	(Fischer et al., 2003)

Table 1.3. A summary of some important outer surface proteins from *Borrelia*. The name of each protein is given, followed by an overview of its function and the whether the protein is upregulated in the tick vector or human host. Adapted from (Pulzova & Bhide, 2014).

1.3.3 Integral β-barrel proteins

Outer membrane β-barrel proteins are located in the outer membranes of Gram-negative bacteria, chloroplasts and mitochondria (Schulz, 2000). These β-barrels can exist in both monomeric and oligomeric form and vary in size, generally having between eight and twenty two β-strands spanning the OM (Schulz, 2000). They follow a set of construction principles outlined by Schulz which are listed in table 4. The circular barrel is formed by hydrogen bonding between the edge strands of a β-sheet and is stabilised by hydrogen bonds between the transmembrane anti-parallel β-strands (Schulz, 2002; Wimley, 2003). In contrast to water soluble proteins which have a hydrophobic core, transmembrane β-barrels contain a polar internal core with non-polar hydrophobic residues on the exterior of the barrel facing the lipid membrane and many smaller barrels contain water in their core (Schulz, 2002). They can therefore be described as inverted micelles. β-barrel proteins serve a variety of important functions in the bacterial OM, including porins, specific transport channels, signal transduction and aiding in adhesion and immune evasion (Koebnik et al., 2000).

1.	The number of β strands is even* and the N and C termini are at the periplasmic barrel end.
2.	The β-strand tilt is always around 45° and corresponds to the common β-sheet twist. Only one of the two possible tilt directions is assumed, the other one is an energetically disfavoured mirror image.
3.	The shear number of an n-stranded barrel is positive and around n+2, in agreement with the observed tilt.
4.	All β strands are antiparallel and connected locally to their next neighbours along the chain, resulting in a maximum neighbourhood correlation.
5.	The strand connections at the periplasmic barrel end are short turns of a couple of residues named T1, T2 and so on.
6.	At the external barrel end, the strand connections are usually long loops named L1, L2 and so on.
7.	The β-barrel surface contacting the nonpolar membrane interior consists of aliphatic sidechains forming a nonpolar ribbon with a width of about 22 Å.
8.	The aliphatic ribbon is lined by two girdles of aromatic side chains, which have intermediate polarity and contact the two nonpolar–polar interface layers of the membrane.
9.	The sequence variability of all parts of the β barrel during evolution is high when compared with soluble proteins.
10.	The external loops show exceptionally high sequence variability and they are usually mobile.

Table 1.4. Schultz's set of construction principles for β-barrel membrane proteins. (Schulz, 2000) *An exception to this principle is voltage-gated anion channel (VDAC), which has 19 strands rather than an even number.

1.3.3.1 *E. coli* OMPs

The OMPs of *E.coli* have been studied extensively. Knowledge of the structure and biochemistry of these model systems can be extended to other Gram negative bacteria, including the distantly related spirochetes. *E. coli* encodes approximately 26 integral outer membrane proteins and even fewer lipoproteins (~5) (Molloy et al., 2000). One of the best characterised outer membrane proteins is outer membrane protein A (OmpA) from *E. coli* (Smith et al., 2007) and this family of proteins is found in most Gram negative bacteria. OmpA is a heat modifiable protein and has an apparent molecular weight on SDS-PAGE dependent on the temperature conditions it is in, ranging between 28 and 36 kDa (Confer & Ayalew, 2013). An eight-stranded β-barrel with short periplasmic turns and 4 external loops make up

the highly conserved transmembrane (TM) domain of this protein (figure 1.11) (Krishnan & Prasadarao, 2012; Pautsch & Schulz, 2000). The interior of the barrel is polar and the exterior surface is hydrophobic, so the domain resembles an inverse micelle (Koebnik et al., 2000). It is thought that OmpA contributes towards bacterial surface stability, linking the outer membrane with the peptidoglycan layer (Koebnik et al., 2000). Bacterial OmpA proteins also have roles in bacterial invasion and adhesion, as well as survival in a host and immune evasion (Confer & Ayalew, 2013).

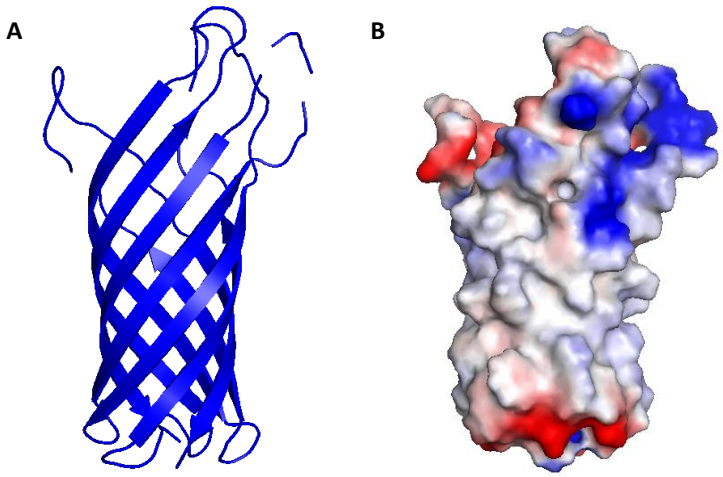


Figure 1.11. Crystal structure of the transmembrane domain of *E. coli* OmpA (PDB: 1BXW). A. Cartoon representation of OmpA transmembrane domain **B.** Surface electrostatic potential of OmpA transmembrane domain. Images produced using PyMol (Schrödinger, 2010).

OmpX is similar to OmpA in that its structure consists of 8 antiparallel β-strands making up the barrel (figure 1.12), but differs enough that attempting to solve the structure using OmpA for molecular replacement failed (Vogt & Schulz, 1999). Similar to other OMPs, it has aromatic residues forming two girdles and has non-polar residues attaching it to the interior of the membrane (Vogt & Schulz, 1999). OmpX resembles an inverse micelle and is not thought to form a pore. The barrel of OmpX is long in comparison to that of OmpA and it is thought to have a section that protrudes out of the exterior, leaving the edge of the β-sheet exposed to

bind to other proteins with a surface β-strand (Schulz, 2000). It is therefore thought to have a role in adhesion and invasion of cells, as well evasion of the complement response.

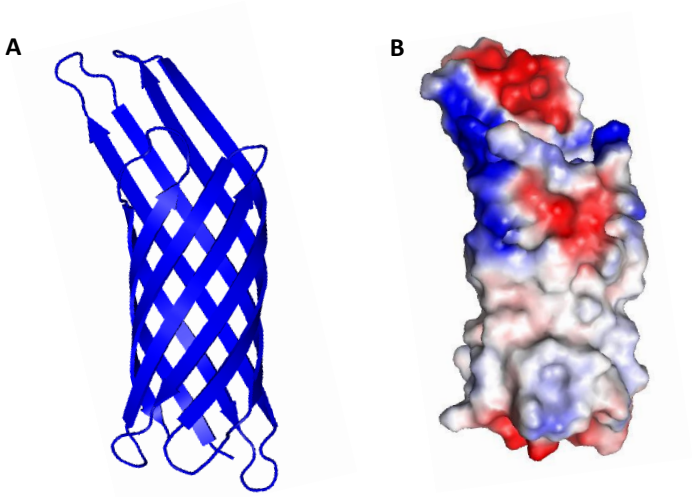


Figure 1.12. Crystal structure of *E. coli* OmpX (PDB: 1QJ8). A. Cartoon representation of OmpX **B.** Surface electrostatic potential of OmpX. Images produced using PyMol (Schrödinger, 2010).

OmpW is another small 8-stranded β-barrel protein found in the OM of *E. coli* (figure 1.13). It forms a long channel, differing from other small OMPs with its unusual hydrophobic core, suggesting a role in transport of small hydrophobic molecules across the OM (Hong et al., 2006). It consists of short turns on one side and long loops on the other, with the longer loops thought to be on the extracellular surface as with similar proteins (Hong et al., 2006). It is thought to have versatile functionality, protecting bacteria against environmental stresses such as temperature, osmosis and oxidation. *E. coli* OmpW has also been shown to have a role in bacterial immune evasion, as it was found to bind to human factor-H, contributing to resistance to complement-mediated killing via the alternative pathway (Li et al., 2016).

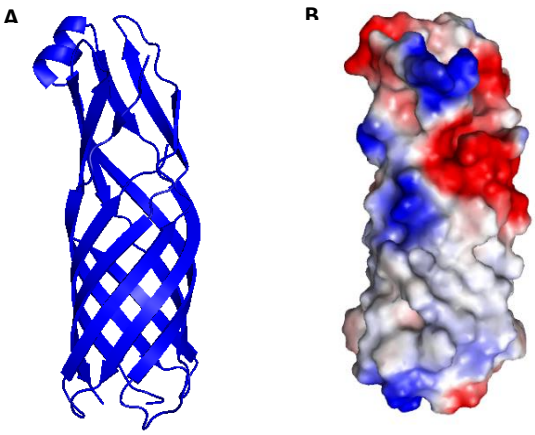


Figure 1.13. Crystal structure of *E. coli* OmpW (PDB: 2F1V). A. Cartoon representation of OmpW **B.** Surface electrostatic potential of OmpW. Images produced using PyMol (Schrödinger, 2010).

1.3.3.2 *Borrelia* OMPs

In contrast to *E. coli* and other Gram negative bacteria, *Borrelia* encodes a high number of lipoproteins (>100) and very few integral β-barrel outer membrane proteins (~10), most of which remain uncharacterised (Fraser et al., 1997; Kenedy et al., 2012). Some potential OmpA-like proteins have been identified in *Borrelia* but identification of these integral OMPs is difficult, likely due to their low abundance, their smaller membrane spanning region compared to other β-barrels and the lack of sequence conservation between orthologues due to variable external loops (Dyer et al., 2015; Kenedy et al., 2012). Unlike the outer surface lipoproteins, which are encoded on various different plasmids making them highly variable proteins, *Borrelia* transmembrane OMPs are encoded on the chromosome and are highly conserved (Fraser et al., 1997).

OMPs have a variety of important roles including transporting nutrients into the cell via porins. Few porins have been identified in *Borrelia* and include P66 (BB0603, Oms66) (Kenedy et al., 2014; Skare et al., 1997), P13 (BB0034) (Noppa et al., 2001), DipA (BB0418, Oms38) (Thein et al., 2012), BesC (BB0142) (Bunikis et al., 2008) and BamA (BB0795) (Lenhart & Akins, 2010) and are predicted to be β-barrels of between 16-24 strands. The most well studied *Borrelia* porin is p66, a 66 kDa protein encoded on the open reading frame BB0603 (Skare et al., 1997). Structural models predict p66 to have a β-barrel structure made up of 22-24 β-strands, supported by circular dichroism analysis indicating that the protein is made up of 48% β-sheet (Kenedy et al., 2014). As well as being found to form pores in artificial membranes, P66 was also shown to potentially have a role in adhesion by binding to β3-chain integrins (Coburn et al., 1999; Coburn & Cugini, 2003). Oms28 is a *Borrelia* protein that was originally thought to be an integral OMP with porin activity (Skare et al., 1996), but later research disputed this after circular dichroism data indicated the protein was 78% α-helix and it was suggested that Oms28 is periplasmic and associated with the OM (Mulay et al., 2007). BB0838 is predicted to be an amphiphilic OMP which is co-transcribed with two components of the nucleotide excision repair pathway, but the pore forming ability of this protein is unknown (Kenedy et al., 2016).

Other notable *B. burgdorferi* OMPs include the paralogous gene family BB_0405 and BB_406 and predicted OMPs BB_0027 and BB_0562. These are predicted to be OmpA-like integral membrane proteins and may be involved in immune evasion (Bhide et al., 2009; Dyer et al., 2015). One of these OmpA-like proteins from *B. afzelii*, BAPKO_0422 (BB0405 in *B. burgdorferi*) was analysed using small angle X-ray scattering and was shown to have a shape consistent with that of an 8-stranded β-barrel (Dyer et al., 2015). Functional studies found BAPKO_0422 binds to human factor-H, suggesting this protein potentially has a role in complement response regulation (Bhide et al., 2009; Dyer et al., 2015). However, a more recent study presents conflicting data, as it did not identify any factor-H binding by BAPKO_0422 (Shrestha et al., 2017).

Hypothetical proteins BB0405 and BB0406 share 59% sequence similarity and were both found to be amphiphilic, localised to the *B. burgdorferi* membrane and able to form pores in large unilamellar vesicles, suggesting a role as porins (Kenedy et al., 2016). Both proteins were shown to be immunogenic but only BB0405 was shown to be essential for infectivity in mouse models. Further investigations into the role of these two OMPs found that they did not bind human factor-H, but further studies are needed to assess their role in virulence (Shrestha et al., 2017). Most recently, a study investigating the function of BB0406 found that it binds to laminin, a component of basal membranes such as the vascular basal membrane, suggesting this protein aids *Borrelia* adhesion and dissemination (Bista et al., 2020). Other predicted *Borrelia* OMPs such as BB0562 and BB0027 remain uncharacterised and further work to try and elucidate the structure and function of these proteins could contribute to our understanding of *Borrelia* pathogenesis and perhaps identify more proteins with a role in immune evasion. The surface exposed nature of these proteins, along with their potential involvement in immune evasion, could also make them future vaccine target candidates.

1.4 Immune evasion strategies

1.4.1 Overview of *Borrelia* immune evasion

Like many bacteria, *Borrelia* have adapted numerous strategies to overcome the host immune responses in order to establish infection and survive. These mechanisms are well reviewed by (Berndtson, 2013) and include;

- Exploitation of tick salivary proteins to overcome early immune responses.
- Masking of surface antigens to avoid the alternative pathway of complement.
- Hijacking of host plasminogen activation.
- Antigenic variation to avoid humoral immune responses.
- Utilisation of chemotactic response and unique motility to avoid deleterious molecules.
- Engaging in biofilm-like behaviour and quorum sensing.
- Can assume alternate morphologies other than spirochaetal.

This research will focus on the role of *Borrelia* proteins in evasion of the complement response.

1.4.2 Evasion of the complement response

The complement system forms part of both the innate and adaptive immune response and is a first line of defence against microbial infection. It is a tightly regulated system that comprises over 30 proteins that are found in serum and associated with cell membranes (Fujita, 2002). These proteins become activated sequentially in an enzymatic cascade, whereby activation of a protein results in cleavage and activation of the following protein. The complement response can be activated through three different pathways: the classical, the lectin and the alternative pathway (figure 1.14). All three of these pathways result in the activation of C3 which leads to; the production of C3a and eventually C5a, anaphylatoxins that induce the inflammatory response, opsonisation and attraction of macrophages and neutrophils (Lambris et al., 2008). The activation products of the complement response form the membrane attack complex, a transmembrane channel that leads to cell lysis.

The classical pathway is triggered by complexes of IgG or IgM antibodies with non-self antigens (Merle et al., 2015). These antibody-antigen complexes activate the C1-complex, which is made of molecules of C1q, C1r and C1s. This activation occurs when C1q binds to a region of the IgG or IgM antibodies in the immune complex (Merle et al., 2015). The C1-complex cleaves C4 and C2 into C4a, C4b, C2a, and C2b, allowing C4b and C2a to form the C4b2a complex, which is a C3 convertase (Sarma & Ward, 2010). The pathway then converges with the lectin and alternative, cleaving C3 into C3a and C3b, which combines with the C4b2a complex to form the C5 convertase, C4b2a3b. This cleaves C5 into its products C5a and C5b and C5b is able to combine with C6, C7, C8 and C9 to form the C5b-9 membrane attack complex (Sarma & Ward, 2010).

The lectin pathway is similar to the classical pathway, but does not involve C1q and is instead activated by mannose-binding lectin (MBL) or ficolin binding to carbohydrates on the surface of a pathogen (Merle et al., 2015). This binding activates MBL-associated proteins known as MASPs and activation of MASP-2 results in cleavage of C4 into C4a and C4b and C2 into C2a and C2b (Fujita, 2002). As in the classical pathway, C4b and C2a combine to form the C3 convertase C4b2a and the pathway converges with the others.

The alternative pathway is initiated by binding of C3b to proteins, lipids and carbohydrates on the surface of an invading pathogen. C3b is produced at a low level continuously due to unstable C3 being broken down (Sarma & Ward, 2010). When C3b is bound to a pathogenic surface, it is able to bind factor B to form C3bB, which is then cleaved by factor D into Ba and Bb and allows Bb to complex with Cb to form the C3 convertase C3bBb (Lambris et al., 2008; Sarma & Ward, 2010). The C3bBb complex is stabilised by properdin (factor P), which binds to C3b and allows the complex to break down more C3 (Kemper et al., 2010). This pathway is regulated by a large glycoprotein known as factor-H, which allows differentiation between invading pathogens and host cells. Factor-H can bind to host C3b, preventing formation of the C3bBb C3 convertase and further cleavage of C3 (Wu et al., 2009). On the pathogen surface, C3bBb can bind to C3b to from the C5 convertase C3bBb3b, ultimately resulting in the formation of the membrane attack complex and apoptosis of the cell.

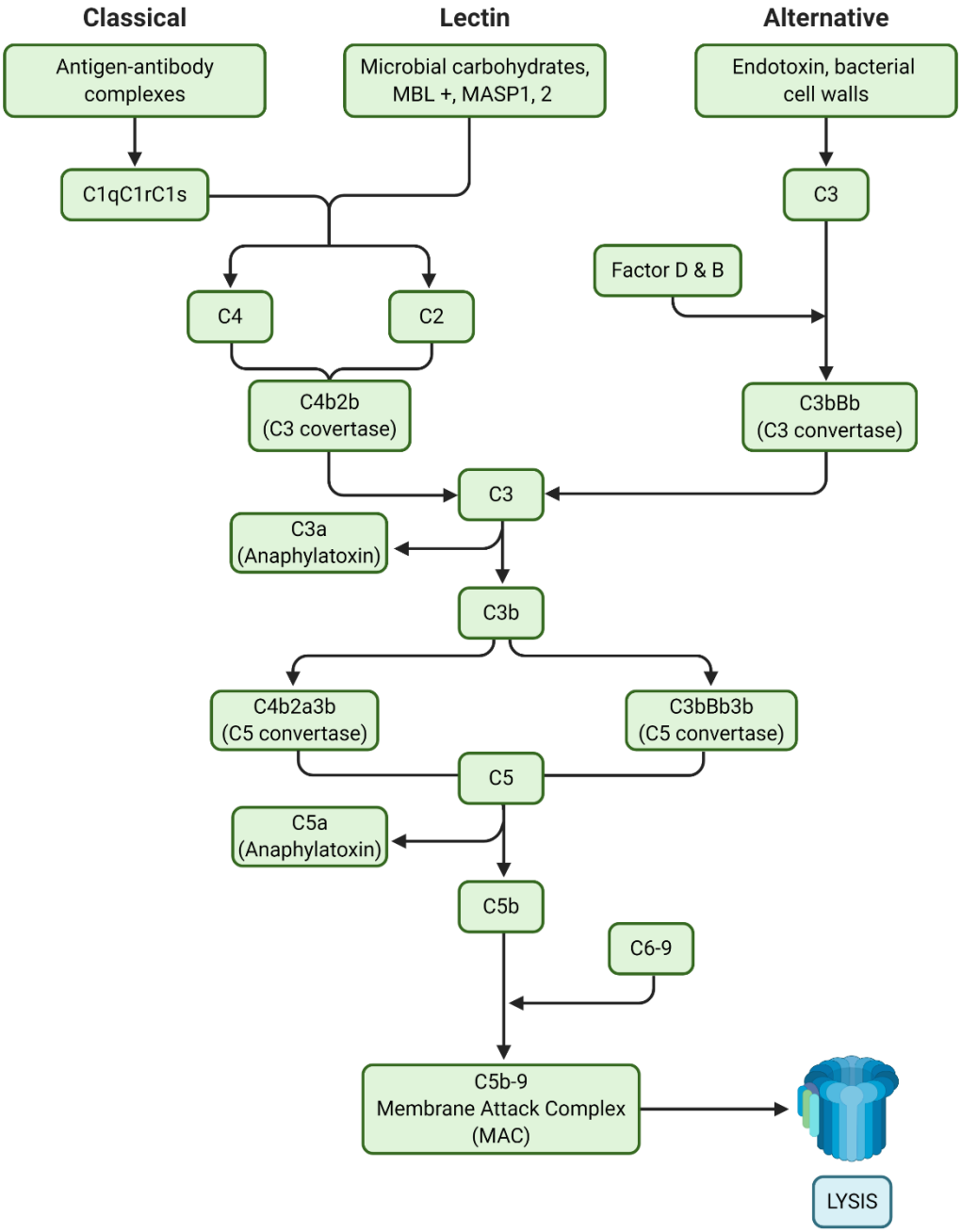


Figure 1.14. Overview of the complement activation pathways. Complement activation via the three pathways; classical, lectin and alternative. Made in BioRender.

As the complement pathways provide early protection against infection, it is vital for invading bacteria to be able to evade this innate immune response. Bacteria employ various mechanisms that target different stages of the cascade, in order to evade complement mediated killing. The alternative pathway is continually active and evasion of killing via this pathway involves utilisation of the complement regulatory proteins that protect the host's own cells, such as human factor-H (FH) and factor-H like-1 (FHL-1). Factor-H is a 155 kDa

glycoprotein found in human plasma and it is composed of 20 short consensus repeat (SCR) domains, each containing ~60 residues (figure 1.15) (Ripoche et al., 1988). Factor-H modulates the complement response by binding to the major complement component C3b via the SCR domains 1-4 (Gordon et al., 1995). This complex of FH and C3b recruits factor I, a protease which cleaves C3b thereby inactivating it. FH also regulates formation of C3b convertases as it competes with factor B to bind to C3b and can accelerate decay of existing C3b convertases, likely by electrostatic repulsion of factor B (Wu et al., 2009).

Domains 19-20 are utilised for attachment to cell surfaces and are important for differentiation between self and non-self cells, with SCR 19 mainly binding deposited C3b to allow complement regulation and SCR 20 simultaneously binding glycosaminoglycans, which are not found on bacterial cells (Kajander et al., 2011). Domains 19 and 20 are also the binding site for microbial attachment to FH, with bacteria able to bind domain 20 causing an enhancement in C3b binding by domain 19 and a downregulation of the complement pathway (Meri et al., 2013).

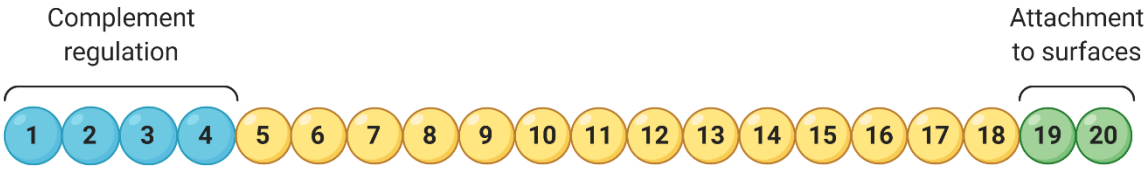


Figure 1.15. The SCR domains of human complement factor-H. Human factor-H is composed of 20 short consensus repeat (SCR) domains. The major functional regions of the protein are highlighted, with SCR 1-4 a main C3b binding site for complement regulation and SCR 19-20 important for attachment of factor-H to host cells. Adapted from (Kopp et al., 2012) and created in BioRender.

In *Borrelia*, the outer surface proteins aid not only in infectivity and dissemination, but also in immune evasion. One of the major *Borrelia* outer surface proteins involved in the evasion of the alternative complement pathway via FH binding is OspE. OspE is upregulated in response to a temperature shift during a blood meal and forms a tripartite complex with FH, binding to FH domain 20 with C3b simultaneously bound to domain 19 (Bhattacharjee et al., 2013; Hefty et al., 2002; Kolodziejczyk et al., 2017) (figure 1.16). This results in complement downregulation. The five *Borrelia* CRASPs (section 1.3.2) are also involved in evasion of the complement response, binding to various members of the FH family; factor H (FH), factor H

like-1 (FHL1) and complement factor H-related proteins (CFHR) 1-5. As well inhibiting the complement response through FH and FHL1 binding, CRASP-1 binds C7 and C9, inhibiting the formation of the membrane attack complex (Hallström et al., 2013). The binding partners of CRASPs from *B. burgdorferi* B31 are summarised in table 1.5. While *Borrelia* CRASPs all have a role in evasion of complement mediated killing, they are upregulated at different stages of transmission and infection, suggesting separate roles for these proteins (Bykowski et al., 2007; Lin et al., 2020).

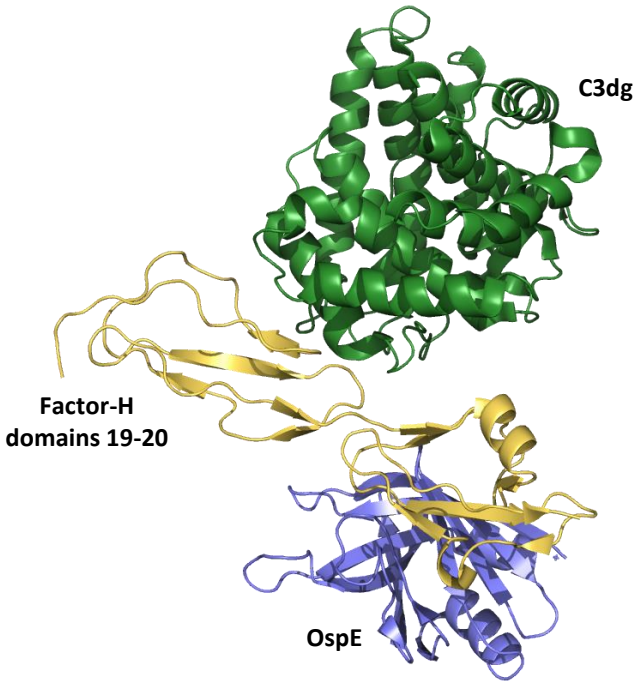


Figure 1.16. The crystal structure of OspE, FH domain 19-20 and C3dg (PDB: 5NBQ). OspE is shown in purple, factor-H domains 19-20 are shown in yellow and C3dg is shown in green. The image is shown as a cartoon representation and was produced using PyMol (Schrödinger, 2010).

	CRASP-1	CRASP-2	CRASP-3	CRASP-4	CRASP-5
Alternative names	BbCRASP-1 CspA BBA68	BbCRASP-2 CspZ BBH06	BbCRASP-3 ErpP BBN38 OspE	BbCRASP-4 ErpC	BbCRASP-5 ErpA BBP38 BBL39
Gene name	*cspA*	*cspZ*	*erpA*	*erpC*	*erpP*
Gene location in *B. burgdorferi B31*	lp54	lp28-3	cp32-1 cp32-5 cp32-8	cp32-2	cp32-9
Gene expression in unfed tick	No	No	No	No	No
Gene expression in feeding tick	Yes	No	Yes	Yes	Yes
Gene expression in skin at tick feeding site	Yes (high)	Yes (low)	Yes	Yes	Yes
Gene expression in mice/ disseminated infection	No	Yes (high)	Yes	Yes	Yes
Complement regulator binding	CFH FHL1	CFH FHL1	CFH CFHR1 CFHR2 CFHR5	CFHR1 CFHR2	CFH CFHR1 CFHR2 CFHR5
PDB code	1W33	4CBE	4BOB	4BXM	

Table 1.5. BbCRASP expression and FH binding partners. A summary of BbCRASPS 1-5, highlighting which stages of transmission and infection they are expressed in and their FH binding partners. Adapted from (Kraiczy & Stevenson, 2013) and (Lin et al., 2020).

Borrelia is also able to evade the lectin pathway of complement mediated killing during the early stages of infection. This is via TSLPI (Tick Salivary Lectin Pathway Inhibitor), a tick protein which inhibits mannose binding lectin from binding to its ligand, preventing activation of the lectin complement pathway (de Taeye et al., 2013; Schuijt et al., 2011). When this protein was added to serum sensitive strains of *B. garinii*, a decrease in complement mediated killing was seen, highlighting the role of TSLPI and inactivation of the lectin pathway in overall complement evasion (de Taeye et al., 2013).

1.5 Research aims

Borrelia genomes encode many lipoproteins (>100) and these have been the focus of much of the research into *Borrelia* and Lyme disease, leading to characterisation of many of these proteins and an understanding of their role in infection and immune evasion (Kenedy et al., 2012). These lipoproteins are often encoded on the plasmids which vary between *Borrelia* genospecies and lipoproteins such as OspA have been the main candidate for vaccine development. Many of these lipoproteins form large paralogous gene families (such as the Erps), potentially hindering the effectiveness of any vaccine that targets one particular variant. In contrast to other Gram negative bacteria such as *E. coli*, the *B. burgdorferi* OM contains a low abundance of integral β-barrel outer membrane proteins and these are encoded on the stable chromosome (Fraser et al., 1997). Very few integral OMPs have been characterised in *B. burgdorferi*, with previous research exclusively exploring the *Borrelia* chromosome for identification of novel OMPs using a computational framework. This project aims to expand on this work and explore the plasmid proteomes of *Borrelia* to search for β-barrel genes.

Predicted OmpA-like integral membrane proteins have been identified in *B. burgdorferi s.l* using a Hidden Markov Model (HMM) approach (Dyer et al., 2015) and these are summarised in table 1.6. Previous research focused on a single full length OM protein from *B. afzelii*, potentially limiting the ability to crystallise this protein (Dyer et al., 2015). The aim of this research is to produce a series of truncation of paralogues in this gene family, to try to increase the chance of successful crystallisation.

B. burgdorferi s.s	***B. afzelii***	***B. garinii***
BB_0027	BAPKO_0026	BG_0027
BB_0405	BAPKO_0422	BG_0407
BB_0406	BAPKO_0423	BG_0408
BB_0562	BAPKO_0591	BG_0572

Table 1.6. Potential OmpA-like proteins in *Borrelia*. Gene names of orthologues of potential OmpA-like proteins in three main *Borrelia* species are given. Proteins were identified using Hidden Markov Model (HMM) searches and analysis of signal sequences (Dyer et al., 2015).

BB_0027 and BB_0562 remain uncharacterised and while some information regarding the function of BB_0405 and BB_406 has been established, a high resolution structure is not available for any of these proteins. Further research to try and elucidate the structure and function of these proteins could develop our understanding of *Borrelia* pathogenesis and perhaps identify more proteins with a role in infection, dissemination and immune evasion. These proteins could provide further targets for diagnostic immunoblots and may serve as novel vaccine targets. Therefore, the aim of this work is to clone, express and purify both full length and a series of truncated recombinant versions of BB_0027, BB_0405, BB_0406 and BB_0562. As crystallography of membrane proteins presents many challenges, this research also aims to identify, clone and express a soluble protein target to increase the chance of obtaining a protein crystal structure.

Summary of research aims- *Borrelia* outer membrane proteins (OMPs):

- Using protein secondary structure prediction algorithms, design full-length and truncated constructs for four *B. burgdorferi s.s* OmpA-like proteins; BB_0027, BB_0405, BB_0406 and BB_0562.
- Identify a potentially soluble target for crystallisation experiments.
- Clone multiple OMP and soluble target constructs using Ligation Independent Cloning (LIC) techniques.
- Produce recombinant OMPs for structural and functional studies.
- Attempt protein crystallisation of truncated OMPs and a soluble protein target.
- Apply a computational framework approach to identify any putative OMPs in the *Borrelia* plasmid proteome.

Chapter 2 Materials and methods

2.1 Materials

2.1.1 DNA and primers

B. burgdorferi B31 DNA isolated from *Ixodes dammini* (DSM No. 4680) was purchased from Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures (DSMZ) and was replicated using the REPLI-g Mini Kit (Qiagen) according to the manufacturer's instructions to use as a template for PCR. Primers were designed to include extensions for Ligation Independent Cloning (table 2.1) and primer parameters were analysed using the Sigma Aldrich OligoEvaluator™. Primer sequences were produced by Eurofins Scientific and were provided as lyophilised samples, which were then diluted to 100 pmol/µL using nuclease free water.

Target	Primer Name	Primer Sequence (5'-3')
BB0562	BB0562-f0	**TACTTCCAATCCATG**AAAAAAATTTTTATATTG
	BB0562-f1	**TACTTCCAATCCATG**ACAAATGGTTTTACAAAAG
	BB0562-f2	**TACTTCCAATCCATG**AGAGGAATTGGCTTTGGAGC
	BB0562-f3	**TACTTCCAATCCATG**CCAATTATTAACTTAATAATGTC
	BB0562-r0	**TATCCACCTTTACTGTCA**AAAGATATAGTACTTAGC
BB0027	BB0027-f0	**TACTTCCAATCCATG**AGAAAGTATATTTTTATAATAC
	BB0027-f1	**TACTTCCAATCCATG**AAGTATATTTTTATAATAC
	BB0027-f2	**TACTTCCAATCCATG**AACATAAAAAAAATTGC
	BB0027-f3	**TACTTCCAATCCATG**AACTCCACTTTAGGAATAG
	BB0027-r0	**TATCCACCTTTACTGTCA**CAATTTATTTACATTCACTG
BB0405	BB0405-f0	**TACTTCCAATCCATG**AGAATGCTATTAGCAAC
	BB0405-f1	**TACTTCCAATCCATG**TCCAAAAGCAAAAGTATGACTG
	BB0405-f2	**TACTTCCAATCCATG**GACTTTGATTTTGATAAACTTC
	BB0405-f3	**TACTTCCAATCCATG**CGTTTATTTGGCATAGGTTTTGG
	BB0405-r0	**TATCCACCTTTACTGTCA**TATATATATTTTTATAAAGC
BB0406	BB0406-f0	**TACTTCCAATCCATG**ATAAAAATTTTTAAAAAAATATAC
	BB0406-f1	**TACTTCCAATCCATG**TCTTTTGCATCTGACAATTATATGG
	BB0406-f2	**TACTTCCAATCCATG**AAGCTTAAAGAAATAAAAG
	BB0406-f3	**TACTTCCAATCCATG**GACTTATTTTCAATGGGCATTGG
	BB0406-r0	**TATCCACCTTTACTGTCA**ATTAAATGCAAATTTTATG
BB0095	BB0095-f0	**TACTTCCAATCCATG**TTAGATTCATATTATTATG
	BB0095-f1	**TACTTCCAATCCATG**TCATATTATTATGTTTTATCTTC
	BB0095-f2	**TACTTCCAATCCATG**AGTGTTTCAGATTTTTTAAATAATG
	BB0095-f3	**TACTTCCAATCCATG**AGTAAAAAGGACTTCAATTTTTTAAAAG
	BB0095-r0	**TATCCACCTTTACTGTCA**AACTTTTTCATTAGTTTTATC
	BB0095-r1	**TATCCACCTTTACTGTCA**AAGATCAGTTAAAAATTGC

	BB0095-r2	**TATCCACCTTTACTGTCA**AAACAAATTATGCCTTAAAAATAACC
BB0040	BB0040-f0	**TACTTCCAATCCATG**AACACGAATCAAAACAAATTCAACC
	BB0040-f1	**TACTTCCAATCCATG**ACTAAGGACGAACTTTCAAG
	BB0040-r0	**TATCCACCTTTACTGTCA**ATTCTTTTTATAGGTGGCAGG
PCNA	PCNA for	**TACTTCCAATCCATG**TTCGAGGCGCGCCTGGTCC
	PCNA rev	**TATCCACCTTTACTGTCA**AGATCCTTCTTCATCCTCGATC
Vector	pLIC for	TGTGAGCGGATAACAATTCC
	pLIC rev	AGCAGCCAACTCAGCTTCC

Table 2.1. Details of primers used for Ligation Independent Cloning. The name of the target is given in the first column, followed by the name given to the primer. The primer sequence is shown from 5' to 3' and extensions added to each primer for Ligation Independent Cloning are highlighted in bold.

Two pET28-a based vectors were used for cloning, pNIC28-Bsa4 and pNH-TrxT, both of which were kindly provided by Dr Chris Cooper at The University of Huddersfield. Both vectors contain a T7 promotor and terminator, followed by a region coding for just a 6x Histidine tag in the pNIC28-Bsa4 vector and a 6x Histidine-Thioredoxin tag for the pNH-TrxT vector. Expression tag regions are followed by a TEV protease cleavage site to allow for tag removal and both vectors contain kanamycin resistance. Full vector maps can be seen in appendix 1.

2.1.2 Reagents and buffers

Unless stated otherwise, chemicals and reagents used throughout this research were purchased from Fisher Scientific (Loughborough, UK) or Sigma-Aldrich (Gillingham, UK) (table 2.2). Kits required for replication or purification of DNA were purchased from Qiagen or Bioline. Reagents and kits for crystallography experiments were purchased from Molecular Dimensions.

Reagent/Chemical	Abbreviation	Supplier
Acrylamide (40% solution)	-	Fisher
Agar	-	Fisher
Ammonium persulfate	APS	Fisher
Benzamidine	-	Sigma
Blue dextran	-	Sigma
β-mercaptoethanol	βME	Sigma
Calcium chloride dihydrate	$CaCl_2$	Fisher
Chloramphenicol	Chl	Fisher
Lauryldimethylamine oxide	LDAO	Sigma
n-Dodecyl-B-D-Maltoside	DDM	Molecular Dimensions
Dithiothreitol	DTT	Sigma

Ethylenediaminetetraacetic acid	EDTA	Sigma
Glycerol	-	Fisher
HEPES	HEPES	Fisher
Imidazole	-	Fisher
Isopropyl β-D-1-thiogalactopyranoside	IPTG	Fisher
Kanamycin	Kan	Fisher
Lysozyme	-	Fisher
Manganese (II) chloride	MnCl2	Sigma
Sodium chloride	NaCl	Fisher
Nickel	Ni	Fisher
Phenylmethylsulfonyl fluoride	PMSF	Fisher
Potassium acetate	-	Fisher
Rubidium chloride	RbCl	Fisher
Sodium dodecyl sulfate	SDS	Fisher
Sucrose	-	Fisher
SYBR safe	-	Sigma
Tetramethylethylenediamine	TEMED	Fisher
Tris (hydroxymethyl)aminomethane-base	Tris base	Fisher
Triton X-100	-	Fisher
Tryptone	-	Fisher
Tween-20	-	Sigma
Urea	-	Fisher
Yeast extract	-	Fisher

Table 2.2. General reagents used throughout this research project. The name of the reagent is given in the first column, followed by the common abbreviation of the reagent if applicable. The final column gives the name of the supplier the reagent was purchased from.

Buffer/Solution	Components
5x Lysis Buffer	2.5 M NaCl, 250 mM HEPES, 25% (v/v) glycerol, pH 7.5
Coomassie Destain	20% (v/v) methanol, 10% (v/v) glacial acetic acid
Coomassie Stain	40% (v/v) methanol, 10% (v/v) glacial acetic acid, 0.1% (w/v) Coomassie R250
Denaturing Buffer	50 mM Tris, 300 mM NaCl, 8 M urea, pH 8.0
Dialysis Buffer	20 mM HEPES, 500 mM NaCl, 5% (v/v) glycerol, pH 7.5
Elution Buffer	1/5 5x Lysis Buffer, 300 mM imidazole, 4 mM βME
Lysis Buffer 0	1/5 5x Lysis Buffer, 10 mM imidazole, 4 mM βME
Lysis Buffer 1	1/5 5x Lysis Buffer, 10 mM imidazole, 4 mM βME, 1 mM PMSF, 1 mM benzamidine
Lysogeny Broth (LB)	1% (w/v) NaCl, 1% (w/v) tryptone, 0.5% (w/v) yeast extract
OMP Elution Buffer	50 mM Tris, 300 mM NaCl, 300 mM imidazole, 0.1% LDAO, pH 8.0
OMP Lysis Buffer	50 mM Tris, 300 mM NaCl, 0.1% Triton X-100, pH 8.0
OMP SEC Buffer 1	50 mM Tris, 300 mM NaCl, 0.1% (w/v) DDM, pH 8.0
OMP SEC Buffer 2	50 mM Tris, 300 mM NaCl, 0.1% (v/v) LDAO, pH 8.0
OMP Wash Buffer	50 mM Tris, 300 mM NaCl, 20 mM imidazole, 0.1% LDAO, pH 8.0

Pellet Wash Buffer 1	50 mM Tris, 300 mM NaCl, 1 mM EDTA, 4 mM βME, 0.5% (v/v) Triton X-100, pH 8.0
Pellet Wash Buffer 2	50 mM Tris, 300 mM NaCl, 0.5% (v/v) Triton X-100, pH 8.0
Pellet Wash Buffer 3	50 mM Tris, 300 mM NaCl, pH 8.0
Refolding Buffer	50 mM Tris, 300 mM NaCl, 1 M urea, 0.1% LDAO, pH 8.0
SEC Buffer	20 mM HEPES, 500 mM NaCl, 5% (v/v) glycerol, pH 7.5
Super Optimal Broth (SOB)	1.5% (w/v) yeast extract, 1% (w/v) tryptone, 10 mM NaCl, 2 mM KCl
TBS Buffer	Tris base, NaCl, pH 7.6
TBST Buffer	Tris base, NaCl, 0.1% Tween-20, pH 7.6
TFB1	30 mM potassium acetate, 100 mM RbCl, 10 mM $CaCl_2$., 50 mM $MnCl_2$, 15% (v/v) glycerol, pH 5.85
TFB2	10 mM MOPS free acid, 10 mM RbCl, 75 mM $CaCl_2$, 15% (v/v) glycerol, pH 6.5
Transfer Buffer	10% (v/v) methanol, 1x Bolt Transfer Buffer
Wash Buffer	1/5 5x Lysis Buffer, 30 mM imidazole, 4 mM βME

Table 2.3. Buffers and solutions used throughout this project. The name given to the buffer/solution is given in the first column, followed by details of the components of the buffer/solution in the second column.

2.2 Methods

2.2.1 Identification of plasmid encoded β-barrel proteins

2.2.1.1 Filtering of sequences using a computational framework

The plasmid proteomes of the *Borrelia burgdorferi s.l* complex (table 2.4) were obtained from Uniprot and a computational framework was used to predict putative OMPs. The framework used was based on previous methodology used on proteomes of *Borrelia* and *Treponema* (Cox et al., 2010; Kenedy et al., 2016). A custom R-script (written by Dr Jarek Byrk at The University of Huddersfield, appendix 5) was used to analyse and filter sequences based on the results from prediction algorithms. A total of 5923 protein sequences from the *B. burgdorferi s.l* plasmid proteomes were analysed. SpLip (Setubal et al., 2006) and LipoP (Juncker et al., 2003) were used to identify potential lipoproteins. Proteins were categorised as either 'not lipoprotein', 'possible lipoprotein' or 'probable lipoprotein' by SpLip and any proteins considered 'probable' or 'possible' lipoproteins were excluded. Remaining proteins were then analysed using LipoP, which identifies potential transmembrane helical proteins and cleavage sites for Signal Pepidase I or II. If these are not identified, the protein is classified as cytoplasmic. Sequences with a Signal Pepidase II site were excluded and proteins with a Signal Peptidase I site, transmembrane helical proteins and those considered cytoplasmic were subject to further analysis.

TMHMM-2.0 (Krogh et al., 2001) was used to predict inner membrane proteins and only proteins with zero predicted transmembrane spanning regions were analysed further. To predict cytoplasmic proteins, PSORTb 3.0 (Yu et al., 2010) was used and separated proteins into 4 categories; 'cytoplasmic', 'cytoplasmic membrane', 'outer membrane' and 'unknown'. Predicted 'cytoplasmic' proteins were removed and all others continued to the next stage of analysis. SignalP 5.0 (Armenteros et al., 2019) was used for prediction of signal peptidase I-type signal sequences and only proteins with a predicted signal sequence were retained. The final analysis stage used PRED-TMBB2 (Tsirigos et al., 2016) to predict potential β-barrel membrane proteins, where a score between 0 to 1 is given, with >0.43 a positive result.

Organism (strain)	Proteome ID	Taxon ID	Number of plasmid genes
B. afzelii **(PKo)**	UP000005216	390236	566
B. bavariensis **(ATCC BAA-2496 / DSM 23469 / PBi)**	UP000002276	290434	430
B. bissettii **(CO275)**	UP000183624	64897	0
B. bissettii **(DN127)**	UP000001634	521010	587
B. burgdorferi **(297)**	UP000243802	521009	453
B. burgdorferi **(64b)**	UP000006162	498740	684
B. burgdorferi **(ATCC 35210 / B31 / CIP 102532 / DSM 4680)**	UP000001807	224326	512
B. burgdorferi **(ZS7)**	UP000006901	445985	359
B. chilensis **(VA1)**	UP000030940	1245910	93
B. finlandensis **(SV1)**	UP000006166	498741	409
B. garinii **(PBr)**	UP000006103	498743	515
B. japonica **(ATCC 51557)**	UP000199262	34095	0
B. mayonii **(MN14-1539)**	UP000185492	1674146	328
B. spielmanii **(A14S)**	UP000003481	498742	323
B. valaisiana **(VS116)**	UP000006163	445987	517
Candidatus B. tachyglossi **(Bc-F10-1268)**	UP000244655	1964448	147

Table 2.4. Non-redundant proteomes of *B. burgdorferi s.l.* Plasmid information was obtained from UniProtKB.

2.2.1.2 Phylogenetic analysis of BBJ25 homologs

BBJ25 from *B. burgdorferi* B31 was used as a query sequence for PSI-BLAST searches (NCBI) and two iterations were used to yield 91 full-length sequences. Expasy Decrease Redundancy was then used and sequences with >95% identity to other sequences were removed. The subsequent 30 sequences were aligned using MUSCLE (Edgar, 2004) and a phylogenetic tree was generated in MEGAX using the Maximum Likelihood method and a JTT matrix-based model. The JTT+G model was chosen based on a ranking of 60 models by prottest3 (Darriba et al., 2011). A bootstrap method was applied using 1000 replications.

2.2.1.3 Phylogenetic analysis of the BBJ25 operon

Genomic DNA corresponding to the predicted operon were extracted from the NCBI Nucleotide database. The precise region was determined by the first/last codon of the first/last gene (either BBJ23 to BBJ29 for Group 1 or BBJ23 to BBJ25 for Group 2). Reverse

Complement (Stothard, 2000) was used as necessary to convert all sequences to the forward strand.

Accession Code	Region	Species	Plasmid Name
CP001309.1	13943-20412	*B. garinii PBr*	PBr_lp17
CP002258.1	13122-19592	*B. burgdorferi 297*	297_lp38
AE000787.1	16520-23008	*B. burgdorferi B31*	lp38
NC_012185.1	55156-61633	*B. valaisiana VS116*	VS116_lp28-3
NZ_QBLM01000004.1	31228-37699 (reverse)	*B. turdi TPT2017*	lp30 contig4
CP005719.1	2259-9343	*B. hermsii YBT*	Contig0014
CP005838.1	1-8253 (reverse)	*B. anserina BA2*	Contig0010
CP028884.1:	947190-953518	*B. turcica IST7*	chromosome
CP001442.1	5941-11964 (reverse)	*B. valaisiana VS116*	VS116_lp28-8
NZ_QBLN01000011.1	6179-12518	*B. turdi strain T1990A*	contig11
CP002947.1	5392-11724 (reverse)	*B. afzelii PKo*	lp28-8
CP001465.1	9552-15884 (reverse)	*B. spielmanii A14S*	lp28-8
FMTE01000008.1	10240-16576	*B. japonica ATCC 51557*	contig: Ga0052881_scaffold00007.7
NZ_CP025786.1	76006-78423 (reverse)	*Candidatus B. tachyglossi Bc-F10-1268*	pl78
MDCO01000010.1	171623-177612	*Brachyspira hampsonii P280/1*	chromosome

Table 2.5. GenBank accession codes and genomic regions used for phylogenetic analysis of the putative operon.

DNA sequences were aligned using MUSCLE (Edgar, 2004) using the default parameters in MEGAX (penalties for gap open -400, gap extend 0, maximum iterations 16, cluster method UPGMA, Max Diagonal Length 16). The sequence alignment was exported in NEXUS format (number of sequences: 15, number of sites: 10158). In order to use Maximum Likelihood algorithms, jModelTest (Posada, 2008), was used to select a nucleotide substitution model. The initial base tree was inferred by Maximum Likelihood. The models were then ranked according to Akaike Information Criterion (AIC) (Akaike, 1974), and Akaike Information Criterion corrected for small sample sizes (AICc) (Hurvich & Tsai, 1989). Phylogenetic trees were inferred using the bootstrap method (1000 replicates) and the General Time Reversible Model (GTR) with Gamma Distributed rates among sites (number of discrete gamma categories = 5) using MEGAX (Kumar et al., 2018). Neighbouring genes and their locations were identified from complete genome sequences in GenBank (NCBI) and displayed using the gggenes package in R (R Core Team, 2021). Predicted operons were identified using Operon-Mapper (Taboada et al., 2018).

2.2.2 Secondary structure prediction

DNA sequences of genes of interest were obtained from the National Centre for Biotechnology Information (NCBI) database and were translated using the ExPASy Translate tool. Protein sequences were first subject to protein BLAST analysis to identify any similar sequences in the protein databank. The NCBI BLAST tool was used for BLAST searches and default algorithm parameters were used, with BLOSUM62 as the scoring matrix. Outer membrane protein sequences were then analysed using SignalP 5.0 to identify signal sequence sites for removal when designing constructs for recombinant expression in inclusion bodies. Gram negative was selected as the organism group and default parameters were used. All target sequences were then submitted to the secondary structure prediction server PSI-PRED to identify any potential structural elements and disordered regions, which was essential for designing truncations where sequences shared low homology to known sequences in the Protein Data Bank (PDB).

2.2.3 Preparation of competent *E. coli*

E. coli cells were streaked onto LB agar plates (no antibiotics) from either a Mach1 (Invitrogen™) or Rosetta glycerol stock and were grown overnight at 37°C. A single colony was used to inoculate 5 mL of SOB and the culture was grown overnight at 37°C/250 rpm. The following day, 200 mL of SOB medium was inoculated with 5 mL of overnight culture and was grown until an OD_{600} of 0.45-0.55 was reached. Aliquots of culture were dispensed into centrifuge tubes and cells were centrifuged at 2500 g for 5 minutes. The supernatant was discarded and the pellets were gently resuspended in ice cold TFB1 and centrifuged as before. The buffer was discarded and the cells were resuspended in ice cold TFB2 and split into 500 µL aliquots, which were frozen in liquid nitrogen.

2.2.4 High throughput Ligation Independent Cloning (LIC)

Ligation Independent Cloning (LIC) techniques were used to allow a high-throughput approach to produce multiple constructs. A graphical summary of the process can be seen in figure 2.1.

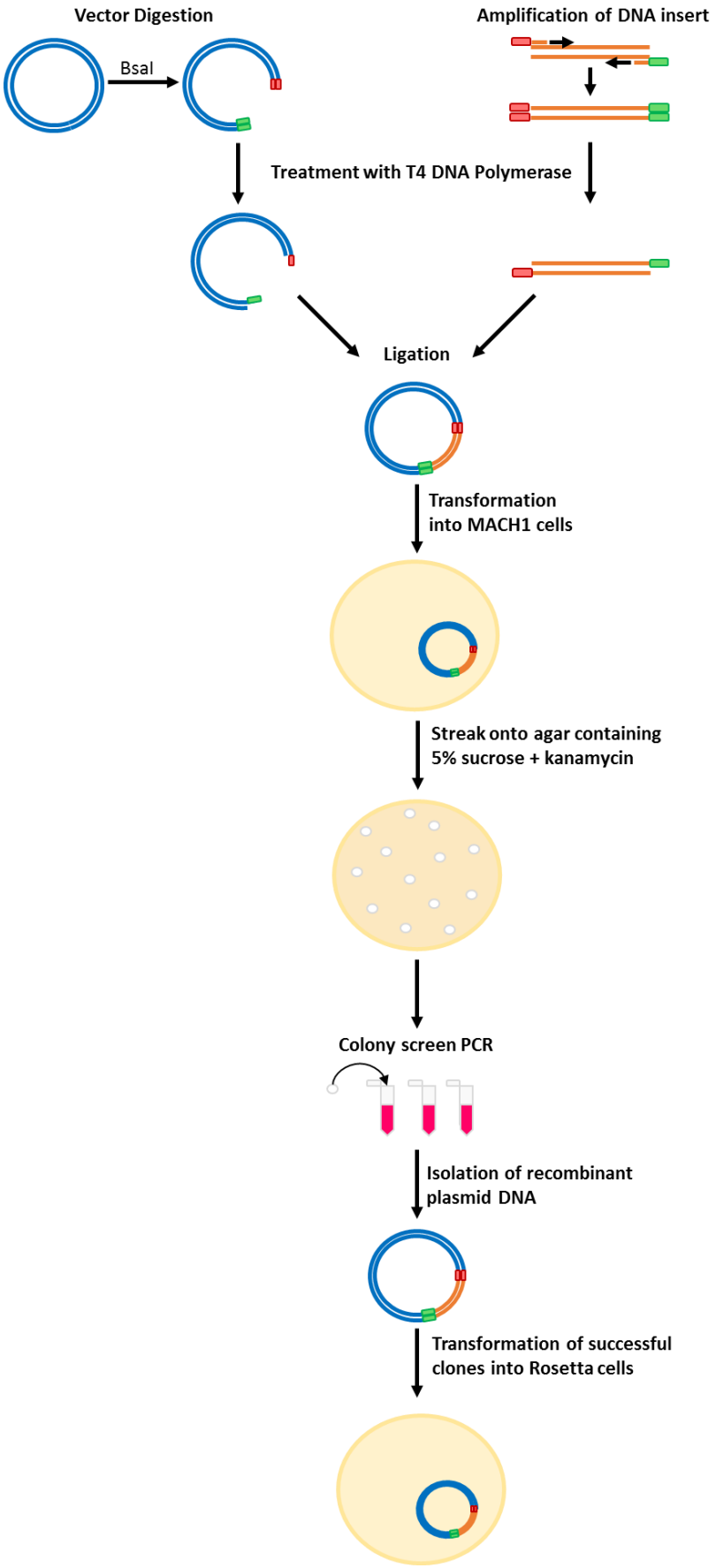


Figure 2.1. Schematic diagram of the Ligation Independent Cloning method.

2.2.4.1 PCR amplification of genomic DNA

The appropriate forward and reverse primers for each construct (table 2.1) were mixed to a concentration of 2.5 µM each. A master mix was prepared consisting of the following; 5x Phusion HF Buffer, dNTP set (Bioline) mixed to a final concentration of 10 mM, sterile nuclease free H_2O (Invitrogen), Phusion® High-Fidelity DNA Polymerase (New England Biolabs®), template DNA (30 ng/µL). 1x master mix was added to each tube, followed by primer mix. Standard PCR was carried out in a thermocycler using the following conditions: 94°C for 5 minutes, 25 cycles at [94°C for 30 seconds; 50°C for 30 seconds; 68°C for 1 minute], 72°C for 10 minutes. PCNA was used as the template DNA for the positive control and the negative control substituted the DNA for sterile nuclease free water. For optimisation where a low yield was obtained, PCR was carried out using 30 ng/µL of the *B. burgdorferi* DNA from DSMZ as a template, rather than the genomic DNA obtained from the replication kit.

2.2.4.2 Preparation of vectors and inserts

Digestion of vectors was carried out by mixing sterile nuclease free water (Invitrogen), CutSmart® Buffer (New England Biolabs®), 5 µg of plasmid and *Bsa*I restriction enzyme (New England Biolabs®). The mixture was incubated at 50°C for 2 hours. Both the amplified inserts and digested plasmids were then purified using the QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's instructions. Following purification, the *Bsa*I- digested vectors and the inserts were treated with T4 DNA polymerase to create cohesive ends. The mixture for the inserts contained: 10x NEB Buffer 2.1 (New England Biolabs®), 25 mM dCTPs, 100 mM DTT, sterile nuclease free water (Invitrogen), purified insert and T4 DNA Polymerase (New England Biolabs®). The vector mixture was the same except the dCTPs were substituted for dGTPs. The mixtures were then incubated in a thermocycler at 22°C for 30 minutes and then 75°C for 20 minutes.

2.2.4.3 Ligation and transformation into Mach1 competent *E. coli*

For ligating the T4 treated vectors and inserts, a ratio of 1:2 was used of vector to insert and the mixture was incubated at room temperature for 30 minutes. 30 µL of Mach1 *E. coli* cells (Invitrogen) were then added to the plasmid/insert mixture and this was incubated on ice for 30 minutes. A heat shock was carried out in a water bath at 42°C for 45 seconds and the tubes

were placed back on ice for 2 minutes. 90 µL of 2x LB was added to the tubes and this was incubated at 37°C, 200 rpm for at least 60 minutes. 100 µL of each transformation was then spread onto LB agar plates containing 5% sucrose and kanamycin (50 µg/mL) and these were incubated overnight at 37°C.

2.2.4.4 Colony screen PCR

Colonies present on sucrose selection plates were screened for the presence of the correct construct. The reaction mixture was prepared using 5x MyTaq™ Reaction Buffer Red (Bioline), DMSO, sterile nuclease free H_2O (Invitrogen), 10 µM pLIC primer mix and MyTaq™ Red DNA Polymerase (Bioline). A colony was picked from the plate using a sterile pipette tip and was dipped into the reaction mixture before being used to inoculate 5 mL of LB containing kanamycin (50 µg/mL). The thermocycler was pre-heated to 95°C and the colony PCR was carried out using the following conditions: 95°C for 10 minutes, 30 cycles at (95°C for 30 seconds; 50°C for 30 seconds; 72°C for 1 minute), 72°C for 5 minutes. Cultures were grown at 37°C/180 rpm overnight.

2.2.4.5 Plasmid miniprep and transformation into Rosetta cells

Cells containing successful clones identified through DNA agarose gel analysis (section 2.2.4.7) were used to prepare overnight cultures for a plasmid miniprep. The plasmids were isolated using the ISOLATE II Plasmid Mini Kit (Bioline) according to the manufacturer's instructions. Purified plasmids were then transformed into *E. coli* Rosetta cells using methods described in section 2.2.4.3 and were grown on LB agar plates containing kanamycin (50 µg/mL) and chloramphenicol (34 µg/mL).

2.2.4.6 DNA quantification

To assess the concentration and purity of the DNA samples, a NanoDrop™ 2000/2000c Spectrophotometer (Thermo Scientific™) was used. Water was used as a blank and then 2 µL of sample was loaded on to the pedestal. The concentration given was noted and the A_{260}/A_{280} value was used to assess purity.

2.2.4.7 DNA agarose gel electrophoresis

DNA agarose gel electrophoresis was used to analyse PCR products. 1x TBE buffer was prepared and a 1.5% agarose gel was poured containing SYBR® Safe DNA Gel Stain (Invitrogen™). Samples were diluted with ddH_2O and 6x DNA Gel Loading Dye (New England Biolabs®) was added. Colony screen samples were loaded directly onto the gel. The gel was ran at 80 V for 1 hour in 1x TBE buffer.

2.2.5 Small scale protein expression

Successful clones identified from DNA agarose gel electrophoresis of colony screen samples were transformed into *E. coli* Rosetta cells and streaked on LB agar containing kanamycin (50 µg/mL) and chloramphenicol (34 µg/mL). Colonies were scraped from the plate and used to inoculate 5 mL of LB containing the appropriate antibiotics. Cultures were grown overnight at 37°C/180 rpm in an orbital shaker. The following day, 50 mL of LB (kanamycin, 50 µg/mL) was inoculated with 500 µL of overnight culture and grown at 37°C/180 rpm until an OD_{600} of ~0.6 was reached. Cells were then induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a working concentration of 1 mM.

Protein expression was carried out overnight at 18°C/180 rpm and cells were then pelleted using centrifugation at 2500 g, 4°C for 30 minutes. The supernatant was discarded and the pellets resuspended in 3 mL of Lysis Buffer 1. To aid lysis and remove contaminating DNA, 0.5 mg/mL lysozyme was added along with Thermo Scientific™ Pierce™ Universal Nuclease for Cell Lysis and samples were incubated at 4°C for 1 hour. The cells were placed on ice and further disrupted by pulsed sonication at an amplitude of 50%, 5 seconds on 10 seconds off for 2 minutes. Lysates were then split into microfuge tubes and centrifuged in a benchtop centrifuge at 13,500 rpm, 4°C for 30 minutes. Following centrifugation, the supernatant was retained as the soluble fraction and the insoluble pellet resuspended in Denaturing Buffer for SDS-PAGE analysis. The soluble fraction was then transferred to a 15 mL conical tube and diluted 2-fold with Lysis Buffer 1 to reduce viscosity.

2.2.6 Soluble protein screening

For equilibration and to remove residual ethanol, 1 volume of Nickel-Nitrilotriacetic acid (Ni-NTA) beads were centrifuged at 2500 g for 2 minutes and the ethanol poured off. The beads were then resuspended in 1 volume of ddH_2O and centrifuged as before. This process was repeated a further 2 times. Finally, 1 volume of Lysis Buffer 0 was added to the beads and centrifuged as before 3 times.

500 µL of equilibrated 50% Ni-NTA beads were added to the soluble fractions and the tubes were incubated at 4°C/100 rpm overnight to bind the protein to the beads. The following day the beads were pelleted by centrifuging samples at 2500 g, 4°C for 5 minutes and the supernatant was poured off. The soft pellet was then resuspended in 1 mL of Lysis Buffer 1 and transferred to a microfuge tube. This was incubated for 10 minutes as an initial wash step and centrifuged at 2500 g for 2 minutes, with the supernatant removed following centrifugation. A further 5 washes were performed by resuspending the beads in 1 mL of Wash Buffer and centrifuging as before. To elute the bound protein, 200 µL of Elution Buffer was added to each pellet and incubated for 30 minutes at 4°C/100 rpm. Samples were centrifuged as previously described and the supernatant was retained and stored at 4°C. Eluates were analysed using SDS-PAGE to determine whether soluble protein was present.

2.2.7 Chemotaxis protein methyltransferase (BbCheR)

2.2.7.1 Recombinant soluble expression

E. coli Rosetta cells containing the recombinant plasmid were streaked on an LB agar plate containing kanamycin (50 µg/mL) and chloramphenicol (34 µg/mL) and overnight cultures were prepared as previously described. The following day, 6 Erlenmeyer flasks containing 500 mL of LB (kanamycin, 50 µg/mL) were each inoculated with 5 mL of overnight culture and were grown at 37°C/180 rpm until an OD_{600} of 0.6 was reached. Cells were induced with 1 mM IPTG and protein expression carried out at 18°C/180 rpm overnight. Cultures were centrifuged at 10,000 g, 4°C for 30 minutes to pellet the cells. Cell pellets were then resuspended in 50 mL of Lysis Buffer 1 per litre of culture and were incubated on ice with 0.5 mg/mL lysozyme and Thermo Scientific™ Pierce™ Universal Nuclease for Cell Lysis for 1 hour. Following incubation, cells were split into 50 mL batches and were further disrupted by

sonication at an amplitude of 80%, 10 seconds on 10 seconds off for 10 minutes. Lysates were centrifuged at 33,000 g, 4°C for 60 minutes to separate soluble and insoluble material. The supernatant containing the soluble protein was retained for purification and the pellet resuspended in Denaturing Buffer for SDS-PAGE analysis.

2.2.7.2 Ni-NTA Immobilised Metal Affinity Chromatography (IMAC)

The Ni-NTA IMAC column was prepared by pipetting 4 mL of 50% Nickel-Nitrilotriacetic acid (Ni-NTA) beads (Thermo Scientific™) into a 30 mL Econo-Pac column and the storage ethanol was allowed to drip through, leaving a 2 mL bed volume. The resin was equilibrated with 10-column volumes of H_2O and then 10-column volumes of Lysis Buffer 0. The clarified lysate was applied to the equilibrated resin and the flow through collected. The resin was then washed with 10-column volumes of Wash Buffer and the flow through collected. Finally, the protein was eluted with 10-column volumes of Elution Buffer and 2-column volume fractions were collected.

2.2.7.3 Expression tag removal

The protein concentration of the pooled fractions was estimated using a NanoDrop™ 2000/2000c Spectrophotometer (Thermo Scientific™). To cleave the tag, TEV protease was added to the sample at a minimum mass ratio of 1:20 (TEV:Protein). The sample was then pipetted into pre-wetted Thermo Scientific™ SnakeSkin™ 3.5K MWCO Dialysis Tubing, with one end sealed. The other end was sealed after removing air and the tubing was attached to a float and left in Dialysis Buffer overnight. The following day, the sample was passed through equilibrated Ni-NTA resin in a 30 mL Econo-Pac column twice and the flow through retained. To remove any remaining protein, the TEV protease and the bound expression tag, 2-column volume elutions of 20 mM, 40 mM, 100 mM and 300 mM imidazole were collected. Samples containing the cleaved protein were pooled and concentrated using a Pierce™ Protein Concentrator PES, 10K MWCO (Thermo Scientific™).

2.2.7.4 Size Exclusion Chromatography (SEC)

A GE Healthcare Superdex 75/300 GL column was equilibrated with 2 column volumes of SEC Buffer and 1 mL of sample was injected into a 1 mL flow loop. The sample was applied to

column and proteins eluted with 1-column volume of SEC Buffer. Fractions under peaks were collected and analysed using SDS-PAGE. Fractions containing the protein of interest were pooled, concentrated and snap frozen in liquid nitrogen in preparation for structural studies. The void volume of the column was determined using blue dextran and a calibration curve was prepared using BioRad Gel Filtration Standards (figure 2.2).

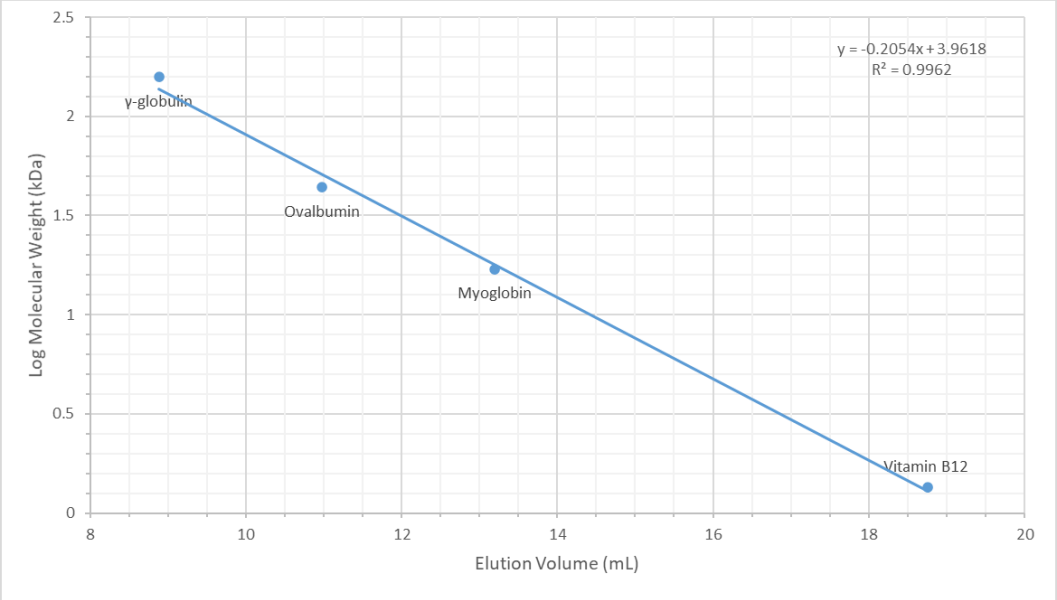


Figure 2.2. Size exclusion calibration curve of known molecular weight protein standards. The log molecular weight of the protein standards is plotted against the volume they eluted in.

2.2.8 Outer Membrane Proteins (OMPs)

2.2.8.1 Harvesting of OMPs from inclusion bodies

OMP constructs were expressed following the same procedure as for the soluble protein. Cell pellets were resuspended in OMP Lysis Buffer, incubated with 0.5 mg/mL lysozyme and Thermo Scientific™ Pierce™ Universal Nuclease for Cell Lysis for an hour and were then sonicated as previously described. After centrifuging the lysate to separate the soluble and insoluble components, a sample of supernatant was taken for SDS-PAGE and the pellet retained for further purification. The pellet was resuspended in 10 mL of Pellet Wash Buffer 1 per gram of pellet and was sonicated at an amplitude of 50%, 5 seconds on, 10 seconds off for 30 seconds. The sample was centrifuged at 10,000 g, 4°C for 20 minutes and the supernatant discarded. Wash 1 was repeated. The pellet was then resuspended in Pellet Wash

Buffer 2 and centrifuged as previously described. The final two washes were carried out using Pellet Wash Buffer 3. The washed pellet was resuspended in Denaturing Buffer and incubated at room temperature overnight. The solubilised pellet was then clarified by centrifugation at 20,000 g for 40 minutes.

2.2.8.2 IMAC and on-column refolding

A 5 mL GE Healthcare HiTrap HP Ni-NTA column was attached to an AKTA Prime FPLC and was equilibrated with Denaturing Buffer. The sample was loaded by cycling it through the column at 0.1 mL/min overnight. After loading, the column was re-equilibrated with 10-column volumes of Denaturing Buffer until a baseline was reached. Protein refolding was performed on the column, using a gradient from Denaturing Buffer to Refolding Buffer at 0.5 mL/min over 180 mL. On completion of the gradient, a further 5-column volumes of Refolding Buffer was ran through. This was followed by 10-column volumes of OMP Wash Buffer to remove any non-specific proteins and other contaminants. Finally, OMP Elution Buffer was ran through while collecting 1 mL fractions until the target protein had eluted. Fractions under the peak were collected for SDS-PAGE analysis and pooled. Pooled samples were then concentrated using a Pierce™ Protein Concentrator PES, 10K MWCO (Thermo Scientific™).

2.2.8.3 Size exclusion chromatography and detergent exchange

Following protein concentration, ~500 µL of protein was injected onto a GE Healthcare Superdex 75/300 GL column equilibrated with OMP SEC Buffer 1 or OMP SEC Buffer 2. As OMP SEC Buffer 1 contained a different detergent to the one present in the protein solution, size exclusion chromatography enabled exchange of the detergent without the need for dialysis. The sample was applied to column using a 1 mL flow loop and was eluted with 1-column volume of buffer. Fractions under peaks were collected for SDS-PAGE analysis and those containing the protein of interest were pooled, concentrated and snap frozen in liquid nitrogen.

2.2.9 Protein concentration determination

Protein concentrations were estimated using a NanoDrop™ 2000/2000c Spectrophotometer (Thermo Scientific™). An appropriate buffer was used as a blank and then 2 µL of protein

sample was loaded on to the pedestal. The concentration was estimated using the absorbance of the sample at 280 nm and the theoretical extinction coefficient of the sample as calculated by the online ExPASy ProtParam tool.

2.2.10 SDS-PAGE

Samples were prepared by mixing NuPAGE™ LDS Sample Buffer (4X), NuPAGE™ Sample Reducing Agent (10X) and an appropriate amount of protein sample. These were then heated at 75°C for 10 minutes and were loaded onto either a precast or homemade gel. 10 µL of either Invitrogen™ Novex™ Sharp Pre-stained Protein Standard or Precision Plus Protein™ Dual Xtra Prestained Protein Standard was also loaded. For precast gels, samples were loaded onto an Invitrogen™ Bolt™ 4-12% Bis-Tris Plus Gels, 12-Well gel and were ran in Invitrogen™ Novex™ Bolt™ MES SDS Running Buffer at 165 V constant for 45 minutes. When not using precast gels, 12% gels were cast using the Invitrogen™ SureCast™ Gel Handcast System. A resolving gel was prepared using SureCast™ Acrylamide (40%), SureCast™ Resolving Buffer, ddH_2O, SDS, 10% SureCast™ APS and SureCast™ TEMED. A 4% stacking gel was prepared with the same components, replacing the SureCast™ Resolving Buffer for SureCast™ Stacking Buffer. Prepared gels were then ran in Invitrogen™ Novex™ Tris-Glycine SDS Running Buffer at 125 V constant for 90 minutes. After gels had been removed from the cassette, they were stained in Coomassie Blue R250 stain and then destained until background stain had been removed.

2.2.11 Western blotting

Following SDS-PAGE, the wells and foot of the gel were removed ready for transfer. A PVDF membrane was activated in methanol, followed by H_2O and finally transfer buffer. The sponges and filter paper were also pre-wetted in transfer buffer. The transfer sandwich was prepared as shown in figure 2.3, with the gel placed on the PVDF membrane and a sponge and two pieces of filter paper each side of the gel/membrane. The sandwich was then placed into an Invitrogen™ Mini Blot Module, with the gel closest to the module cathode. A wet transfer was carried out in transfer buffer at 20 V for 60 minutes. The membrane was blocked overnight in 5% milk TBS and was then washed 3 times in 0.1% TBST. Primary and secondary antibodies were made up in 5% milk TBST, as described in table 2.5. Primary antibody

incubation was carried out for 1 hour and was followed by 3 washes in 0.1% TBST. The membrane was then incubated with a secondary antibody for 1 hour and was washed 3 times in 0.01% SDS TBST. After a final wash in TBS, the membrane was imaged at 680nm using a LICOR Odyssey infrared imaging device (LI-COR Biosciences).

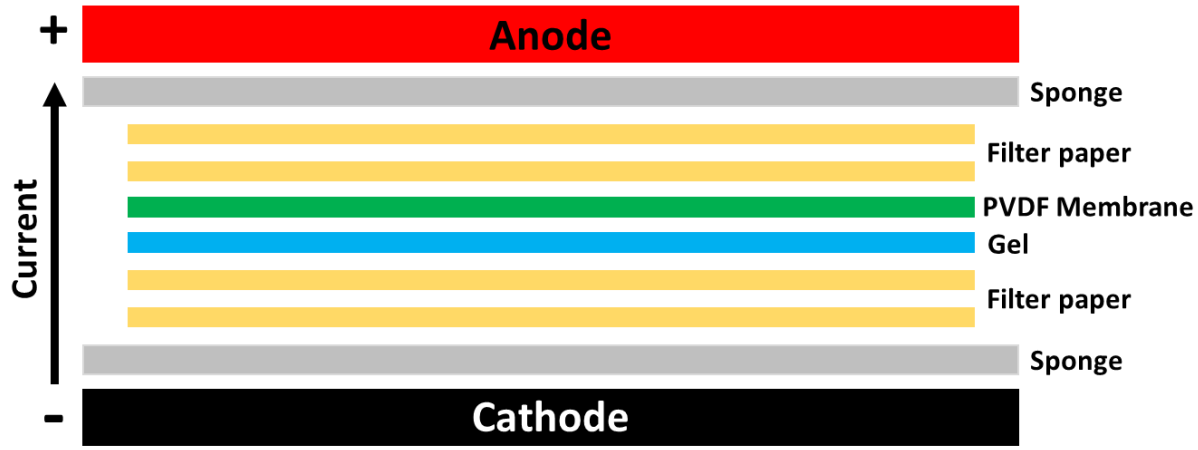


Figure 2.3. Arrangement of the blot sandwich. The blot sandwich is arranged with a sponge and two filter papers either side of the gel and PVDF membrane. The gel is placed underneath the membrane, closest to the module cathode. The current during transfer passes from the cathode to the anode.

Antibody	Description	Ratio (in 5% milk TBST)
Mouse anti-HisTag Monoclonal Antibody (ProteinTech)	Primary antibody detecting the histidine tag of recombinant proteins	1:10000
Mouse anti Human Complement Factor H Antibody (BioRad)	Primary antibody detecting the human complement factor H	1:5000
IgG (H+L) Cross-Adsorbed Goat anti-Mouse, Alexa Fluor® 680 (Invitrogen™)	Secondary antibody with fluorescence at 680nm	1:15000

Table 2.6. Antibodies used for Western blotting. The name and supplier of the antibody is given in the first column, followed by a description including whether it used as a primary or secondary antibody and what it detects. The final column shows the concentration of antibody used, given as a ratio, in 5% milk TBST.

2.2.12 Affinity ligand binding immunoblot (ALBI)

In order to identify if any of the OMP targets bound to human complement factor H (FH), an ALBI was carried out using the dot blot method. Each protein was prepared in OMP SEC Buffer

2 at a concentration of 1 mg/mL. A grid was marked on a PVDF membrane and proteins were pipetted directly onto the membrane, increasing from 1 µL dots up to 4 µL dots (1 µg – 4 µg). The membrane was blocked overnight in 5% milk TBS and was then washed 3 times in 0.1% TBST. Following this, the membrane was incubated overnight with 30 µg/mL of complement factor H from human plasma (Sigma Aldrich) in TBS. Washes were carried out as before in 0.1% TBST and antibody incubations were carried out as described above, using 1:5000 mouse anti human complement factor H antibody (BioRad) as the primary antibody and 1:15000 IgG (H+L) cross-adsorbed goat anti-mouse, Alexa Fluor® 680 (Invitrogen™) as the secondary antibody.

2.2.13 Crystallography screens

2.2.13.1 Hanging drop vapour diffusion

Membrane protein crystal trials were carried out using hanging drop vapour diffusion. Trays were set up using CELLSTAR® 24 well culture plates and MemPlus™ reagents were purchased from Molecular Dimensions (Newstead et al., 2008). The edge of each well was lined with high vacuum grease and 500 µL of reagent was added to each well. The protein solution (~5 mg/mL) was centrifuged at 13,500 rpm for 10 minutes to remove any precipitates formed when thawing and 1 µL of sample was pipetted in the centre of an 18 mm x 18 mm coverslip. The protein drop was mixed with 1 µL of well solution and the coverslip inverted with the drop facing into the well. The plate was then analysed under a microscope to assess the initial appearance of drops and then it was placed in a 20°C incubator. The plate was viewed after 3 and 7 days and then once per week thereafter to screen for crystals.

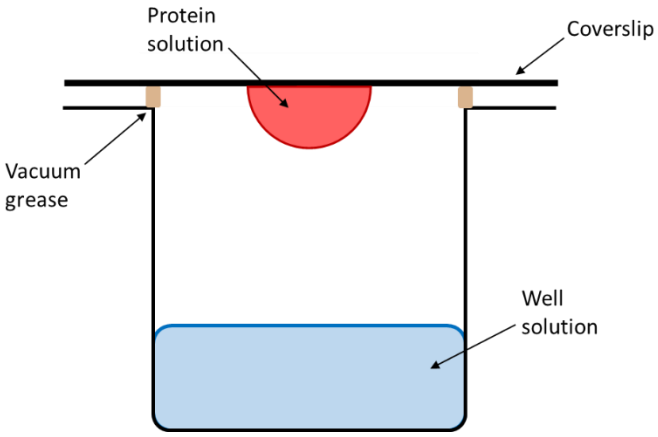


Figure 2.4. A diagram summarising the hanging drop vapour diffusion well setup. 24 well plates were used and each well was set up as demonstrated in the diagram. 500 µL of well solution was used and protein drops contained 1 µL of protein mixed with an equal amount of well solution.

2.2.13.2 Sitting drop vapour diffusion

High throughput protein crystal screens for soluble BbCheR protein were set up using sitting drop vapour diffusion. Trays were set up using JCSG-plus™ and PACT premier™ sparse matrix screens from Molecular Dimensions (Newstead et al., 2008) and 50 µL of reservoir solution was pipetted into the wells of a Swissci 96-well 2 drop MRC-Crystallization Plate (Molecular Dimensions) using a multichannel pipette. Protein solution (~8.9 mg/mL) was rapidly thawed from a frozen stock and was centrifuged at 13,500 rpm for 10 minutes to remove any precipitate present from thawing. Two plates were set up using the JCSG-plus™ screen and one using the PACT premier™ screen. For the first JCSG-plus™ screen and the PACT premier™ screen, 1 µL of protein was pipetted onto the raised conical wells and was mixed with 1 µL of reservoir solution. The JCSG-plus™ screen utilised both drop wells, with the first containing 0.75 µL of protein and 1.5 µL of reservoir solution and the second containing 1.5 µL of protein and 0.75 µL of reservoir solution. Plates were then sealed with a sealing film and initial drop observations were done under a microscope. The plates were incubated at 20°C and monitored for crystal formation at 3 and 7 days and then once per week thereafter. For co-crystallisation attempts, protein in 20 mM HEPES pH 7.5, 125 mM NaCl and 5% glycerol at a concentration of ~8.9 mg/mL was mixed with S-adenosyl-L-homocysteine (SAH) to 1.25 mM. The BbCheR-SAH solution was used to prepare trays as previously described.

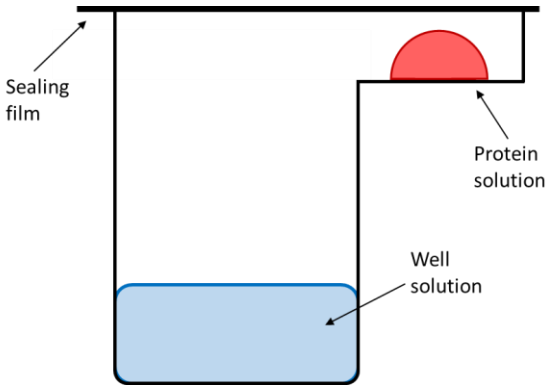


Figure 2.5. A diagram summarising the sitting drop vapour diffusion well setup. Swissci 96-well 2 drop MRC-Crystallization Plates were used and each well was set up as demonstrated in the diagram. 50 µL of reservoir solution was used and protein drops contained varied amounts of protein to reservoir solution depending on the screening conditions.

Chapter 3 Target identification and cloning of *Borrelia* constructs

3.1 Target identification and construct design

3.1.1 Identification of *Borrelia* OMPs and potential soluble proteins

Six protein targets coded on the *B. burgdorferi* B31 chromosome were identified for this project (table 3.1 and 3.2) which included 4 predicted OmpA-like β-barrel proteins and 2 potentially soluble targets. Predicted β-barrel outer membrane proteins from *Borrelia* were selected based on previous work by Dr Adam Dyer (Dyer et al., 2015). OmpA-like proteins were identified using a Hidden Markov Model (HMM) approach, as BLAST searches using *E. coli* OMPs such as OmpA, OmpX and OmpW as a reference were not sensitive enough to identify homologues based on sequence similarity due to the high sequence variability of OmpA proteins (Schulz, 2000). While previous research focused on a single protein from *Borrelia afzelii*, BAPKO_0422, this project widened the approach to investigate 4 paralogous genes in *Borrelia burgdorferi s.s*; BB_0027, BB_0405, BB_0406 and BB_0562.

Name	Accession Number (NCBI)	Protein (B31 Chromosome)	No of Amino Acids	BLAST (PDB)
BB_0027	NP_212161	Hypothetical protein	212	No putative conserved domains
BB_0405	NP_212539	Hypothetical protein	203	No putative conserved domains
BB_0406	NP_212540	Hypothetical protein	203	No putative conserved domains
BB_0562	NP_212696	Hypothetical protein	180	No putative conserved domains

Table 3.1. Overview of the outer membrane protein targets. The gene/protein name is given in the first column followed by the NCBI accession number, the function of the protein, the size in amino acids and the result from a protein BLAST sequence search against the Protein Data Bank (PDB).

Potentially soluble targets were selected due to the challenges of crystallising membrane proteins and soluble proteins are more likely to produce quality crystals. The search for potentially soluble target proteins was narrowed down to the *B. burgdorferi* B31 chromosome and then proteins under positive selection were selected. This was defined by a Ka/Ks (number of non-synonymous substitutions per non-synonymous site (Ka)/the number of synonymous substitutions per synonymous site (Ks)) value above 1, based on supplementary data from (Mongodin et al., 2013). After sequences were analysed for homology to other known structures in the Protein Data Bank (PDB), domains predicted and signal sequences

identified, any proteins predicted to be membrane proteins were ruled out as potential targets.

Name	Accession Number (NCBI)	Protein (B31 Chromosome)	Ka/Ks Value	No of Amino Acids	BLAST (PDB)
BB_0095	NP_212229	Hypothetical protein	2.03	181	No putative conserved domains
BB_0040	NP_212174	Chemotaxis protein methyltransferase	2.01	283	Putative conserved domains

Table 3.2. Overview of the potentially soluble protein targets. The gene/protein name is given in the first column followed by the NCBI accession number, the function of the protein, the Ka/Ks value (Mongodin et al., 2013), the size in amino acids and the result from a protein BLAST sequence search against the Protein Data Bank (PDB).

3.1.2 Signal sequence analysis

The four membrane proteins (BB_0027, BB_0405, BB_0406 and BB_0562) were analysed using SignalP 5.0 to predict their signal sequences (figure 3.1). While full length proteins were cloned, primers were also designed in order to exclude the signal sequence and avoid translocation of the protein into the membrane of *E. coli* during expression, allowing inclusion body formation. All four OMPs; BB_0027, BB_0405, BB_0406 and BB_0562, showed a predicted signal sequence score equal to or above the cut-off probability of 0.5.

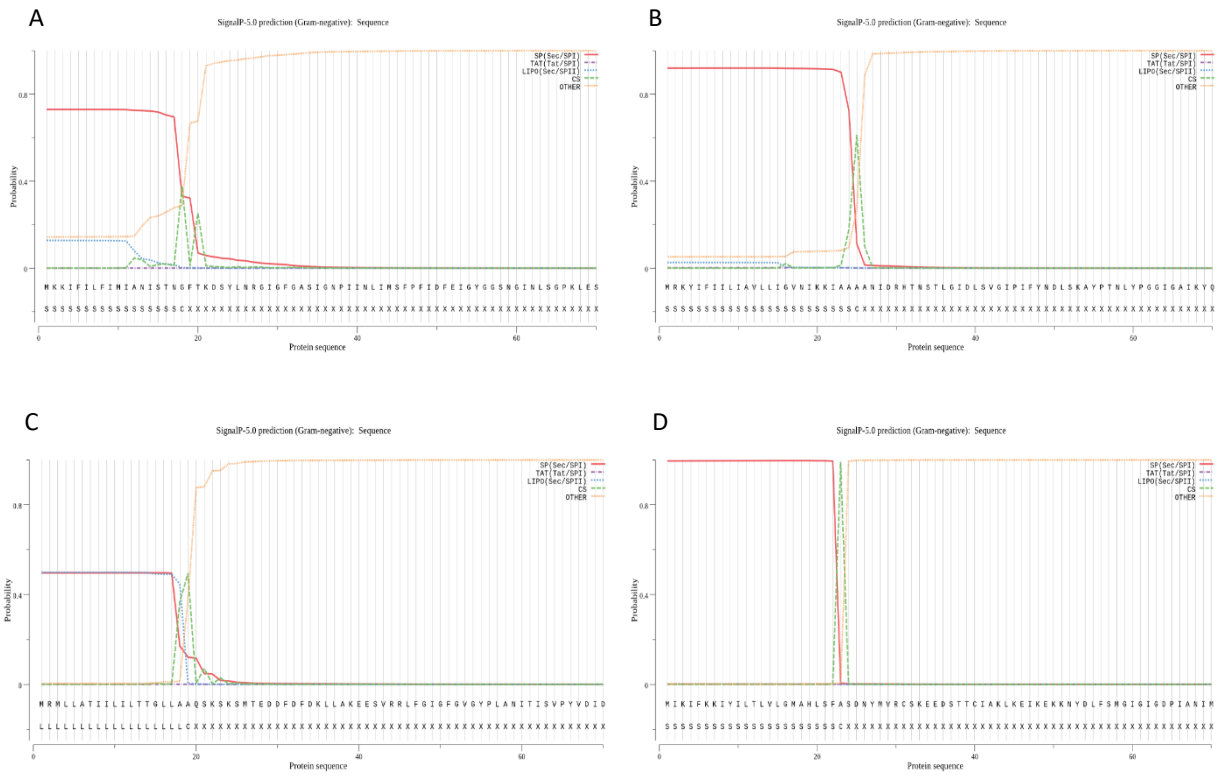


Figure 3.1. Outputs from SignalP 5.0 analysis of *Borrelia* OMP targets. A) BB_0562. B) BB_0027. C) BB_0405. D) BB_0406. The amino acid sequence of each protein is shown along the bottom of the graph, with S representing the signal sequence, C showing the cleavage site and L referring to a lipoprotein signal sequence. The red line represents the probability of each amino acid being part of the signal sequence, with a score above 0.5 considered significant. The green dashed line shows the probability of the cleavage site location and the blue dotted line represents the probability of a lipoprotein signal sequence.

3.1.3 Multiple sequence alignments

Sequences of the predicted OmpA-like proteins; BB_0027, BB_0405, BB_0406 and BB_0562 from *B. burgdorferi s.s* were aligned using MUSCLE (Edgar, 2004) (figure 3.2). The areas of similarity are proposed to be the membrane spanning regions separated by loop/turn regions which have much lower similarity (Dyer et al., 2015). Paralogues BB_0405 and BB_0406 share 52% sequence identity, but BB_0027 appears more distantly related, with similarity only seen in the proposed membrane spanning regions.

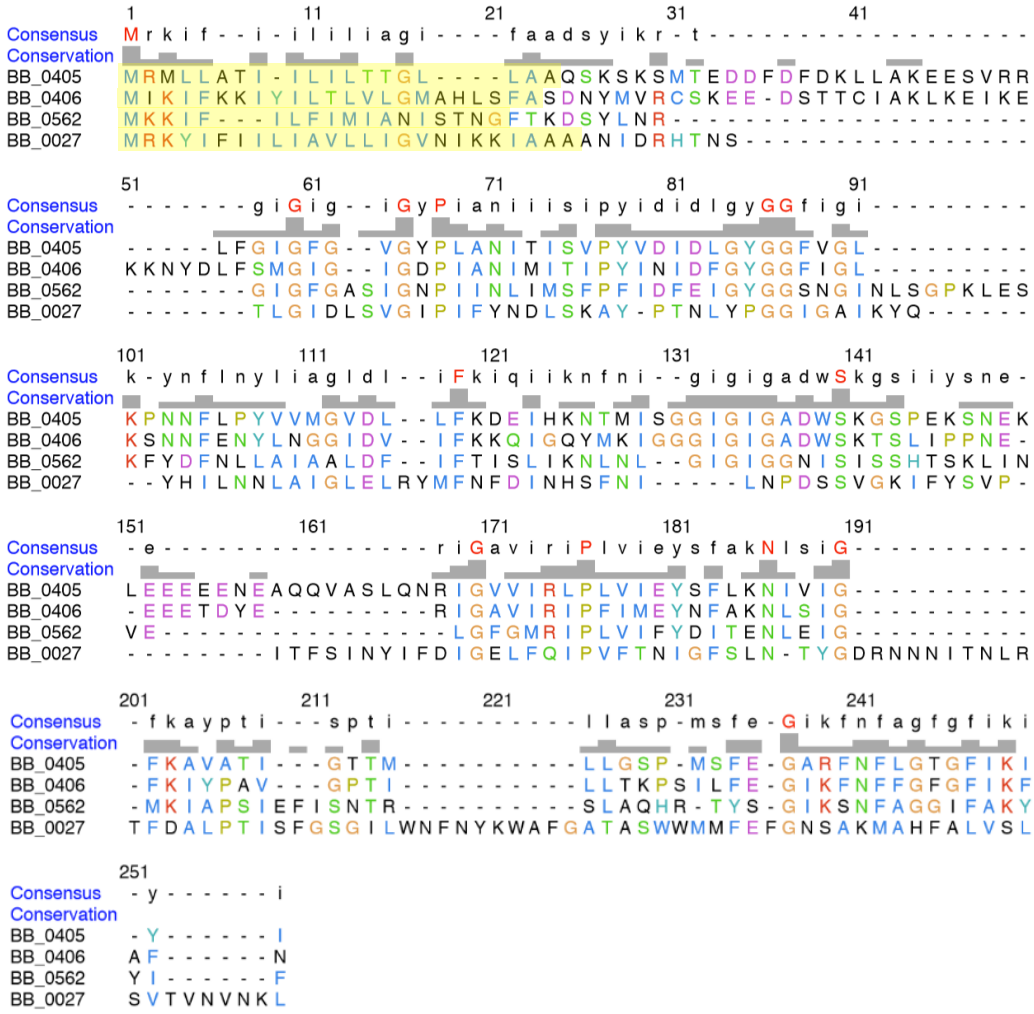
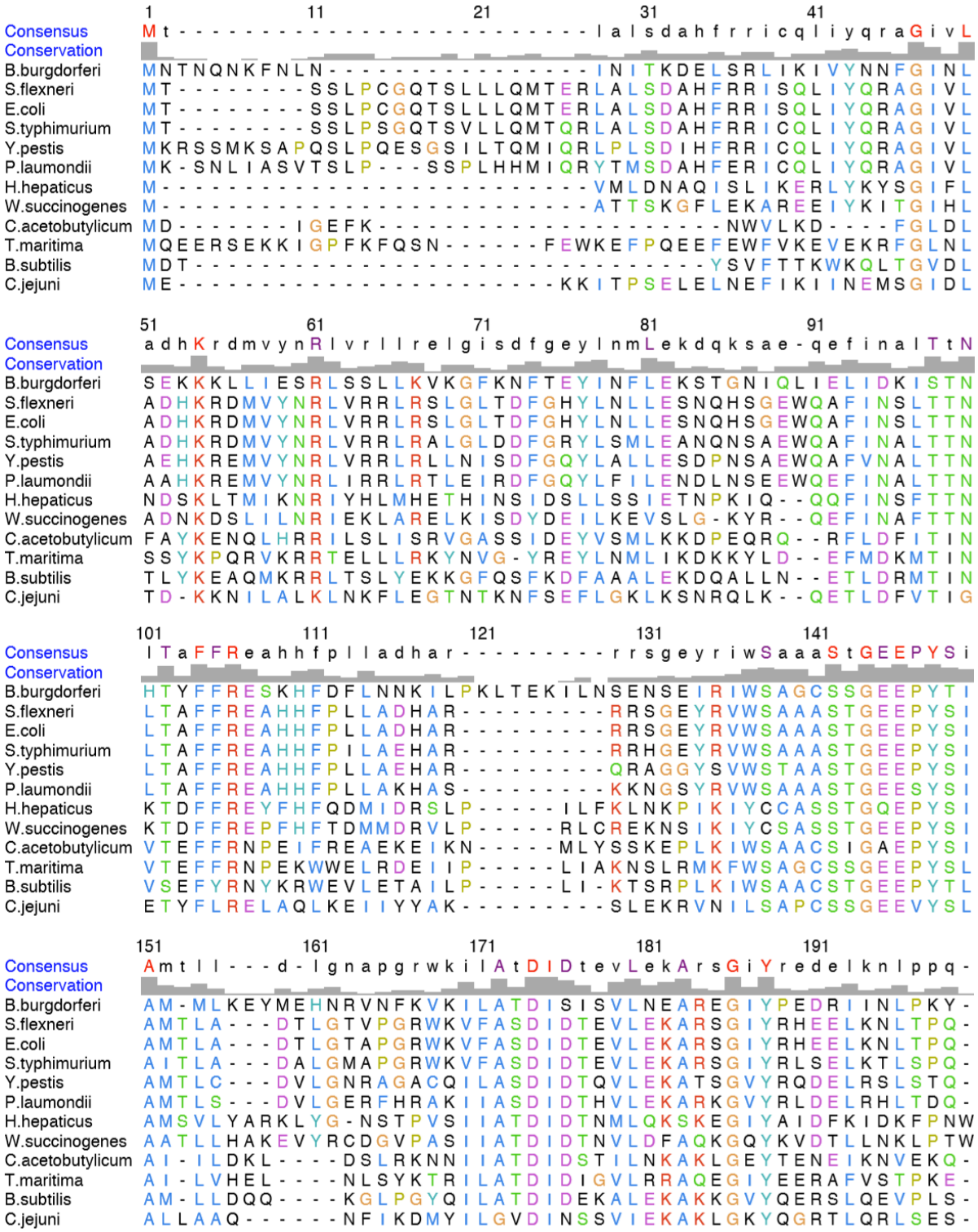


Figure 3.2. Multiple sequence alignment of *Borrelia* OMPs. The sequences BB_0405, BB_0406 and predicted OMPs BB_0562 and BB_0027 were aligned using MUSCLE (Edgar, 2004). The signal sequence as predicted by SignalP 5.0 is highlighted.

A BLAST search of the protein sequence for BB_0040 against sequences in the PDB showed yielded 3 chemotaxis protein methyltransferases from the CheR family, with BB_0040 sharing 36% sequence identity to the PICheR from *Bacillus subtilis* (BsCheR), 34% identity to CheR1

from *Pseudomonas aeruginosa* and 33% identity to the PDCheR from *Salmonella typhimurium* (StCheR). Using MUSCLE (Edgar, 2004), a multiple sequence alignment of the *Borrelia burgdorferi* CheR (BbCheR) with 11 other CheR proteins from different Gram negative bacteria (figure 3.3) identified that BbCheR contains the conserved structural motif thought be involved in catalytic activity and the binding of S-adenosylmethionine (SAM) and S-adenosyl-L-homocysteine (SAH). BbCheR lacks the conserved residues found in the β-subdomain of PDCheRs that are involved in pentapeptide binding and is therefore likely a PICheR.


1 11 21 31 41
Consensus
Conservation
B.burgdorferi
S.flexneri
E.coli
S.typhimurium
Y.pestis
P.laumondii
H.hepaticus
W.succinogenes
C.acetobutylicum
T.maritima
B.subtilis
C.jejuni
51 61 71 81 91
101 111 121 131 141
151 161 171 181 191

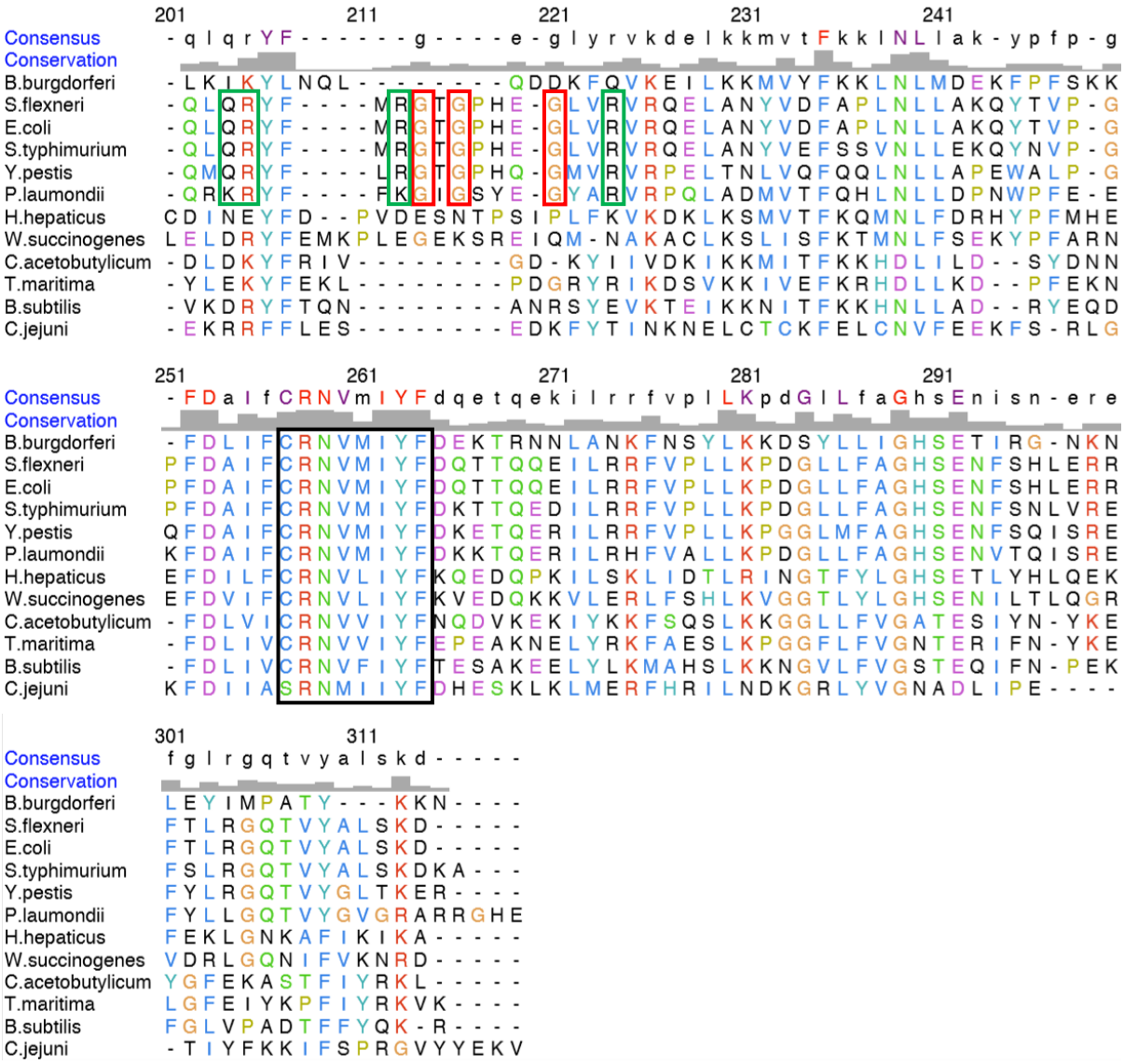


Figure 3.3. Multiple sequence alignment of CheR sequences from various Gram negative bacteria. The amino acid sequence for *B. burgdorferi* CheR was aligned with PDCheRs (*S. flexneri, Escherichia coli, S. typhimurium, Y. pestis, P. laumondii*) and PICheRs (*H. hepaticus, W. succinigenes, C. acetobutylicum, T. maritima, B. subtilis, C. jejuni*). The black box highlights a conserved sequence motif thought to be involved in catalytic activity. The green boxes highlight residues in the β-subdomain that are important in PDCheRs for binding to the pentapeptide sequence. The red boxes highlight three conserved glycine residues in the β-subdomain of PDCheRs. Adapted from (Batra et al., 2016).

3.1.4 Secondary structure predication and primer design

Truncations made to the N- and C- terminus of proteins have been found to increase recombinant protein expression levels and could help crystal formation (Gräslund et al., 2008). For this reason, each protein target sequence was cloned full length and with a range of truncations made to either the N-terminal, C-terminal or a combination of both (figure 3.5). In order to determine truncation sites for each protein target, amino acid sequences obtained through NCBI were analysed using BLAST and SMART to identify any known homology to other proteins in the Protein Data Bank (PDB) and domain boundaries. PSI-PRED was used to predict secondary structure elements and disordered regions. The output from SignalP analysis was used to design constructs with just the signal sequence removed, but full length OMPs with the predicted signal sequence included were also designed. Two further N-terminal truncations were also made for each OMP using the PSI-PRED outputs, to avoid removing major structural elements which could be detrimental to protein production and to remove any potential disordered regions. The PSI-PRED output for BB_0406 can be seen in figure 3.4 and outputs for the remaining proteins are provided in appendix 2.

Numerous truncations of BB_0095 were made to both the N and C-terminus primarily using secondary structure predications from PSI-PRED, as a BLAST search of this protein did not find any conserved domains or significant similarities to other proteins in the PDB. Only one truncation was designed for BbCheR and this was at the N-terminus, removing the short region before the large SAM domain. Primers were designed starting from the first codon of each truncation and included appropriate sequence extensions for Ligation Independent Cloning, which was used to amplify each sequence and clone the constructs into the desired expression vector.

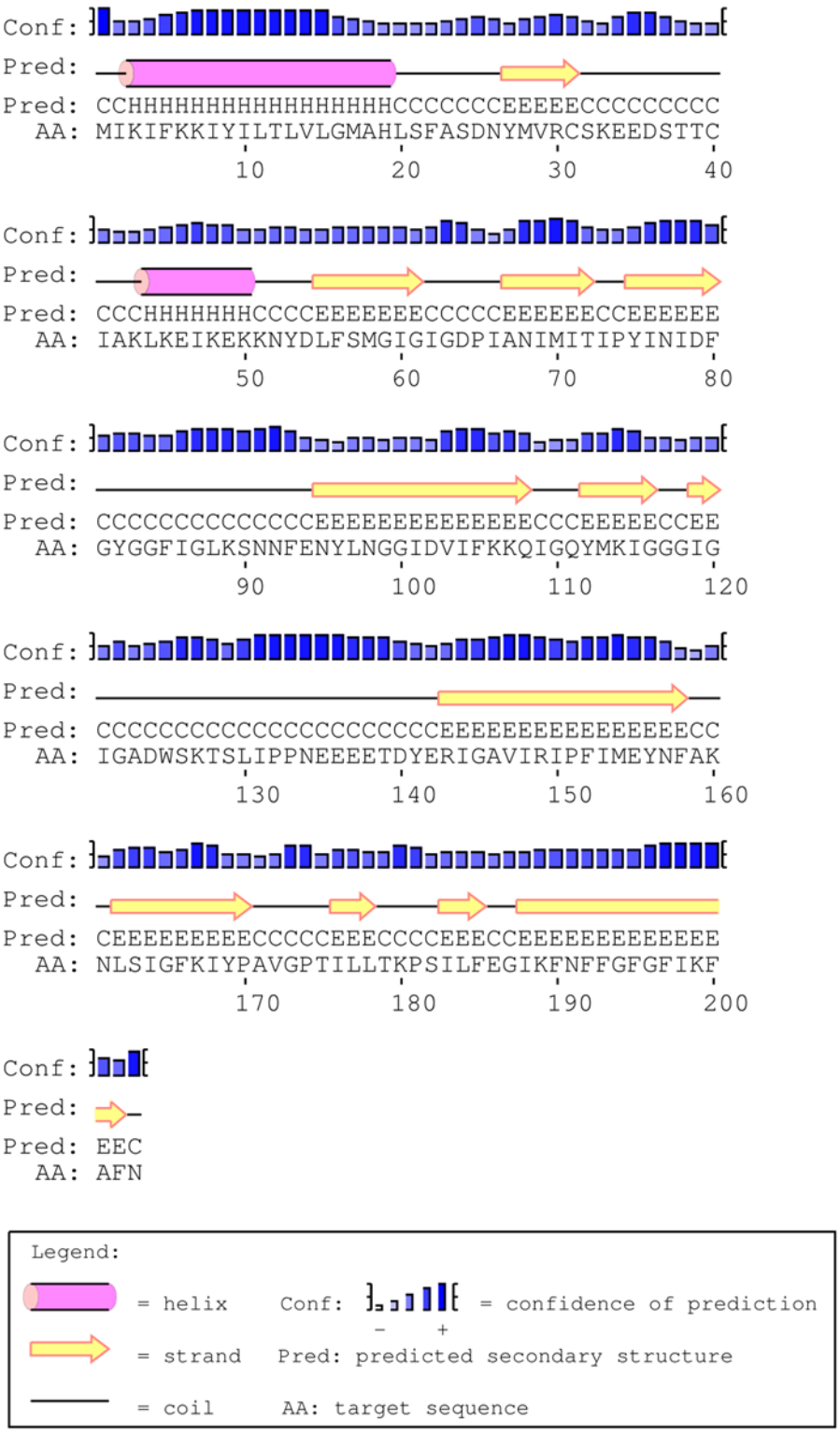


Figure 3.4. PSI-PRED output for BB_0406. The protein sequence for BB_0406 was obtained from NCBI and PSI-PRED was used to predict protein secondary structure elements.

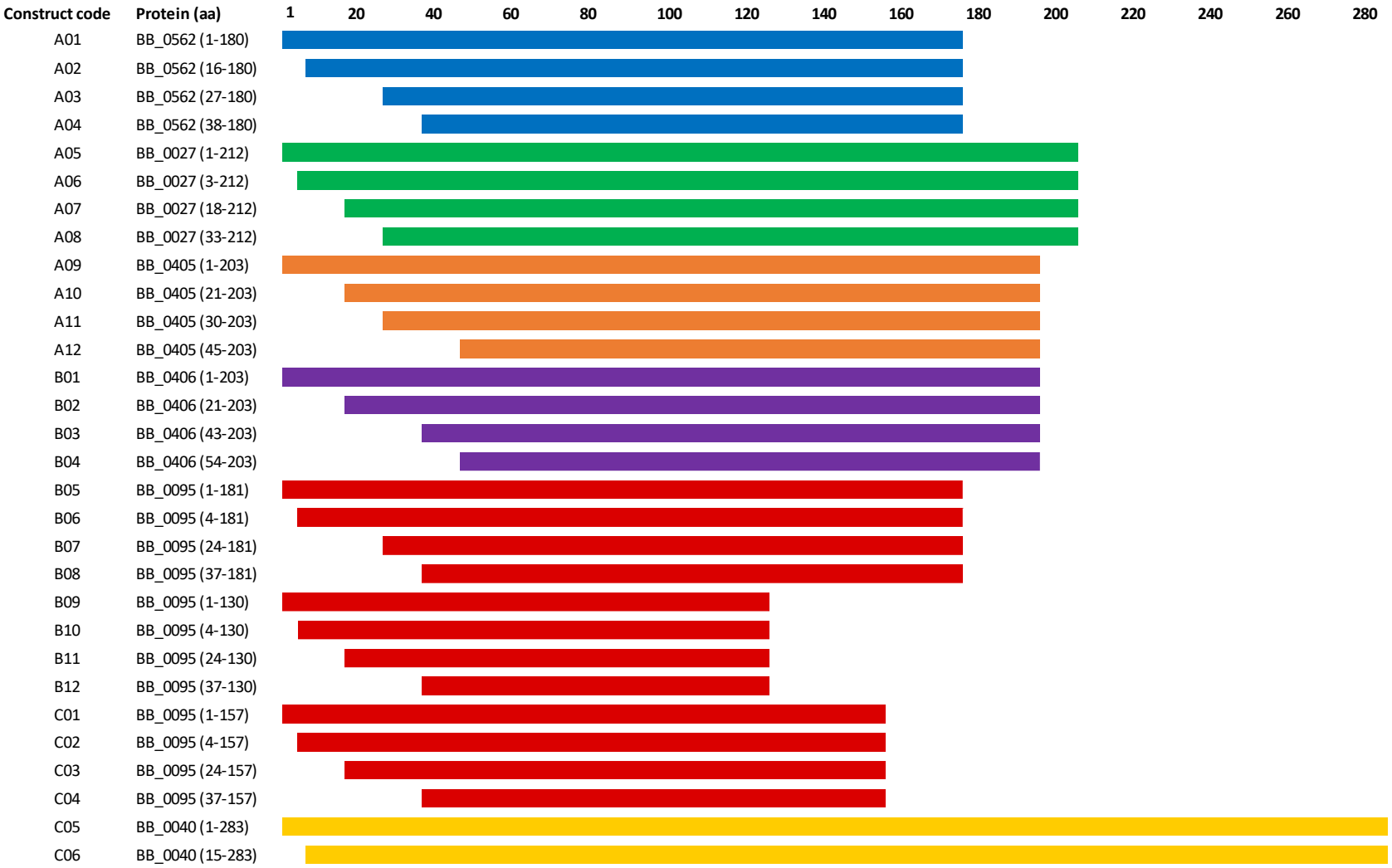


Figure 3.5. *Borrelia* protein constructs designed for LIC. Amino acid number is shown along the top. Blue bars: BB_0562 constructs, green bars: BB_0027 constructs, orange bars: BB_0405 constructs, purple bars: BB_0406 constructs, red bars: BB_0095 constructs, yellow bars: BB_0040 constructs. The first bar for each protein represents the full-length construct.

Code	Construct (aa)	Primer (forward)	Primer (reverse)	Length (bp)	Vector	PCR screen expected size
A01	BB_0562 (1-180)	BB0562-f0	BB0562-r0	543	pNIC28-Bsa4	828
A02	BB_0562 (16-180)	BB0562-f1	BB0562-r0	498	pNIC28-Bsa4	783
A03	BB_0562 (27-180)	BB0562-f2	BB0562-r0	465	pNIC28-Bsa4	750
A04	BB_0562 (38-180)	BB0562-f3	BB0562-r0	432	pNIC28-Bsa4	717
A05	BB_0027 (1-212)	BB0027-f0	BB0027-r0	639	pNIC28-Bsa4	924
A06	BB_0027 (3-212)	BB0027-f1	BB0027-r0	633	pNIC28-Bsa4	918
A07	BB_0027 (18-212)	BB0027-f2	BB0027-r0	588	pNIC28-Bsa4	873
A08	BB_0027 (33-212)	BB0027-f3	BB0027-r0	543	pNIC28-Bsa4	828
A09	BB_0405 (1-203)	BB0405-f0	BB0405-r0	612	pNIC28-Bsa4	897
A10	BB_0405 (21-203)	BB0405-f1	BB0405-r0	552	pNIC28-Bsa4	837
A11	BB_0405 (30-203)	BB0405-f2	BB0405-r0	525	pNIC28-Bsa4	810
A12	BB_0405 (45-203)	BB0405-f3	BB0405-r0	480	pNIC28-Bsa4	765
B01	BB_0406 (1-203)	BB0406-f0	BB0406-r0	612	pNIC28-Bsa4	897
B02	BB_0406 (21-203)	BB0406-f1	BB0406-r0	552	pNIC28-Bsa4	837
B03	BB_0406 (43-203)	BB0406-f2	BB0406-r0	486	pNIC28-Bsa4	771
B04	BB_0406 (54-203)	BB0406-f3	BB0406-r0	453	pNIC28-Bsa4	738
B05	BB_0095 (1-181)	BB0095-f0	BB0095-r0	546	pNH-TrxT	1158
B06	BB_0095 (4-181)	BB0095-f1	BB0095-r0	537	pNH-TrxT	1149
B07	BB_0095 (24-181)	BB0095-f2	BB0095-r0	477	pNH-TrxT	1089
B08	BB_0095 (37-181)	BB0095-f3	BB0095-r0	438	pNH-TrxT	1050
B09	BB_0095 (1-130)	BB0095-f0	BB0095-r1	393	pNH-TrxT	1005
B10	BB_0095 (4-130)	BB0095-f1	BB0095-r1	384	pNH-TrxT	996
B11	BB_0095 (24-130)	BB0095-f2	BB0095-r1	324	pNH-TrxT	936
B12	BB_0095 (37-130)	BB0095-f3	BB0095-r1	285	pNH-TrxT	897
C01	BB_0095 (1-157)	BB0095-f0	BB0095-r2	474	pNH-TrxT	1086
C02	BB_0095 (4-157)	BB0095-f1	BB0095-r2	465	pNH-TrxT	1077
C03	BB_0095 (24-157)	BB0095-f2	BB0095-r2	405	pNH-TrxT	1017
C04	BB_0095 (37-157)	BB0095-f3	BB0095-r2	366	pNH-TrxT	978
C05	BB_0040 (1-283)	BB0040-f0	BB0040-r0	852	pNH-TrxT	1464
C06	BB_0040 (15-283)	BB0040-f1	BB0040-r0	810	pNH-TrxT	1422

Table 3.3. Details of the selected *B. burgdorferi* B31 gene targets. The code for each construct is shown in the first column, followed by the protein/gene name. The numbers under the name indicate which amino acid the construct begins and ends with. The forward and reverse primers used for amplification are indicated, followed by the length of the construct sequence in base pairs. The plasmid the construct was attempted to be cloned into is shown, followed by the expected colony screen size of the cloned construct in base pairs.

3.2 PCR amplification of target genes

Genomic DNA from *B. burgdorferi* (DSMZ) was replicated using the REPLI-g Mini Kit (Qiagen) and 30 ng/µL was used as a template to amplify variations of six *Borrelia* genes from the *B. burgdorferi* chromosome. Each target was amplified using standard PCR conditions and a combination of the primers shown in table 2.1. Primer combinations for each construct are shown in table 3.3. A standard PCR reaction consists of an initial denaturing step, amplification cycles and a final extension step. The amplification cycles have three main stages:

- Denaturing - Similar to the initial denaturing step, this involves heating the DNA to 95°C to separate the double stranded DNA helix into single strands, to allow primers and enzymes to bind for amplification.
- Annealing - The temperature is lowered to allow binding of primers to the DNA strand. This temperature is usually determined by the lowest melting temperature (Tm) of the primers.
- Extension - The temperature is raised to 72°C to allow synthesis of the new DNA strand by Taq polymerase. The extension time is determined by the longest DNA template.

A general annealing temperature of 50°C was used for all of the constructs as both forward and reverse primers are designed with long extensions for LIC, which affects the primer Tm. The forward primer extension and reverse primer extension sequences are highlighted in table 2.1. Initially, a low yield was obtained for inserts C5 and C6, so 30 ng/µL of the original non-replicated *B. burgdorferi* DNA purchased from DSMZ was used as a template, rather than the DNA product from the kit, producing a much higher yield. All constructs, apart from C5 and C6, amplified first time and did not require optimisation of conditions. Both C5 and C6 were optimised as previously mentioned. All the bands seen in figure 3.6 appear to be in the correct relative positions (sizes shown in table 3.3) and a pattern of decreasing band size can be seen for each gene, corresponding to a smaller length in base pairs due to truncations. The expected size for the PCNA positive control is 786bp and a band can be seen in around this size on each of the gels. PCNA was used as the template DNA for the positive control as it is considered robust for PCR applications.

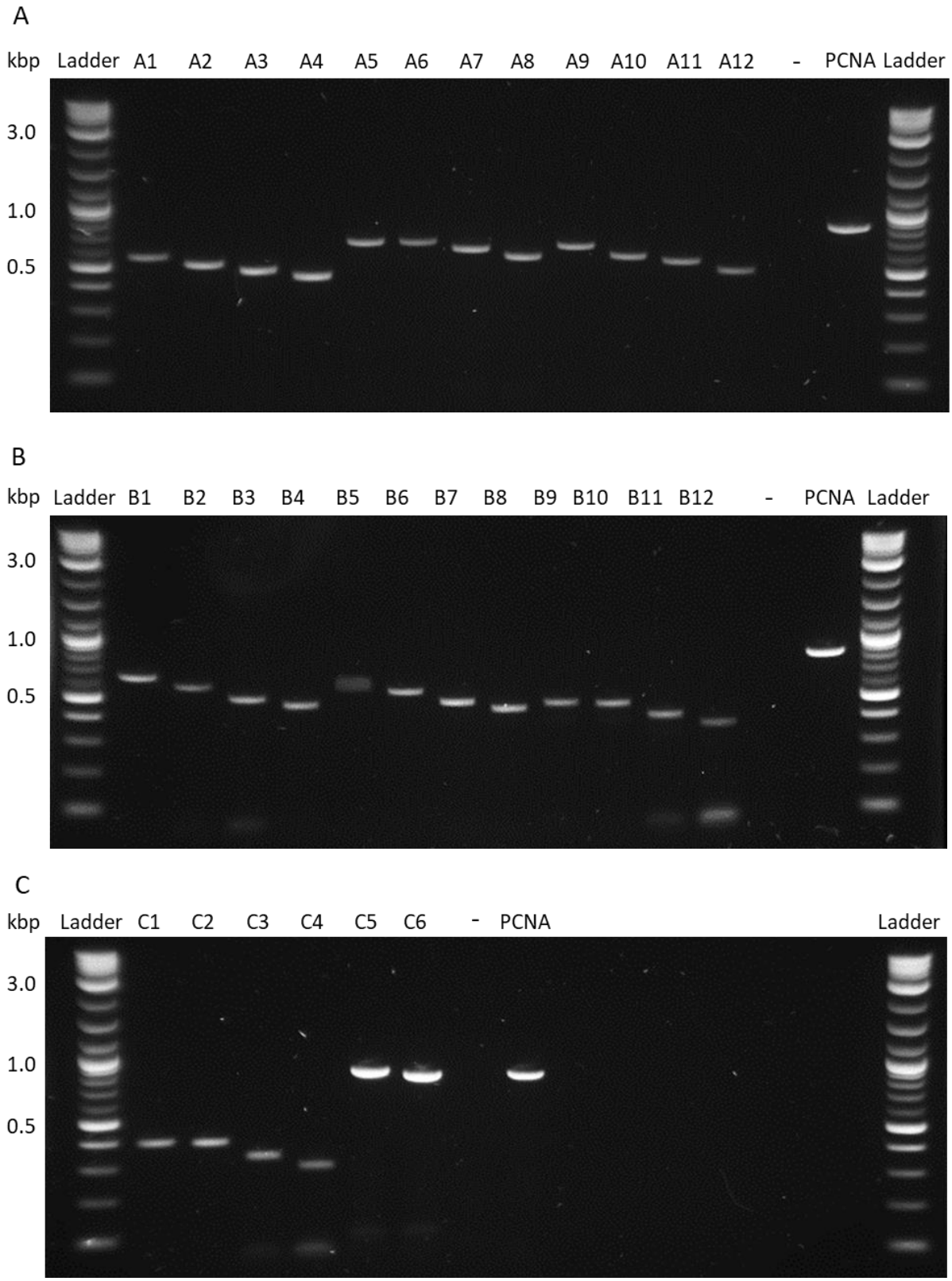


Figure 3.6. A 1.5% DNA agarose gel of the amplified genomic DNA inserts A1-C6 from the *B. burgdorferi* B31 chromosome. **A** Lane 1: NEB 2-Log DNA Ladder (0.1-10.0 kb). Lanes 2-13: Inserts A1-A12. Lane 14: Negative Control. Lane 15: PCNA positive control. Lane 16: NEB 2-Log DNA Ladder (0.1-10.0 kb). **B** Lane 1: NEB 2-Log DNA Ladder (0.1-10.0 kb). Lanes 2-13: Inserts B1-B12. Lane 14: Negative control. Lane 15: PCNA positive control. Lane 16: NEB 2-Log DNA Ladder (0.1-10.0 kb). **C** Lane 1: NEB 2-Log DNA Ladder (0.1-10.0 kb). Lanes 2-7: Inserts C1-C6. Lane 8: Negative control. Lane 9: PCNA positive control. Lane 16: NEB 2-Log DNA Ladder (0.1-10.0 kb).

3.3 Preparation of expression vectors

Two pET28a expression vectors, pNIC28-Bsa4 and pNH-TrxT, were selected for cloning (appendix 1). For soluble targets, pNH-TrxT was used as it not only contains a 6x-Histidine tag but also thioredoxin, which can aid protein solubility. For OMP constructs, which are expected to be insoluble, pNIC28-Bsa4 was used as it only contains the 6x-His tag without a solubility partner. Both vectors are under the T7 promotor and *lac* operon system, allowing for tight control of expression using IPTG, and contain a TEV cleavage site for expression tag removal using a TEV protease. Vectors were digested by incubation with *Bsa*I restriction enzyme and samples were ran on a 1.5% agarose gel to ensure digestion. Both vectors are a similar size, with pNIC28-Bsa4 being 7284bp and pNH-TrxT being 7602bp and despite the bands of the ladder not being clear above 3000bp, the bands appear to be the correct relative size as they are near the highest ladder band of 10,000bp (figure 3.7). The smaller band visible at ~2000bp is a SacB stuffer region that is released during digestion. The SacB region is used as a negative selection marker, coding for levansucrase which breaks down sucrose into products that are toxic to *E. coli*, meaning that cells containing uncut vector are unable to grow when streaked on 5% sucrose plates.

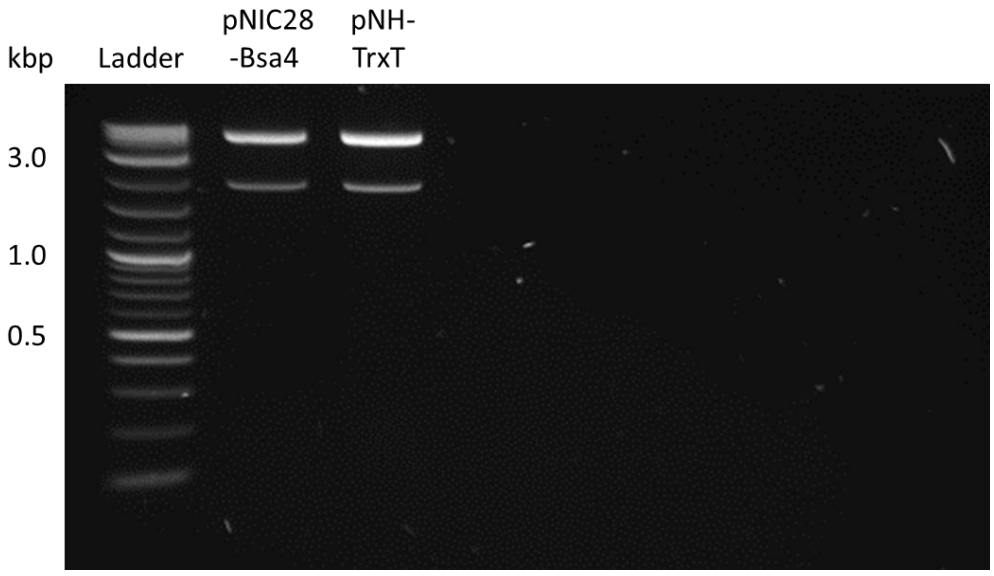


Figure 3.7. A 1.5% DNA agarose gel of the two cloning vectors after digestion with *Bsa*I. Lane 1: NEB 2-Log DNA Ladder (0.1-10.0 kb). Lane 2: 10 µL of digested pNIC28-Bsa4 vector. Lane 3: 10 µL of digested pNH-TrxT vector.

3.4 Colony PCR of cloned constructs

The digested vectors and PCR products from insert amplification were purified and then treated with T4 DNA polymerase to create cohesive ends for ligation and transformation into Mach1™ *E. coli* cells. T4 DNA polymerase is used in the presence of dCTP only for inserts and dGTP only for vectors. As T4 DNA pol has 3'-5' exonuclease activity and no 5'-3' exonuclease activity, it will remove bases in the 3'-5' direction only. Bases will be removed up to the first complementary base to the nucleotide added to the reaction mixture, where the polymerase activity will incorporate the nucleotide. When using LIC techniques, long overhangs are created so a ligase is not required when combining the vector and insert, producing an annealed but nicked product. During transformation into *E. coli*, the host cell ligase is able to repair the nicks in the phosphodiester backbone of the new cloned construct.

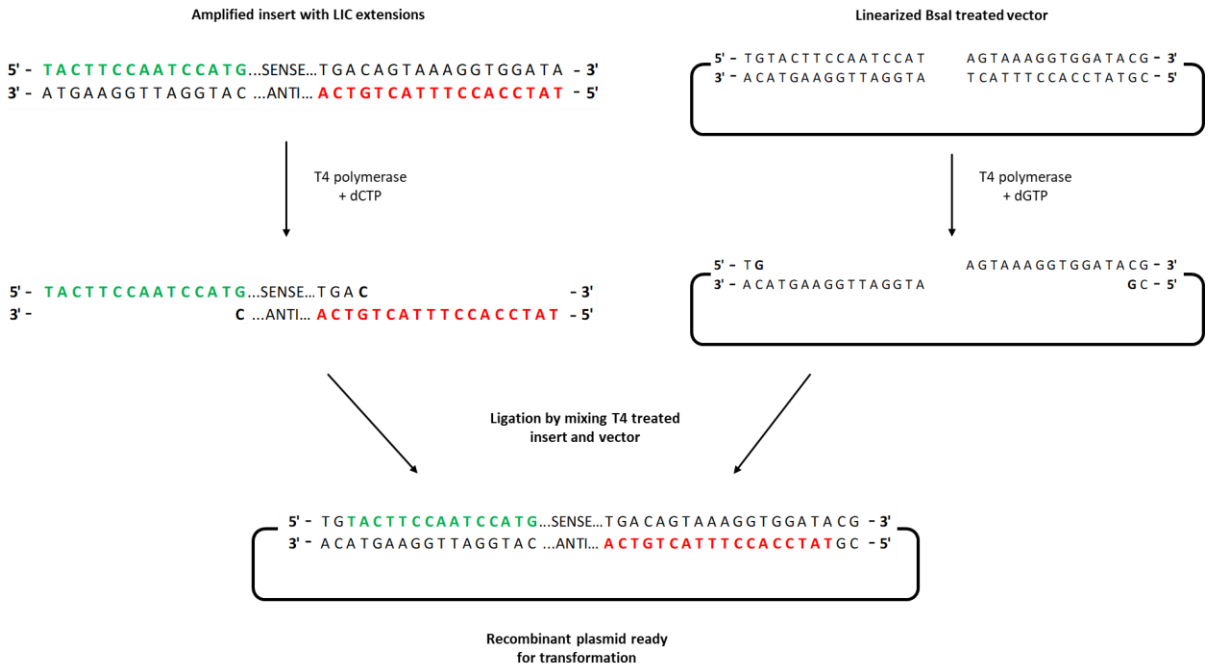


Figure 3.8. Diagram showing the generation of cohesive ends and ligation of vectors and inserts during LIC. The amplified sequence is shown with the LIC extension from the forward primer in green and the LIC extension from the reverse primer in red. Sense represents the target gene sequence sense strand and 'anti' represents the complementary target gene sequence on the anti-sense strand. The added nucleotide during T4 treatment is shown in bold.

For transformation, the initial ligation ratio of 1:2 vector to insert did not produce any colonies, so a ratio of 2:3 was used and this produced colonies on all plates. Colonies from the

transformation of ligated vectors and inserts into Mach1 competent *E. coli* were then screened for the presence of successful clones. Colony PCR was performed using pLIC primers (table 2.1) which are specific to the vector sequence either side of the insert site and PCR products were analysed using DNA agarose gel electrophoresis. Numerous colonies were screened for each construct and a final summary gel of all the successful constructs can be seen in figure 3.9. The bands present in each lane, including the PCNA, appear around the expected size with the exception of construct C6 (sizes are given in table 3.3). Construct C6 (BB_0040 residues 15-283) did not successfully clone even after multiple attempts with different vector: insert ratios. While the band for C5 is present, it is fainter than the other bands, possibly due to an error in the amount of DNA added to the PCR reaction mix for this construct. Constructs A3, A8, A10, B4, B6 and C5 were all confirmed to be in frame by DNA sequencing. Construct plasmids were isolated and transformed into Rosetta™ cells for expression.

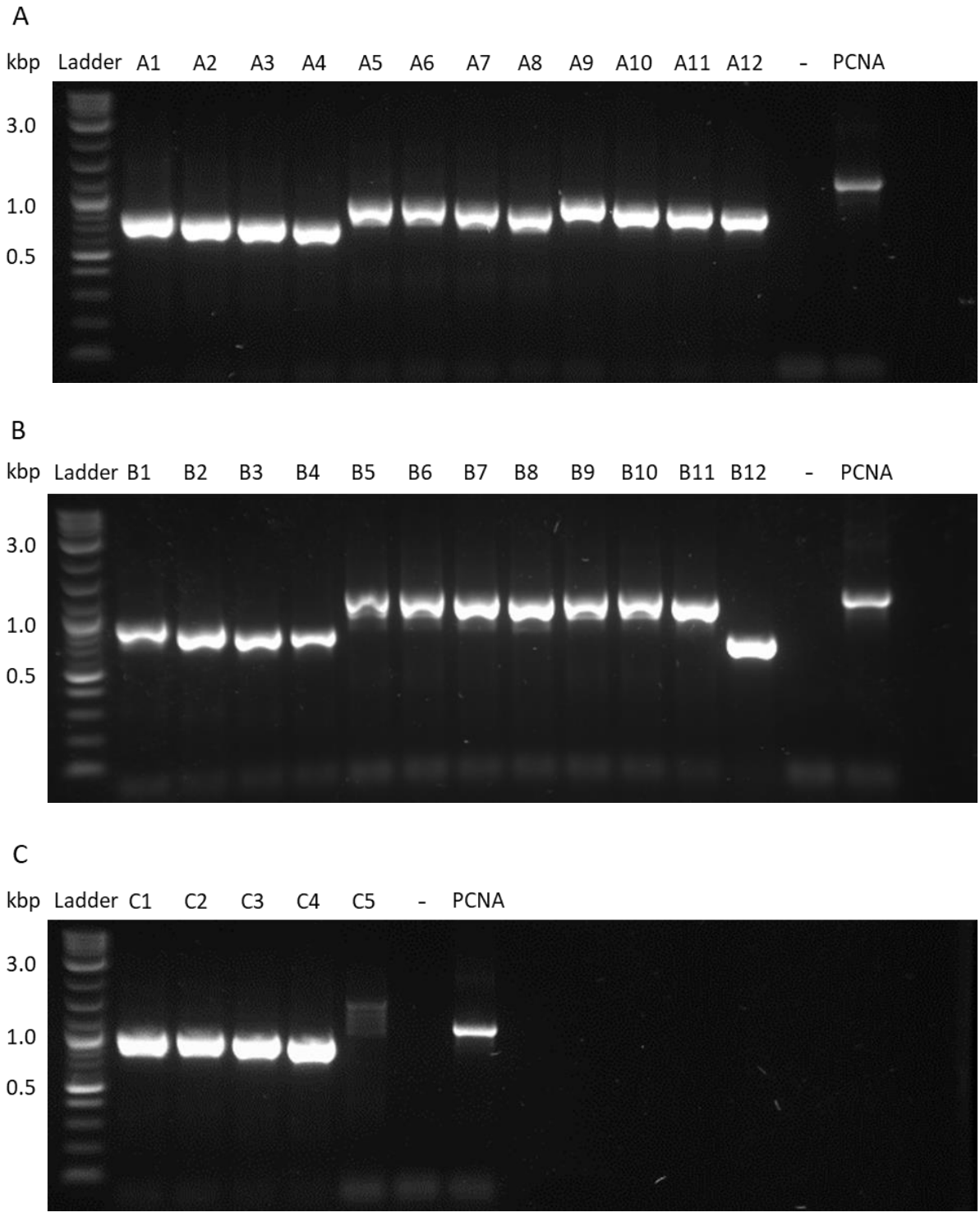


Figure 3.9. DNA agarose gel electrophoresis of successful clones from colony PCR screens. Construct plasmids were diluted to 10 ng/µL and were used for a final colony PCR. The first lane of each gel contains 6 µL of prepared NEB 2-Log DNA Ladder (0.1-10.0 kb). For each construct 10 µL of PCR mixture was loaded into each lane and the lanes are labelled with the construct code. Negative controls (-) contained nuclease free water instead of DNA as a template and the positive control (+) contained PCNA.

3.5 Summary

High Throughput Ligation Independent Cloning was used to allow production of numerous truncations of each protein target, to increase the targets available for structural experiments. This method is an alternative to traditional cloning techniques, negating the need for restriction enzymes, DNA ligases and alkaline phosphatases and allows screening of multiple genes simultaneously (Aslanidis & de Jong, 1990). Constructs were cloned into 2 different pET based expression vectors, with membrane protein genes cloned into pNIC28-Bsa4 and potentially soluble targets cloned into pNH-TrxT. Both of these vectors contained the sequence for an N-terminal 6x Histidine tag, but pNH-TrxT also contained the sequence for thioredoxin, an attached fusion protein known for increasing the solubility of recombinant protein. Truncations were made exclusively to the N-terminus of BB0027, BB0405, BB0406 and BB0562, as the C-terminal region is predicted to form the terminal strand of the β-barrel without any significant C-terminal extension beyond this. As the N-terminal of these proteins was predicted to be disordered, removal of this flexible region may facilitate better crystal packing.

Chapter 4 Recombinant protein expression and purification

4.1 Test expression and protein solubility screening

Small scale protein test expression was carried out using 50 mL of LB to assess whether the successfully cloned constructs produced soluble or insoluble protein (figure 4.1). Proteins were expressed in Rosetta™ *E. coli* cells and both the soluble and insoluble fractions were analysed using SDS-PAGE, as OMPs are expected to be insoluble. Proliferating cell nuclear antigen (PCNA) has an expected size of ~36kDa and was used as a robust soluble protein control. For BB_0095 constructs and BbCheR, the soluble fraction was purified using Ni-NTA resin only if no obvious protein was present in the insoluble fraction (figure 4.2) as these are not predicted to be membrane proteins. The expected protein sizes can be seen in table 4.1. Only one of the full-length OMPs (A9) expressed, whereas the other 3 OMPs had an absence of a band of the expected size in the soluble and insoluble fractions. This is expected of outer membrane proteins, as they contain a signal sequence which will result in translocation of the protein to the membrane of *E. coli* after expression, which can disrupt the membrane and inhibit cell growth and further expression. Neither the full length nor any of the truncated constructs expressed for BB_0027 (A5-A8). All other OMP constructs appeared to successfully express as bands are visible around the correct expected size for each construct in the insoluble fractions, meaning all of these constructs were potential proteins for large scale production.

BB_0095 was a potentially soluble hypothetical protein and the full-length protein, as well as 11 truncated versions, were successfully cloned. Expression was attempted for 7 of the truncations, as B12 and C2 were successfully cloned at a later date. While 7 of the BB_0095 constructs appeared to express protein, this was present in the insoluble fraction and no obvious bands could be seen in the soluble fraction. The remaining 3 BB_0095 constructs (C1, C3, C4) did not show any obvious bands in the insoluble fraction, so the soluble fractions were purified to see if they were expressed soluble at low levels (figure 4.2). Bands were not present for C1, C3 or C4 after Ni-NTA, so these constructs were not expressed. As BB_0095 was only of interest as a potentially soluble target, this protein was not continued as neither the full-length construct or truncated constructs produced soluble protein. The only BbCheR construct to be test expressed was the full-length protein (C5), as the truncated construct did

not clone. The soluble fraction for this protein showed a large band around the expected size, so the soluble fraction was purified using Ni-NTA to confirm this was not an *E. coli* protein. The purified soluble fraction showed a strong band around the expected size of 47.57 kDa (figure 4.2), so BbCheR was carried forward as a soluble protein target. The other bands present in the soluble fractions are likely soluble *E. coli* proteins, which are still expected to be seen at this stage.

Code	Construct (aa)	Accession number	Protein length (aa)	Vector	Expected size (kDa)	Expected size + tag (kDa)
A01	BB_0562 (1-180)	AAC66924.1	180	pNIC28-Bsa4	20	22
A02	BB_0562 (16-180)	AAC66924.1	165	pNIC28-Bsa4	18	21
A03	BB_0562 (27-180)	AAC66924.1	154	pNIC28-Bsa4	17	19
A04	BB_0562 (38-180)	AAC66924.1	143	pNIC28-Bsa4	16	18
A05	BB_0027 (1-212)	O51058.1	212	pNIC28-Bsa4	24	26
A06	BB_0027 (3-212)	O51058.1	210	pNIC28-Bsa4	24	26
A07	BB_0027 (18-212)	O51058.1	195	pNIC28-Bsa4	22	25
A08	BB_0027 (33-212)	O51058.1	180	pNIC28-Bsa4	21	23
A09	BB_0405 (1-203)	AAC66795.1	203	pNIC28-Bsa4	22	25
A10	BB_0405 (21-203)	AAC66795.1	183	pNIC28-Bsa4	20	23
A11	BB_0405 (30-203)	AAC66795.1	174	pNIC28-Bsa4	19	22
A12	BB_0405 (45-203)	AAC66795.1	159	pNIC28-Bsa4	18	20
B01	BB_0406 (1-203)	AAC66794.1	203	pNIC28-Bsa4	23	25
B02	BB_0406 (21-203)	AAC66794.1	183	pNIC28-Bsa4	20	23
B03	BB_0406 (43-203)	AAC66794.1	161	pNIC28-Bsa4	18	21
B04	BB_0406 (54-203)	AAC66794.1	150	pNIC28-Bsa4	17	19
B05	BB_0095 (1-181)	O51122.1	181	pNH-TrxT	22	36
B06	BB_0095 (4-181)	O51122.1	178	pNH-TrxT	22	35
B07	BB_0095 (24-181)	O51122.1	158	pNH-TrxT	19	33
B08	BB_0095 (37-181)	O51122.1	145	pNH-TrxT	18	32
B09	BB_0095 (1-130)	O51122.1	130	pNH-TrxT	16	29
B10	BB_0095 (4-130)	O51122.1	127	pNH-TrxT	15	29
B11	BB_0095 (24-130)	O51122.1	107	pNH-TrxT	13	27
B12	BB_0095 (37-130)	O51122.1	94	pNH-TrxT	11	25
C01	BB_0095 (1-157)	O51122.1	157	pNH-TrxT	19	33
C02	BB_0095 (4-157)	O51122.1	154	pNH-TrxT	19	33
C03	BB_0095 (24-157)	O51122.1	134	pNH-TrxT	16	30
C04	BB_0095 (37-157)	O51122.1	121	pNH-TrxT	15	29
C05	BB_0040/CheR (1-283)	O51069.1	283	pNH-TrxT	34	48
C06	BB_0040/CheR (15-283)	O51069.1	269	pNH-TrxT	32	45

Table 4.1. Details of the selected *B. burgdorferi* B31 protein targets. The code for each construct is shown in the first column, followed by the protein/gene name. The numbers next to the name indicate which amino acid number the construct begins and ends with. The NCBI accession number for the protein is given, followed by the length of the protein in number of amino acids. The plasmid the construct was cloned into is shown, followed by the ExPASY predicted protein size without the expression tag and with the tag.

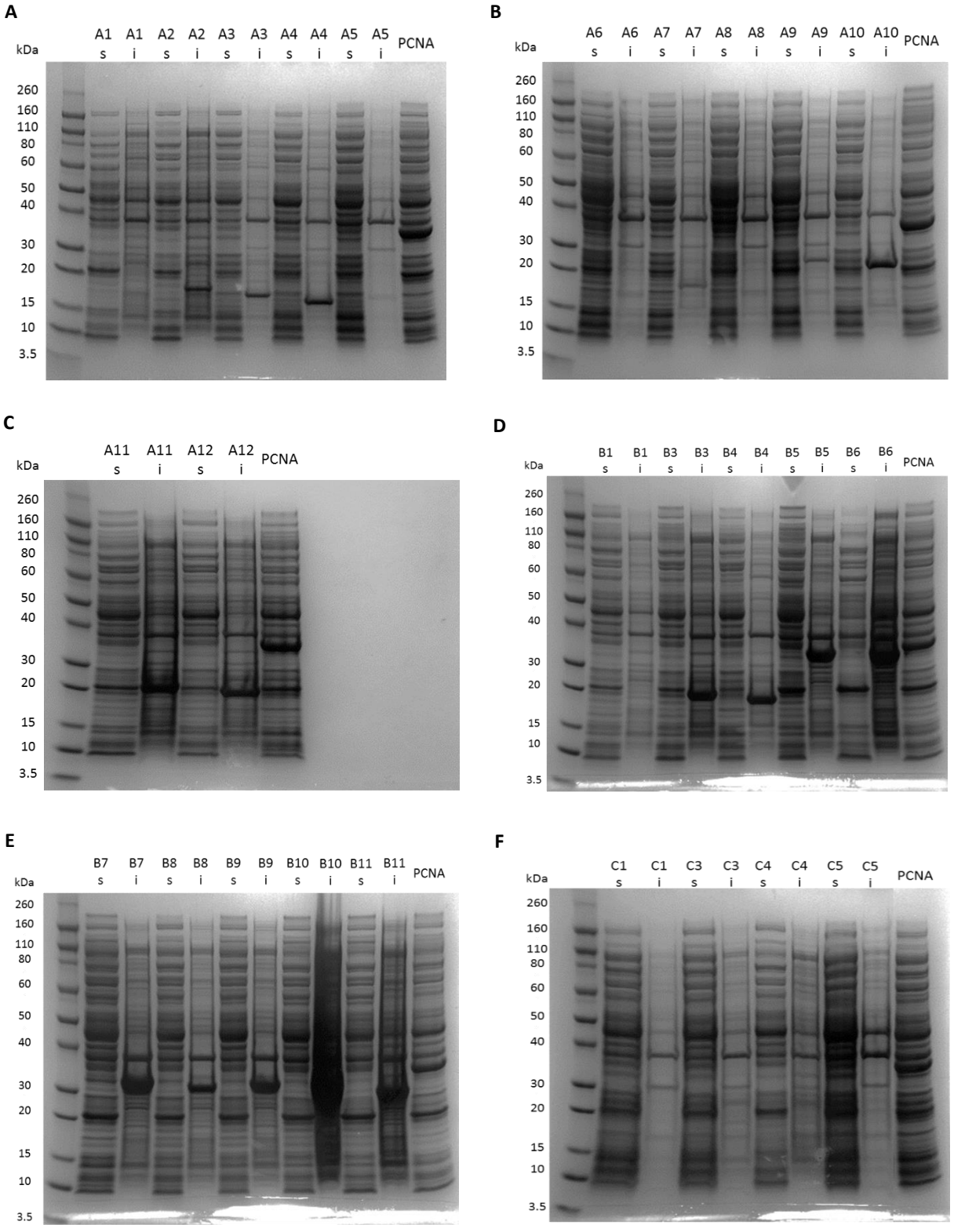


Figure 4.1. SDS-PAGE analysis of *Borrelia* construct test expressions. Constructs were expressed in *E. coli* Rosetta cells and soluble (s) and insoluble (i) samples analysed on Bolt™ 4-12% Bis-Tris Plus Gels (Invitrogen). The first lane in each gel shows 10 µL of Novex™ Sharp Pre-stained Protein Standard (Invitrogen) and the last lane is the PCNA control. **A** Constructs A1-A5. **B** Constructs A6-A10. **C** Constructs A11-A12. **D** Constructs B1 and B3-B6. **E** Constructs B7-B11. **F** Constructs C1 and C3-C5.

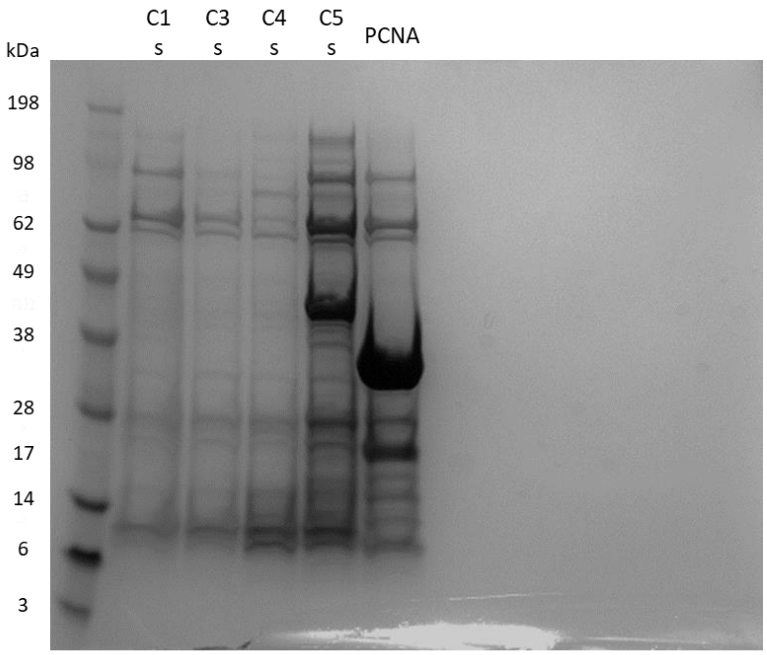


Figure 4.2. SDS-PAGE analysis of Ni-NTA purified soluble fractions. Soluble fractions were purified using Ni-NTA resin and samples analysed on Bolt™ 4-12% Bis-Tris Plus Gels (Invitrogen). Lane 1: 10 µL of SeeBlue® Plus2 Pre-Stained Protein Standard. Lane 2-4: Purified C1-C4 (BB_0095) soluble fractions. Lane 5: Purified C5 (BB_0040) soluble fraction. Lane 6: Purified PCNA control soluble fraction.

4.2 Large scale expression, purification and refolding of *Borrelia* OMP constructs

Membrane protein constructs were selected for large scale expression based on estimated test expression yield. For structural studies, such as crystal screens, a truncated BB_406 construct (B4) was expressed in 3 L of LB overnight using 1 mM IPTG for induction. For binding studies, a construct was expressed for each OMP target; BB_562, BB_405 and BB_406. These were: A3, A10 and B4 as these were successfully expressed and purified on a large scale and were as close to full-length as possible. BAPKO_0422 and superoxide dismutase A (SodA) from *B. afzelii* were produced to provide control proteins for binding studies.

4.2.1 Purification of inclusion bodies

While expression in the form of inclusion bodies is not ideal, due to the need to subsequently denature and refold the protein, it did produce a high expression yield. After lysis of the cells, the soluble and insoluble material was separated using centrifugation and the insoluble pellet was subject to a series of washes described in section 2.2.8.1. This multi-step washing procedure allows higher purity to be achieved prior to any chromatography techniques, as it removes some contaminating insoluble *E. coli* proteins and cell debris. The supernatant from each wash was analysed using SDS-PAGE and shows removal of contaminating proteins with minimal loss of the target protein. The soluble fraction appears dark as it contains soluble *E. coli* proteins and is not purified. Analysis of B4 inclusion body preparation can be seen as an example in figure 4.3 below. After washing, the pellet was solubilised in buffer containing 8 M urea (Denaturing Buffer) and was clarified to remove insoluble debris.

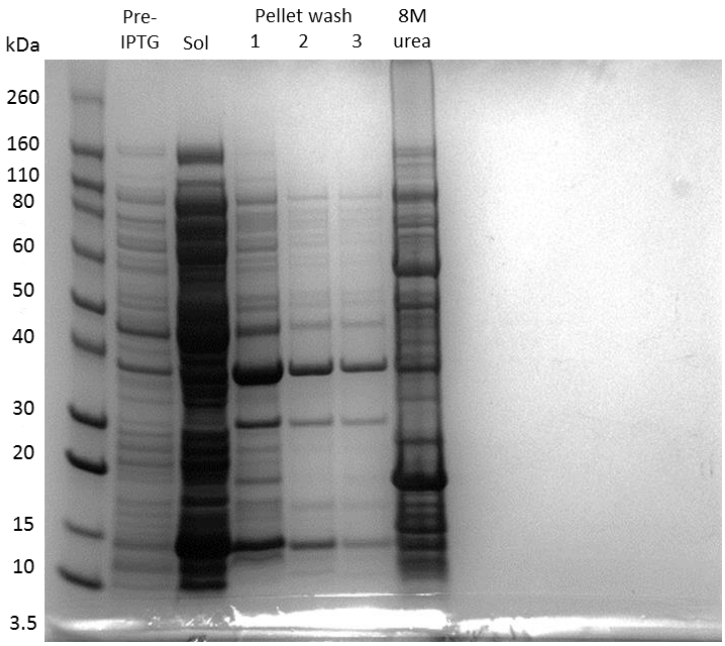


Figure 4.3. SDS-PAGE analysis of B4 inclusion body preparation. A multi-step washing procedure was used to increase inclusion body purity prior to chromatography purification. Samples were analysed on a Bolt™ 4-12% Bis-Tris Plus Gel (Invitrogen). Lane 1: 10 µL of Novex™ Sharp Pre-stained Protein Standard (Invitrogen). Lane 2: A sample taken before inducing protein expression with IPTG. Lane 3: Soluble fraction. Lanes 4-6: Supernatant from inclusion body washes. Lane 7: B4 solubilised in 8 M urea denaturant.

4.2.2 IMAC and on-column refolding

Denatured membrane proteins were loaded onto a Ni-NTA pre-packed column using an AKTA Prime FPLC and were refolded using a gradual removal of denaturant and an addition of detergent (0.1% DDAO). The flow rate and length of the refolding gradient was optimised by (Dyer et al., 2015) to improve protein yield, allow proper refolding of proteins and to potentially prevent misfolding and aggregation. The column containing the bound refolded protein was then washed with a buffer containing a low concentration (20 mM) of imidazole to remove any weakly bound contaminants and the protein of interest was eluted using a high imidazole concentration (300 mM). Figure 4.4 shows a screenshot of the chromatogram produced by the PrimeView software for the purification and refolding of B4. The first peak corresponds to protein loading and the final sharp peak shows elution of B4 from the column. Fractions were collected and analysed using SDS-PAGE, along with the flow-through from loading to identify any unbound protein (figure 4.5). The approximate size of B4 is 19 kDa and bands can be seen around this size in the majority of the elution fractions, confirming that the peak corresponds to the elution of B4. In comparison to the protein prior to purification shown in figure 4.3, many contaminants have been removed by the IMAC purification step.

However, a high level of purity is required for crystallography and fractions were pooled and concentrated for further purification.

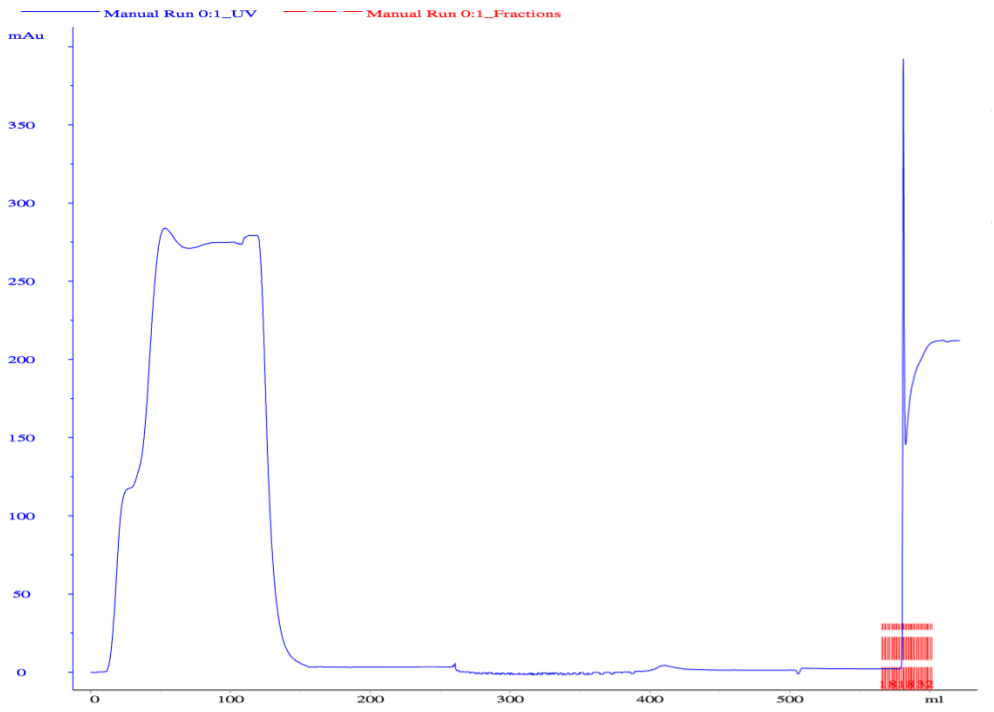


Figure 4.4. Screenshot of the AKTA Prime chromatogram of B4 IMAC. Denatured B4 was loaded onto a 5 mL GE Healthcare HiTrap HP Ni-NTA column and was refolded by gradual removal of denaturant. The refolded protein was then washed and eluted with imidazole . The absorbance at 280 nm in mAU is given on the Y-axis and the volume in mL is given on the X-axis. The blue line represents the UV absorbance trace and the red lines show fractions collected.

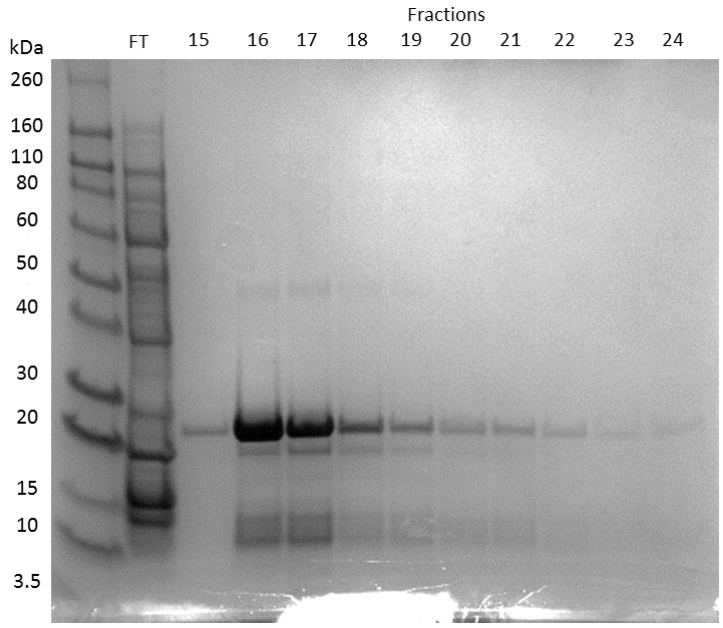


Figure 4.5. SDS-PAGE analysis of Immobilised Metal Affinity Chromatography purification of B4. Fractions collected from the elution of B4 were analysed on a Bolt™ 4-12% Bis-Tris Plus Gel (Invitrogen). Lane 1: 10 µL of Novex™ Sharp Pre-stained Protein Standard (Invitrogen). Lane 2: The flow-through from protein loading. Lane 3-12: Fractions collected from the elution of B4.

4.3.3 SEC of refolded protein

After IMAC, proteins were concentrated and subject to Size Exclusion Chromatography as a final purification step. A Superdex 75/300 column was equilibrated with buffer containing DDM detergent, as this is the most successful detergent used for membrane protein crystal structures in the Protein Data Bank (Parker & Newstead, 2016). The concentrated protein (~3 mg/mL) was injected onto the column using a flow loop and was eluted at a flow rate of ~0.5 mL/min, depending on column pressure. Figure 4.6 shows the chromatogram produced from the elution of B4. Fractions were collected during elution and fractions under the peak were collected and analysed using SDS-PAGE (figure 4.7). While a calibration graph was produced as seen in section 2.2.7.4, molecular weight estimation of membrane proteins using SEC is not accurate as detergent molecules mimic the lipid membrane and bind to the hydrophobic core of the protein, forming a protein-detergent complex (PDC). The number of detergent molecules bound is unknown and dependent on the detergent concentration and micelle properties. Therefore, the PDC molecular weight and shape is different to that of the protein alone and can affect separation during SEC (Anandan & Vrielink, 2016). A double peak can be

seen in figure 4.6 and fractions under the second peak were pooled and the protein concentrated to ~5 mg/mL for use in initial crystal screens.

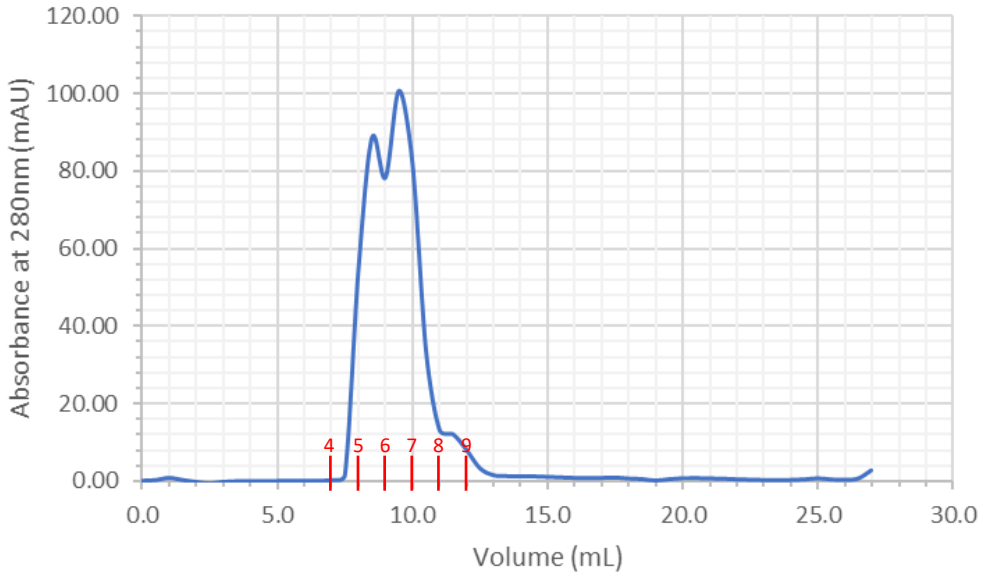


Figure 4.6. Size Exclusion Chromatography chromatogram of B4. B4 was injected onto a GE Healthcare Superdex 75/300 GL column equilibrated with buffer containing 0.1% DDM and elution fractions were collected. The absorbance at 280 nm in mAU is given on the Y-axis and the elution volume in mL is given on the X-axis.

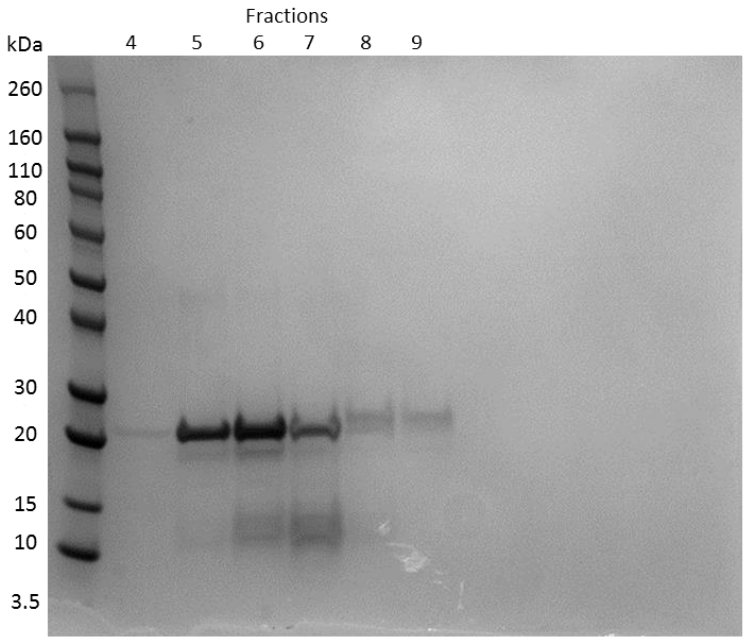


Figure 4.7. SDS-PAGE analysis of B4 elution fractions after Size Exclusion Chromatography. B4 was injected onto a GE Healthcare Superdex 75/300 GL column equilibrated with buffer containing 0.1% DDM and elution fractions were collected. Samples were analysed on a Bolt™ 4-12% Bis-Tris Plus Gel (Invitrogen). Lane 1: 10 µL of Novex™ Sharp Pre-stained Protein Standard (Invitrogen). Lane 2-7: Fractions from B4 elution.

4.4 Expression and purification of a chemotaxis protein methyltransferase

4.4.1 Large scale soluble protein expression and purification

Following the successful test expression of full-length BbCheR (BB_0040, C5), the protein expression procedure was scaled up to use 3 L of culture as large quantities of protein are required for structural studies. Expression was induced in *E. coli* Rosetta cells by the addition of IPTG, the cells were lysed and the soluble fraction was retained for purification using IMAC. Soluble and insoluble fractions were analysed using SDS-PAGE to confirm that the protein had been expressed (figure 4.8). Fractions were collected from a Ni-NTA column, following elution of the protein with imidazole, and analysed to identify which contained the protein of interest and to assess the purity (figure 4.8). In the soluble fraction prior to purification, a large band can be seen around 48 kDa, which is the predicted size of BbCheR with its expression tag, but isn't clear due to the large amount of *E. coli* contaminants present. After purification using a Ni-NTA column, this band becomes more prominent as the protein becomes more concentrated upon elution and the level of contaminants decreases. Despite purification, many other bands are still present in the elution fractions and these contaminants are due to *E. coli* proteins that have interacted with the Ni-NTA resin. As IMAC is used an initial purification step, this is expected.

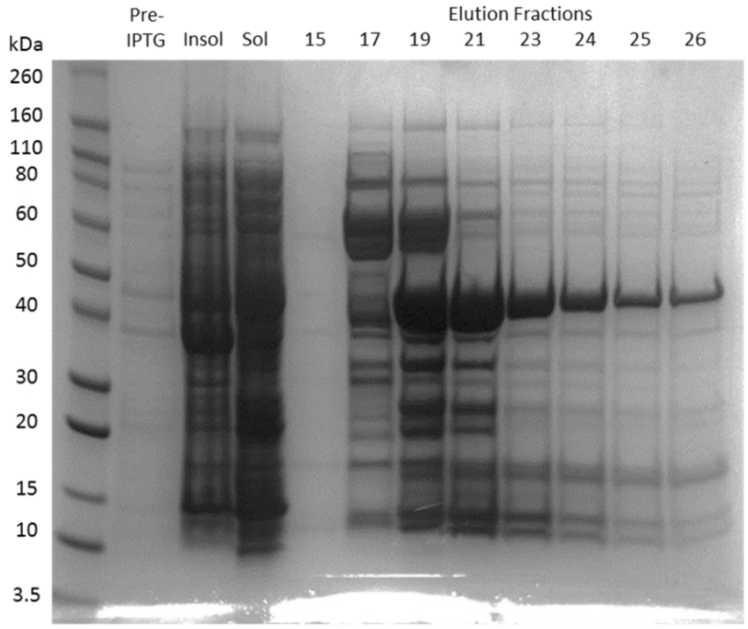


Figure 4.8. SDS-PAGE analysis of large-scale soluble expression and Ni-NTA purification of CheR. Recombinant BbCheR was expressed in *E. coli* Rosetta cells and was purified using Ni-NTA resin. Samples were analysed on a Bolt™ 4-12% Bis-Tris Plus Gel (Invitrogen). Lane 1: 10 µL of Novex™ Sharp Pre-stained Protein Standard (Invitrogen). Lane 2: A sample of cells before induction with IPTG. Lane 3: Insoluble fraction. Lane 4: Soluble fraction prior to purification. Lane 5-12: Elution fractions from Ni-NTA purification.

4.4.2 Expression tag removal and IMAC rebind

BB_0040 (BbCheR, C5) was cloned into a vector containing a TEV protease cleavage site followed by a 6xHis-Thiorodoxin expression tag, which is a large tag with the potential to interfere with downstream experiments such as protein crystallography. For this reason, complete removal of the tag was attempted. The protein was dialysed to remove imidazole and was incubated with a TEV protease. The TEV treated protein was then passed back through Ni-NTA resin as without the 6x His tag, the protein should not bind to the column and should elute in the flow-through. Increasing concentrations of imidazole were used to ensure all of the target protein was eluted from the column, as well as the protease and the tag. Samples were analysed using SDS-PAGE (figure 4.9). The size of the protein with and without the tag can be seen in table 4.1. A sample of uncut protein was ran to allow size comparison after TEV incubation and to assess whether tag cleavage was successful.

Following tag cleavage, the 47.57 kDa band of the uncut protein is minimal in the flow-through and imidazole elution fractions, suggesting successful removal of the tag. Some of the cleaved protein did appear in the flow-through, around the expected size of 34 kDa, but the majority can be seen in the 20 mM and 40 mM imidazole elution. These concentrations of imidazole are used as wash steps to remove weakly bound proteins, so cleaved protein can still be expected to elute in these fractions. Some of the BbCheR protein does not elute until 100 mM and 300 mM imidazole concentrations, which is likely due to charged surface regions on the protein interacting with the nickel resin. The TEV protease is ~27 kDa and a band of this size is present in the 300 mM imidazole elution. Finally, the 6x His-Thiorodoxin tag (~13 kDa) elutes at 300 mM imidazole as the tag binds strongly to the Ni-NTA resin. The 20 mM and 40 mM elution fractions were combined and the protein was concentrated in preparation for further purification.

A Western blot was performed on uncut protein and the fraction following treatment with the TEV protease, using an anti-His antibody to confirm tag removal (figure 4.10). A strong band can be seen for the uncut protein, which would be expected as the protein still contains the 6x histidine tag. The uncut protein band is not present in any of the other lanes and bands are absent for the 20 mM and 40 mM imidazole elution fractions, suggesting removal of the

tag was achieved. For the 100 mM and 300 mM imidazole elution fractions, a band is present in each lane and does not appear to be the same molecular weight as the uncut protein (figure 4.10).

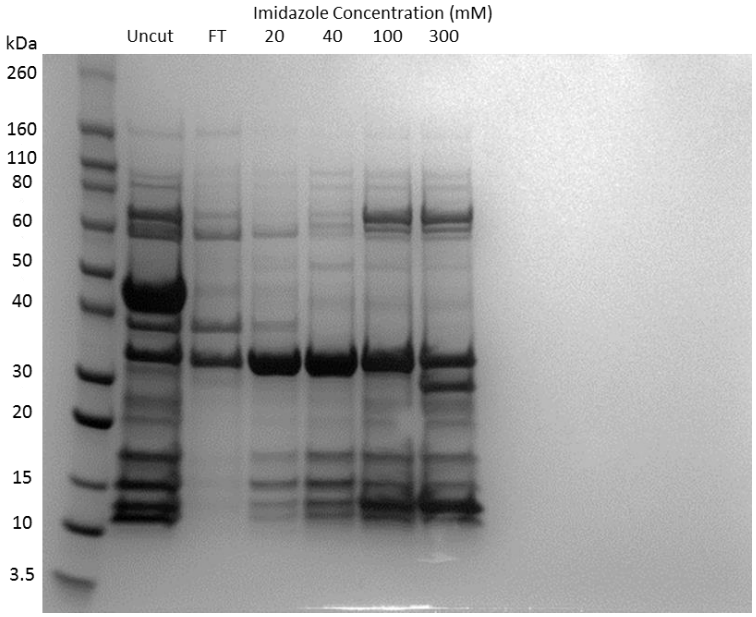


Figure 4.9. SDS-PAGE analysis of Ni-NTA rebind of CheR after TEV cleavage of the expression tag. Purified BbCheR was incubated with TEV protease to remove the His-Thioredoxin tag and was passed through Ni-NTA resin. Samples were analysed on a Bolt™ 4-12% Bis-Tris Plus Gel (Invitrogen). Lane 1: 10 µL of Novex™ Sharp Pre-stained Protein Standard (Invitrogen). Lane 2: A sample of protein prior to TEV cleavage. Lane 3: The flow-through from loading the protein onto the resin. Lanes 4-7: Fractions from elution with increasing concentrations of imidazole (given in mM).

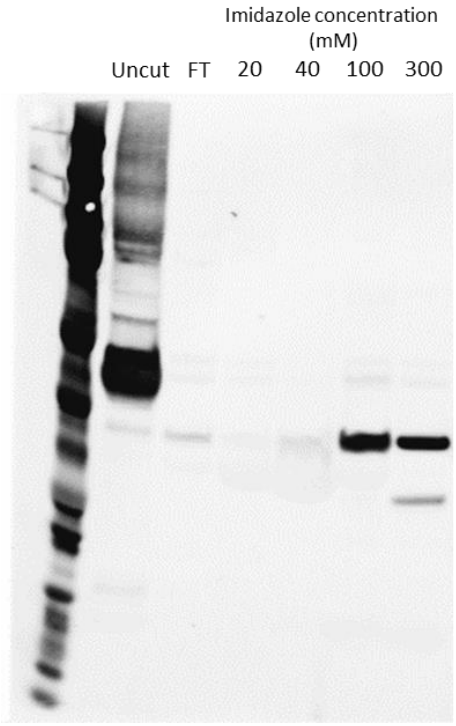


Figure 4.10. Western blot analysis of Ni-NTA rebind of CheR after TEV cleavage of the expression tag. The membrane was incubated with 1:10000 Mouse anti-HisTag Monoclonal Antibody (ProteinTech) as the primary antibody and 1:15000 IgG (H+L) Cross-Adsorbed Goat anti-Mouse, Alexa Fluor® 680 (Invitrogen™) was used as the secondary antibody. The membrane was imaged at 680nm using a LICOR Odyssey infrared imaging device (LI-COR Biosciences). Lane 1: Precision Plus Protein™ Dual Xtra Prestained Protein Standards (BioRad). Lane 2: A sample of protein prior to TEV cleavage. Lane 3: The flow-through from loading the protein onto the resin. Lanes 4-7: Fractions from elution with increasing concentrations of imidazole (given in mM).

4.4.3 SEC of BbCheR

A Superdex 75/300 column was equilibrated and calibrated as described in section 2.2.6.4 and concentrated BbCheR (~10 mg/mL) was injected onto the column using a flow loop at a flow rate of ~0.5 mL/min, depending on column pressure. Figure 4.11 shows the chromatogram produced from the elution of BbCheR and the protein peak is seen at an elution volume of ~12 mL, which corresponds to a molecular weight of approximately 32 kDa. The expected size of BbCheR with the expression tag removed was 35 kDa, so this suggests the eluted protein is monomeric and not aggregated. Fractions were collected during elution and fractions under the peak were collected and analysed using SDS-PAGE (figure 4.12). A high yield of protein was obtained after gel filtration and fractions 8-10 were pooled and concentrated to ~8.89 mg/mL for use in crystal screen set up.

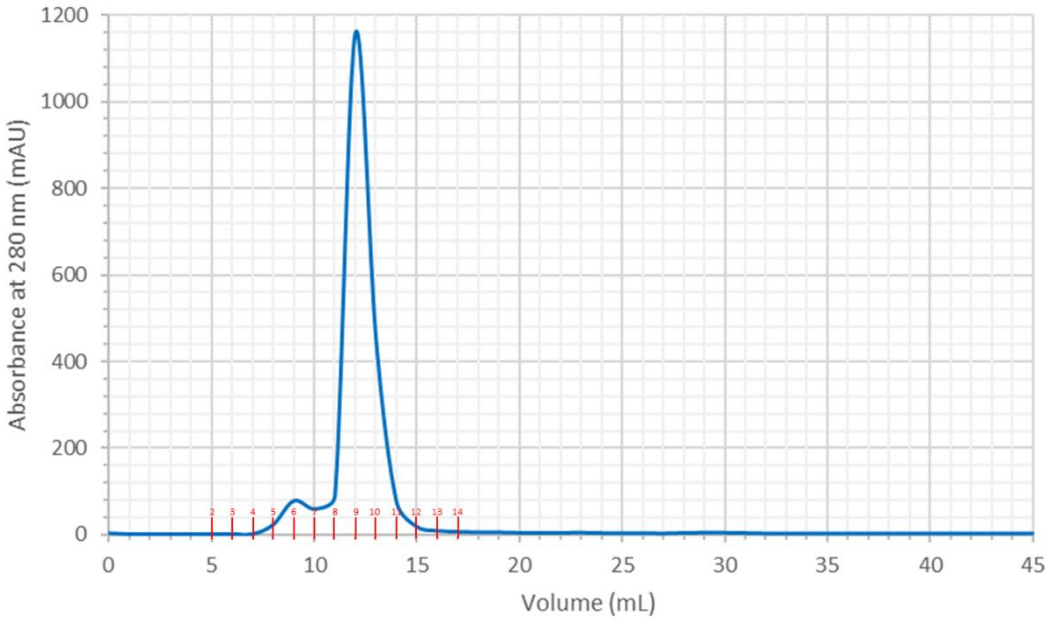


Figure 4.11. Size Exclusion Chromatography chromatogram of BbCheR. Cleaved BbCheR protein was injected onto a GE Healthcare Superdex 75/300 GL column and elution fractions were collected. The absorbance at 280 nm in mAU is given on the Y-axis and the elution volume in mL is given on the X-axis.

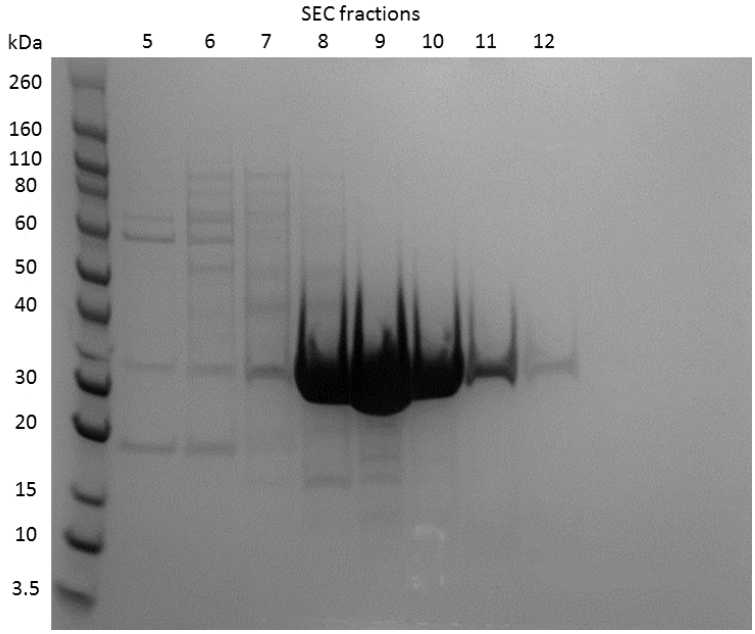


Figure 4.12. SDS-PAGE analysis of BbCheR fraction after Size Exclusion Chromatography. Cleaved BbCheR protein was injected onto a GE Healthcare Superdex 75/300 GL column and elution fractions were collected. Samples were analysed on a Bolt™ 4-12% Bis-Tris Plus Gel (Invitrogen). Lane 1: 10 µL of Novex™ Sharp Pre-stained Protein Standard (Invitrogen). Lane 2-9: Fractions from BbCheR elution.

4.5. Vapour diffusion crystallography trials of a BB_0406 construct

Hanging drop vapour diffusion screens were set up as outlined in section 2.2.13.1 using ~5 mg/mL of truncated BB_0406 (B4) protein. DDM was used as the detergent in the protein buffer due to its success in the crystallisation of membrane proteins in the PDB. Prior to preparation of B4 in buffer containing DDM, tetraethylene glycol monooctyl ether (C8E4) was used as this detergent was used for crystallisation of OmpA from *E. coli*, as well as other integral β-barrel OMPs. During buffer exchange into the C8E4 buffer, the protein solution precipitated. Many of the drops showed immediate precipitation upon mixing of the reservoir solution with the protein. Trays were checked consistently over a period of 3 months, with one drop containing 0.2 M ammonium sulfate, 0.1 M CHES 10.0, 14% w/v PEG 4000 showing formation of small crystals measuring approximately 10-20 micrometres after this time (figure 4.13). Attempts to collect the crystals for x-ray analysis proved challenging, as they were difficult to manipulate due to the formation of a membrane-like layer over the drop. Cryo-solution containing the reservoir solution with the addition of 10% glycerol was added to the drop to try and dislodge the crystals, but this resulted in the crystals dissolving. Optimisation trays of this condition were set up using thawed protein from the same batch, but these were unsuccessful.

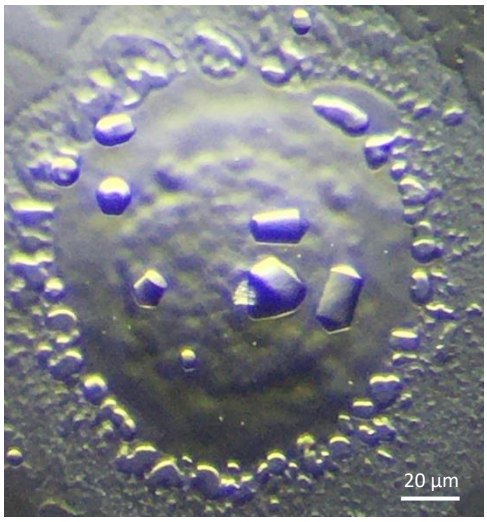


Figure 4.13. Vapour diffusion crystallisation trials of truncated BB_0406. Photograph of a protein drop after 3 months of incubation at 20°C. The drop was prepared using 1 µL of protein solution and 1 µL of reservoir solution, which contained 0.2 M ammonium sulfate, 0.1 M CHES 10.0, 14% w/v PEG 4000. Small crystals measuring approximately 10-20 micrometres can be seen.

4.6. ALBI of *Borrelia* OMPs with human factor H

Due to time limitations, an affinity ligand binding immunoblot was carried out using the dot blot method, as protein can be applied directly to a PVDF membrane without the requirement of a native PAGE and a blot transfer. All protein samples used, including the FH control, were at a concentration of ~1 mg/mL and 1-4 µL (1-4 µg) was applied to the membrane. Complement factor-H from human plasma was used as a control for the anti-FH antibody. *Borrelia* SodA was used a negative control, as it is not known to bind human factor-H. A positive control is not present and a suitable factor-H binding protein, such as a BbCRASP, would be needed for future binding studies. The membrane was incubated with 30 µg/mL of complement factor-H, which is approximately 10 times lower than the concentration found in human plasma and this was due to the expense of purchasing factor-H. No binding can be detected for any of the *Borrelia* OMPs (figure 4.14), including BAPKO_0422, which has previously been found to bind to FH (Dyer et al., 2015). However, these results do agree with the most recent literature, where BB_0405, BB_0406 and BAPKO_0422 were found not to bind FH (Shrestha et al., 2017). A dot can be seen for 1 µL of SodA likely due to contamination with factor H pipetted on the row below, as no dots are seen with the increasing concentrations of the protein.

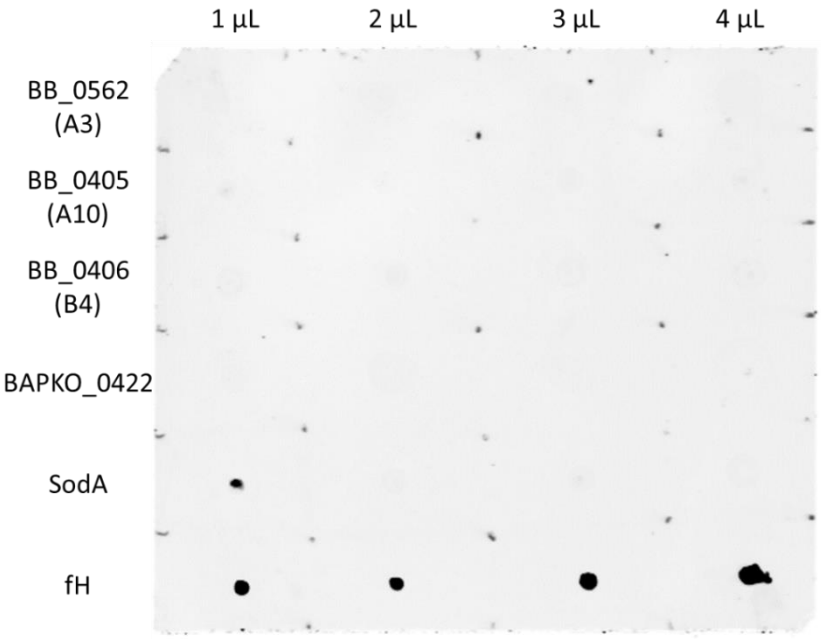


Figure 4.14. Affinity ligand binding dot-blot of *Borrelia* OMPs with complement human factor H. A PVDF membrane was marked with a grid and was activated with methanol. Each protein was prepared to 1 mg/mL and 1-4 µL of protein solution was pipetted onto the membrane. *Borrelia* SodA was used as a negative control as it is not known to bind human factor-H. Complement human factor-H was used as a control for the anti-FH antibody.

4.7 Discussion

4.7.1 *B. burgdorferi* β-barrel outer membrane proteins

Borrelia contain many lipoproteins (>100) but very few integral OMPs, with only around 9 identified in *Borrelia burgdorferi s.l* including: P66 (BB0603, Oms66), P13 (BB0034), BesC (BB0142), BamA (BB0795), DipA (BB0418, Oms38), BB0838, BB0405 and BB0406 (Bunikis et al., 2008; Dyer et al., 2015; Kenedy et al., 2016; Kenedy et al., 2014; Lenhart & Akins, 2010; Noppa et al., 2001; Skare et al., 1997; Thein et al., 2012). This project focused on 4 predicted small β-barrel OMPs from *Borrelia burgdorferi s.s*; BB_0027, BB_0405, BB_0406 and BB_0562, all predicted to be similar in structure to *E. coli* OmpA. As small angle x-ray scattering (SAXS) has previously been used to acquire some structural information on a BB_0405 homologue, BAPKO_0422, as well as BB_0406 and BB_0562 (Brown, 2016), the main aim of this project was to successfully crystallise these proteins, attempts of which had previously failed. SAXS data allowed a molecular envelope for these *Borrelia* proteins to be generated and were consistent with a small β-barrel structure, with homology modelling fitting that of an 8-stranded β-barrel (Brown, 2016).

Paralogues BB_0405 and BB_0406 have been the focus of a few studies, while BB_0027 and BB_0562 remain uncharacterised. Both BB_0405 and BB_0406 have been found to be amphiphilic proteins, with localisation to the membrane of *B. burgdorferi* and an ability to form pores in large unilamellar vesicles (Kenedy et al., 2016). These OMPs have also been shown to be immunogenic as antibodies against these proteins demonstrated to be borreliacidal, but only BB_0405 has been shown to be essential for infectivity in mouse models (Kung et al., 2016; Shrestha et al., 2017). A study found that BB_0405 and BB_0406 do not bind to human factor H (Shrestha et al., 2017) and recent research investigating the role of BB_0406 identified that this OMP binds to the basal membrane component laminin, suggesting a role for this protein in later infection and dissemination, rather than immune evasion (Bista et al., 2020). The importance of this protein in later infection is also suggested by previous findings showing a weak antibody response to BB_0406 in the first 6 weeks of infection in baboon sera, with a robust response seen following this time (Shrestha et al., 2017).

Test expressions were carried out in *E. coli* Rosetta cells for each full length OMP, as well as for each truncated protein. Both the soluble and insoluble fractions were analysed and expression was seen for truncations of BB_0405 and BB_0406. The full length proteins still contained the signal sequence and were therefore not expected to express, as they will likely translocate to the *E. coli* outer membrane, potentially disrupting the membrane and inhibiting further cell growth and protein expression. This was seen by an absence of bands of the expected size in both the soluble and insoluble fractions of the full length OMPs. No expression was seen for any of the BB_0027 truncations and this was therefore discarded as a target for large scale expression. Optimisation of expression conditions could result in protein expression of this target, such as testing differing IPTG concentrations and expression temperatures. Although many targets were available, large scale production of recombinant protein as insoluble inclusion bodies is time consuming and therefore one target was initially selected for expression. A BB_0406 construct with the first 53 amino acids removed, labelled as B4, was chosen due to its visually high expression yield on SDS-PAGE from test expression attempts and its absence of the two cysteine residues at positions 31 and 40, which caused issues with formation of high order oligomers during previous research (Brown, 2016).

Expression using 3 L of *E. coli* culture yielded around 5 mg of recombinant BB_0406 (B4) following refolding and SEC, providing limited quantities for use in crystallography experiments. Protein was refolded over a long gradient of decreasing urea concentration, based on previously successful methods for refolding these *Borrelia* proteins. However, confirmation of successful refolding would be useful as misfolded protein could form aggregates and impact crystallisation experiments. This would be especially useful due to the truncations made based on predicted secondary structure, as essential structural components could have been removed and prevented proper refolding of the protein. Circular dichroism is useful for rapid determination of protein secondary structure and is defined by the disproportionate absorption of left-handed and right-handed circularly polarised light, which occurs in the presence of chiral chromophores (Greenfield, 2006). This technique generates a spectra which allows the percentage of secondary structure elements, such as β-strand, to be calculated, providing an indication on whether proper folding was present.

Size exclusion chromatography (SEC) was used as a final purification step, separating proteins based on their size and molecular weight. A calibration curve of proteins with known sizes was produced (figure 2.2, section 2.2.7) and allows the molecular weight of the protein of interest to be calculated based on its elution volume. For the production of the *Borrelia* OMPs, SEC was only used for the purpose of purification and not for molecular weight estimation. This is due to the addition of detergent to the protein buffers, which is required to maintain membrane protein solubility by binding to the hydrophobic core and mimicking the proteins natural environment in the lipid bilayer. As an unknown number of detergent molecules bind to the protein, the molecular weight of the protein-detergent complex is difficult to determine and protein separation can be affected (Anandan & Vrielink, 2016). During SEC, the detergent used was exchanged from DDAO to DDM when protein was to be used for crystallography trials. This was because DDM is the most common detergent used for membrane protein crystal structures in the Protein Data Bank (Parker & Newstead, 2016). However, the majority of membrane proteins in the PDB are transmembrane α-helical proteins, rather than β-barrels. For this reason, detergent exchange from DDAO to tetraethylene glycol monooctyl ether (C8E4) was also attempted, as this detergent was used in the crystallisation of OmpA from *E. coli* (PDB: 1BXW). This resulted in immediate precipitation of the protein sample, so this detergent was not used for crystallisation trials.

Another common detergent used for β-barrel membrane protein crystallisation is DDAO, which was used as the detergent for purification using IMAC, or for proteins intended for binding studies. While DDAO has been successfully used for crystallisation of β-barrel proteins in the PBD, it was not used for crystallisation attempts during this research. This is because previous attempts to crystallise a BB_0406 homolog, BAPKO_0422, using DDAO as the detergent, were unsuccessful. This research therefore focussed on alternative detergents. DDM was not used until the crystallography stage, as it is a much more expensive detergent than DDAO. The truncated BB_0406 protein (B4) elution showed a poorly defined double peak, potentially due to differing amounts of bound detergent to protein molecules or formation of oligomers. Analysis using SDS-PAGE did not show evidence of oligomerisation, but this was under denaturing conditions and a native-PAGE would be useful to identify if this

had occurred. Samples following SEC were centrifuged at high speeds to try to remove any potential aggregates.

Crystallisation trays were set up using the vapour diffusion method for the BB_0406 truncation, B4, which had the first 53 amino acids removed. As crystallography had previously been attempted using full length *Borrelia* OMPs, truncated protein was used in this research to potentially remove disordered, flexible regions and loops that could prevent crystal formation. MemPlus™ reagents from Molecular Dimensions were used for crystal screens as they were developed based on successful conditions used for β-barrel membrane proteins in the PDB (Newstead et al., 2008). Further screens could be used for future experiments as this only provided a limited number of conditions. Upon mixing of the well solution to the protein drop, immediate precipitation was seen for many of the conditions tested. Some crystal formation was seen for one condition containing 0.2 M ammonium sulfate, 0.1 M CHES 10.0, 14% w/v PEG 400 after 3 months. As crystal formation was slow, difficulties were encountered when attempting to pick the crystals due to a membrane-like layer on the drop. This made manipulation of the crystals difficult and as they were fragile, some broke. Addition of reservoir solution to the drop was attempted to try loosen the crystals, but dilution of the drop resulted in the remaining crystals dissolving, so x-ray diffraction could not be performed. This means it is still unclear as to whether protein crystals were formed or if these were crystals of detergent or salt.

For membrane protein crystallography, detergent choice is a critical factor in the formation of high quality crystals (Parker & Newstead, 2016). For crystallisation trials of B4, DDM was the only detergent used. While this detergent has been shown to be most popular in membrane protein crystallography, it forms large micelles which could potentially hinder protein-protein interactions needed for formation of well diffracting crystals (Newstead et al., 2008). Future success of crystallising this protein, as well as other OMP targets, will likely require screening of a wide range of different detergents. The polyhistidine tag was not removed prior to crystallisation experiments and this, along with the TEV protease cleavage site sequence, could have resulted in a flexible region hindering crystal packing. For future experiments, the expression tag could be removed via cleavage with a TEV protease or

untagged protein could be produced, but the latter would not allow purification and refolding using immobilised metal affinity chromatography.

The *B. afzelii* OmpA-like protein BAPKO_0422 was identified as a factor-H binding protein by Dyer et al, using affinity binding immunoblot assays (Dyer et al., 2015). However, more recent research by Shrestha et al did not find any factor-H binding by BAPKO_0422, or its *B. burgdorferi* homologue BB_0405 and BB_0406 (Shrestha et al., 2017). Due to this conflicting data, this project attempted FH binding using an affinity ligand binding dot-blot on *B. burgdorferi* OMPs; BB_0562, BB_0405 and BB_0406 and included the *B. afzelii* OMP BAPKO_0422. The research undertaken by Dyer et al (2015) identifying BAPKO_0422 as a FH binding protein loaded a much higher concentration of protein onto the SDS-PAGE gel used for blot transfer, using up to 50 µg for ALBI assays. In contrast, for the ALBI assays carried out by Shrestha et al (2017), only 1 µg of protein was used, which is around the concentration recommended for Western blotting. A similar concentration of target protein was applied to the membrane in the dot-blot included in this research. The concentration of protein used for the FH binding studies could therefore be significant for whether binding can be detected or not, as the use of high concentrations showed positive binding. The FH used by Dyer et al was not pure and contained traced of contaminating proteins. It is possible that BAPKO_0422 binds to one of these contaminating proteins, which are likely to bind to FH as they co-purified with it. The concentration of FH used by Dyer et al was higher than the concentration used for this research at 73 µg/ml and it may be that the binding is only seen when using higher concentrations of FH, as the concentration of FH binding contaminants will also be increased.

Another major difference in these two studies was the use of a positive control, as Shrestha et al included *B. burgdorferi* CRASP-1 in their experiments, a well-known FH binding protein that showed positive bands. The binding studies by Dyer et al and in this research project did not include a positive control and this would be essential for any future binding studies. Due to time limitations, a dot-blot was used for the FH binding study in this research project meaning proteins were not separated via SDS-PAGE prior to application on the membrane. To try and reduce the presence of contaminating proteins, pure protein was used. However, existing pure protein stocks had to be used and only truncated versions of BB_0405, BB_0406

and BB_0562 available for use in the dot blot. If these proteins were factor H binding, the binding site is unknown and could have been removed by the truncation made. BAPKO_0422 and sodA were full length pure protein. Ideally, the full length proteins would have been separated using native-PAGE and Western Blotting used to transfer the proteins onto the membrane. While all current literature agrees that BB_0405 and BB_0406 do not bind human factor-H, further studies could be necessary to confirm whether BAPKO_0422 from *B. afzelii* indeed binds human factor-H or binds to a factor-H binding protein. As BB_ 0406 was recently found to bind the ECM component laminin (Bista et al., 2020), future binding studies for other small *Borrelia* OMPs could focus on other components of the ECM such as heparin, fibronectin and collagen.

In summary, this project aimed to acquire structural and functional information on a small group of predicted *Borrelia* OMPs; BB_0405, BB_0406, BB_0027 and BB_0562, with truncations made to these proteins for crystallography attempts. BB_0406 with the first 53 amino acids removed (B4) was soluble in DDM to a concentration of 5 mg/mL and protein crystal trials produced small crystals in 0.2 M ammonium sulfate, 0.1 M CHES 10.0, 14% w/v PEG 400. These were fragile and difficult to manipulate and x-ray diffraction data was unable to be collected to establish if these were protein crystals. An ALBI was performed as a dot blot for BB027, BB0562, BB405, BB406 and BAPKO_0422 to assess factor-H binding, with no binding identified for any of these proteins. This agrees with current literature, but conflicting data is available for the FH binding ability of BAPKO_0422.

Due to time restraints, many experiments could not be completed and there is extensive scope for future work. Despite the difficulties membrane protein crystallography presents, a high resolution structure of these small *B. burgdorferi* OMPs would be invaluable and further experiments could be attempted to try to crystallise these proteins. While BB_0027 and BB_0562 are thought to be β-barrel OMPs, these remain uncharacterised and structural data could establish whether these do indeed belong to the OmpA-like family of proteins. While evidence suggests that the four proteins in this project do not bind human factor H, recent research identifying BB_0406 as a laminin binding protein could suggest a role for other *Borrelia* β-barrels in establishing infection and dissemination, rather than evasion of the

complement immune response. The surface exposed nature of these proteins mean they have potential in improving current diagnostic tests for *Borrelia* infection and may provide targets for vaccine development.

4.7.2 *B. burgdorferi* chemotaxis protein methyltransferase (BbCheR)

As membrane protein crystallography presents many challenges, a soluble protein target was selected for this project in order to increase the likelihood of a crystal structure. Using data from Mongodin et al (2013), two potentially soluble targets were identified; BB_0095 and BB_0040. As BB_0095 did not produce any soluble protein during protein test expressions, it was ruled out as a target. BB_0040 was identified as a chemotaxis protein methyltransferase (CheR), an enzyme involved in the chemotactic response in Gram negative bacteria. The CheRs from *Salmonella typhimurium* and *Bacillus subtilis* are well studied, but BbCheR and many of the other chemotaxis proteins of *Borrelia* remain uncharacterised.

A multiple sequence alignment of BbCheR with both pentapeptide dependent and pentapeptide independent CheRs showed a conserved structural motif involved in the binding of S-adenosylmethionine (SAM) and S-adenosyl-L-homocysteine (SAH), which was expected. As both PD and PI CheRs were used, it could be determined which of these groups BbCheR likely belonged to, with three conserved glycine residues involved in pentapeptide binding found in the β-subdomain of PDCheRs (Perez & Stock, 2007). As BbCheR lacked these residues, it is likely a PICheR, the most common type of CheR employed by bacteria and the addition of a pentapeptide sequence was therefore not required (Perez et al., 2004).

Expression of BbCheR produced reasonable yields of soluble protein, with 3 L of *E. coli* culture providing ~10 mg of protein after tag removal and purification steps. A 6x histidine thioredoxin tag was used for BbCheR expression, as thioredoxin is known to enhance the solubility of recombinant proteins. This large tag (~ 13 kDa) was removed using a TEV protease prior to crystallisation experiments, to avoid any flexible regions interfering with crystal formation. Tag removal appeared to be successful, with SDS-PAGE analysis showing no bands of the tagged protein size in any of the elution samples following Ni-NTA rebinding. Western blot analysis using an anti-His antibody confirmed this, with a strong band present in the uncut sample where the His tag is still present, and no bands of this size seen in the elution samples following rebind. In the 100 mM and 300 mM imidazole elution fractions, a strong band can be seen at a lower molecular weight than the uncut protein. As this appeared at a lower molecular weight, it is not likely to be uncut protein still containing the 6x His tag. As it

is unknown what these two bands correspond to, further analysis would need to be carried out by slicing the band from the gel and using this for mass spectrometry. This would determine the amino acid sequence of the protein and therefore determine what this band is. Following tag removal, ion exchange chromatography was attempted, but the lower salt concentration required for this resulted in precipitation of the protein. Size exclusion chromatography was performed and showed a single peak corresponding to monomeric BbCheR.

Initial sitting drop vapour diffusion crystal screens were set up as described in section 2.2.13.2 using ~8.86 mg/mL of BbCheR without the addition of SAH. The protein solution for these initial screens contained 500 mM NaCl, which is high and not ideal for protein crystallography. Reduction of salt concentration was attempted in previous protein preparations prior to size exclusion chromatography and resulted in rapid precipitation of the protein. The initial screens of BbCheR did not produce any crystals. As problems with protein aggregation were noted by Batra et al when the *Bacillus subtilis* CheR was prepared without the presence of SAM or SAH, as well as a lack of crystals with CheR alone or in complex with SAM (Batra et al., 2016), co-crystallisation of BbCheR in complex with SAH was attempted. Due to solubility limitations when preparing a stock solution of SAH, 1.25 mM SAH was added to 8.86 mg/mL of BbCheR instead of the desired 2.5 mM SAH, resulting in 5 fold ligand compared to protein rather than 10 fold. The salt concentration of the protein solution was reduced from 500 mM to 125 mM for co-crystallisation attempts and trays were set up as previously described. This co-crystallisation attempt was unsuccessful and no crystals were formed under any of the conditions attempted.

Due to time limitations, the binding affinity of BbCheR for SAH could not be determined using a suitable method such as isothermal titration calorimetry (ITC). For this reason, SAH concentration was determined based on previous methodology for the crystallisation of the pentapeptide independent CheR from *Bacillus subtilis* (BsCheR), which used a similar concentration of CheR protein (Batra et al., 2016). Future work attempting to crystallise this protein should focus on determining the concentration of SAH to add to pure BbCheR, as the

instability of the protein without a co-factor makes it difficult to work with. A wider range of crystallisation conditions could be used to increase the probability of crystal formation.

Chapter 5 A computational framework to search for β-barrel genes in the plasmid proteomes of *Borrelia*

5.1 Introduction

The genome of the *Borrelia* species is unique in its structure, comprised of a linear chromosome approximately 1 Mbp in size and multiple variable circular and linear plasmids, ranging in size from ~5-56 kilo base pairs (Beaurepaire & Chaconas, 2005; Fraser et al., 1997). This unique composition is not shared by other spirochete families under the phylum *Spirochaetes*. Like *Borrelia*, *Treponema* is an obligate parasite with a small genome, but its genome consists only of a circular chromosome of ~ 1.1 Mbp (Fraser et al., 1998). *Brachyspira*, a parasitic spirochete able to transiently survive in soil and faeces, has a larger circular chromosome of ~ 3 Mbp and a circular plasmid of 36 kbp (Bellgard et al., 2009). The genome of *Leptospira interrogans*, an obligate aerobic spirochete capable of surviving in aquatic environments, is significantly larger than that of the previously mention spirochetes, consisting of a large circular chromosome of ~4.3 Mbp and a smaller circular chromosome of ~ 3.5 kbp (Jia et al., 2003). The linear and circular plasmids of *Borrelia* make up ~47% of the genome and are variable between different species and strains, with the occurrence of recombination events and horizontal gene transfer (Casjens, 2000; Casjens et al., 2017). Many of the plasmids encode proteins important for virulence, such as outer surface proteins A, B and C and loss of plasmids can result in loss of infectivity in mice models (Norris et al., 2011).

While *B. burgdorferi s.l* is considered Gram negative, the components of its outer membrane differ to other Gram negative bacteria in that it contains an abundance of lipoproteins and relatively few integral β-barrel outer membrane proteins (Fraser et al., 1997; Kenedy et al., 2012). Of the OMPs that have been characterised, many are porins and are listed by their given name, followed by the gene locus and any alternative names; P66 (BB0603, Oms66) (Kenedy et al., 2014; Skare et al., 1997), P13 (BB0034) (Noppa et al., 2001), DipA (BB0418, Oms38) (Thein et al., 2012), BesC (BB0142) (Bunikis et al., 2008), BamA (BB0795) (Lenhart & Akins, 2010), BB0405 and BB0406 (Dyer et al., 2015; Kenedy et al., 2016). Other uncharacterised OMPs include BB0838 (Kenedy et al., 2016), BB0027 and BB0562 (Dyer et al., 2015).

The analysis of whole proteomes for *ab initio* prediction of β-barrel membrane spanning proteins relies on more than a single prediction algorithm, due to the high rate of false positives. For this reason, a computational framework is generally used where sequences are successively eliminated based on the results from a range of different prediction algorithms (Cox et al., 2010; Kenedy et al., 2016). This computational framework has been used on the *B. burgdorferi s.l* chromosome, identifying the β-barrel OMPs BB405, BB406 and BB0838 (Kenedy et al., 2016). However, the plasmid proteomes of *Borrelia* have not been subject to analysis to identify potential β-barrel OMPs. Therefore, this project aimed to develop a computational framework based on previous work to comprehensively search the linear and circular plasmids of 14 *Borrelia* genomes for putative β-barrel sequences.

5.2 Identification of a plasmid encoded β-barrel outer membrane protein

The non-redundant proteomes of 16 *Borrelia* genospecies from the *B. burgdorferi s.l* complex were obtained from UniProtKB and the plasmid components of 14 of these were extracted to give a total of 5923 protein sequences. A computational framework was used to filter out sequences after analysis by various prediction algorithms, to identify any potential β-barrel outer membrane proteins (figure 5.1).

Potential lipoproteins can be reliably identified based on identification of an N-terminal signal sequence and a conserved cysteine residue within the lipobox motif (Wilson 2015). However, the spirochete C-terminal lipobox differs to that of other Gram negative bacteria and so the SpLip 1.0 predication algorithm was used for more specific prediction of spirochaetal lipoproteins (Setubal et al., 2006). Results from SpLip 1.0 filtered out 795 lipoproteins, with 5129 proteins progressing to the next stage. LipoP was used to further filter out lipoproteins, removing 123 proteins from the data set. As SpLip 1.0 identifies lipoproteins based on spirochetes and LipoP is based on Gram negative bacteria, using both algorithms may result in false positive identification of lipoproteins in the spirochaetal proteomes. However, these proteins are removed from further analysis and may reduce false positive β-barrel identification.

TMHMM – 2.0 was used to predict proteins with transmembrane helices based on a Hidden Markov model (Krogh et al., 2001) and identified 704 sequences predicted to form 1 or more transmembrane helices, which were removed. PSORTb 3.0 was then used to predict bacterial protein subcellular localisation, with 5 localisations predicted for Gram-negative bacteria; extracellular, outer membrane, periplasmic, inner membrane and cytoplasmic (Yu et al., 2010). Only 45% of the *Borrelia* plasmid proteome sequences submitted received a prediction in any category, with the remaining 55% considered ‘unknown’. A total of 1817 sequences were predicted to be cytoplasmic proteins and were removed from the process.

The final stages of the framework focused on identification of outer membrane proteins, starting with prediction of the presence of a signal peptide using SignalP 5.0 (Armenteros et al., 2019). Proteins localised to the bacterial outer membrane possess a short sequence of

amino acids at the N-terminus, a signal peptide, in order to allow processing and translocation to the outer membrane. SignalP 5.0 identifies proteins with a Signal Peptidase 1-type (SP1) signal sequence and 2421 proteins negative for this were removed. This left 63 sequences which were submitted to PRED-TMBB2, which utilises Hidden Markov models to predict the topology of β-barrel proteins located in the outer membrane (Tsirigos et al., 2016). This resulted in 13 sequences predicted to be β-barrel outer membrane proteins (table 5.1). As the initial sequences were from the plasmid proteomes of various genospecies of *B. burgdorferi s.l*, some of which were different plasmids from the same isolate, the 13 putative β-barrel sequences were not likely to be unique.

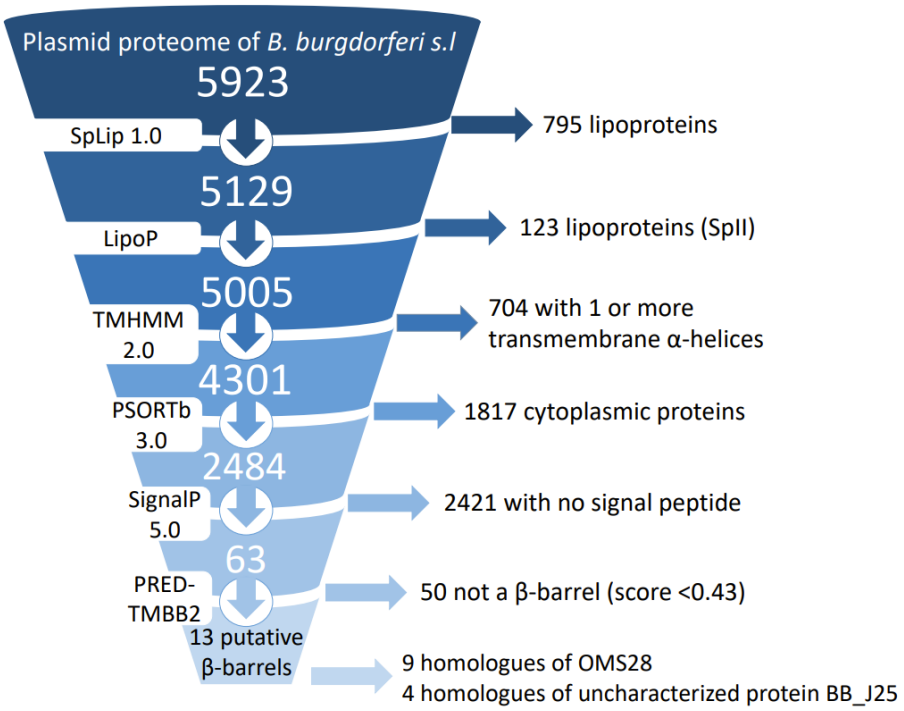


Figure 5.1. Computational Framework to identify potential OM β-barrels encoded by plasmid genes of *B. burgdorferi s.l.* The workflow is based on similar frameworks used previously on *Borrelia* and *Treponema* proteomes (Cox et al., 2010; Kenedy et al., 2016). The number of protein sequences eliminated at each stage is indicated.

	RefSeq	UniProt Accession number	Sequence length	Species (strain)	Taxon id	Plasmid
Outer membrane porin OMS28	WP_011703833.1	Q0SLT3	273	*B. afzelii* (PKo)	390236	lp54
	WP_010257191.1	O50963	257	*B. burgdorferi* (B31)	224326	lp54
	WP_011187217.1	Q6ASF1	257	*B. bavariensis* (PBi)	290434	lp54
	WP_015939517.1	B8F1P7	264	*B. garinii* (PBr)	498743	lp54
	WP_010257191.1	A0A0E0STG3	257	*B. burgdorferi* (64b)	498740	lp54
	WP_012665701.1	C0R8A8	272	*B. valaisiana* (VS116)	445987	lp54
	WP_012672322.1	C0RQK4	257	*B. finlandensis*	498741	lp54
	WP_010257191.1	A0A0H3C0J6	257	*B. burgdorferi* (ZS7)	445985	lp54
	WP_014023254.1	G0APJ4	273	*B. bissettii* (DN127)	521010	lp54
Uncharacterised protein	WP_010890368.1	O50780	346	*B. burgdorferi* (B31)	224326	lp38
	WP_010890368.1	A0A0E0ST89	346	*B. burgdorferi* (64b)	498740	lp38
	WP_015899315.1	C0R918	346	*B. valaisiana* (VS116)	445987	lp28-3
	ACN53112.1	C0R9H1	332	*B. valaisiana* (VS116)	445987	lp28-8

Table 5.1. List of hits identified from the computational search of *Borrelia* plasmid sequences. The NCBI Reference Sequences (RefSeq) are given, along with the Uniprot Accession numbers, taxonomic identifiers (NCBI) and plasmid name. The hits fall into two main groups, nine with homology to OMS28 (Ordered Locus Name BB_A74), and four uncharacterized proteins with homology to BB_J25, two of which are paralogs from the same organism (*B. valaisiana*).

5.3 Phylogenetic analysis of BBJ25 homologs

A computational search of the sequences found 9 to be orthologous to each other, annotated as OMS28 from 9 different *Borrelia* species. The remaining 4 sequences were all sequences with homology to the uncharacterised protein BBJ25. Two of these BBJ25 sequences were found to be from two different strains of *B. burgdorferi s.s*, B31 and 64b. The other two (COR918 and COR9H1) are homologous sequences sharing 54% sequence identity from the same species, *B. valaisiana* (VS116), but are located on different plasmids, lp28-3 and lp28-8 respectively. BBJ25 from *B. burgdorferi* B31 was used as a query sequence for PSI-BLAST searches, to identify homologous sequences from other families in the Spirochaetes phylum. The initial search yielded 75 sequences from *Borrelia* species and *Brachyspira*, with a second iteration yielding a total of 98 sequences from *Borrelia, Brachyspira* and *Treponema*. No homologous sequences were found in *Leptospira*. The redundancy of these sequences was decreased, leaving a total of 30 BBJ25 homologues with a sequence identity around 20-24% between the *Borrelia, Brachyspira* and *Treponema* families (appendix 6). All potential homologues were confirmed by reciprocal BLAST (Ward & Moreno-Hagelsieb, 2014). Phylogenetic analysis demonstrated clustering into three distinct groups corresponding to each spirochete family (figure 5.2).

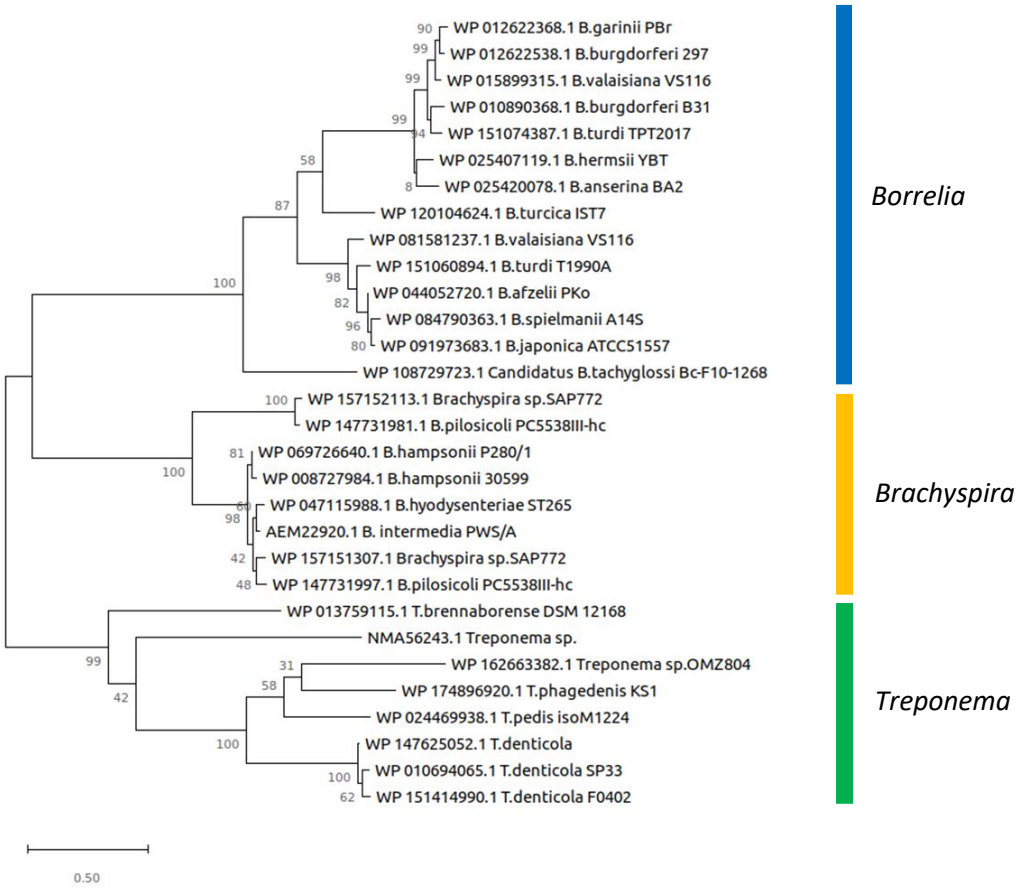


Figure 5.2. Maximum likelihood phylogenetic tree of BBJ25 homologs. The phylogenetic tree was generated in MEGAX using the Maximum Likelihood method and a JTT matrix-based model (Kumar et al., 2018). A bootstrap method was applied using 1000 replications. Branch lengths are measured in number of substitutions per site. This tree is unrooted.

Clustering within the *Borrelia* genus did not correlate with relapsing fever or Lyme disease causing species and did not appear to be correlated with host vectors, geographical location or plasmid type. Horizontal gene transfer has historically occurred between *Borrelia* linear plasmids and the genetic content of plasmids can vary between *Borrelia* genospecies (Casjens et al., 2017). Evidence of gene transfer may be available from analysis of neighbouring genes and the location of BBJ25 homologs within each plasmid was therefore investigated. A putative operon was identified and tested using Operon Mapper (Taboada et al., 2018) (table 5.3). The operon is predicted to be involved in lipoprotein release, but this would need experimental validation. The BBJ25 open reading frame (ORF) was found on a range of linear plasmids within *Borrelia*: lp38, lp17, lp28-3 and lp28-8 and in *B. turcica* IST7, BBJ25 was located near the end of the linear chromosome. Analysis of the genes surrounding BBJ25

showed separate clusters of two distinct groups in the phylogenetic analysis, labelled Group 1 and Group 2 (figure 5.4). As the phylogenetic analysis was originally based only on the protein sequence of BBJ25 (figure 5.3), a phylogenetic tree was generated using the DNA sequence of the whole predicted operon (figure 5.4). jModelTest (Posada, 2008), was used to select an appropriate nucleotide substitution model for phylogenetic analysis of the BBJ25 operon. A total of 88 models were tested and ranked according to both Akaike Information Criterion (AIC) and AIC corrected for small sample sizes (AICc) (Hurvich & Tsai, 1989). The top four nucleotide substitution models are shown in table 5.2.

Gene ID	Type	COGgene	PosLeft	PosRight	Strand	Function
BKHGOCJI_00001	CDS	COG0457	521	1330	+	FOG: TPR repeat
BKHGOCJI_00002	CDS	COG0457	1389	2171	+	FOG: TPR repeat
BKHGOCJI_00003	CDS	NA	2215	3255	+	NA
BKHGOCJI_00004	CDS	COG4181	3314	4009	+	Predicted ABC-type transport system involved in lysophospholipase L1 biosynthesis
BKHGOCJI_00005	CDS	COG4591	4110	5231	+	[M] ABC-type transport system, involved in lipoprotein release
BKHGOCJI_00006	CDS	NA	5221	5949	+	NA
BKHGOCJI_00007	CDS	NA	5972	7009	+	NA

Table 5.2. Prediction of the putative operon containing BBJ25 from *B. burgdorferi* B31. The operon prediction was produced using Operon Mapper (Taboada et al., 2018). The gene ID is given, followed by gene type (CDS = protein coding gene), clusters of orthologous group (COG) assignment if available, the position of the gene on the operon, the strand the gene is on and the predicted function of the gene if known.

Name	Description	Weight	deltaAICc
TVM +G	**transversion model with gamma distribution:** variable base frequencies, variable transversion rates, transition rates equal, G= gamma distributed rate variation among sites	0.4609	0
GTR + G	**general time reversible with gamma distribution:** variable base frequencies, symmetrical substitution matrix, G = gamma distributed rate variation among sites	0.2771	1.01
TVM +I +G	**transversion model with gamma distribution):** variable base frequencies, variable transversion rates, transition rates equal, G= gamma distributed rate variation among sites I = with a proportion of invariable sites	0.1642	2.06
GTR +I +G	**general time reversible with gamma distribution:** variable base frequencies, symmetrical substitution matrix, G = gamma distributed rate variation among sites I = with a proportion of invariable sites	0.0979	3.10

Table 5.3. jModelTest recommended nucleotide substitution models for the operon dataset. From 88 nucleotide substitution models, the four highest scoring models ranked by both AIC and AICc are shown (the top four models from both AIC and AICc were the same). DeltaAICc shows the decrease in AICc score from the highest ranked model. The DeltaAICc scores ranged from 0 for the best to 8531 for the worst scoring model. The GTR +G model was chosen for phylogenetic analysis as this can be implemented in MEGAX.

The overall grouping of the tree based on the whole operon is very similar to the tree based on BBJ25 alone with only minor differences in the terminal branches. Within both Group 1 and Group 2, the predicted operon consists of 7 genes: a Tetratricopeptide repeat domain protein (BBJ23), an ABC transporter with both ATP-binding (BBJ26) and permease domains (BBJ27), BBJ25 and three hypothetical proteins (BBJ23, BBJ28 and BBJ29). A conserved gene order is seen among the *Borrelia* species, with the exception of BBJ25 which is located in the third position on the predicted operon or the last position. BBJ25 is seen immediately upstream or downstream of the ABC-transporter genes in both groups. In Group 2, the predicted operon containing BBJ25 is followed by a Streptolysin S-like (SLS) operon (Molloy et al., 2015). The order of chromosomal genes in *Brachyspira* is similar to that of Group 2, with collinearity of the two ABC transporter genes and BBJ25 homolog, but these are separated by two pseudogenes. The gene structure in *Treponema* genospecies appears to be

quite different and the BBJ25 homologues are surrounded by unrelated genes (results not shown).

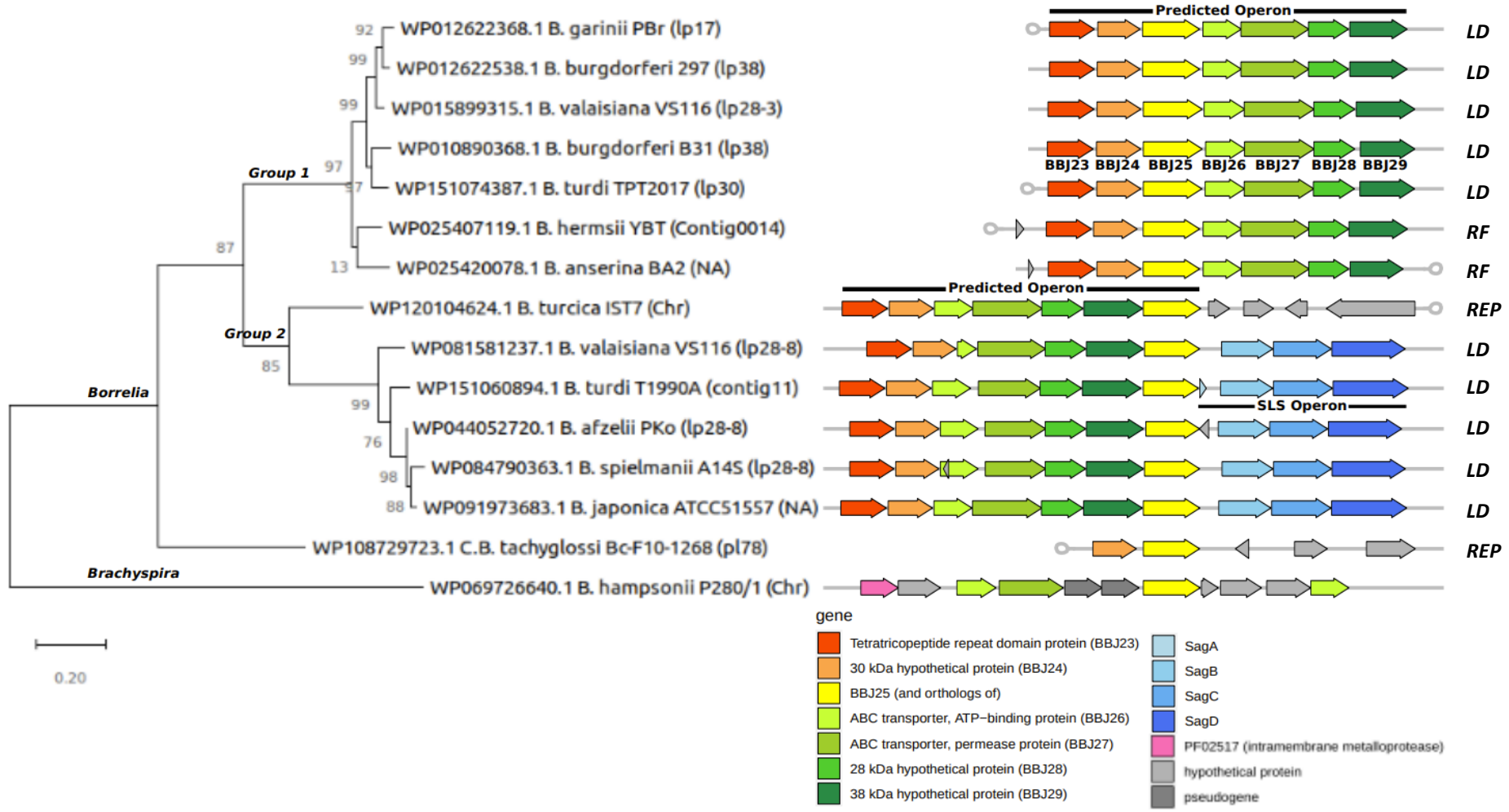


Figure 5.3. Phylogenetic analysis of BBJ25 protein sequences from *Borrelia* species and *Brachyspira*. The phylogenetic tree was generated in MEGAX using the Maximum Likelihood method and a JTT matrix-based model (Kumar et al., 2018). The JTT+G model was chosen based on a ranking of 60 models by prottest3 (Darriba et al., 2011). A bootstrap method was applied using 1000 replications. Branch lengths are measured in number of substitutions per site. A graphical representation of the predicted operon is shown and the group each species belongs to is indicated as either Lyme disease (LD), Relapsing Fever (RF) or reptile (REP) associated. Graphical representation generated using the gggenes package in R (R Core Team, 2021) and nucleotide sequence lengths in GenBank (NCBI).

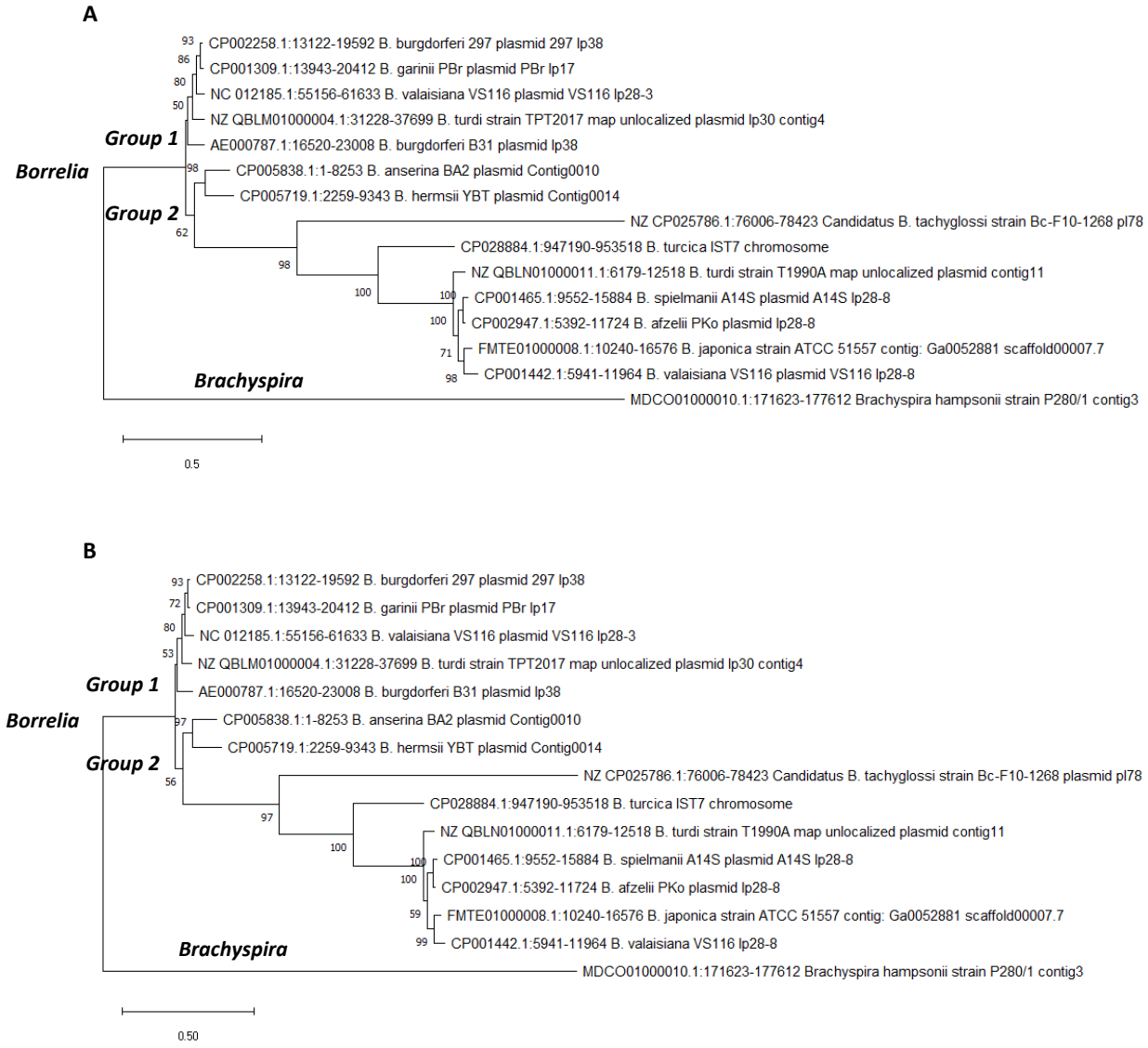


Figure 5.4. Phylogenetic analysis of predicted BBJ25 operon DNA sequences from *Borrelia* species and *Brachyspira*. A) Analysis of the whole operon based on genomic regions shown in table 2.5 with 10158 positions. **B)** Analysis of the equivalent regions with omission of the BBJ25 coding sequence with a total of 7845 positions. DNA sequences were aligned using MUSCLE (Edgar, 2004) as described in the Methods section. The evolutionary history was inferred by using the Maximum Likelihood method and General Time Reversible model (Nei & Kumar, 2000). The trees with the highest log likelihood is shown (A = -46915.60, B = -39097.67). The percentage of trees in which the associated taxa clustered together is shown next to the branches based on 1000 bootstrap replications (Felsenstein, 1985). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites with 5 categories (+G, parameter = 0.8601 for panel A and 0.9021 for panel B). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 15 nucleotide sequences. Evolutionary analyses were conducted in MEGA X (Kumar et al., 2018).

As the framework used in this research only utilised one algorithm for identification of β-barrel proteins, PRED-TMBB2, additional algorithms were used to improve confidence of topology predications. BBJ25 sequences from *Borrelia*, *Treponema* and *Brachyspira* species (appendix 6) were analysed using BOMP (Berven et al., 2004) and BOCTOPUS2 (Hayat et al., 2016). Both algorithms are consistent with PRED-TMBB2 predication of a β-barrel topology, but there are discrepancies between the number of β-strands. PRED-TMBB2 predicted a β-barrel topology spanning the whole BBJ25 sequence with 16-18 β-strands. BOCTOPUS2 predicted a smaller β-barrel at the N-terminus between 8-10 strands, with the C-terminus containing a non-barrel domain.

5.4 Homology modelling of BBJ25 from *B. burgdorferi* B31

Homology models were produced for BBJ25 from *B. burgdorferi* B31 using the structural prediction algorithms I-TASSER, MODELLER, PHYRE2 and Robetta. The model produced using I-TASSER lacked a complete β-barrel structure, with the N-terminal region not forming any ordered secondary structure (figure 5.5). The surface electrostatics do not fit what would be expected of a β-barrel OM protein, as the barrel region shows charged regions where non-polar residues would be expected to interface with the non-polar glycolipids of the *Borrelia* membrane.

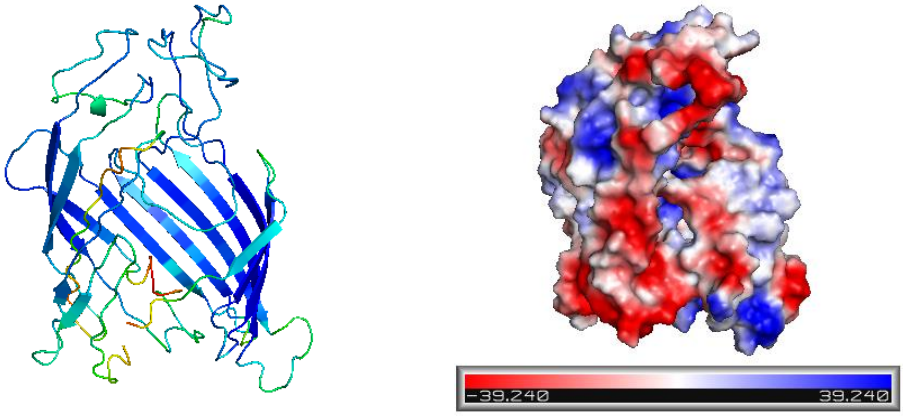


Figure 5.5. BBJ25 homology model using I-TASSER. Left: Cartoon representation with B-factors to show prediction confidence. This is shown as a rainbow spectrum from red (low confidence) to dark blue (high confidence). Right: Model showing surface electrostatic potential. Images produced using PyMol (Schrödinger, 2010).

HHsearch utilises hidden Markov models (HMMs) and is noted as the method with the highest sensitivity for detecting distant evolutionary relationship that other methods generally do not detect (Pereira & Alva, 2021). Sequences are aligned and submitted to MODELLER (Šali & Blundell, 1993). The first 10 available templates were selected for homology modelling (table 5.4). The topology of this model is more consistent with that of other known β-barrels in Gram negative bacteria, with an even number of strands, both the N and C-termini on the same side and all of the strands in an anti-parallel conformation. However, some of the strands in this model show low confidence in the prediction. While the surface electrostatics are more like that expected of a β-barrel OMP in comparison to the I-TASSER model, charged regions are still seen around the centre of the barrel.

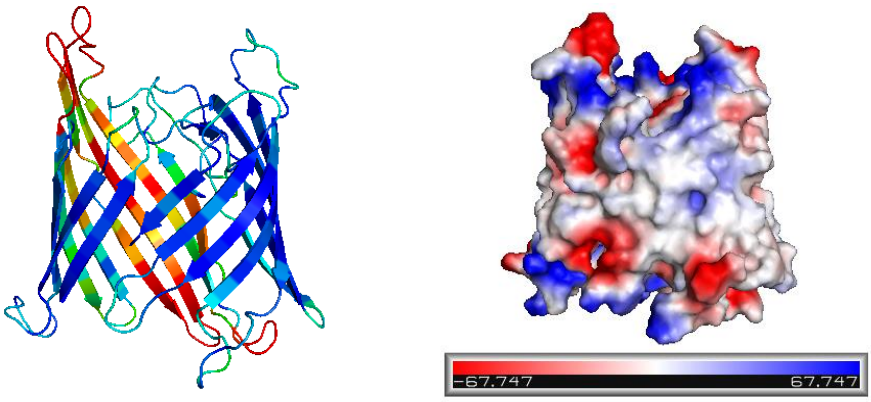


Figure 5.6. BBJ25 homology model using MODELLER. Left: Cartoon representation with B-factors to show prediction confidence. This is shown as a rainbow spectrum from red (low confidence) to dark blue (high confidence). Right: Model showing surface electrostatic potential. Images produced using PyMol (Schrödinger, 2010).

Template	Sequence identity	Probability
BamA – *E.coli* (6QGW)	15%	94.94%
Outer membrane transporter CdiB – *E.coli* (6WIM)	10%	94.46%
Substrate-engaged Bam complex – *E. coli* (6V05)	15%	94.22%
BamA with one POTRA domain – *E. coli* (4C4V)	15%	93.56%
BamABCDE complex – *E. coli* (5D0O)	15%	92.07%
Filamentous hemagglutinin transporter protein FhaC - *Bordetella pertussis* (4QL0)	9%	91.69%
Translocation and assembly module TamA - *E. coli* (4C00)	14%	91.45%
Hemolysin activator protein CdiB - *Acinetobacter baumannii* (6WIL)	11%	91.23%
BamA lacking POTRA domains 1-3 - *Haemophilus ducreyi* (4K3C)	14%	90.27%
BamA - *Neisseria gonorrhoeae* (4K3B)	17%	89.71%

Table 5.4. MODELLER templates for homology modelling. The name of the template and the corresponding PDB is given, followed by the percentage sequence identity to BBJ25.

PHYRE2 selected 3 templates for homology modelling (table 5.5) and 55% of residues modelled at >90% confidence. The highest scoring template was an outer membrane channel from *E. coli*, but this only shared 11% sequence identity with BBJ25 (table 5.5). Low sequence identity is also seen for the templates used by MODELLER (table 5.4). This model shows two domains, a β-barrel domain and a C-terminal non-barrel domain.

Template	Sequence identity	Confidence	Coverage
Probable n-acetylneuraminic acid outer membrane channel – *E. coli* (2WJQ)	11%	92.1	54%
BamA - *Haemophilus ducreyi* (4K3C)	15%	73.1	22%
BamA- *E. coli* (4C4V)	18%	68.3	21%

Table 5.5. PHYRE2 templates for homology modelling. The name of the template and the corresponding PDB is given, followed by the percentage sequence identity to BBJ25, the prediction confidence and the sequence coverage.

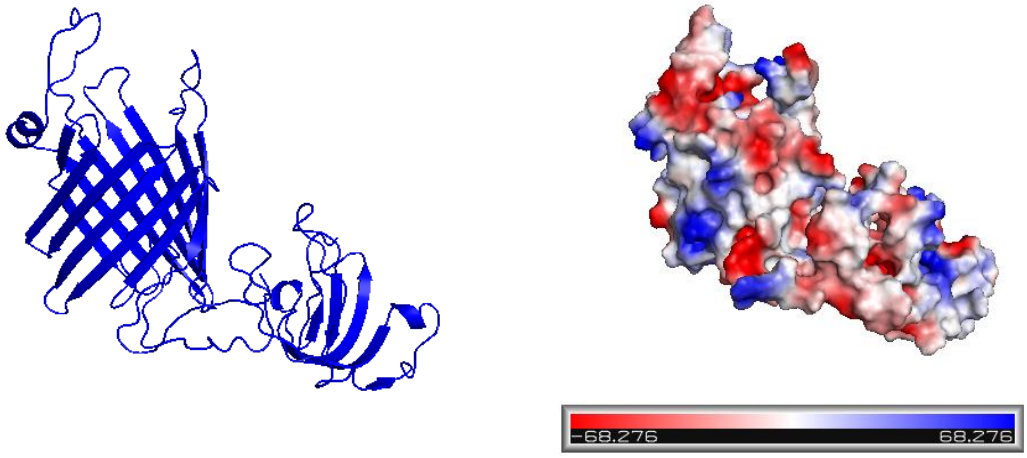


Figure 5.7. BBJ25 homology model using PHYRE2. Left: Cartoon representation. Right: Model showing surface electrostatic potential. Images produced using PyMol (Schrödinger, 2010).

The BBJ25 sequence was submitted to the Robetta server and RoseTTAfold modelling method was selected. RoseTTAfold is a deep modelling method that predicts protein structure based on a 3-track network and is suggested as the most accurate method available on the Robetta server (Baek et al., 2021). The five models produced from this method show a 16 stranded β-barrel topology, consistent with Schulz's rules for OM β-barrels (table 1.4, section 1.3.3). The

membrane spanning strands are predicted with high confidence, with the exception of the first and last strand where the barrel closes. The terminal strand should likely be further extended to form a more closed barrel, as is seen for other β-barrel OMPs. There is low confidence in the prediction of the loop regions, but this is expected due to the high variability in the loops of β-barrel proteins, which share little homology with one another and are frequently mobile. All five models have a loop region between β-strands 11 and 12, that occupies the centre of the barrel. This loop region forms one or more short α-helices in 4 out of the 4 models. Surface electrostatics for these models fit that expected of a β-barrel OMP, with a non-polar core region and charged loop regions. Alignment of the five models (figure 5.10) shows consistent orientation of the 16 β-strands, whereas the orientation of the eight external loops show some variation.

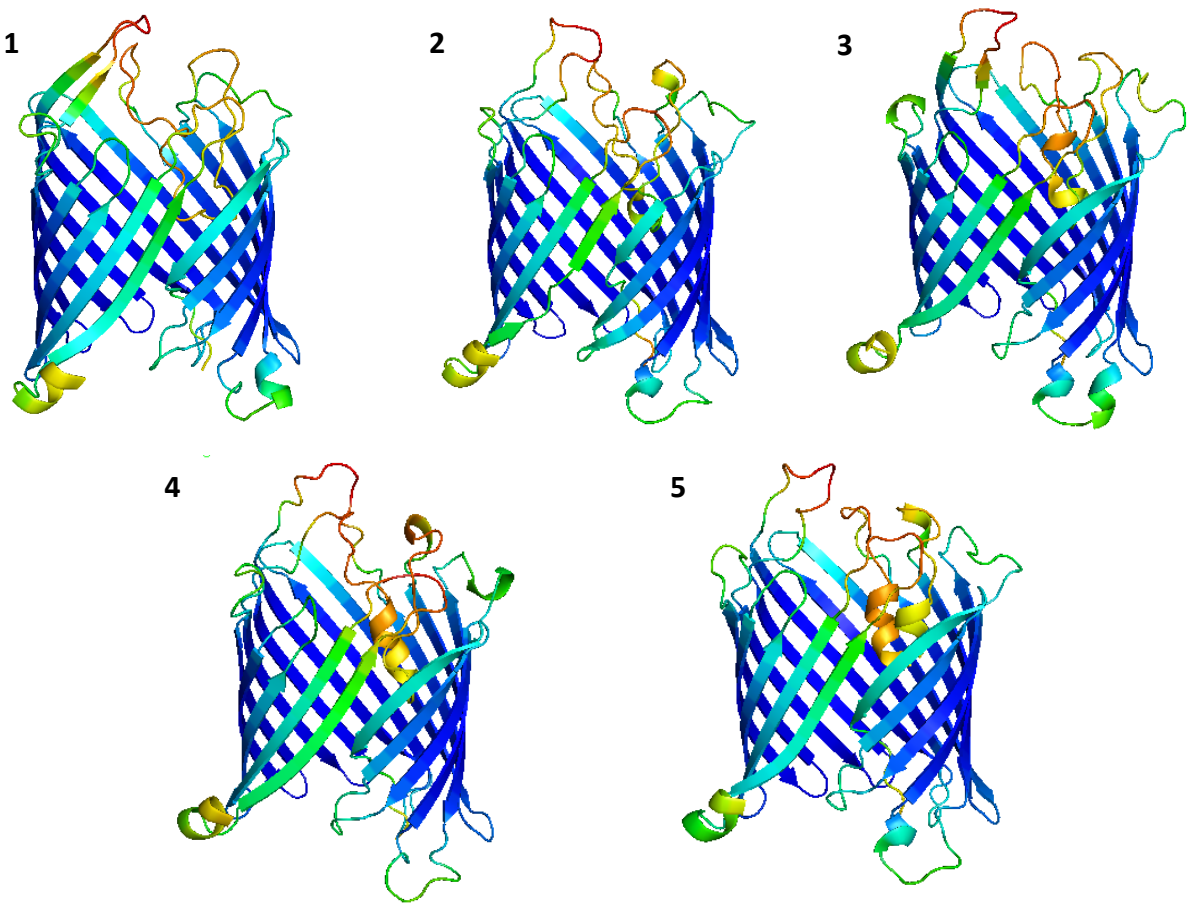


Figure 5.8. BBJ25 homology models using Robetta RoseTTAfold. Models are shown as a cartoon representation with B-factors to show prediction confidence. This is shown as a rainbow spectrum from red (low confidence) to dark blue (high confidence). Images produced using PyMol (Schrödinger, 2010).

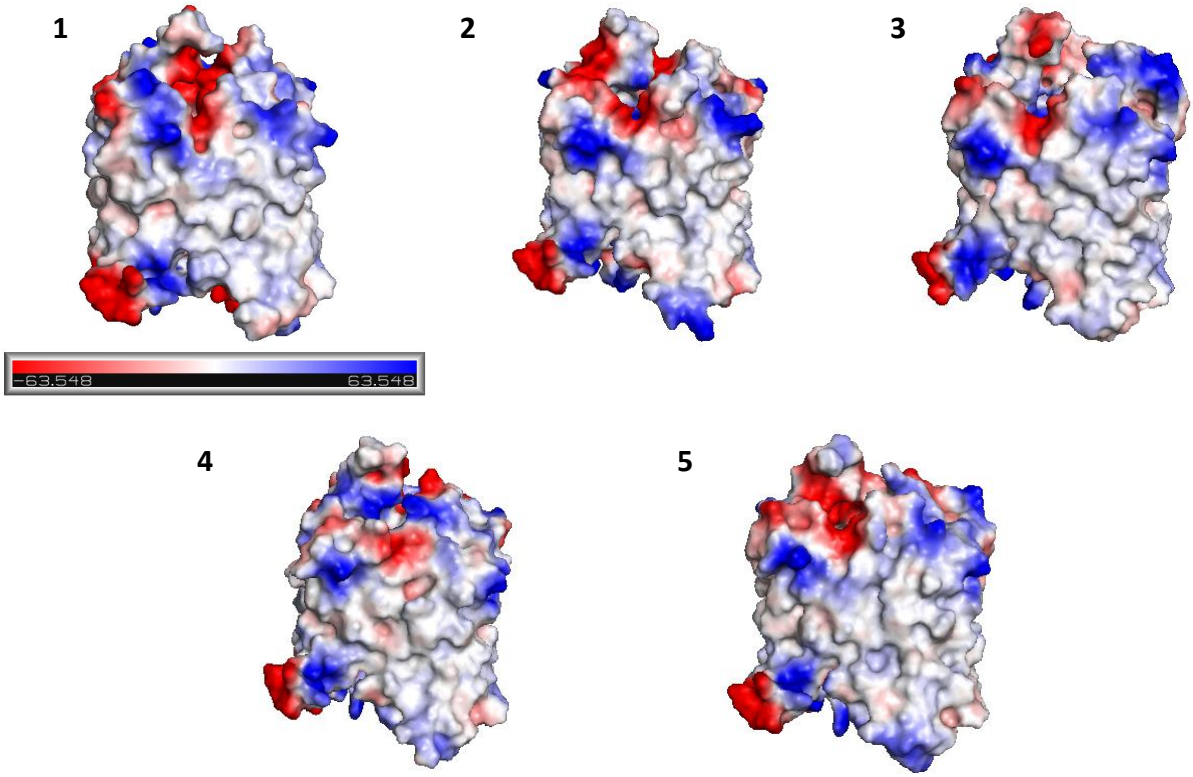


Figure 5.9. BBJ25 homology models using Robetta RoseTTAFold. Models showing surface electrostatic potential. Images produced using PyMol (Schrödinger, 2010).

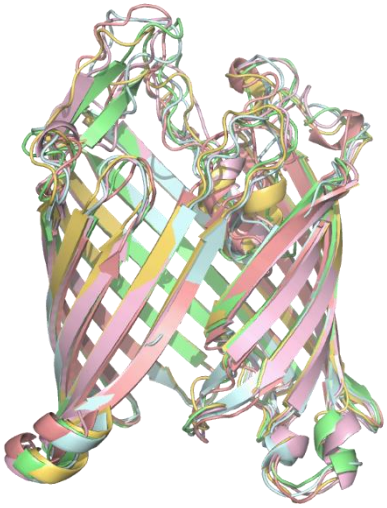


Figure 5.10. Alignment of BBJ25 RoseTTAfold models. The 5 models produced by RoseTTAfold were aligned in PyMol (Schrödinger, 2010) and superimposed. Model 1 = green, Model 2 = light blue, Model 3 = pink, Model 4 = gold, Model 5 = salmon.

The computational search of the *Borrelia* plasmid proteomes yielded 9 homologues of Oms28, a protein originally identified as a predicted integral OM protein with porin activity (Skare et al., 1996). Homology models of Oms28 were generated using PHYRE2 and Robetta RoseTTAfold, with both servers predicting this protein as a predominantly helical and therefore soluble (non-membrane) protein (figure 5.11). This is consistent with circular dichroism data identifying Oms28 as 78% α-helix and a periplasmic protein associated with the OM (Mulay et al., 2007). The PHYRE2 model showed low confidence, with 0% of residues modelled at >90% confidence. Despite this helical topology, Oms28 was identified as a potential β-barrel protein after filtering through a series of prediction algorithms and highlights a limitation of this method. The computational framework used could be modified, perhaps by the addition of a secondary structure prediction algorithm such as PSI-PRED, to prevent periplasmic proteins appearing as false positives.

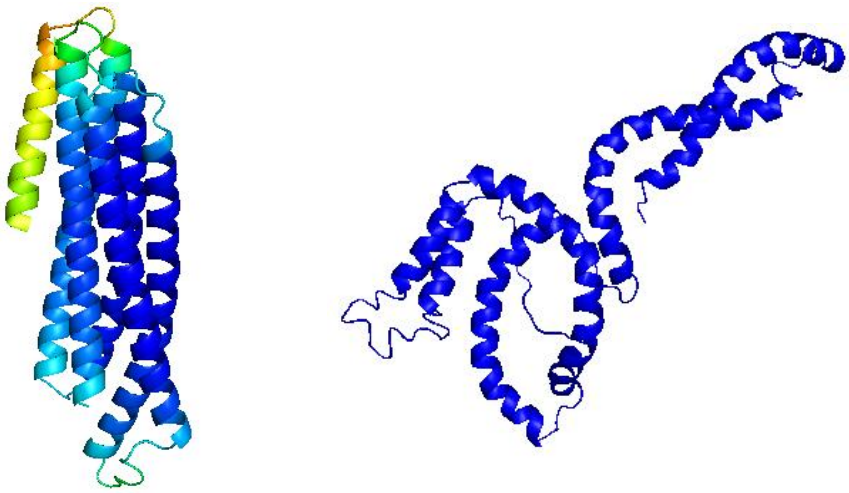


Figure 5.11. **Oms28 homology models**. Left: Robetta RoseTTAFold model as a cartoon representation with B-factors to show prediction confidence. This is shown as a rainbow spectrum from red (low confidence) to dark blue (high confidence). Right: PHYRE2 model as a cartoon representation. Confidence is not shown. Images produced using PyMol (Schrödinger 2010).

5.5 Discussion

This project developed a computational framework to search for β-barrel OM proteins, based on previous work which filtered sequences through a series of prediction algorithms to identify those predicted to form transmembrane β-barrels (Kenedy et al., 2016). As this previous framework focussed on identification of OMPs on the *B. burgdorferi* B31 chromosome, we extended this search to the plasmid proteomes of 14 *Borrelia* species. After sequential elimination of sequences through various prediction algorithms, we identified only two potential OM-β-barrel proteins, OMS28 which is annotated as a porin, and the uncharacterised protein, BBJ25. This low number was expected, as *Borrelia* contain few integral β-barrel OM proteins compared to other Gram-negative bacteria and currently characterised *Borrelia* OMPs are coded for on the chromosome. Analysis of the location of BBJ25 homologs within each plasmid showed that BBJ25 is associated with a 7-gene operon, neighboured by a Tetratricopeptide repeat domain protein (BBJ23), an ABC transporter with both ATP-binding (BBJ26) and permease domains (BBJ27) and three hypothetical proteins (BBJ23, BBJ28 and BBJ29).

Phylogenetic analysis of the BBJ25 operon

The phylogenetic analysis showed that the evolutionary history of this region does not match what would be expected from vertical transmission alongside the chromosome, possibly because of a combination of horizontal gene/whole-plasmid transfer followed by recombination. The clustering seen is unusual as LD and RF *Borrelia* are closely related and share a large common subset of genes (Margos et al., 2018), generally forming two distinct groups even if a phylogenetic tree is constructed from a single chromosomal gene (appendix 8). Throughout the *Borrelia* genus, the BBJ25-operon is located on linear genetic elements, either the smaller linear plasmids (lp38, lp17, lp28-3, lp28-8) or close to the end of the linear chromosome in the case of *B. turcica*. These smaller linear plasmids are relatively unstable, and it is well documented that there have been several inversions, recombination events and lateral gene transfer both within and between different *Borrelia* species (Casjens et al., 2018). Even whole plasmids can, on occasion be exchanged, although this is relatively rare (Casjens et al., 2018). In addition, it has been shown that there is extensive horizontal transfer of small sections of DNA (<2000bp) within sympatric populations of *Borrelia* (Haven et al., 2011). This

recombination is so frequent, it is 3-times more common than the point mutation rate. However, the BBJ25-operon is ~6500bp long, and RF and LD-type *Borrelia* are generally not sympatric, therefore greatly reducing the impact of this type of horizontal transfer.

The conservation of this operon possibly results from the selective advantage of controlling expression using a single promoter, allowing simpler gene regulation and therefore a rapid response to changes in environment, such as the change from tick to mammalian environment (Zhao et al., 2019). Furthermore, there may be a selective advantage in maintaining accurate stoichiometry even after lateral gene transfer of these regions. The phylogenetic analysis of the operon revealed a genealogy with two distinct groups, each containing many closely related sequences. This pattern is consistent with previous studies, as phylogenetic analysis of *Borrelia* isolates often reveal distinct groups separated by long internal branches but with short terminal branches within each group. This pattern is seen for a range of different genetic markers, such as the well-studied OspC (Margos et al., 2008; Qiu et al., 2008; Travinsky et al., 2010). This genealogy is seen in genomes under frequency dependent selection (FDS), the effects of which would be expected to be seen genome wide in *B. burgdorferi s.l* where the recombination frequency exceeds the mutation rate (Haven et al., 2011).

The computational framework identified two homologous copies of BBJ25 in the plasmid proteome of *B. valaisiana* (COR918 and COR9H1, 54% sequence identity, table 5.2). Extensive BLAST searches of *Borrelia* proteomes only revealed a single copy of BBJ25 in all other genospecies. An initial explanation is that the *B. valaisiana* sequences are paralogs and arose by a gene duplication event that has only occurred in this genospecies. However, subsequent analysis of the surrounding genes revealed that the two BBJ25 variants are found within two different allelic versions of the BBJ25-operon with a different gene order (Groups 1 and 2, figure 5.4). As the two versions of this operon are found distributed across many different genospecies, and the two copies of BBJ25 from *B. valaisiana* cluster with other sequences from the same allelic group (not with each other), the most parsimonious explanation is that *B. valaisiana* has acquired the second copy by horizontal transfer of the whole operon.

There is a possibility that *B. valaisiana* does not stably maintain both copies, and their detection results from either a transient event or from contamination with another genospecies. However, multiple isolates of *B. valaisiana* taken at different times and from different locations have been found to contain both copies. A search of the NCBI database for Identical Proteins revealed that *B. valaisiana* strains IPT14, IPT160, IPT111 and IPT116 all have both versions of BBJ25 (identical protein groups WP_015899315.1 and WP_012665424.1). These strains have been isolated from *Ixodes ricinus* ticks at different times (from 1997 to 2005) and at a range of locations across Europe. This suggests the presence of purifying selection to maintain both copies of the operon within this genospecies, which could be explored using Ka/Ks tests. The sequence identity between BBJ25 in group 1 and group 2 is ~54%, suggesting there is sufficient divergence for a shift in function (such as differential binding to different substrates). There may perhaps be a selective advantage in maintaining both copies in *B. valaisiana* as they act on different substrates.

Homologues of BBJ25 were found in all major families of spirochaetes with the exception of the *Leptospira* and *Brevinematales*. Phylogenetic analysis of the branching order of the spirochetes based on 16s rRNA sequences (Gupta et al., 2013) implies that the *Brachyspira* lineage diverged prior to the emergence of *Borrelia*, *Leptospira* and *Treponema*. Therefore, the existence of BBJ25 in the *Brachyspira* implies a role for this protein in the last common ancestor of all extant spirochetes. Evidence from comparative genomic analysis suggests that the most recent common ancestor of the spirochetes was most likely an invertebrate gut symbiont (Subramanian et al., 2000). Subsequent evolution towards obligate parasitic lifestyles has been associated with extensive lineage-specific gene-loss in the *Borrelia* and *Treponema* lineages. Functional characterisation of BBJ25 may reveal something about the biology of the early evolution of the spirochetes.

A possible role in nutrient acquisition in the mammalian host

While the precise role(s) of BBJ25 and the associated operon remain unknown, it is possible to establish certain constraints on the possible functions based on clones lacking the host plasmid. Within the background of *B. burgdorferi B31*, BBJ25 is found on lp38, a plasmid that has been shown to be dispensable for *in vitro* growth and the establishment of infection in

mouse by either needle inoculation or transmission by *I. scapularis* (Dulebohn et al., 2011). The absence of any detectable impact of pathogenicity suggests that BBJ25 does not have a critical role in the transmission from ticks, establishment of infection and dissemination to other tissues. The efficient *in vitro* growth of clones lacking lp38 was established using BSK medium, a rich medium which may potentially mask a phenotype if the protein is involved in nutrient acquisition or metabolism. This disagrees with previous research which demonstrated reduced mouse infectivity in a clone lacking lp38 (Botkin et al., 2006). A *B. burgdorferi* N40 strain was found to be no longer infectious in mice after 120 passages and was identified as having lost lp38 and lp-28-1 (Thomas et al., 2001). As lp-28-1 contains the *vls* locus associated with infectivity, no conclusions can be made from this study as to the role of lp38 in infectivity. However, as *B. burgdorferi* lacking the lp-38 plasmid was still able to grow and divide in BSK media, BBJ25 and other proteins encoded on this plasmid cannot be involved in cell structure or viability.

The genes in this operon have been included in studies analysing the regulation of *Borrelia* genes in response to environmental factors, to assess whether they are upregulated or downregulated in the mammalian host. Using DNA microarrays and a dialysis membrane chamber in a rat model, BBJ23, BBJ26 and BBJ27 from *B. burgdorferi* B31 were found to be upregulated in response to mammalian signals (Brooks et al., 2003). A 5.48 fold upregulation was seen for BBJ26, 3.59 fold upregulation for BBJ27 and 1.85 fold upregulation for BBJ23. None of the genes in the operon were found to be downregulated. Brooks et al found that the majority of proteins located on the plasmids were downregulated during cultivation in the mammalian model, whereas the majority of proteins located on the chromosome were upregulated. This is consistent with a previous study which noted that *Borrelia* outer surface lipoproteins are downregulated in mouse models during chronic *Borrelia* infection (Liang et al., 2002). A previous study by Revel et al also found upregulation of BBJ23 in response to mammalian signals, using similar methods to the above (Revel et al., 2002). BBJ23 was upregulated 4.09 fold in dialysis membrane chambers implanted in rats compared to conditions of unfed ticks (23°C, pH 7.5) and is upregulated 5.12 fold in fed ticks (response to both temperature and pH). However, this study used uncloned *B. burgdorferi* isolates, which are known to contain heterogeneity based on phenotype, plasmid content and growth rates

(Elias et al., 2002). A clonal isolate is therefore recommended for genetic studies using *B. burgdorferi* strains.

Upregulation of BBJ27 and BBJ28 has also been shown by Tokarz et al, which found a 4.8 and 5.5 fold upregulation of these proteins respectively in response to the addition of blood to spirochete cultures (Tokarz et al., 2004). A significant upregulation of BBJ25 was also seen, with 12.5 fold upregulation noted in response to a blood meal. This ABC transporter system was not expressed in unfed ticks and an upregulation was seen upon tick feeding. Furthermore, real time PCR showed that BBJ25 was only expressed in feeding ticks. It is likely that the operon remains expressed for the duration of mammalian infection, as this study shows upregulation during tick feeding and Brooks et al found upregulation during later infection (Brooks et al., 2003; Tokarz et al., 2004).

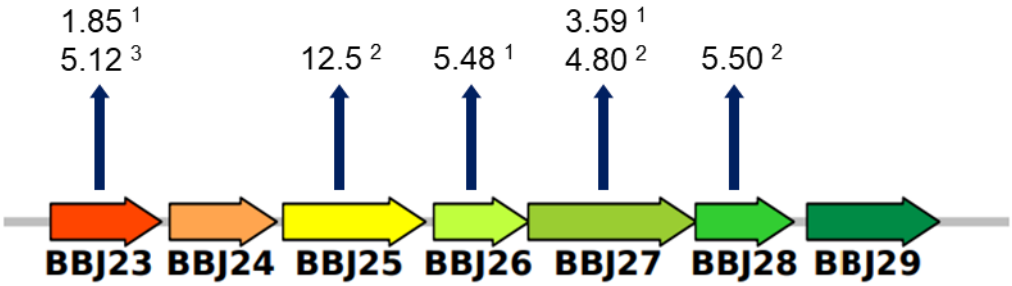


Figure 5.12. A summary of gene upregulation on the *B. burgdorferi* B31 lp38 ABC transporter operon. The numbers given refer to the fold upregulation of the corresponding gene to mammalian signals. [1] (Brooks et al., 2003), [2] (Tokarz et al., 2004), [3] (Revel et al., 2002)

Homology modelling of BBJ25 showed a β-barrel porin-like structure with 14-16 strands, with the majority showing a 16 stranded barrel, fitting with Schulz's rules for OM β-barrels (table 1.4, section 1.3.3). PHYRE2 was the only homology model to show 2 domains, with a 14 stranded β-barrel domain and a C-terminal non-barrel domain, which was shown by BOCTOPUS2 prediction. However, PHYRE2 prediction was low in confidence and the 16 stranded β-barrel modelled by RoseTTAfold had higher confidence. The templates used by MODELLER and PHYRE2 showed low sequence identity to BBJ25, but sequence identity is often low between β-barrel OM proteins despite a similar structure (Schulz, 2000). This has

been shown by the crystal structures of two porins which were superimposable despite only sharing moderate sequence identity (Koebnik et al., 2000).

Out of all of the modelling software used in this project, the RoseTTAfold algorithm produced a model most consistent with Schultz's rules for β-barrel proteins and predicted this with high confidence. The RoseTTAfold algorithm differs from traditional *ab initio* and homology modelling methods, and is based on the DeepMind framework used by AlphaFold. Homology modelling relies on the use of a known 3D structure as a template and where low homology to known structures is present, models produced by servers such as PHYRE2 are very unreliable (Kelley et al., 2015). As β-barrel OM proteins share low homology with one another, predicting protein folding and structure of these proteins using traditional modelling methods is not ideal. This problem may be overcome with the recent development of AlphaFold, an AI based algorithm that is capable of predicting a proteins 3D structure from its amino acid sequence with an accuracy near that of experimental structures (Jumper et al., 2021). RoseTTAfold is based on the neural architectures used by AlphaFold and produces structures with accuracies nearing those of AlphaFold, making it the most accurate modelling software used in this research (Baek et al., 2021).

Limitations were seen when using this framework to identify plasmid encoded β-barrel outer membrane proteins. As previously mentioned, an α-helical periplasmic protein (Oms28) was identified as a β-barrel outer membrane protein, despite the final filtering step using PRED-TMBB2. This may be overcome by the addition of a secondary structure prediction software, such as PSI-PRED, or by the utilisation of multiple β-barrel prediction algorithms as used by Kenedy et al. When the sequence for Oms28 is inputted in the β-barrel prediction software BOCTOPUS, BOMP and PRED-TMBB, it is not identified as a β-barrel protein and would therefore not have been seen in the final filtering results had multiple algorithms been used at the final stage. This framework also failed to identify some homologs of BBJ25, which were later identified using BBJ25 from *B. burgdorferi* B31 as a search sequence in PSI-BLAST. This included homologs from closely related species of *Borrelia* such as *B. afzelii* and *B. garinii*, which were filtered out prior to β-barrel prediction using PRED-TMBB2.

As this framework failed to identify obvious homologs, alternative methods of identifying β-barrel OM proteins based on the amino acid sequence could be explored. Given the advances in protein structure prediction by DeepMind algorithms such as AlphaFold, sequences could be submitted straight to these servers to determine if they are a β-barrel protein. However, the R-code used for the framework in this project allowed 5923 protein sequences to be filtered through numerous prediction algorithms in a matter of minutes. Submitting such a large number of sequences to a server like AlphaFold is a much slower process and is also likely to require manual viewing of each result to determine which are the protein type of interest (such as a β-barrel).

Chapter 6 Future work and conclusions

The aim of this research was to produce a series of truncated and full-length predicted β-barrel membrane proteins from *B. burgdorferi s.s*; BB_0027, BB_0405, BB_0406 and BB_0562, towards structural and functional characterisation. As these proteins and known *Borrelia* β-barrel OM proteins are encoded on the chromosome, this project also aimed to develop a framework to search the plasmid proteomes of *Borrelia*, to identify any plasmid encoded β-barrel OM proteins.

The computational framework developed in this project utilised an R-script to filter 14 plasmid proteomes of *Borrelia* through a series of prediction algorithms, in order to identify predicted β-barrel OM proteins. Although this framework did fail to identify obvious homologs of BBJ25, the use of an R-script provides a reproducible, less labour intensive method of filtering sequences to identify β-barrel proteins and could be reliably applied to other genomes. As the framework was only applied to plasmid proteomes, it could be used to analyse whole *Borrelia* genomes and results compared to those produced by Kenedy et al using a previous framework (Kenedy et al., 2016). While a false positive was identified, the framework can be modified to include a prediction algorithm more suitable for filtering out periplasmic proteins. Further analysis at the DNA level could be carried out to identify potential promoter sequences that may allow differential expression of each gene. As two copies of BBJ25 are maintained in *B. valaisiana*, future work should explore if other genospecies of *Borrelia* maintain two copies of this gene. The R-script used for this framework could be made more widely available, by publication in a repository such as GitHub. It could also be developed as a webserver, but this would be challenging as it is heavily reliant on the results from multiple third-party algorithms.

The identification of BBJ25 on the linear plasmids of *Borrelia* species opens up future research on this protein. Cloning and expression techniques utilised for *Borrelia* OMPs discussed in Chapter 3 and 4 could be applied for BBJ25, to produce protein suitable for structural studies. As this is predicted to be a β-barrel between 8-18 strands, structural data is essential to resolve this discrepancy in topology prediction. While high resolution X-ray data would be ideal for structure determination of BBJ25, in the absence of diffraction quality crystals any

lower resolution data such as SAXS or circular dichroism would provide experimental evidence to support or refute the homology models presented. Furthermore, it has been shown that for *de novo* protein structure prediction Google's DeepMind Alphafold outperforms the RoseTTAfold algorithm used in this work and therefore may provide a more robust prediction of BBJ25. As the BBJ25-operon is uncharacterised and several of the proteins have been shown to be upregulated in the mammalian host, the work presented here highlights the need to investigate the function of the whole operon.

The recently developed DeepMind modelling algorithms, RoseTTAfold and AlphaFold, provide more accurate predictions of protein structures in comparison to traditional homology modelling algorithms such as PHYRE2 and MODELLER. RoseTTAfold could be used to produce accurate models of BB_0027, BB_0405, BB_0406 and BB_0562, as they have previously only been modelled using traditional methods. This is especially useful given the low sequence identity between β-barrel OM proteins, as many traditional modelling methods rely on homology to proteins of known structure. Furthermore, the AI algorithm utilised by AlphaFold results in highly accurate structure predictions when compared against the experimental structure and may be helpful for membrane protein structure determination due to the difficulties crystallising this class of proteins, as were faced during the experimental work in this project.

Chapter 7 References

Adeolu, M., & Gupta, R. S. (2014). A phylogenomic and molecular marker based proposal for the division of the genus *Borrelia* into two genera: The emended genus *Borrelia* containing only the members of the relapsing fever *Borrelia*, and the genus *Borreliella* gen. nov. containing the members of the Lyme disease *Borrelia* (*Borrelia burgdorferi sensu lato* complex). *Antonie Van Leeuwenhoek, 105*(6), 1049-1072. https://doi.org/10.1007/s10482-014-0164-x

Adler, J. (1966). Chemotaxis in bacteria. *Science, 153*(3737), 708-716. https://doi.org/10.1126/science.153.3737.708

Aguero-Rosenfeld, M. E., Wang, G., Schwartz, I., & Wormser, G. P. (2005). Diagnosis of Lyme borreliosis. *Clinical Microbiology Reviews, 18*(3), 484-509. https://doi.org/10.1128/CMR.18.3.484-509.2005

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723. https://doi.org/10.1109/TAC.1974.1100705

Alao, O. R., & Decker, C. F. (2012). Lyme disease. *Disease-a-Month, 58*(6), 335-345. https://www.sciencedirect.com/science/article/pii/S0011502912000466?via%3Dihub

Anandan, A., & Vrielink, A. (2016). Detergents in membrane protein purification and crystallisation. *Advances in Experimental Medicine and Biology, 922*, 13. https://www.ncbi.nlm.nih.gov/pubmed/27553232

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology, 37*(4), 420-423. https://doi.org/10.1038/s41587-019-0036-z

Arvikar, S. L., MD, & Steere, A. C., MD. (2015). Diagnosis and treatment of Lyme arthritis. *Infectious Disease Clinics of North America, 29*(2), 269-280. https://doi.org/10.1016/j.idc.2015.02.004

Aslanidis, C., & de Jong, P. J. (1990). Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Research, 18*(20), 6069-6074. https://doi.org/10.1093/nar/18.20.6069

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., Dijk, A. A. v., Ebrecht, A. C., . . . Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track

neural network. *Science, 373*(6557), 871-876. https://www.science.org/doi/abs/10.1126/science.abj8754

Batra, M., Sharma, R., Malik, A., Dhindwal, S., Kumar, P., & Tomar, S. (2016). Crystal structure of pentapeptide-independent chemotaxis receptor methyltransferase (CheR) reveals idiosyncratic structural determinants for receptor recognition. *Journal of Structural Biology, 196*(3), 364-374. https://doi.org/10.1016/j.jsb.2016.08.005

Beaurepaire, C., & Chaconas, G. (2005). Mapping of essential replication functions of the linear plasmid lp17 of *B. burgdorferi* by targeted deletion walking. *Molecular Microbiology, 57*(1), 132-142. https://doi.org/10.1111/j.1365-2958.2005.04688.x

Becker, M., Bunikis, J., Lade, B. D., Dunn, J. J., Barbour, A. G., & Lawson, C. L. (2005). Structural investigation of *Borrelia burgdorferi* OspB, a bactericidal fab target. *The Journal of Biological Chemistry, 280*(17), 17363. https://www.ncbi.nlm.nih.gov/pubmed/15713683

Bellgard, M. I., Wanchanthuek, P., La, T., Ryan, K., Moolhuijzen, P., Albertyn, Z., Shaban, B., Motro, Y., Dunn, D. S., Schibeci, D., Hunter, A., Barrero, R., Phillips, N. D., & Hampson, D. J. (2009). Genome sequence of the pathogenic intestinal spirochete *Brachyspira hyodysenteriae* reveals adaptations to its lifestyle in the porcine large intestine. *Plos One, 4*(3), e4641. https://doi.org/10.1371/journal.pone.0004641

Bergström, S., & Normark, J. (2018). Microbiological features distinguishing Lyme disease and relapsing fever spirochetes. *Wiener Klinische Wochenschrift, 130*(15), 484-490. https://doi.org/10.1007/s00508-018-1368-2

Berndtson, K. (2013). Review of evidence for immune evasion and persistent infection in Lyme disease. *International Journal of General Medicine, 6*(default), 291-306. https://doi.org/10.2147/ijgm.s44114

Berven, F. S., Flikka, K., Jensen, H. B., & Eidhammer, I. (2004). BOMP: A program to predict integral β -barrel outer membrane proteins encoded within genomes of Gram-negative bacteria. *Nucleic Acids Research, 32*(Web Server), W394-W399. https://doi.org/10.1093/nar/gkh351

Bhate, C., & Schwartz, R. A. (2011). Lyme disease: Part I. advances and perspectives. *Journal of the American Academy of Dermatology, 64*(4), 619. http://www.ncbi.nlm.nih.gov/pubmed/21414493

Bhattacharjee, A., Oeemig, J. S., Kolodziejczyk, R., Meri, T., Kajander, T., Lehtinen, M. J., Iwaï, H., Jokiranta, T. S., & Goldman, A. (2013). Structural basis for complement evasion by

Lyme disease pathogen *Borrelia burgdorferi. The Journal of Biological Chemistry, 288*(26), 18685-18695. https://doi.org/10.1074/jbc.M113.459040

Bhide, M. R., Escudero, R., Camafeita, E., Gil, H., Jado, I., & Anda, P. (2009). Complement factor H binding by different Lyme disease and relapsing fever *Borrelia* in animals and human. *BMC Research Notes, 2*(1), 134. https://doi.org/10.1186/1756-0500-2-134

Bista, S., Singh, P., Bernard, Q., Yang, X., Hart, T., Lin, Y., Kitsou, C., Singh Rana, V., Zhang, F., Linhardt, R. J., Zhnag, K., Akins, D. R., Hritzo, L., Kim, Y., Grab, D. J., Dumler, J. S., & Pal, U. (2020). A novel laminin-binding protein mediates microbial-endothelial cell interactions and facilitates dissemination of Lyme disease pathogens. *The Journal of Infectious Diseases, 221*(9), 1438-1447. https://doi.org/10.1093/infdis/jiz626

Borchers, A. T., Keen, C. L., Huntley, A. C., & Gershwin, M. E. (2014). Lyme disease: A rigorous review of diagnostic criteria and treatment. *Journal of Autoimmunity, 57*, 82-115. https://doi.org/10.1016/j.jaut.2014.09.004

Botkin, D. J., Abbott, A. N., Stewart, P. E., Rosa, P. A., Kawabata, H., Watanabe, H., & Norris, S. J. (2006). Identification of potential virulence determinants by Himar1 transposition of infectious *Borrelia burgdorferi* B31. *Infection and Immunity, 74*(12), 6690-6699. https://doi.org/10.1128/IAI.00993-06

Brissette, C. A., Haupt, K., Barthel, D., Cooley, A. E., Bowman, A., Skerka, C., Wallich, R., Zipfel, P. F., Kraiczy, P., & Stevenson, B. (2009). *Borrelia burgdorferi* infection-associated surface proteins ErpP, ErpA, and ErpC bind human plasminogen. *Infection and Immunity, 77*(1), 300-306. https://doi.org/10.1128/IAI.01133-08

Brisson, D., Drecktrah, D., Eggers, C. H., & Samuels, D. S. (2012). Genetics of *Borrelia burgdorferi. Annual Review of Genetics, 46*, 515-536. https://doi.org/10.1146/annurev-genet-011112-112140

Brooks, C. S., Hefty, P. S., Jolliff, S. E., & Akins, D. R. (2003). Global analysis of *Borrelia burgdorferi* genes regulated by mammalian host-specific signals. *Infection and Immunity, 71*(6), 3371-3383. https://doi.org/10.1128/IAI.71.6.3371-3383.2003

Brown, G. (2016). Characterisation and structural studies of a superoxide dismutase and OmpA-like proteins from *Borrelia burgdorferi sensu lato* http://eprints.hud.ac.uk/31870/

Brown, G., Broxham, A. H., Cherrington, S. E., Thomas, D. C., Dyer, A., Stejskal, L., & Bingham, R. J. (2019). Expression, purification and metal utilization of recombinant

SodA from *Borrelia burgdorferi*. *Protein Expression and Purification, 163*, 105447. https://doi.org/10.1016/j.pep.2019.105447

Bunikis, I., Denker, K., Östberg, Y., Andersen, C., Benz, R., & Bergström, S. (2008). An RND-type efflux system in *Borrelia burgdorferi* is involved in virulence and resistance to antimicrobial compounds. *PLoS Pathogens, 4*(2), e1000009. https://doi.org/10.1371/journal.ppat.1000009

Burgdorfer, W., Barbour, A. G., Hayes, S. F., Benach, J. L., Grunwaldt, E., & Davis, J. P. (1982). Lyme disease-a tick-borne spirochetosis? *Science, 216*(4552), 1317-1319. https://doi.org/10.1126/science.7043737

Bykowski, T., Woodman, M. E., Cooley, A. E., Brissette, C. A., Brade, V., Wallich, R., Kraiczy, P., & Stevenson, B. (2007). Coordinated expression of *Borrelia burgdorferi* complement regulator-acquiring surface proteins during the Lyme disease spirochete's mammal-tick infection cycle. *Infection and Immunity, 75*(9), 4227-4236. https://doi.org/10.1128/IAI.00604-07

Byram, R., Stewart, P. E., & Rosa, P. (2004). The essential nature of the ubiquitous 26-kilobase circular replicon of *Borrelia burgdorferi*. *Journal of Bacteriology, 186*(11), 3561-3569. https://doi.org/10.1128/JB.186.11.3561-3569.2004

Cadavid, D., Bai, Y., Hodzic, E., Narayan, K., Barthold, S. W., & Pachner, A. R. (2004). Cardiac involvement in non-human primates infected with the Lyme disease spirochete *Borrelia burgdorferi*. *Laboratory Investigation, 84*(11), 1439-1450. https://doi.org/10.1038/labinvest.3700177

Cardenas de la Garza, Jesus Alberto, De la Cruz-Valadez, E., Ocampo-Candiani, J., & Welsh, O. (2019). Clinical spectrum of Lyme disease. *European Journal of Clinical Microbiology & Infectious Diseases, 38*(2), 201-208. https://doi.org/10.1007/s10096-018-3417-1

Casjens, S. (2000). *Borrelia* genomes in the year 2000. *Journal of Molecular Microbiology and Biotechnology, 2*(4), 401-410.

Casjens, S. R., Di, L., Akther, S., Mongodin, E. F., Luft, B. J., Schutzer, S. E., Fraser, C. M., & Qiu, W. (2018). Primordial origin and diversification of plasmids in Lyme disease agent bacteria. *BMC Genomics, 19*(1), 218. https://doi.org/10.1186/s12864-018-4597-x

Casjens, S. R., Gilcrease, E. B., Vujadinovic, M., Mongodin, E. F., Luft, B. J., Schutzer, S. E., Fraser, C. M., & Qiu, W. (2017). Plasmid diversity and phylogenetic consistency in the Lyme disease agent *Borrelia burgdorferi*. *BMC Genomics, 18*(1), 165. https://doi.org/10.1186/s12864-017-3553-5

Casjens, S. R., Mongodin, E. F., Qiu, W., Luft, B. J., Schutzer, S. E., Gilcrease, E. B., Huang, W. M., Vujadinovic, M., Aron, J. K., Vargas, L. C., Freeman, S., Radune, D., Weidman, J. F., Dimitrov, G. I., Khouri, H. M., Sosa, J. E., Halpin, R. A., Dunn, J. J., & Fraser, C. M. (2012). Genome stability of Lyme disease spirochetes: Comparative genomics of *Borrelia burgdorferi* plasmids. *PloS One, 7*(3), e33280. https://doi.org/10.1371/journal.pone.0033280

Casjens, S., Palmer, N., Van Vugt, R., Mun Huang, W., Stevenson, B., Rosa, P., Lathigra, R., Sutton, G., Peterson, J., Dodson, R. J., Haft, D., Hickey, E., Gwinn, M., White, O., & M. Fraser, C. (2000). A bacterial genome in flux: The twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi. Molecular Microbiology, 35*(3), 490-516. https://doi.org/10.1046/j.1365-2958.2000.01698.x

Centers for Disease Control and Prevention. (1995). Recommendations for test performance and interpretation from the second national conference on serologic diagnosis of Lyme disease. *MMWR. Morbidity and Mortality Weekly Report, 44*(31), 590-591.

Centers for Disease Control and Prevention. (2018a). Lyme disease charts and figures: Historical data. https://www.cdc.gov/lyme/stats/graphs.html

Centers for Disease Control and Prevention. (2018b). Lyme disease rashes and look-alikes | CDC. https://www.cdc.gov/lyme/signs_symptoms/rashes.html

Cerar, T., Korva, M., Avšič-Županc, T., & Ružić-Sabljić, E. (2015). Detection, identification and genotyping of *Borrellia* spp. in rodents in Slovenia by PCR and culture. *BMC Veterinary Research, 11*(1), 188. https://doi.org/10.1186/s12917-015-0501-y

Coburn, J., & Cugini, C. (2003). Targeted mutation of the outer membrane protein P66 disrupts attachment of the Lyme disease agent, *Borrelia burgdorferi*, to integrin αvβ3. *Proceedings of the National Academy of Sciences - PNAS, 100*(12), 7301-7306. https://doi.org/10.1073/pnas.1131117100

Coburn, J., Chege, W., Magoun, L., Bodary, S. C., & Leong, J. M. (1999). Characterization of a candidate *Borrelia burgdorferi* beta3-chain integrin ligand identified using a phage display library. *Molecular Microbiology, 34*(5), 926-940. https://doi.org/10.1046/j.1365-2958.1999.01654.x

Confer, A. W., & Ayalew, S. (2013). The OmpA family of proteins: Roles in bacterial pathogenesis and immunity. *Veterinary Microbiology, 163*(3-4), 207-222. https://doi.org/10.1016/j.vetmic.2012.08.019

Cordes, F. S., Kraiczy, P., Roversi, P., Simon, M. M., Brade, V., Jahraus, O., Wallis, R., Goodstadt, L., Ponting, C. P., Skerka, C., Zipfel, P. F., Wallich, R., & Lea, S. M. (2006). Structure–function mapping of BbCRASP-1, the key complement factor H and FHL-1 binding protein of *Borrelia burgdorferi. International Journal of Medical Microbiology, 296*, 177-184. https://doi.org/10.1016/j.ijmm.2006.01.011

Cordes, F. S., Roversi, P., Kraiczy, P., Simon, M. M., Brade, V., Jahraus, O., Wallis, R., Skerka, C., Zipfel, P. F., Wallich, R., & Lea, S. M. (2005). A novel fold for the factor H-binding protein BbCRASP-1 of *Borrelia burgdorferi. Nature Structural & Molecular Biology, 12*(3), 276-277. https://doi.org/10.1038/nsmb902

Cox, D. L., Luthra, A., Dunham-Ems, S., Desrosiers, D. C., Salazar, J. C., Caimano, M. J., & Radolf, J. D. (2010). Surface immunolabeling and consensus computational framework to identify candidate rare outer membrane proteins of treponema pallidum. *Infection and Immunity, 78*(12), 5178-5194. https://doi.org/10.1128/IAI.00834-10

Crowley, J. T., Toledo, A. M., LaRocca, T. J., Coleman, J. L., London, E., & Benach, J. L. (2013). Lipid exchange between *Borrelia burgdorferi* and host cells. *PLoS Pathogens, 9*(1), e1003109. https://doi.org/10.1371/journal.ppat.1003109

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2011). ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics (Oxford, England), 27*(8), 1164-1165. https://doi.org/10.1093/bioinformatics/btr088

de Taeye, S. W., Kreuk, L., van Dam, A. P., Hovius, J. W., & Schuijt, T. J. (2013). Complement evasion by *Borrelia burgdorferi*: It takes three to tango. *Trends in Parasitology, 29*(3), 119-128. https://doi.org/10.1016/j.pt.2012.12.001

Djordjevic, S., & Stock, A. M. (1997). Crystal structure of the chemotaxis receptor methyltransferase CheR suggests a conserved structural motif for binding S-adenosylmethionine. *Structure, 5*(4), 545-558. https://doi.org/10.1016/S0969-2126(97)00210-4

Dulebohn, D. P., Bestor, A., Rego, R. O. M., Stewart, P. E., & Rosa, P. A. (2011). *Borrelia burgdorferi* linear plasmid 38 is dispensable for completion of the mouse-tick infectious cycle. *Infection and Immunity, 79*(9), 3510-3517. https://doi.org/10.1128/IAI.05014-11

Dworkin, M. S., Schwan, T. G., Anderson, D. E., & Borchardt, S. M. (2008). Tick-borne relapsing fever. *Infectious Disease Clinics of North America, 22*(3), 449-viii. https://doi.org/10.1016/j.idc.2008.03.006

Dyer, A., Brown, G., Stejskal, L., Laity, P. R., & Bingham, R. J. (2015). The *Borrelia afzelii* outer membrane protein BAPKO_0422 binds human factor-H and is predicted to form a membrane-spanning β-barrel. *Bioscience Reports, 35*(4), e00240. https://doi.org/10.1042/BSR20150095

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research, 32*(5), 1792-1797. https://doi.org/10.1093/nar/gkh340

Elias, A. F., Stewart, P. E., Grimm, D., Caimano, M. J., Eggers, C. H., Tilly, K., Bono, J. L., Akins, D. R., Radolf, J. D., Schwan, T. G., & Rosa, P. (2002). Clonal polymorphism of *Borrelia burgdorferi* strain B31 MI: Implications for mutagenesis in an infectious strain background. *Infection and Immunity, 70*(4), 2139-2150. https://doi.org/10.1128/IAI.70.4.2139-2150.2002

Federizon, J., Lin, Y., & Lovell, J. F. (2019). Antigen engineering approaches for Lyme disease vaccines. *Bioconjugate Chemistry, 30*(5), 1259-1272. https://doi.org/10.1021/acs.bioconjchem.9b00167

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution, 39*(4), 783-791. https://doi.org/10.2307/2408678

Fikrig, E., Pal, U., Chen, M., Anderson, J. F., & Flavell, R. A. (2004). OspB antibody prevents *Borrelia burgdorferi* colonization of *Ixodes scapularis*. *Infection and Immunity, 72*(3), 1755-1759. https://doi.org/10.1128/IAI.72.3.1755-1759.2004

Fikrig, E., Feng, W., Barthold, S. W., Telford, S. R., III, & Flavell, R. A. (2000). Arthropod- and host-specific *Borrelia burgdorferi* BBK32 expression and the inhibition of spirochete transmission. *The Journal of Immunology, 164*(10), 5344-5351. http://www.jimmunol.org/cgi/content/abstract/164/10/5344

Fischer, J. R., Parveen, N., Magoun, L., & Leong, J. M. (2003). Decorin-binding proteins A and B confer distinct mammalian cell type-specific attachment by *Borrelia burgdorferi*, the Lyme disease spirochete. *Proceedings of the National Academy of Sciences of the United States of America, 100*(12), 7307-7312. https://doi.org/10.1073/pnas.1231043100

Fischer, J. R., LeBlanc, K. T., & Leong, J. M. (2006). Fibronectin binding protein BBK32 of the Lyme disease spirochete promotes bacterial attachment to glycosaminoglycans. *Infection and Immunity, 74*(1), 435-441. https://doi.org/10.1128/IAI.74.1.435-441.2006

Fish, A. E., Pride, Y. B., & Pinto, D. S. (2008). Lyme carditis. *Infectious Disease Clinics of North America, 22*(2), 275-288, vi. https://doi.org/10.1016/j.idc.2007.12.008

Franke, J., Hildebrandt, A., & Dorn, W. (2013). Exploring gaps in our knowledge on Lyme borreliosis spirochaetes – updates on complex heterogeneity, ecology, and pathogenicity. *Ticks and Tick-Borne Diseases, 4*(1-2), 11-25. https://doi.org/10.1016/j.ttbdis.2012.06.007

Fraser, C. M., Norris, S. J., Weinstock, G. M., White, O., Sutton, G. G., Dodson, R., Gwinn, M., Hickey, E. K., Clayton, R., Ketchum, K. A., Sodergren, E., Hardham, J. M., McLeod, M. P., Salzberg, S., Peterson, J., Khalak, H., Richardson, D., Howell, J. K., Chidambaram, M., . . . Venter, J. C. (1998). Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science (New York, N.Y.), 281*(5375), 375-388. https://doi.org/10.1126/science.281.5375.375

Fraser, C. M., Casjens, S., Gocayne, J., Salzberg, S., Hanson, M., McDonald, L., Hickey, E. K., van Vugt, R., Tomb, J., Smith, H. O., Gwinn, M., Quackenbush, J., Sutton, G. G., Clayton, R., White, O., Lathigra, R., Bowman, C., Cotton, M. D., Hatch, B., . . . Dodson, R. (1997). Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature, 390*(6660), 580-586. https://doi.org/10.1038/37551

Fraser, C. M., Fleischmann, R. D., Sutton, G. G., van Vugt, R., Clayton, R., Artiach, P., McDonald, L., Adams, M. D., Dodson, R., Ketchum, K. A., Utterback, T., Palmer, N., Horst, K., Richardson, D., Quackenbush, J., Bowman, C., Garland, S., Gocayne, J., Peterson, J., . . . Cotton, M. (1997). Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature, 390*(6660), 580-586. https://doi.org/10.1038/37551

Fuchs, H., Wallich, R., Simon, M. M., & Kramer, M. D. (1994). The outer surface protein A of the spirochete *Borrelia burgdorferi* is a plasmin(ogen) receptor. *Proceedings of the National Academy of Sciences of the United States of America, 91*(26), 12594-12598. https://doi.org/10.1073/pnas.91.26.12594

Fujita, T. (2002). Evolution of the lectin–complement pathway and its role in innate immunity. *Nature Reviews Immunology, 2*(5), 346-353. https://doi.org/10.1038/nri800

García-Fontana, C., Reyes-Darias, J. A., Muñoz-Martínez, F., Alfonso, C., Morel, B., Ramos, J. L., & Krell, T. (2013). High specificity in CheR methyltransferase function CheR2 of *Pseudomonas putida* is essential for chemotaxis, whereas CheR1 is involved in biofilm formation. *Journal of Biological Chemistry, 288*(26), 18987-18999. https://doi.org/10.1074/jbc.M113.472605

Gofton, A. W., Margos, G., Fingerle, V., Hepner, S., Loh, S., Ryan, U., Irwin, P., & Oskam, C. L. (2018). Genome-wide analysis of *Borrelia turcica* and '*Candidatus Borrelia tachyglossi*' shows relapsing fever-like genomes with unique genomic links to Lyme disease *Borrelia*. *Infection, Genetics and Evolution, 66*, 72-81. https://doi.org/10.1016/j.meegid.2018.09.013

Gordon, D. L., Kaufman, R. M., Blackmore, T. K., Kwong, J., & Lublin, D. M. (1995). Identification of complement regulatory domains in human factor H. *The Journal of Immunology (1950), 155*(1), 348-356. http://www.jimmunol.org/cgi/content/abstract/155/1/348

Gräslund, S., Sagemark, J., Berglund, H., Dahlgren, L., Flores, A., Hammarström, M., Johansson, I., Kotenyova, T., Nilsson, M., Nordlund, P., & Weigelt, J. (2008). The use of systematic N- and C-terminal deletions to promote production and structural studies of recombinant proteins. *Protein Expression and Purification, 58*(2), 210-221. https://doi.org/10.1016/j.pep.2007.11.008

Greenfield, N. J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols, 1*(6), 2876-2890. https://doi.org/10.1038/nprot.2006.202

Grimm, D., Tilly, K., Byram, R., Stewart, P. E., Krum, J. G., Bueschel, D. M., Schwan, T. G., Policastro, P. F., Elias, A. F., Rosa, P. A., & Falkow, S. (2004). Outer-surface protein C of the Lyme disease spirochete: A protein induced in ticks for infection of mammals. *Proceedings of the National Academy of Sciences - PNAS, 101*(9), 3142-3147. https://doi.org/10.1073/pnas.0306845101

Gupta, R. S., Mahmood, S., & Adeolu, M. (2013). A phylogenomic and molecular signature based approach for characterization of the phylum spirochaetes and its major clades: Proposal for a taxonomic revision of the phylum. *Frontiers in Microbiology, 4*, 217. https://doi.org/10.3389/fmicb.2013.00217

Hallström, T., Haupt, K., Kraiczy, P., Hortschansky, P., Wallich, R., Skerka, C., & Zipfel, P. F. (2010). Complement Regulator—Acquiring surface protein 1 of *Borrelia burgdorferi* binds to human bone morphogenic protein 2, several extracellular matrix proteins, and plasminogen. *The Journal of Infectious Diseases, 202*(3), 490-498. https://doi.org/10.1086/653825

Hallström, T., Siegel, C., Mörgelin, M., Kraiczy, P., Skerka, C., & Zipfel, P. F. (2013). CspA from *Borrelia burgdorferi* inhibits the terminal complement pathway. *mBio, 4*(4), 13. https://doi.org/10.1128/mBio.00481-13

Haupt, K., Kraiczy, P., Wallich, R., Brade, V., Skerka, C., & Zipfel, P. F. (2007). Binding of human factor H-related protein 1 to serum-resistant *Borrelia burgdorferi* is mediated by borrelial complement regulator-acquiring surface proteins. *The Journal of Infectious Diseases, 196*(1), 124-133. https://doi.org/10.1086/518509

Haven, J., Vargas, L. C., Mongodin, E. F., Xue, V., Hernandez, Y., Pagan, P., Fraser-Liggett, C. M., Schutzer, S. E., Luft, B. J., Casjens, S. R., & Qiu, W. (2011). Pervasive recombination and sympatric genome diversification driven by frequency-dependent selection in *Borrelia burgdorferi*, the Lyme disease bacterium. *Genetics (Austin), 189*(3), 951-966. https://doi.org/10.1534/genetics.111.130773

Hayat, S., Peters, C., Shu, N., Tsirigos, K. D., & Elofsson, A. (2016). Inclusion of dyad-repeat pattern improves topology prediction of transmembrane β-barrel proteins. *Bioinformatics, 32*(10), 1571-1573. https://doi.org/10.1093/bioinformatics/btw025

Hefty, P. S., Jolliff, S. E., Caimano, M. J., Wikel, S. K., & Akins, D. R. (2002). Changes in temporal and spatial patterns of outer surface lipoprotein expression generate population heterogeneity and antigenic diversity in the Lyme disease spirochete, *Borrelia burgdorferi. Infection and Immunity, 70*(7), 3468-3478. https://doi.org/10.1128/IAI.70.7.3468-3478.2002

Hong, H., Patel, D. R., Tamm, L. K., & van den Berg, B. (2006). The outer membrane protein OmpW forms an eight-stranded β-barrel with a hydrophobic channel. *The Journal of Biological Chemistry, 281*(11), 7568-7577. https://doi.org/10.1074/jbc.M512365200

Hu, L. T. (2016). In the clinic: Lyme disease. *Annals of Internal Medicine, 164*(9), ITC65. https://search.proquest.com/docview/1788553342

Hurvich, C. M., & Tsai, C. (1989). Regression and time series model selection in small samples. *Biometrika, 76*(2), 297-307. https://doi.org/10.1093/biomet/76.2.297

Jia, J., Lu, L., Huang, H., Somerville, R. L., Zeng, R., Miao, Y., Shen, Y., Zhang, Y., Ma, W., Chen, Z., Fu, G., Gu, W., Cai, Z., Zhu, G., Xia, Q., Tu, Y., Yao, Z., Guo, X., Xu, H., . . . Yin, H. (2003). Unique physiological and pathogenic features of *Leptospira interrogans* revealed by whole-genome sequencing. *Nature (London), 422*(6934), 888-893. https://doi.org/10.1038/nature01597

John, T. M., & Taege, A. J. (2019). Appropriate laboratory testing in Lyme disease. *Cleveland Clinic Journal of Medicine, 86*(11), 751-759.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A.

J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature (London), 596*(7873), 583-589. https://doi.org/10.1038/s41586-021-03819-2

Juncker, A. S., Willenbrock, H., von Heijne, G., Brunak, S., Nielsen, H., & Krogh, A. (2003). Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Science, 12*(8), 1652-1662. https://doi.org/10.1110/ps.0303703

Kajander, T., Lehtinen, M. J., Hyvarinen, S., Bhattacharjee, A., Leung, E., Isenman, D. E., Meri, S., Goldman, A., & Jokiranta, T. S. (2011). Dual interaction of factor H with C3d and glycosaminoglycans in host-nonhost discrimination by complement. *Proceedings of the National Academy of Sciences - PNAS, 108*(7), 2897-2902. https://doi.org/10.1073/pnas.1017087108

Kamp, H. D., Swanson, K. A., Wei, R. R., Dhal, P. K., Dharanipragada, R., Kern, A., Sharma, B., Sima, R., Hajdusek, O., Hu, L. T., Wei, C., & Nabel, G. J. (2020). Design of a broadly reactive Lyme disease vaccine. *Npj Vaccines, 5*(1), 1-10. https://doi.org/10.1038/s41541-020-0183-8

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols, 10*(6), 845-858. https://doi.org/10.1038/nprot.2015.053

Kemper, C., Atkinson, J. P., & Hourcade, D. E. (2010). Properdin: Emerging roles of a pattern-recognition molecule. *Annual Review of Immunology, 28*, 131-155. https://doi.org/10.1146/annurev-immunol-030409-101250

Kenedy, M. R., Vuppala, S. R., Siegel, C., Kraiczy, P., & Akins, D. R. (2009). CspA-mediated binding of human factor H inhibits complement deposition and confers serum resistance in *Borrelia burgdorferi. Infection and Immunity, 77*(7), 2773-2782. https://doi.org/10.1128/IAI.00318-09

Kenedy, M. R., Luthra, A., Anand, A., Dunn, J. P., Radolf, J. D., & Akins, D. R. (2014). Structural modeling and physicochemical characterization provide evidence that P66 forms a β-barrel in the *Borrelia burgdorferi* outer membrane. *Journal of Bacteriology, 196*(4), 859-872. https://doi.org/10.1128/JB.01236-13

Kenedy, M. R., Lenhart, T. R., & Akins, D. R. (2012). The role of *Borrelia burgdorferi* outer surface proteins. *FEMS Immunology & Medical Microbiology, 66*(1), 1-19. https://doi.org/10.1111/j.1574-695X.2012.00980.x

Kenedy, M. R., Scott, E. J., Shrestha, B., Anand, A., & Iqbal, H. (2016). Consensus computational network analysis for identifying candidate outer membrane proteins from *Borrelia* spirochetes. *BMC Microbiology, 16*(1), 141. https://doi.org/10.1186/s12866-016-0762-z

Kingry, L. C., Anacker, M., Pritt, B., Bjork, J., Respicio-Kingry, L., Liu, G., Sheldon, S., Boxrud, D., Strain, A., Oatman, S., Berry, J., Sloan, L., Mead, P., Neitzel, D., Kugeler, K. J., & Petersen, J. M. (2018). Surveillance for and discovery of *Borrelia* species in US patients suspected of tickborne illness. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America, 66*(12), 1864-1871. https://doi.org/10.1093/cid/cix1107

Kobryn, K., & Chaconas, G. (2002). ResT, a telomere resolvase encoded by the Lyme disease spirochete. *Molecular Cell, 9*(1), 195-201. https://doi.org/10.1016/S1097-2765(01)00433-6

Koebnik, R., Locher, K. P., & Van Gelder, P. (2000). Structure and function of bacterial outer membrane proteins: Barrels in a nutshell. *Molecular Microbiology, 37*(2), 239-253. https://doi.org/10.1046/j.1365-2958.2000.01983.x

Kolodziejczyk, R., Mikula, K. M., Kotila, T., Postis, V. L. G., Jokiranta, T. S., Goldman, A., & Meri, T. (2017). Crystal structure of a tripartite complex between C3dg, C-terminal domains of factor H and OspE of *Borrelia burgdorferi. Plos One, 12*(11), e0188127. https://doi.org/10.1371/journal.pone.0188127

Kopp, A., Hebecker, M., Svobodová, E., & Józsi, M. (2012). Factor H: A complement regulator in health and disease, and a mediator of cellular interactions. *Biomolecules (Basel, Switzerland), 2*(1), 46-75. https://doi.org/10.3390/biom2010046

Kraiczy, P., & Stevenson, B. (2013). Complement regulator-acquiring surface proteins of *Borrelia burgdorferi*: Structure, function and regulation of gene expression. *Ticks and Tick-Borne Diseases, 4*(1-2), 26-34. https://doi.org/10.1016/j.ttbdis.2012.10.039

Kraiczy, P., Skerka, C., Kirschfink, M., Brade, V., & Zipfel, P. F. (2001). Immune evasion of *Borrelia burgdorferi* by acquisition of human complement regulators FHL-1/reconectin and factor H. *European Journal of Immunology, 31*(6), 1674-1684. https://doi.org/AID-IMMU1674>3.0.CO;2-2

Krishnan, S., & Prasadarao, N. V. (2012). Outer membrane protein A and OprF: Versatile roles in Gram-negative bacterial infections. *FEBS Journal, 279*(6), 919-931. https://doi.org/10.1111/j.1742-4658.2012.08482.x

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a Hidden Markov model: Application to complete genomes. *Journal of Molecular Biology, 305*(3), 567-580. https://doi.org/10.1006/jmbi.2000.4315

Krupka, M., Raska, M., Belakova, J., Horynova, M., Novotny, R., & Weigl, E. (2007). Biological aspects of Lyme disease spirochetes: Unique bacteria of the *Borrelia burgdorferi* species group. *Biomedical Papers of the Medical Faculty of the University Palacky, Olomouc, Czechoslovakia, 151*(2), 175-186. https://doi.org/10.5507/bp.2007.032

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution, 35*(6), 1547-1549. https://doi.org/10.1093/molbev/msy096

Kumaran, D., Eswaramoorthy, S., Luft, B. J., Koide, S., Dunn, J. J., Lawson, C. L., & Swaminathan, S. (2001). Crystal structure of outer surface protein C (OspC) from the Lyme disease spirochete, *Borrelia burgdorferi*. *The EMBO Journal, 20*(5), 971-978. https://doi.org/10.1093/emboj/20.5.971

Kung, F., Kaur, S., Smith, A. A., Yang, X., Wilder, C. N., Sharma, K., Buyuktanir, O., & Pal, U. (2016). A *Borrelia burgdorferi* surface-exposed transmembrane protein lacking detectable immune responses supports pathogen persistence and constitutes a vaccine target. *The Journal of Infectious Diseases, 213*(11), 1786-1795. https://doi.org/10.1093/infdis/jiw013

Kurtenbach, K., Hanincová, K., Tsao, J. I., Margos, G., Fish, D., & Ogden, N. H. (2006). Fundamental processes in the evolutionary ecology of Lyme borreliosis. *Nature Reviews. Microbiology, 4*(9), 660-669. https://doi.org/10.1038/nrmicro1475

Lackum, K. v., Miller, J. C., Bykowski, T., Riley, S. P., Woodman, M. E., Brade, V., Kraiczy, P., Stevenson, B., & Wallich, R. (2005). *Borrelia burgdorferi* regulates expression of complement regulator-acquiring surface protein 1 during the mammal-tick infection cycle. *Infection and Immunity, 73*(11), 7398-7405. https://doi.org/10.1128/IAI.73.11.7398-7405.2005

Lagal, V., Portnoï, D., Faure, G., Postic, D., & Baranton, G. (2006). *Borrelia burgdorferi sensu stricto* invasiveness is correlated with OspC–plasminogen affinity. *Microbes and Infection, 8*(3), 645-652. https://doi.org/10.1016/j.micinf.2005.08.017

Lambris, J. D., Ricklin, D., & Geisbrecht, B. V. (2008). Complement evasion by human pathogens. *Nature Reviews. Microbiology, 6*(2), 132. https://doi.org/10.1038/nrmicro1824

LaRocca, T. J., Crowley, J. T., Cusack, B. J., Pathak, P., Benach, J., London, E., Garcia-Monco, J. C., & Benach, J. L. (2010). Cholesterol lipids of *Borrelia burgdorferi* form lipid rafts and are required for the bactericidal mechanism of a complement-independent antibody. *Cell Host & Microbe, 8*(4), 331-342. https://doi.org/10.1016/j.chom.2010.09.001

Lenhart, T. R., & Akins, D. R. (2010). *Borrelia burgdorferi* locus BB0795 encodes a BamA orthologue required for growth and efficient localization of outer membrane proteins. *Molecular Microbiology, 75*(3), 692-709. https://doi.org/10.1111/j.1365-2958.2009.07015.x

Li, H., Dunn, J. J., Luft, B. J., & Lawson, C. L. (1997). Crystal structure of Lyme disease antigen outer surface protein A complexed with an fab. *Proceedings of the National Academy of Sciences - PNAS, 94*(8), 3584-3589. https://doi.org/10.1073/pnas.94.8.3584

Li, W., Wen, L., Li, C., Chen, R., Ye, Z., Zhao, J., & Pan, J. (2016). Contribution of the outer membrane protein OmpW in *Escherichia coli* to complement resistance from binding to factor H. *Microbial Pathogenesis, 98*, 57-62. https://doi.org/10.1016/j.micpath.2016.06.024

Liang, F. T., Alvarez, A. L., Gu, Y., Nowling, J. M., Ramamoorthy, R., & Philipp, M. T. (1999). An immunodominant conserved region within the variable domain of VlsE, the variable surface antigen of *Borrelia burgdorferi*. *The Journal of Immunology, 163*(10), 5566. http://www.jimmunol.org/cgi/content/abstract/163/10/5566

Liang, F. T., Nelson, F. K., & Fikrig, E. (2002). Molecular adaptation of *Borrelia burgdorferi* in the murine host. *The Journal of Experimental Medicine, 196*(2), 275-280. https://doi.org/10.1084/jem.20020770

Lin, Y., Frye, A. M., Nowak, T. A., & Kraiczy, P. (2020). New insights into CRASP-mediated complement evasion in the Lyme disease enzootic cycle. *Frontiers in Cellular and Infection Microbiology, 10*, 1. https://doi.org/10.3389/fcimb.2020.00001

Lindgren, E., & Jaenson, T. G. T. (2007). Lyme borreliosis in Europe: Influences of climate and climate change, epidemiology, ecology and adaptation measures. World Health Organization.

Marcus, K. (1912). Verhandlungen der dermatologischen Gesellschaft zu Stockholm. *Archiv für Dermatologie und Syphilis, 112*(1), 26-28. doi:10.1007/BF01866988

Margos, G., Marosevic, D., Cutler, S., Derdakova, M., Diuk-Wasser, M., Emler, S., Fish, D., Gray, J., Hunfeldt, K., Jaulhac, B., Kahl, O., Kovalev, S., Kraiczy, P., Lane, R. S., Lienhard, R., Lindgren, P. E., Ogden, N., Ornstein, K., Rupprecht, T., . . . Fingerle, V. (2017). There

is inadequate evidence to support the division of the genus *Borrelia*. *International Journal of Systematic and Evolutionary Microbiology, 67*(4), 1081-1084. https://doi.org/10.1099/ijsem.0.001717

Margos, G., Gatewood, A. G., Aanensen, D. M., Hanincová, K., Terekhova, D., Vollmer, S. A., Cornet, M., Piesman, J., Donaghy, M., Bormane, A., Hurn, M. A., Feil, E. J., Fish, D., Casjens, S., Wormser, G. P., Schwartz, I., & Kurtenbach, K. (2008). MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi*. *Proceedings of the National Academy of Sciences of the United States of America, 105*(25), 8730-8735. https://doi.org/10.1073/pnas.0800323105

Margos, G., Gofton, A., Wibberg, D., Dangel, A., Marosevic, D., Loh, S., Oskam, C., & Fingerle, V. (2018). The genus *Borrelia* reloaded. *PloS One, 13*(12), e0208432. https://doi.org/10.1371/journal.pone.0208432

Mathiesen, D. A., Oliver, J. H., Kolbert, C. P., Tullson, E. D., Johnson, B. J., Campbell, G. L., Mitchell, P. D., Reed, K. D., Telford, S. R., Anderson, J. F., Lane, R. S., & Persing, D. H. (1997). Genetic heterogeneity of *Borrelia burgdorferi* in the United States. *The Journal of Infectious Diseases, 175*(1), 98-107. https://doi.org/10.1093/infdis/175.1.98

Meri, T., Amdahl, H., Lehtinen, M. J., Hyvarinen, S., McDowell, J. V., Bhattacharjee, A., Meri, S., Marconi, R., Goldman, A., & Jokiranta, T. S. (2013). Microbes bind complement inhibitor factor H via a common site. *PLoS Pathogens, 9*(4), e1003308. https://doi.org/10.1371/journal.ppat.1003308

Merle, N. S., Church, S. E., Fremeaux-Bacchi, V., & Roumenina, L. T. (2015). Complement system part I – molecular mechanisms of activation and regulation. *Frontiers in Immunology, 6* https://www.frontiersin.org/article/10.3389/fimmu.2015.00262

Molloy, E. M., Casjens, S. R., Cox, C. L., Maxson, T., Ethridge, N. A., Margos, G., Fingerle, V., & Mitchell, D. A. (2015). Identification of the minimal cytolytic unit for streptolysin S and an expansion of the toxin family. *BMC Microbiology, 15*(1), 141. https://doi.org/10.1186/s12866-015-0464-y

Molloy, M. P., Herbert, B. R., Slade, M. B., Rabilloud, T., Nouwens, A. S., Williams, K. L., & Gooley, A. A. (2000). Proteomic analysis of the *Escherichia coli* outer membrane. *European Journal of Biochemistry, 267*(10), 2871-2881. https://doi.org/10.1046/j.1432-1327.2000.01296.x

Mongodin, E. F., Casjens, S. R., Bruno, J. F., Xu, Y., Drabek, E. F., Riley, D. R., Cantarel, B. L., Pagan, P. E., Hernandez, Y. A., Vargas, L. C., Dunn, J. J., Schutzer, S. E., Fraser, C. M., Qiu,

W., & Luft, B. J. (2013). Inter- and intra-specific pan-genomes of *Borrelia burgdorferi sensu lato*: Genome stability and adaptive radiation. *BMC Genomics, 14*(1), 693. https://doi.org/10.1186/1471-2164-14-693

Motaleb, M. A., Miller, M. R., Li, C., Bakker, R. G., Goldstein, S. F., Silversmith, R. E., Bourret, R. B., & Charon, N. W. (2005). CheX is a phosphorylated CheY phosphatase essential for *Borrelia burgdorferi* chemotaxis. *Journal of Bacteriology, 187*(23), 7963-7969. https://doi.org/10.1128/JB.187.23.7963-7969.2005

Motaleb, M. A., Corum, L., Bono, J. L., Elias, A. F., Rosa, P., Samuels, D. S., & Charon, N. W. (2000). *Borrelia burgdorferi* periplasmic flagella have both skeletal and motility functions. *Proceedings of the National Academy of Sciences of the United States of America, 97*(20), 10899-10904. https://doi.org/10.1073/pnas.200221797

Mulay, V., Caimano, M. J., Liveris, D., Desrosiers, D. C., Radolf, J., D., & Schwartz, I. (2007). *Borrelia burgdorferi* BBA74, a periplasmic protein associated with the outer membrane, lacks porin-like properties. *Journal of Bacteriology, 189*(5), 2063-2068. https://doi.org/10.1128/JB.01239-06

Murray, T. S., & Shapiro, E. D. (2010). Lyme disease. *Clinics in Laboratory Medicine, 30*(1), 311-328.

Nei, M., & Kumar, S. (2000). Molecular evolution and phylogenetics. Oxford University Press.

Newstead, S., Hobbs, J., Jordan, D., Carpenter, E. P., & Iwata, S. (2008). Insights into outer membrane protein crystallization. *Molecular Membrane Biology, 25*(8), 631-638. https://doi.org/10.1080/09687680802526574

Nigrovic, L. E., & Thompson, K. M. (2007). The Lyme vaccine: A cautionary tale. *Epidemiology and Infection, 135*(1), 1-8. https://doi.org/10.1017/S0950268806007096

Noppa, L., Östberg, Y., Lavrinovicha, M., & Bergström, S. (2001). P13, an integral membrane protein of *Borrelia burgdorferi*, is C-terminally processed and contains surface-exposed domains. *Infection and Immunity, 69*(5), 3323-3334. https://doi.org/10.1128/IAI.69.5.3323-3334.2001

Norgard, M. V., Anderson, J. F., Yang, X. F., Fish, D., Koski, R. A., Pal, U., Bao, F., Anguita, J., Kantor, F. S., Fikrig, E., Ramamoorthi, N., & Narasimhan, S. (2005). The Lyme disease agent exploits a tick protein to infect the mammalian host. *Nature, 436*(7050), 573-577. https://doi.org/10.1038/nature03812

Norris, S. J., Howell, J. K., Odeh, E. A., Lin, T., Gao, L., & Edmondson, D. G. (2011). High-throughput plasmid content analysis of *Borrelia burgdorferi* B31 by using luminex

multiplex technology. *Applied and Environmental Microbiology, 77*(4), 1483-1492. https://doi.org/10.1128/AEM.01877-10

Novak, C., Harrison, A., & Aucott, J. (2017). Early disseminated Lyme disease with carditis complicated by posttreatment Lyme disease syndrome. *Case Reports in Infectious Diseases, 2017*https://doi.org/10.1155/2017/5847156

Pal, U., Yang, X., Chen, M., Bockenstedt, L. K., Anderson, J. F., Flavell, R. A., Norgard, M. V., & Fikrig, E. (2004). OspC facilitates *Borrelia burgdorferi* invasion of *Ixodes scapularis* salivary glands. *Journal of Clinical Investigation, 113*(2), 220-230. https://doi.org/10.1172/JCI200419894

Pal, U., de Silva, A. M., Montgomery, R. R., Fish, D., Anguita, J., Anderson, J. F., Lobet, Y., & Fikrig, E. (2000). Attachment of *Borrelia burgdorferi* within *Ixodes scapularis* mediated by outer surface protein A. *The Journal of Clinical Investigation, 106*(4), 561-569. https://doi.org/10.1172/jci9427

Pal, U., Li, X., Wang, T., Montgomery, R. R., Ramamoorthi, N., deSilva, A. M., Bao, F., Yang, X., Pypaert, M., Pradhan, D., Kantor, F. S., Telford, S., Anderson, J. F., & Fikrig, E. (2004). TROSPA, an *Ixodes scapularis* receptor for *Borrelia burgdorferi*. *Cell, 119*(4), 457-468. https://doi.org/10.1016/j.cell.2004.10.027

Parker, J. L., & Newstead, S. (2016). Membrane protein crystallisation: Current trends and future perspectives. *Advances in Experimental Medicine and Biology, 922*, 61. https://doi.org/10.1007/978-3-319-35072-1_5

Paster, B. J., & Dewhirst, F. E. (2000). Phylogenetic foundation of spirochetes. *Journal of Molecular Microbiology and Biotechnology, 2*(4), 341-344.

Pautsch, A., & Schulz, G. E. (2000). High-resolution structure of the OmpA membrane domain. *Journal of Molecular Biology, 298*(2), 273-282. http://www.ingentaconnect.com/content/ap/mb/2000/00000298/00000002/art03671

Pearson, S. (2015). Lyme disease: Cause, symptoms, prevention and treatment. *Nurse Prescribing, 13*(2), 88-93. https://doi.org/10.12968/npre.2015.13.2.88

Penza, P., Moniuszko-Malinowska, A., Czupryna, P., Pancewicz, S., & Zajkowska, J. (2016). *Borrelia burgdorferi* – morphological structure and motility as adaptation for transmission and survival in the habitat of a tick-vertebrate setup. *Przeglad Epidemiologiczny, 70*(3), 420. https://www.ncbi.nlm.nih.gov/pubmed/27883377

Pereira, J., & Alva, V. (2021). How do I get the most out of my protein sequence using bioinformatics tools? *Acta Crystallographica. Section D, Structural Biology, 77*(9), 1116-1126. https://doi.org/10.1107/S2059798321007907

Perez, E., & Stock, A. M. (2007). Characterization of the *Thermotoga maritima* chemotaxis methylation system that lacks pentapeptide-dependent methyltransferase CheR:MCP tethering. *Molecular Microbiology, 63*(2), 363-378. https://doi.org/10.1111/j.1365-2958.2006.05518.x

Perez, E., West, A. H., Stock, A. M., & Djordjevic, S. (2004). Discrimination between different methylation states of chemotaxis receptor tar by receptor methyltransferase CheR. *Biochemistry, 43*(4), 953-961. https://doi.org/10.1021/bi035455q

Poland, G. A. (2011). Vaccines against Lyme disease: What happened and what lessons can we learn? *Clinical Infectious Diseases, 52*(suppl_3), s253-s258. https://doi.org/10.1093/cid/ciq116

Posada, D. (2008). jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution, 25*(7), 1253-1256. https://doi.org/10.1093/molbev/msn083

Posey, J. E., & Gherardini, F. C. (2000). Lack of a role for iron in the Lyme disease pathogen. *Science, 288*(5471), 1651-1653. https://doi.org/10.1126/science.288.5471.1651

Probert, W. S., & Johnson, B. J. B. (1998). Identification of a 47 kDa fibronectin-binding protein expressed by *Borrelia burgdorferi* isolate B31. *Molecular Microbiology, 30*(5), 1003-1015. https://doi.org/10.1046/j.1365-2958.1998.01127.x

Pulzova, L., & Bhide, M. (2014). Outer surface proteins of *Borrelia*: Peerless immune evasion tools. *Current Protein & Peptide Science, 15*(1), 75-88. http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1389-2037&volume=15&issue=1&spage=75

Qiu, W., Bruno, J. F., McCaig, W. D., Xu, Y., Livey, I., Schriefer, M. E., & Luft, B. J. (2008). Wide distribution of a high-virulence *Borrelia burgdorferi* clone in Europe and North America. *Emerging Infectious Diseases, 14*(7), 1097-1104. https://doi.org/10.3201/eid1407.070880

R Core Team. (2021). R: A language and environment for statistical computing. https://www.R-project.org/

Radolf, J. D., Caimano, M. J., Stevenson, B., & Hu, L. T. (2012). Of ticks, mice and men: Understanding the dual-host lifestyle of Lyme disease spirochaetes. *Nature Reviews. Microbiology, 10*(2), 87-99. https://doi.org/10.1038/nrmicro2714

Ramamoorthi, N., Narasimhan, S., Pal, U., Bao, F., Yang, X. F., Fish, D., Anguita, J., Norgard, M. V., Kantor, F. S., Anderson, J. F., Koski, R. A., & Fikrig, E. (2005). The Lyme disease agent exploits a tick protein to infect the mammalian host. *Nature, 436*(7050), 573-577. https://doi.org/10.1038/nature03812

Revel, A. T., Talaat, A. M., & Norgard, M. V. (2002). DNA microarray analysis of differential gene expression in *Borrelia burgdorferi*, the Lyme disease spirochete. *Proceedings of the National Academy of Sciences of the United States of America, 99*(3), 1562-1567. https://doi.org/10.1073/pnas.032667699

Ripoche, J., Day, A. J., Harris, T. J., & Sim, R. B. (1988). The complete amino acid sequence of human complement factor H. *Biochemical Journal, 249*(2), 593-602. https://doi.org/10.1042/bj2490593

Rosa, P. A., Tilly, K., & Stewart, P. E. (2005a). The burgeoning molecular genetics of the Lyme disease spirochaete. *Nature Reviews Microbiology, 3*(2), 129-143. https://doi.org/10.1038/nrmicro1086

Rosa, P. A., Tilly, K., & Stewart, P. E. (2005b). The burgeoning molecular genetics of the Lyme disease spirochaete. *Nature Reviews Microbiology, 3*(2), 129-143. https://doi.org/10.1038/nrmicro1086

Rostoff, P., Gajos, G., Konduracka, E., Gackowski, A., Nessler, J., & Piwowarska, W. (2010). Lyme carditis: Epidemiology, pathophysiology, and clinical features in endemic areas. *International Journal of Cardiology, 144*(2), 328-333. https://doi.org/10.1016/j.ijcard.2009.03.034

Saier, M. H. (2000). Spirochetes: Evolution, genome analyses and physiology. *Journal of Molecular Microbiology and Biotechnology, 2*(339)

Sal, M. S., Li, C., Motalab, M. A., Shibata, S., Aizawa, S., & Charon, N. W. (2008). *Borrelia burgdorferi* uniquely regulates its motility genes and has an intricate flagellar hook-basal body structure. *Journal of Bacteriology, 190*(6), 1912-1921. https://doi.org/10.1128/JB.01421-07

Šali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology, 234*(3), 779-815. https://doi.org/10.1006/jmbi.1993.1626

Samuels, D. S., & Radolf, J. D. (2010). Borrelia molecular biology, host interaction and pathogenesis. Caister Academic Press.

Sanchez, J. L. (2015). Clinical manifestations and treatment of Lyme disease. *Clinics in Laboratory Medicine, 35*(4), 765-778. https://doi.org/10.1016/j.cll.2015.08.004

Sarma, J. V., & Ward, P. A. (2010). The complement system. *Cell and Tissue Research, 343*(1), 227-235. https://doi.org/10.1007/s00441-010-1034-0

Schmidt, J., Müsken, M., Becker, T., Magnowska, Z., Bertinetti, D., Möller, S., Zimmermann, B., Herberg, F. W., Jänsch, L., & Häussler, S. (2011). The *Pseudomonas aeruginosa* chemotaxis methyltransferase CheR1 impacts on bacterial surface sampling. *PLoS One, 6*(3), e18184. https://doi.org/10.1371/journal.pone.0018184

Schrödinger. (2010). The PyMOL molecular graphics system [computer software]

Schuijt, T. J., Coumou, J., Narasimhan, S., Dai, J., DePonte, K., Wouters, D., Brouwer, M., Oei, A., Roelofs, Joris J. T. H., van Dam, A. P., van der Poll, T., van 't Veer, C., Hovius, J. W., & Fikrig, E. (2011). A tick mannose-binding lectin inhibits the vertebrate complement cascade to enhance transmission of the Lyme disease agent. *Cell Host & Microbe, 10*(2), 136-146. https://doi.org/10.1016/j.chom.2011.06.010

Schulz, G. E. (2000). β-barrel membrane proteins. *Current Opinion in Structural Biology, 10*(4), 443-447. https://doi.org/10.1016/S0959-440X(00)00120-2

Schulz, G. E. (2002). The structure of bacterial outer membrane proteins. *Biochimica Et Biophysica Acta. Biomembranes, 1565*(2), 308-317. https://doi.org/10.1016/s0005-2736(02)00577-1

Schwab, J., Hammerschmidt, C., Richter, D., Skerka, C., Matuschka, F. R., Wallich, R., Zipfel, P. F., & Kraiczy, P. (2013). *Borrelia valaisiana* resist complement-mediated killing independently of the recruitment of immune regulators and inactivation of complement components. *PLoS One, 8*(1), e53659. https://doi.org/10.1371/journal.pone.0053659

Schwan, T. G., Piesman, J., Golde, W. T., Dolan, M. C., & Rosa, P. A. (1995). Induction of an outer surface protein on *Borrelia burgdorferi* during tick feeding. *Proceedings of the National Academy of Sciences - PNAS, 92*(7), 2909-2913. https://doi.org/10.1073/pnas.92.7.2909

Seemanapalli, S. V., Xu, Q., McShan, K., & Liang, F. T. (2010). Outer surface protein C is a dissemination-facilitating factor of *Borrelia burgdorferi* during mammalian infection. *PloS One, 5*(12), e15830. https://doi.org/10.1371/journal.pone.0015830

Seriburi, V., Ndukwe, N., Chang, Z., Cox, M. E., & Wormser, G. P. (2012). High frequency of false positive IgM immunoblots for *Borrelia burgdorferi* in clinical practice. *Clinical*

Microbiology and Infection, 18(12), 1236-1240. https://doi.org/10.1111/j.1469-0691.2011.03749.x

Setubal, J. C., Reis, M., Matsunaga, J., & Haake, D. A. (2006). Lipoprotein computational prediction in spirochaetal genomes. *Microbiology (Reading, England), 152*(Pt 1), 113-121. https://doi.org/10.1099/mic.0.28317-0

Shapiro, E. D., & Gerber, M. A. (2000). Lyme disease. *Clinical Infectious Diseases, 31*(2), 533-542. https://doi.org/10.1086/313982

Shrestha, B., Kenedy, M. R., & Akins, D. R. (2017). Outer membrane proteins BB0405 and BB0406 are immunogenic, but only BB0405 is required for *Borrelia burgdorferi* infection. *Infection and Immunity, 85*(2), 803. https://doi.org/10.1128/IAI.00803-16

Siegel, C., Hallström, T., Skerka, C., Eberhardt, H., Uzonyi, B., Beckhaus, T., Karas, M., Wallich, R., Stevenson, B., Zipfel, P. F., & Kraiczy, P. (2010). Complement factor H-related proteins CFHR2 and CFHR5 represent novel ligands for the infection-associated CRASP proteins of *Borrelia burgdorferi*. *PLoS One, 5*(10), e13519. https://doi.org/10.1371/journal.pone.0013519

Sigal, L. H., Zahradnik, J. M., Lavin, P., Patella, S. J., Bryant, G., Haselby, R., Hilton, E., Kunkel, M., Adler-Klein, D., Doherty, T., Evans, J., Malawista, S. E., Molloy, P. J., Seidner, A. L., Sabetta, J. R., Simon, H. J., Klempner, M. S., Mays, J., & Marks, D. (1998). A vaccine consisting of recombinant *Borrelia burgdorferi* outer-surface protein A to prevent Lyme disease. *The New England Journal of Medicine, 339*(4), 216-222. https://doi.org/10.1056/NEJM199807233390402

Silhavy, T. J., Kahne, D., & Walker, S. (2010). The bacterial cell envelope. *Cold Spring Harbor Perspectives in Biology, 2*(5), a000414. https://doi.org/10.1101/cshperspect.a000414

Skare, J. T., Champion, C. I., Mirzabekov, T. A., Shang, E. S., Blanco, D. R., Erdjument-Bromage, H., Tempst, P., Kagan, B. L., Miller, J. N., & Lovett, M. A. (1996). Porin activity of the native and recombinant outer membrane protein Oms28 of *Borrelia burgdorferi*. *Journal of Bacteriology, 178*(16), 4909-4918. https://doi.org/10.1128/jb.178.16.4909-4918.1996

Skare, J. T., Mirzabekov, T. A., Shang, E. S., Blanco, D. R., Erdjument-Bromage, H., Bunikis, J., Bergström, S., Tempst, P., Kagan, B. L., Miller, J. N., & Lovett, M. A. (1997). The Oms66 (p66) protein is a *Borrelia burgdorferi* porin. *Infection and Immunity, 65*(9), 3654-3661. http://iai.asm.org/content/65/9/3654.abstract

Smith, S. G. J., Mahon, V., Lambert, M. A., & Fagan, R. P. (2007). A molecular Swiss army knife: OmpA structure, function and expression. *FEMS Microbiology Letters, 273*(1), 1-11. https://doi.org/10.1111/j.1574-6968.2007.00778.x

Stanek, G., Fingerle, V., Hunfeld, K. -., Jaulhac, B., Kaiser, R., Krause, A., Kristoferitsch, W., O'Connell, S., Ornstein, K., Strle, F., & Gray, J. (2011). Lyme borreliosis: Clinical case definitions for diagnosis and management in Europe. *Clinical Microbiology and Infection, 17*(1), 69-79. https://doi.org/10.1111/j.1469-0691.2010.03175.x

Stanek, G., Wormser, G. P., Gray, J., & Strle, F. (2012). Lyme borreliosis. *Lancet, The, 379*(9814), 461-473. https://doi.org/10.1016/S0140-6736(11)60103-7

Steere, A. C., Broderick, T. F., & Malawista, S. E. (1978). Erythema chronicum migrans and Lyme arthritis: Epidemiologic evidence for a tick vector. *American Journal of Epidemiology, 108*(4), 312. https://www.ncbi.nlm.nih.gov/pubmed/727200

Steere, A. C., Malawista, S. E., Snydman, R., Shope, E., Andiman, A., Ross, M. R., & Steele, F. M. (1977). An epidemic of oligoarticular arthritis in children and adults in three Connecticut communities. *Arthritis & Rheumatism, 20*(1), 7-17. https://doi.org/10.1002/art.1780200102

Steere, A. C., Sikand, V. K., Meurice, F., Parenti, D. L., Fikrig, E., Schoen, R. T., Nowakowski, J., Schmid, C. H., Laukamp, S., Buscarino, C., & Krause, D. S. (1998). Vaccination against Lyme disease with recombinant *Borrelia burgdorferi* outer-surface lipoprotein A with adjuvant. Lyme disease vaccine study group. *The New England Journal of Medicine, 339*(4), 209-215. https://doi.org/10.1056/NEJM199807233390401

Steere, A. C. (1989). Lyme disease. *New England Journal of Medicine, 321*(9), 586-596. https://doi.org/10.1056/NEJM198908313210906

Steere, A. C., Strle, F., Wormser, G. P., Hu, L. T., Branda, J. A., Hovius, J. W. R., Li, X., & Mead, P. S. (2016). Lyme borreliosis. *Nature Reviews. Disease Primers, 2*(1), 16090. https://doi.org/10.1038/nrdp.2016.90

Stevenson, B., Fingerle, V., Wormser, G. P., & Margos, G. (2019). Public health and patient safety concerns merit retention of Lyme borreliosis-associated spirochetes within the genus *Borrelia*, and rejection of the genus novum *Borreliella*. *Ticks and Tick-Borne Diseases, 10*(1), 1-4. https://doi.org/10.1016/j.ttbdis.2018.08.010

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques, 28*(6), 1102-1104. https://doi.org/10.2144/00286ir01

Stübs, G., Fingerle, V., Wilske, B., Göbel, U. B., Zähringer, U., Schumann, R. R., & Schröder, N. W. J. (2009). Acylated cholesteryl galactosides are specific antigens of *Borrelia* causing Lyme disease and frequently induce antibodies in late stages of disease. *The Journal of Biological Chemistry, 284*(20), 13326-13334. https://doi.org/10.1074/jbc.m809575200

Subramanian, G., Koonin, E. V., & Aravind, L. (2000). Comparative genome analysis of the pathogenic spirochetes *Borrelia burgdorferi* and *Treponema pallidum*. *Infection and Immunity, 68*(3), 1633-1648. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC97324/

Taboada, B., Estrada, K., Ciria, R., & Merino, E. (2018). Operon-mapper: A web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics, 34*(23), 4118-4120. https://doi.org/10.1093/bioinformatics/bty496

Takayama, K., Rothenberg, R. J., & Barbour, A. G. (1987). Absence of lipopolysaccharide in the Lyme disease spirochete, *Borrelia burgdorferi. Infection and Immunity, 55*(9), 2311-2313. http://iai.asm.org/content/55/9/2311.abstract

Talagrand-Reboul, E., Boyer, P. H., Bergström, S., Vial, L., & Boulanger, N. (2018). Relapsing fevers: Neglected tick-borne diseases. *Frontiers in Cellular and Infection Microbiology, 8*, 98. https://doi.org/10.3389/fcimb.2018.00098

Talagrand-Reboul, E., Raffetin, A., Zachary, P., Jaulhac, B., & Eldin, C. (2020). Immunoserological diagnosis of human borrelioses: Current knowledge and perspectives. *Frontiers in Cellular and Infection Microbiology, 10*, 241. https://doi.org/10.3389/fcimb.2020.00241

Thein, M., Bonde, M., Bunikis, I., Denker, K., Sickmann, A., Bergström, S., & Benz, R. (2012). DipA, a pore-forming protein in the outer membrane of Lyme disease spirochetes exhibits specificity for the permeation of dicarboxylates. *PloS One, 7*(5), e36523. https://doi.org/10.1371/journal.pone.0036523

Thomas, V., Anguita, J., Samanta, S., Rods, P. A., Stewart, P., Barthold, S. W., & Fikrig, E. (2001). Dissociation of infectivity and pathogenicity in *Borrelia burgdorferi*. *Infection and Immunity, 69*(5), 3507-3509. https://doi.org/10.1128/IAI.69.5.3507-3509.2001

Tokarz, R., Anderton, J. M., Katona, L. I., & Benach, J. L. (2004). Combined effects of blood and temperature shift on *Borrelia burgdorferi* gene expression as determined by whole genome DNA array. *Infection and Immunity, 72*(9), 5419-5432. https://doi.org/10.1128/IAI.72.9.5419-5432.2004

Travinsky, B., Bunikis, J., & Barbour, A. G. (2010). Geographic differences in genetic locus linkages for *Borrelia burgdorferi*. *Emerging Infectious Diseases, 16*(7), 1147-1150. https://doi.org/10.3201/eid1607.091452

Tsirigos, K. D., Elofsson, A., & Bagos, P. G. (2016). PRED-TMBB2: Improved topology prediction and detection of beta-barrel outer membrane proteins. *Bioinformatics, 32*(17), i665-i671. https://doi.org/10.1093/bioinformatics/btw444

Vogt, J., & Schulz, G. E. (1999). The structure of the outer membrane protein OmpX from *Escherichia coli* reveals possible mechanisms of virulence. *Structure (London), 7*(10), 1301-1309. https://doi.org/10.1016/s0969-2126(00)80063-5

Wagemakers, A., Staarink, P. J., Sprong, H., & Hovius, J. W. R. (2015). *Borrelia miyamotoi*: A widespread tick-borne relapsing fever spirochete. *Trends in Parasitology, 31*(6), 260-269. https://doi.org/10.1016/j.pt.2015.03.008

Ward, N., & Moreno-Hagelsieb, G. (2014). Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: How much do we miss? *PloS One, 9*(7), e101850. https://doi.org/10.1371/journal.pone.0101850

Wilson, M. M., & Bernstein, H. D. (2015). Surface-exposed lipoproteins: An emerging secretion phenomenon in Gram-negative bacteria. *Trends in Microbiology, 24*(3), 198-208. https://doi.org/10.1016/j.tim.2015.11.006

Wimley, W. C. (2003). The versatile β-barrel membrane protein. *Current Opinion in Structural Biology, 13*(4), 404-411. https://doi.org/10.1016/S0959-440X(03)00099-X

Wormser, G. P. (2006). Early Lyme disease. *New England Journal of Medicine, 354*(26), 2794-2801. https://doi.org/10.1056/NEJMcp061181

Wormser, G. P., Dattwyler, R. J., Shapiro, E. D., Halperin, J. J., Steere, A. C., Klempner, M. S., Krause, P. J., Bakken, J. S., Strle, F., Stanek, G., Bockenstedt, L., Fish, D., Dumler, J. S., & Nadelman, R. B. (2006). The clinical assessment, treatment, and prevention of Lyme disease, human granulocytic anaplasmosis, and babesiosis: Clinical practice guidelines by the infectious diseases society of America. *Clinical Infectious Diseases, 43*(9), 1089-1134. https://doi.org/10.1086/508667

Wu, J., Wu, Y., Ricklin, D., Janssen, B. J. C., Lambris, J. D., & Gros, P. (2009). Structure of C3b-factor H and implications for host protection by complement regulators. *Nature Immunology, 10*(7), 728-733. https://doi.org/10.1038/ni.1755

Yang, X. F., Pal, U., Alani, S. M., Fikrig, E., & Norgard, M. V. (2004). Essential role for OspA/B in the life cycle of the Lyme disease spirochete. *The Journal of Experimental Medicine, 199*(5), 641-648. https://doi.org/10.1084/jem.20031960

Yeung, C., & Baranchuk, A. (2018). Systematic approach to the diagnosis and treatment of Lyme carditis and high-degree atrioventricular block. *Healthcare (Basel), 6*(4), 119. https://doi.org/10.3390/healthcare6040119

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S. C., Ester, M., Foster, L. J., & Brinkman, F. S. L. (2010). PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics, 26*(13), 1608-1615. https://doi.org/10.1093/bioinformatics/btq249

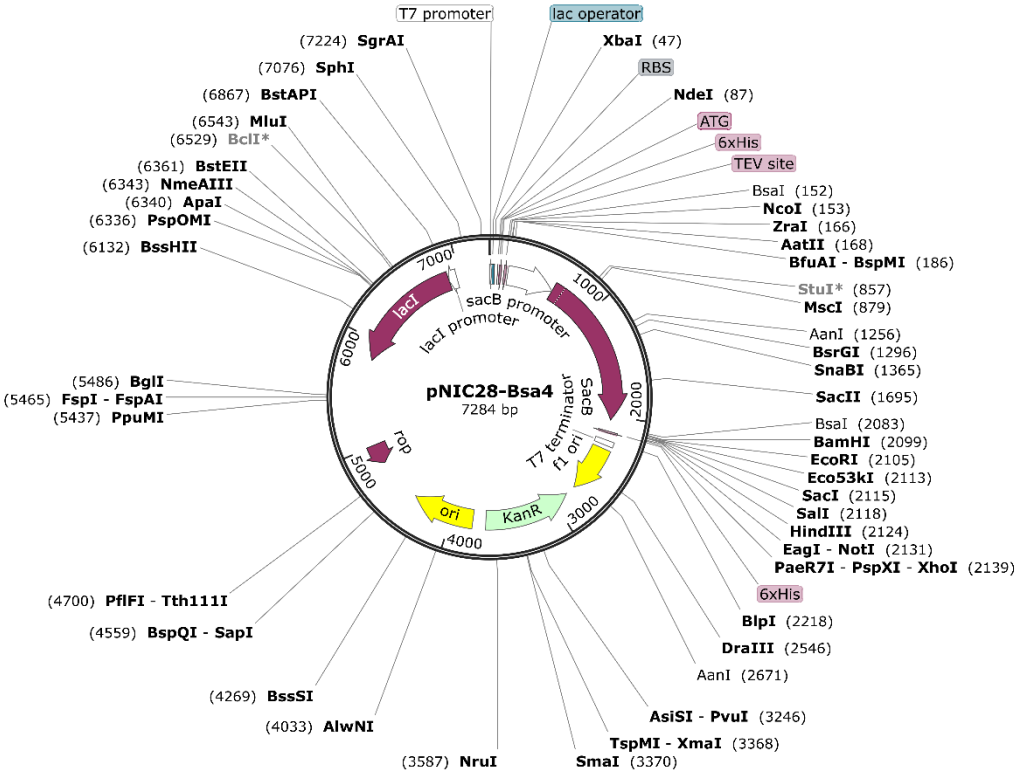
Zhao, J., Zhang, H., Qin, B., Nikolay, R., He, Q., Spahn, C. M. T., & Zhang, G. (2019). Multifaceted stoichiometry control of bacterial operons revealed by deep proteome quantification. *Frontiers in Genetics, 10*, 473. https://doi.org/10.3389/fgene.2019.00473

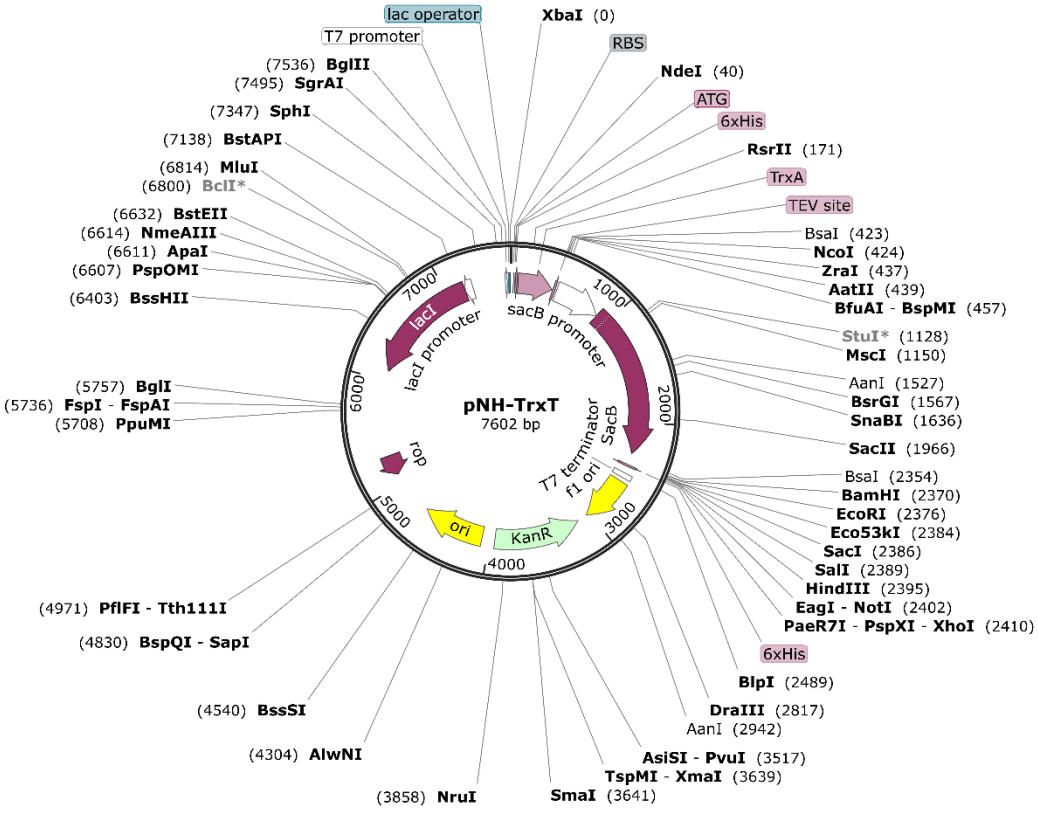
Zückert, W. R. (2014). Secretion of bacterial lipoproteins: Through the cytoplasmic membrane, the periplasm and beyond. *BBA - Molecular Cell Research, 1843*(8), 1509-1516. https://doi.org/10.1016/j.bbamcr.2014.04.022

Chapter 8 Appendix

Appendix 1- Vector maps

Vector maps for pNIC28-Bsa4 and pNH-TrxT were obtained from SnapGene®.

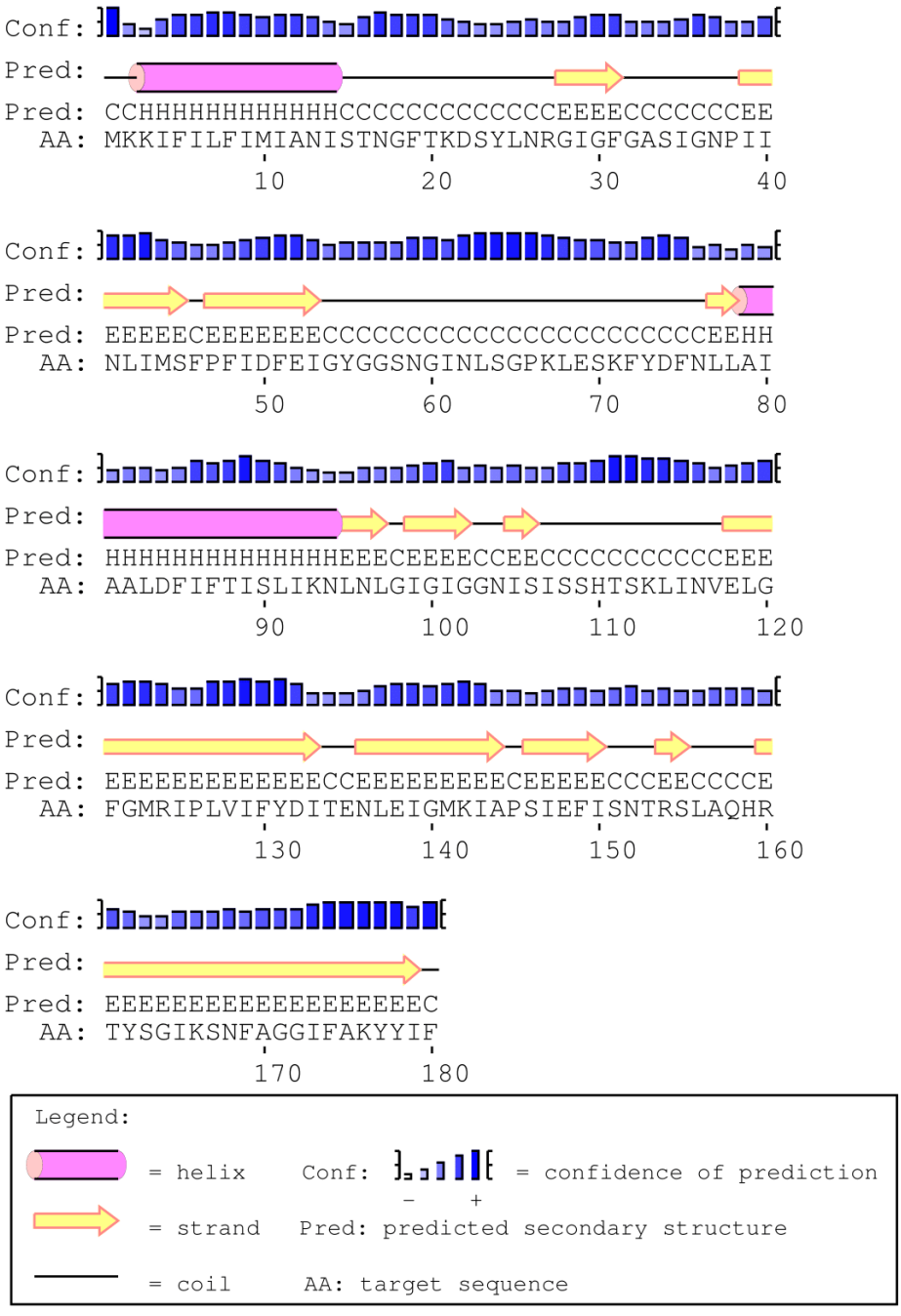




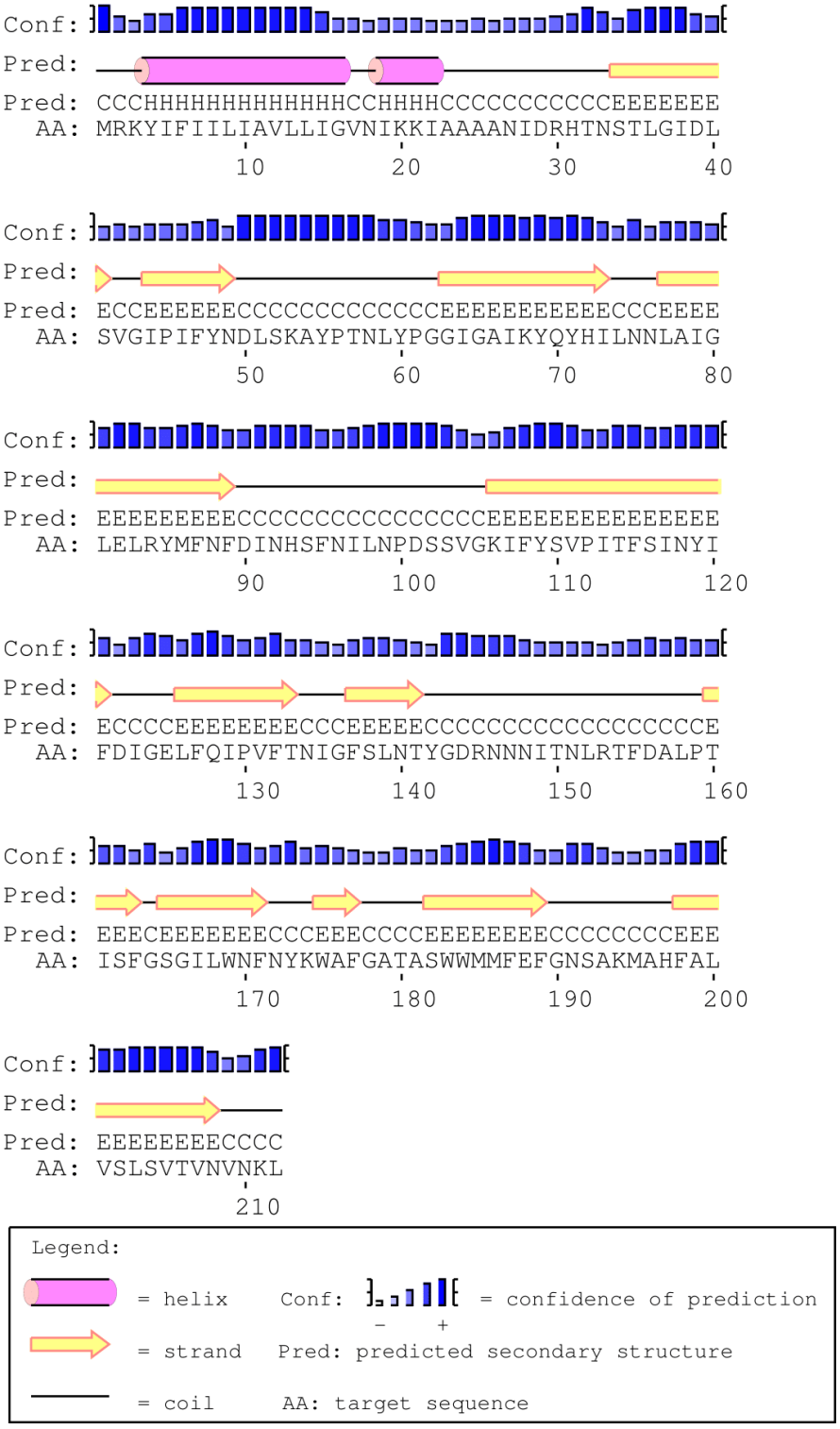
Appendix 2- PSI-PRED outputs

PSI-PRED was used for secondary structure predictions for each protein using amino acid sequences obtained from NCBI as the input.

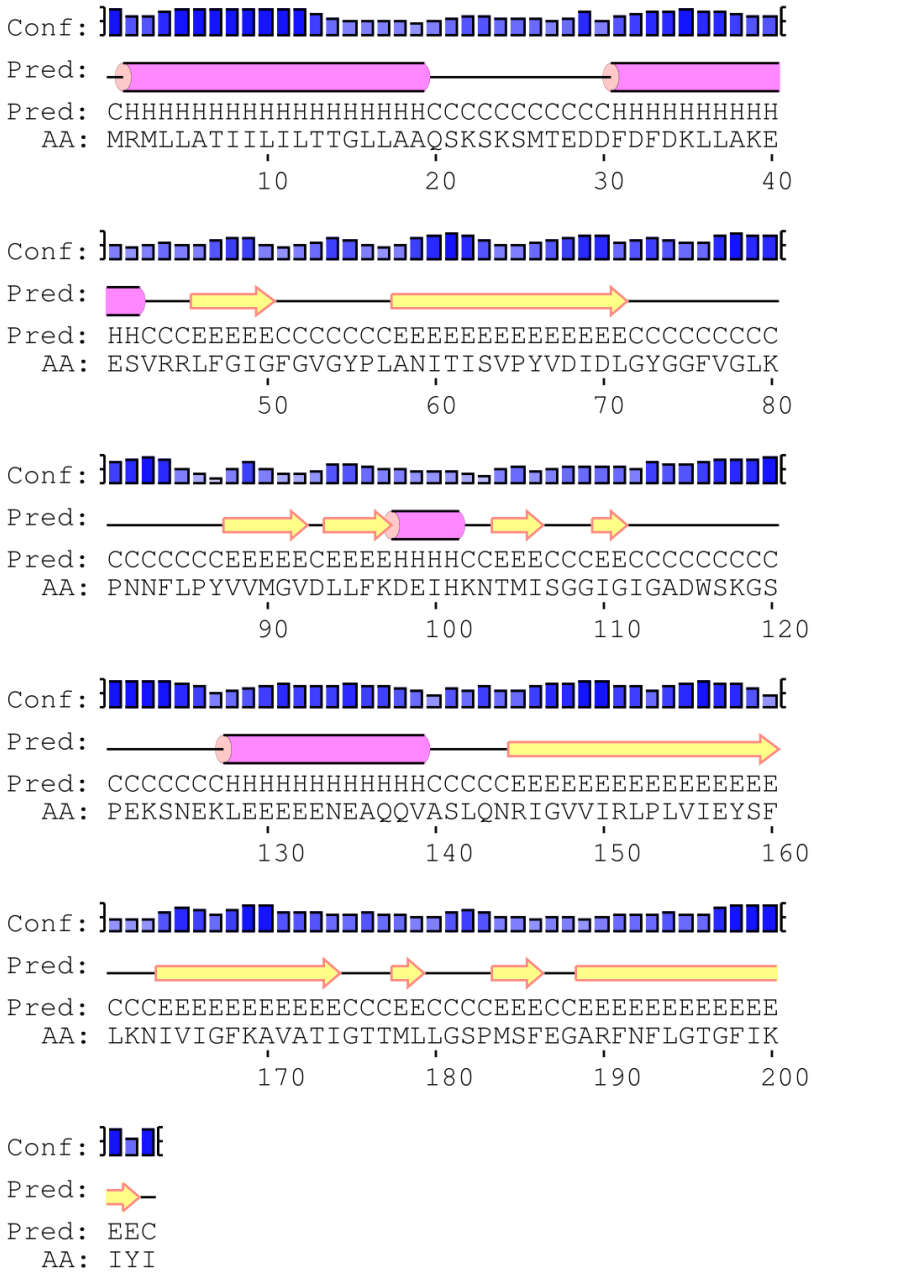
BB_0562



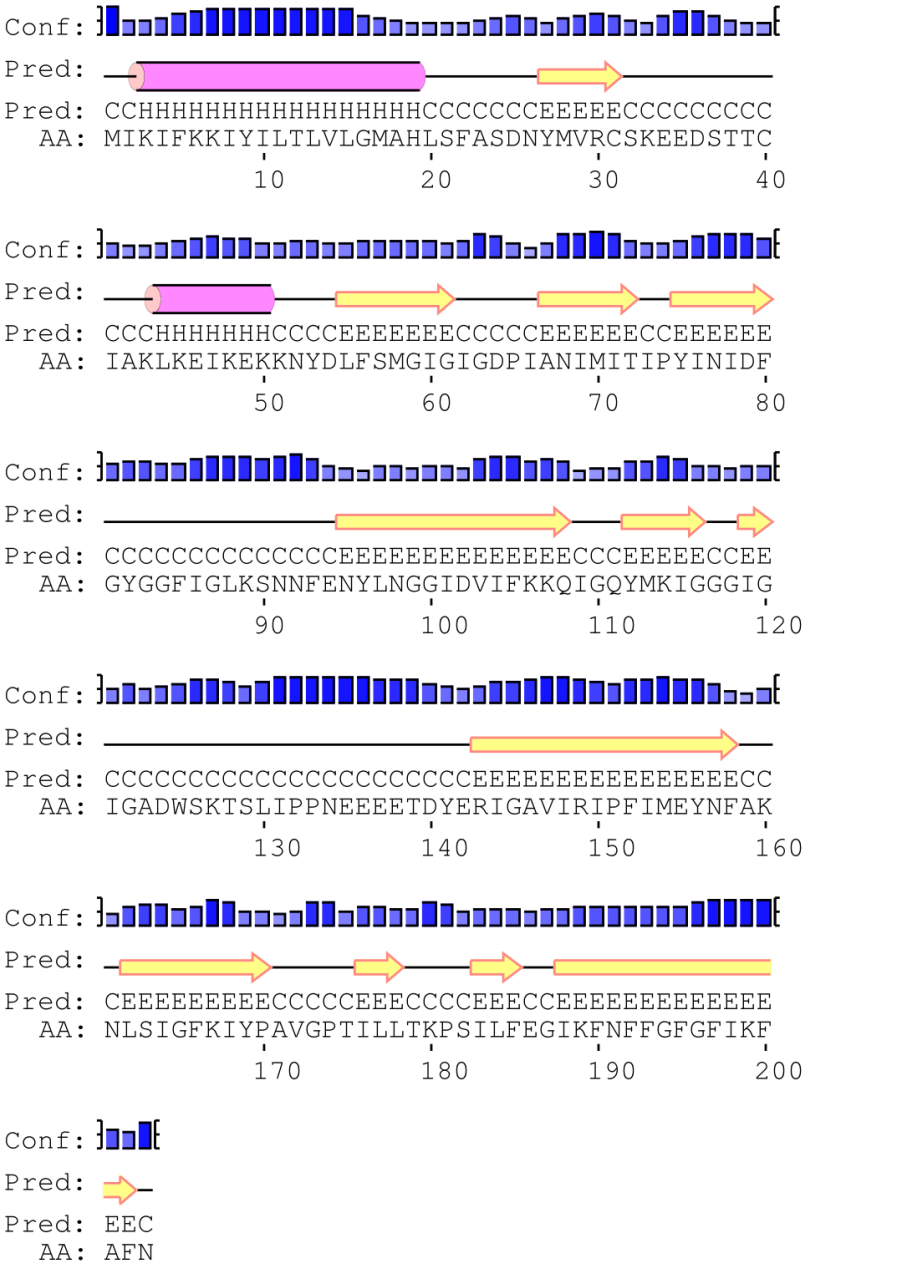
BB_0027


Conf:
Pred:
Pred: CCCHHHHHHHHHHHHHCCHHHHCCCCCCCCCCCCEEEEEEE
AA: MRKYIFIILIAVLLIGVNIKKIAAAANIDRHTNSTLGIDL
10 20 30 40
Conf:
Pred:
Pred: ECCEEEEEECCCCCCCCCCCCCEEEEEEEEEEEECCCEEEE
AA: SVGIPIFYNDLSKAYPTNLYPGGIGAIKYQYHILNNLAIG
50 60 70 80
Conf:
Pred:
Pred: EEEEEEEEECCCCCCCCCCCCCCCCEEEEEEEEEEEEEEE
AA: LELRYMFNFDINHSFNILNPDSSVGKIFYSVPITFSINYI
90 100 110 120
Conf:
Pred:
Pred: ECCCCEEEEEEEECCCEEEEECCCCCCCCCCCCCCCCCCE
AA: FDIGELFQIPVFTNIGFSLNTYGDRNNNITNLRTFDALPT
130 140 150 160
Conf:
Pred:
Pred: EEECEEEEEEECCCEEECCCCEEEEEEEECCCCCCCCEEE
AA: ISFGSGILWNFNYKWAFGATASWWMMFEFGNSAKMAHFAL
170 180 190 200
Conf:
Pred:
Pred: EEEEEEECCCC
AA: VSLSVTVNVNKL
210
Legend:
= helix
= strand
= coil
Conf:
= confidence of prediction
Pred: predicted secondary structure
AA: target sequence

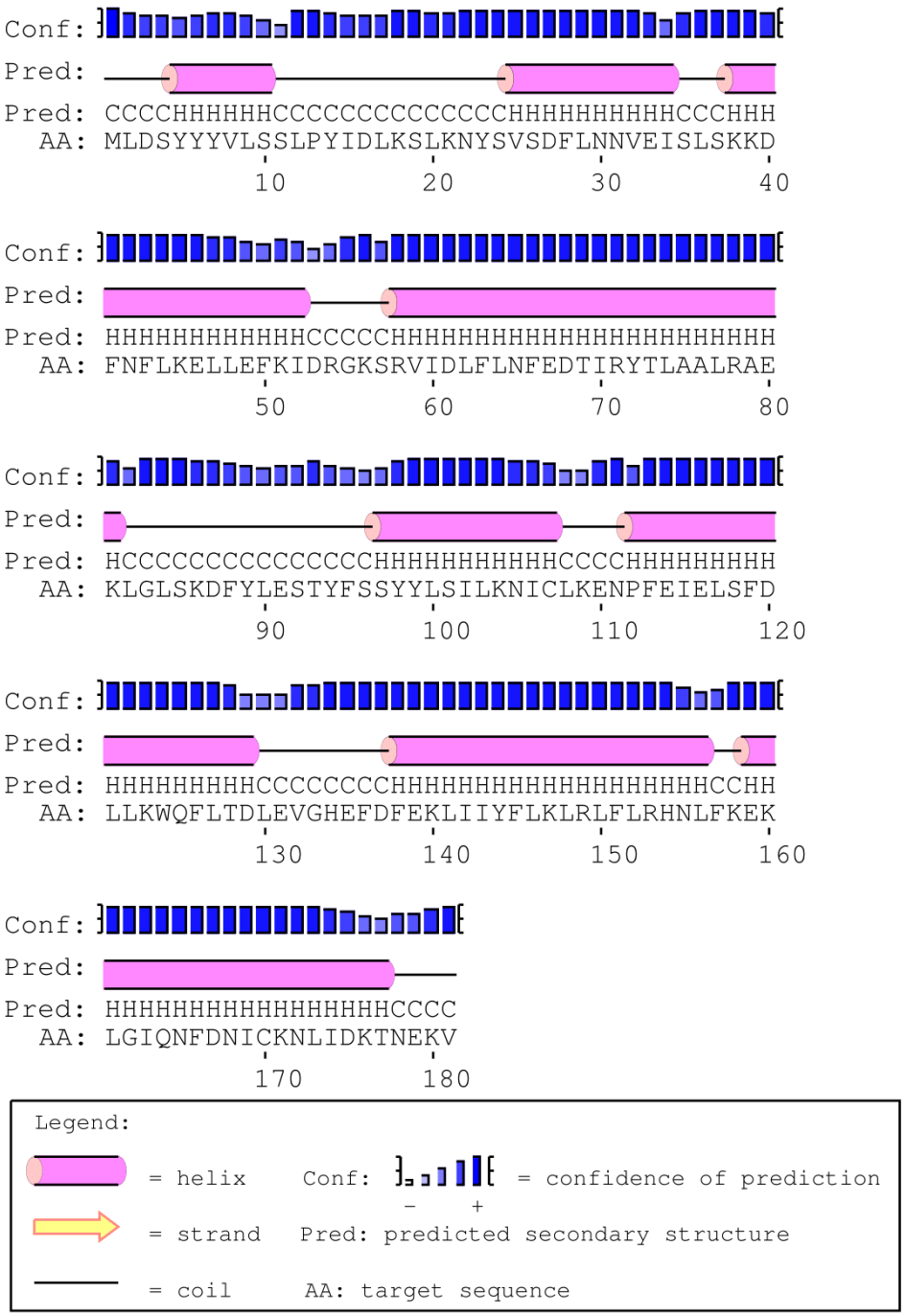
BB_0405


Conf:
Pred:
Pred: CHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCHHHHHHHHHH
AA: MRMLLATIILILTTGLLAAQSKSKSMTEDDFDFDKLLAKE
10 20 30 40
Conf:
Pred:
Pred: HHCCCEEEEECCCCCCCEEEEEEEEEEEEECCCCCCCCC
AA: ESVRRLFGIGFGVGYPLANITISVPYVDIDLGYGGFVGLK
50 60 70 80
Conf:
Pred:
Pred: CCCCCCCEEEEECEEEEHHHHCCEEECCCEECCCCCCCCC
AA: PNNFLPYVVMGVDLLFKDEIHKNTMISGGIGIGADWSKGS
90 100 110 120
Conf:
Pred:
Pred: CCCCCCCHHHHHHHHHHHHCCCCCEEEEEEEEEEEEEEE
AA: PEKSNEKLEEEEENEAQQVASLQNRIGVVIRLPLVIEYSF
130 140 150 160
Conf:
Pred:
Pred: CCCEEEEEEEEEECCCEECCCCEEECCEEEEEEEEEEE
AA: LKNIVIGFKAVATIGTTMLLGSPMSFEGARFNFLGTGFIK
170 180 190 200
Conf:
Pred:
Pred: EEC
AA: IYI
Legend:
= helix
= strand
= coil
Conf:
- +
= confidence of prediction
Pred: predicted secondary structure
AA: target sequence

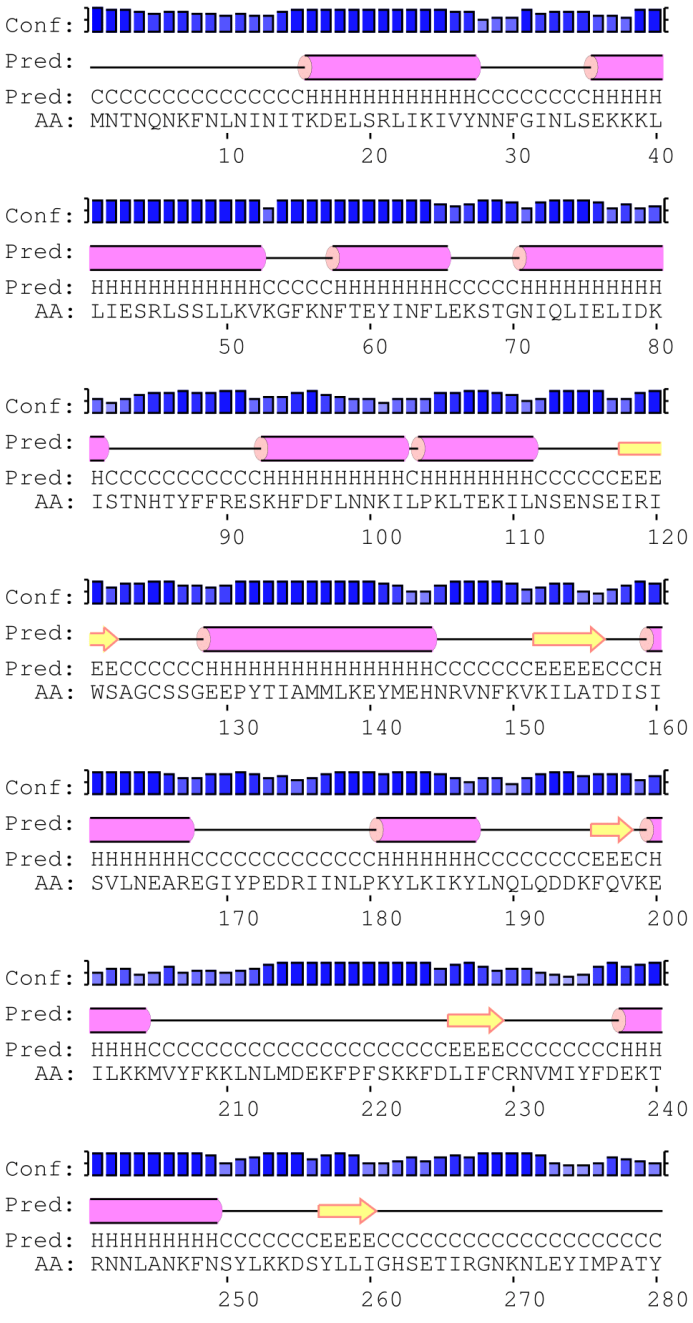
BB_0406


Conf:
Pred:
Pred: CCHHHHHHHHHHHHHHHHCCCCCCCEEEEECCCCCCCCC
AA: MIKIFKKIYILTLVLGMAHLSFASDNYMVRCSKEEDSTTC
10 20 30 40
Conf:
Pred:
Pred: CCCHHHHHHHCCCCEEEEEECCCCCEEEEEECCEEEEEE
AA: IAKLKEIKEKKNYDLFSMGIGIGDPIANIMITIPYINIDF
50 60 70 80
Conf:
Pred:
Pred: CCCCCCCCCCCCCCEEEEEEEEEEEEEECCCEEEEECCEE
AA: GYGGFIGLKSNNFENYLNGGIDVIFKKQIGQYMKIGGGIG
90 100 110 120
Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCEEEEEEEEEEEEEEECC
AA: IGADWSKTSLIPPNEEEETDYERIGAVIRIPFIMEYNFAK
130 140 150 160
Conf:
Pred:
Pred: CEEEEEEEEECCCCCEEECCCCEEECCEEEEEEEEEEEE
AA: NLSIGFKIYPAVGPTILLTKPSILFEGIKFNFFGFGFIKF
170 180 190 200
Conf:
Pred:
Pred: EEC
AA: AFN
Legend:
= helix
= strand
= coil
Conf:
– +
= confidence of prediction
Pred: predicted secondary structure
AA: target sequence

BB_0095


Conf:
Pred:
Pred: CCCCHHHHHHCCCCCCCCCCCCCCHHHHHHHHHHHCCCHHH
AA: MLDSYYYVLSSLPYIDLKSLKNYSVSDFLNNVEISLSKKD
10
20
30
40
Conf:
Pred:
Pred: HHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHHH
AA: FNFLKELLEFKIDRGKSRVIDLFLNFEDTIRYTLAALRAE
50
60
70
80
Conf:
Pred:
Pred: HCCCCCCCCCCCCCCCHHHHHHHHHHHCCCCHHHHHHHHH
AA: KLGLSKDFYLESTYFSSYYLSILKNICLKENPFEIELSFD
90
100
110
120
Conf:
Pred:
Pred: HHHHHHHHHCCCCCCCCHHHHHHHHHHHHHHHHHHHCCHH
AA: LLKWQFLTDLEVGHEFDFEKLIIYFLKLRLFLRHNLFKEK
130
140
150
160
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHCCCC
AA: LGIQNFDNICKNLIDKTNEKV
170
180
Legend:
= helix
= strand
= coil
Conf:
- +
= confidence of prediction
Pred: predicted secondary structure
AA: target sequence

BB_0040


Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCHHHHHHHHHHHHCCCCCCCCHHHHH
AA: MNTNQNKFNLNINITKDELSRLIKIVYNNFGINLSEKKKL
10 20 30 40
Conf:
Pred:
Pred: HHHHHHHHHHHHCCCCCHHHHHHHHCCCCCHHHHHHHHH
AA: LIESRLSSLLKVKGFKNFTEYINFLEKSTGNIQLIELIDK
50 60 70 80
Conf:
Pred:
Pred: HCCCCCCCCCCCCHHHHHHHHHHCHHHHHHHHCCCCCCEEE
AA: ISTNHTYFFRESKHFDFLNNKILPKLTEKILNSENSEIRI
90 100 110 120
Conf:
Pred:
Pred: EECCCCCCHHHHHHHHHHHHHHHHCCCCCCCEEEEECCCH
AA: WSAGCSSGEEPYTIAMMLKEYMEHNRVNFKVKILATDISI
130 140 150 160
Conf:
Pred:
Pred: HHHHHHHCCCCCCCCCCCCCHHHHHHHCCCCCCCCEEECH
AA: SVLNEAREGIYPEDRIINLPKYLKIKYLNQLQDDKFQVKE
170 180 190 200
Conf:
Pred:
Pred: HHHHCCCCCCCCCCCCCCCCCCCCCEEEECCCCCCCCHHH
AA: ILKKMVYFKKLNLMDEKFPFSKKFDLIFCRNVMIYFDEKT
210 220 230 240
Conf:
Pred:
Pred: HHHHHHHHHCCCCCCCEEEECCCCCCCCCCCCCCCCCC
AA: RNNLANKFNSYLKKDSYLLIGHSETIRGNKNLEYIMPATY
250 260 270 280
Conf:
Pred:
Pred: CCC
AA: KKN
Legend:
= helix
= strand
= coil
Conf: = confidence of prediction
Pred: predicted secondary structure
AA: target sequence

Appendix 3- Amino acid sequences of protein targets

The code of the construct is provided in the first column, followed by the protein name and the amino acid sequence length. Finally, the amino acid sequence is given for each full length protein and for truncated proteins. The expression tag sequence is highlighted in yellow and the site for cleavage of the tag is indicated by an Asterix.

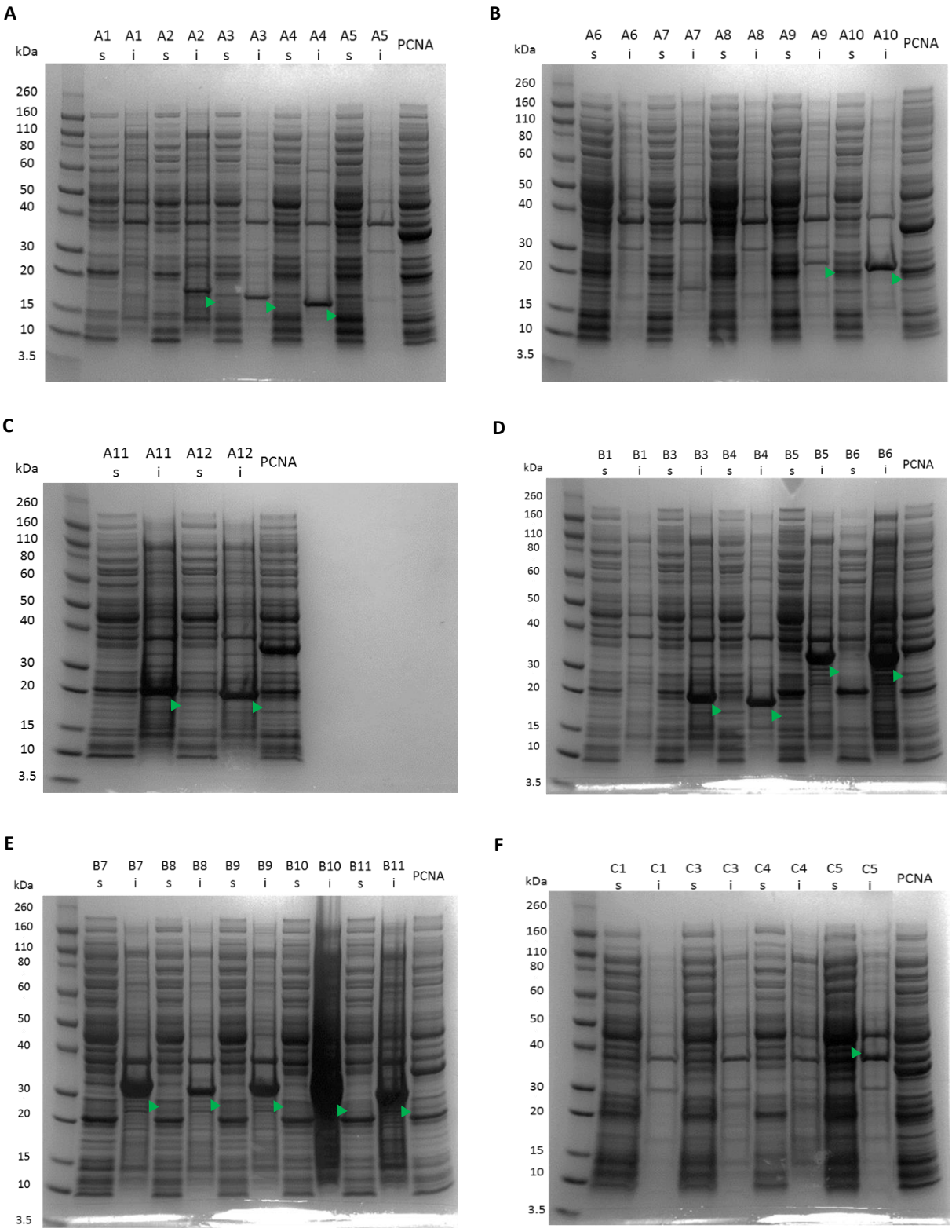
Code	Protein	Amino acid sequence
A1	**BB_0562 (1-180)**	MHHHHHHSSGVDLGTENLYFQ*SMKKIFILFIM IANISTNGFT KDSYLNRGIG FGASIGNPII NLIMSFPFID FEIGYGGSNG INLSGPKLES KFYDFNLLAI AALDFIFTIS LIKNLNLGIG IGGNISISSH TSKLINVELG FGMRIPLVIF YDITENLEIG MKIAPSIEFI SNTRSLAQHR TYSGIKSNFA GGIFAKYYIF
A2	**BB_0562 (16-180)**	MHHHHHHSSGVDLGTENLYFQ*SMTNGFTKDSYLNRGIGFGASIGNPII NLIMSFPFIDFEIGYGGSNGINLSGPKLESKFYDFNLLAIAALDFIFTISLIKN LNLGIGIGGNISISSHTSKLINVELGFGMRIPLVIFYDITENLEIGMKIAPSIE FISNTRSLAQHRTYSGIKSNFAGGIFAKYYIF
A3	**BB_0562 (27-180)**	MHHHHHHSSGVDLGTENLYFQ*SMRGIGFGASIGNPIINLIMSFPFIDFE IGYGGSNGINLSGPKLESKFYDFNLLAIAALDFIFTISLIKNLNLGIGIGGNIS ISSHTSKLINVELGFGMRIPLVIFYDITENLEIGMKIAPSIEFISNTRSLAQH RTYSGIKSNFAGGIFAKYYIF
A4	**BB_0562 (38-180)**	MHHHHHHSSGVDLGTENLYFQ*SMPIINLIMSFPFID FEIGYGGSNG INLSGPKLES KFYDFNLLAI AALDFIFTIS LIKNLNLGIG IGGNISISSH TSKLINVELG FGMRIPLVIF YDITENLEIG MKIAPSIEFI SNTRSLAQHR TYSGIKSNFA GGIFAKYYIF
A5	**BB_0027 (1-212)**	MHHHHHHSSGVDLGTENLYFQ*SMRKYIFIILI AVLLIGVNIK KIAAAANIDR HTNSTLGIDL SVGIPIFYND LSKAYPTNLY PGGIGAIKYQ YHILNNLAIG LELRYMFNFD INHSFNILNP DSSVGKIFYS VPITFSINYI FDIGELFQIP VFTNIGFSLN TYGDRNNNIT NLRTFDALPT ISFGSGILWN FNYKWAFGAT ASWWMMFEFG NSAKMAHFAL VSLSVTVNVN KL
A6	**BB_0027 (3-212)**	MHHHHHHSSGVDLGTENLYFQ*SMKYIFIILIAVLLIGVNIK KIAAAANIDR HTNSTLGIDL SVGIPIFYND LSKAYPTNLY PGGIGAIKYQ YHILNNLAIG LELRYMFNFD INHSFNILNP DSSVGKIFYS VPITFSINYI FDIGELFQIP VFTNIGFSLN TYGDRNNNIT NLRTFDALPT ISFGSGILWN FNYKWAFGAT ASWWMMFEFG NSAKMAHFAL VSLSVTVNVN KL
A7	**BB_0027 (18-212)**	MHHHHHHSSGVDLGTENLYFQ*SMNIKKIAAAANIDR HTNSTLGIDL SVGIPIFYND LSKAYPTNLY PGGIGAIKYQ YHILNNLAIG LELRYMFNFD INHSFNILNP DSSVGKIFYS VPITFSINYI FDIGELFQIP VFTNIGFSLN TYGDRNNNIT NLRTFDALPT ISFGSGILWN FNYKWAFGAT ASWWMMFEFG NSAKMAHFAL VSLSVTVNVN KL

A8	**BB_0027 (33-212)**	MHHHHHHSSGVDLGTENLYFQ*SMNSTLGIDLSVGIPIFYND LSKAYPTNLY PGGIGAIKYQ YHILNNLAIG LELRYMFNFD INHSFNILNP DSSVGKIFYS VPITFSINYI FDIGELFQIP VFTNIGFSLN TYGDRNNNIT NLRTFDALPT ISFGSGILWN FNYKWAFGAT ASWWMMFEFG NSAKMAHFAL VSLSVTVNVN KL
A9	**BB_0405 (1-203)**	MHHHHHHSSGVDLGTENLYFQ*SMRMLLATIIL ILTTGLLAAQ SKSKSMTEDD FDFDKLLAKE ESVRRLFGIG FGVGYPLANI TISVPYVDID LGYGGFVGLK PNNFLPYVVM GVDLLFKDEI HKNTMISGGI GIGADWSKGS PEKSNEKLEE EEENEAQQVA SLQNRIGVVI RLPLVIEYSF LKNIVIGFKA VATIGTTMLL GSPMSFEGAR FNFLGTGFIK IYI
A10	**BB_0405 (21-203)**	MHHHHHHSSGVDLGTENLYFQ*SMSKSKSMTEDD FDFDKLLAKE ESVRRLFGIG FGVGYPLANI TISVPYVDID LGYGGFVGLK PNNFLPYVVM GVDLLFKDEI HKNTMISGGI GIGADWSKGS PEKSNEKLEE EEENEAQQVA SLQNRIGVVI RLPLVIEYSF LKNIVIGFKA VATIGTTMLL GSPMSFEGAR FNFLGTGFIK IYI
A11	**BB_0405 (30-203)**	MHHHHHHSSGVDLGTENLYFQ*SMDFDFDKLLAKE ESVRRLFGIG FGVGYPLANI TISVPYVDID LGYGGFVGLK PNNFLPYVVM GVDLLFKDEI HKNTMISGGI GIGADWSKGS PEKSNEKLEE EEENEAQQVA SLQNRIGVVI RLPLVIEYSF LKNIVIGFKA VATIGTTMLL GSPMSFEGAR FNFLGTGFIK IYI
A12	**BB_0405 (45-203)**	MHHHHHHSSGVDLGTENLYFQ*SMRLFGIGFGVGYPLANI TISVPYVDID LGYGGFVGLK PNNFLPYVVM GVDLLFKDEI HKNTMISGGI GIGADWSKGS PEKSNEKLEE EEENEAQQVA SLQNRIGVVI RLPLVIEYSF LKNIVIGFKA VATIGTTMLL GSPMSFEGAR FNFLGTGFIK IYI
B1	**BB_0406 (1-203)**	MHHHHHHSSGVDLGTENLYFQ*SMIKIFKKIYI LTLVLGMAHL SFASDNYMVR CSKEEDSTTC IAKLKEIKEK KNYDLFSMGI GIGDPIANIM ITIPYINIDF GYGGFIGLKS NNFENYLNGG IDVIFKKQIG QYMKIGGGIG IGADWSKTSL IPPNEEEETD YERIGAVIRI PFIMEYNFAK NLSIGFKIYP AVGPTILLTK PSILFEGIKF NFFGFGFIKF AFN
B2	**BB_0406 (21-203)**	MHHHHHHSSGVDLGTENLYFQ*SMSFASDNYMVR CSKEEDSTTC IAKLKEIKEK KNYDLFSMGI GIGDPIANIM ITIPYINIDF GYGGFIGLKS NNFENYLNGG IDVIFKKQIG QYMKIGGGIG IGADWSKTSL IPPNEEEETD YERIGAVIRI PFIMEYNFAK NLSIGFKIYP AVGPTILLTK PSILFEGIKF NFFGFGFIKF AFN
B3	**BB_0406 (43-203)**	MHHHHHHSSGVDLGTENLYFQ*SMKLKEIKEKKNYDLFSMGI GIGDPIANIM ITIPYINIDF GYGGFIGLKS NNFENYLNGG IDVIFKKQIG QYMKIGGGIG IGADWSKTSL IPPNEEEETD YERIGAVIRI PFIMEYNFAK NLSIGFKIYP AVGPTILLTK PSILFEGIKF NFFGFGFIKF AFN

B4	**BB_0406 (54-203)**	MHHHHHHSSGVDLGTENLYFQ*SMDLFSMGIGIGDPIANIM ITIPYINIDF GYGGFIGLKS NNFENYLNGG IDVIFKKQIG QYMKIGGGIG IGADWSKTSL IPPNEEEETD YERIGAVIRI PFIMEYNFAK NLSIGFKIYP AVGPTILLTK PSILFEGIKF NFFGFGFIKF AFN
B5	**BB_0095 (1-181)**	MHHHHHHSSGMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCK MIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVA ATKVGALSKGQLKEFLDANLAGTENLYFQ*SMLDSYYYVLS SLPYIDLKSL KNYSVSDFLN NVEISLSKKD FNFLKELLEF KIDRGKSRVI DLFLNFEDTI RYTLAALRAE KLGLSKDFYL ESTYFSSYYL SILKNICLKE NPFEIELSFD LLKWQFLTDL EVGHEFDFEK LIIYFLKLRL FLRHNLFKEK LGIQNFDNIC KNLIDKTNEK V
B6	**BB_0095 (4-181)**	MHHHHHHSSGMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCK MIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVA ATKVGALSKGQLKEFLDANLAGTENLYFQ*SMSYYYVLSSLPYIDLKSL KNYSVSDFLN NVEISLSKKD FNFLKELLEF KIDRGKSRVI DLFLNFEDTI RYTLAALRAE KLGLSKDFYL ESTYFSSYYL SILKNICLKE NPFEIELSFD LLKWQFLTDL EVGHEFDFEK LIIYFLKLRL FLRHNLFKEK LGIQNFDNIC KNLIDKTNEK V
B7	**BB_0095 (24-181)**	MHHHHHHSSGMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCK MIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVA ATKVGALSKGQLKEFLDANLAGTENLYFQ*SMSVSDFLNNVEISLSKKD FNFLKELLEF KIDRGKSRVI DLFLNFEDTI RYTLAALRAE KLGLSKDFYL ESTYFSSYYL SILKNICLKE NPFEIELSFD LLKWQFLTDL EVGHEFDFEK LIIYFLKLRL FLRHNLFKEK LGIQNFDNIC KNLIDKTNEK V
B8	**BB_0095 (37-181)**	MHHHHHHSSGMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCK MIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVA ATKVGALSKGQLKEFLDANLAGTENLYFQ*SMSKKDFNFLKELLEF KIDRGKSRVI DLFLNFEDTI RYTLAALRAE KLGLSKDFYL ESTYFSSYYL SILKNICLKE NPFEIELSFD LLKWQFLTDL EVGHEFDFEK LIIYFLKLRL FLRHNLFKEK LGIQNFDNIC KNLIDKTNEK V
B9	**BB_0095 (1-130)**	MHHHHHHSSGMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCK MIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVA ATKVGALSKGQLKEFLDANLAGTENLYFQ*SMLDSYYYVLS SLPYIDLKSL KNYSVSDFLN NVEISLSKKD FNFLKELLEF KIDRGKSRVI DLFLNFEDTI RYTLAALRAE KLGLSKDFYL ESTYFSSYYL SILKNICLKE NPFEIELSFD LLKWQFLTDL
B10	**BB_0095 (4-130)**	MHHHHHHSSGMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCK MIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVA ATKVGALSKGQLKEFLDANLAGTENLYFQ*SMSYYYVLSSLPYIDLKSL KNYSVSDFLN NVEISLSKKD FNFLKELLEF KIDRGKSRVI DLFLNFEDTI RYTLAALRAE KLGLSKDFYL ESTYFSSYYL SILKNICLKE NPFEIELSFD LLKWQFLTDL
B11	**BB_0095 (24-130)**	MHHHHHHSSGMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCK MIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVA

		ATKVGALSKGQLKEFLDANLAGTENLYFQ*SMSVSDFLNNVEISLSKKD FNFLKELLEF KIDRGKSRVI DLFLNFEDTI RYTLAALRAE KLGLSKDFYL ESTYFSSYYL SILKNICLKE NPFEIELSFD LLKWQFLTDL
B12	**BB_0095 (37-130)**	MHHHHHHSSGMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCK MIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVA ATKVGALSKGQLKEFLDANLAGTENLYFQ*SMSKKDFNFLKELLEF KIDRGKSRVI DLFLNFEDTI RYTLAALRAE KLGLSKDFYL ESTYFSSYYL SILKNICLKE NPFEIELSFD LLKWQFLTDL
C1	**BB_0095 (1-157)**	MHHHHHHSSGMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCK MIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVA ATKVGALSKGQLKEFLDANLAGTENLYFQ*SMLDSYYYVLS SLPYIDLKSL KNYSVSDFLN NVEISLSKKD FNFLKELLEF KIDRGKSRVI DLFLNFEDTI RYTLAALRAE KLGLSKDFYL ESTYFSSYYL SILKNICLKE NPFEIELSFD LLKWQFLTDL EVGHEFDFEK LIIYFLKLRL FLRHNLF
C2	**BB_0095 (4-157)**	MHHHHHHSSGMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCK MIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVA ATKVGALSKGQLKEFLDANLAGTENLYFQ*SMSYYYVLSSLPYIDLKSL KNYSVSDFLN NVEISLSKKD FNFLKELLEF KIDRGKSRVI DLFLNFEDTI RYTLAALRAE KLGLSKDFYL ESTYFSSYYL SILKNICLKE NPFEIELSFD LLKWQFLTDL EVGHEFDFEK LIIYFLKLRL FLRHNLF
C3	**BB_0095 (24-157)**	MHHHHHHSSGMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCK MIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVA ATKVGALSKGQLKEFLDANLAGTENLYFQ*SMSVSDFLNNVEISLSKKD FNFLKELLEF KIDRGKSRVI DLFLNFEDTI RYTLAALRAE KLGLSKDFYL ESTYFSSYYL SILKNICLKE NPFEIELSFD LLKWQFLTDL EVGHEFDFEK LIIYFLKLRL FLRHNLF
C4	**BB_0095 (37-157)**	MHHHHHHSSGMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCK MIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVA ATKVGALSKGQLKEFLDANLAGTENLYFQ*SMSKKDFNFLKELLEF KIDRGKSRVI DLFLNFEDTI RYTLAALRAE KLGLSKDFYL ESTYFSSYYL SILKNICLKE NPFEIELSFD LLKWQFLTDL EVGHEFDFEK LIIYFLKLRL FLRHNLF
C5	**BB_0040 (1-283)**	MHHHHHHSSGMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCK MIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVA ATKVGALSKGQLKEFLDANLAGTENLYFQ*SMNTNQNKFNL NINITKDELS RLIKIVYNNF GINLSEKKKL LIESRLSSLL KVKGFKNFTE YINFLEKSTG NIQLIELIDK ISTNHTYFFR ESKHFDFLNN KILPKLTEKI LNSENSEIRI WSAGCSSGEE PYTIAMMLKE YMEHNRVNFK VKILATDISI SVLNEAREGI YPEDRIINLP KYLKIKYLNQ LQDDKFQVKE ILKKMVYFKK LNLMDEKFPF SKKFDLIFCR NVMIYFDEKT RNNLANKFNS YLKKDSYLLI GHSETIRGNK NLEYIMPATY KKN

Appendix 4- Protein test expression labelled SDS-PAGE



Appendix 5- Customised R-code for the computational framework to search for β-barrel genes in the plasmid proteomes of *Borrelia*

The following R-code was written by Dr Jarek Bryk at The University of Huddersfield.

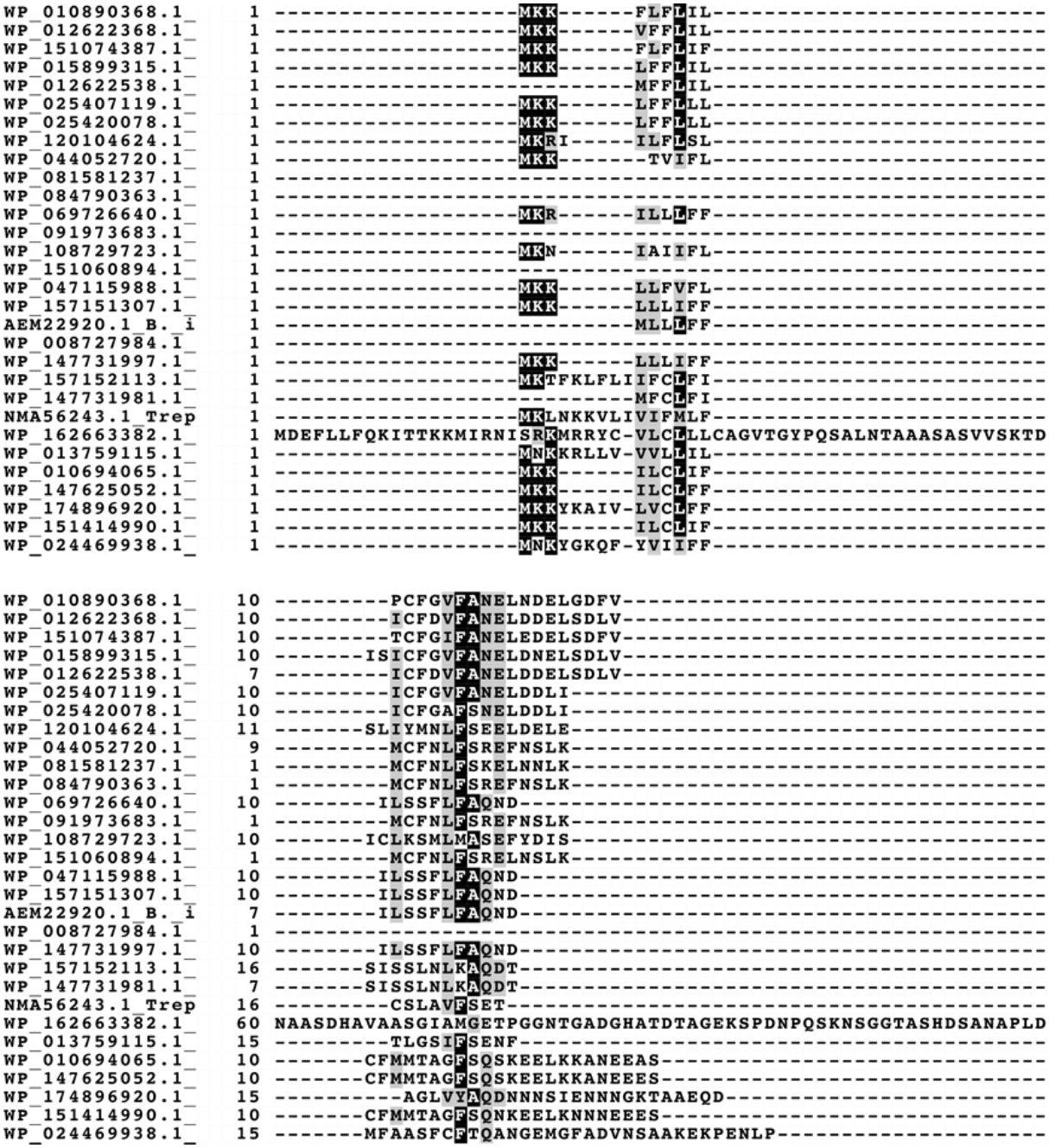
```
#Load packages
install.packages("here")
install.packages("janitor")
install.packages("tidyverse")
if (!requireNamespace("BiocManager", quietly = TRUE))
    install.packages("BiocManager")
BiocManager::install("Biostrings")

library(here)
library(janitor)
library(tidyverse)
library(Biostrings)

# Function to clean the data, parse the columns and remvoe duplicates
# The taxon_id should be a unique number for each proteome.
cleandata <- function(data) {
      data %>%
            separate(names, into = c("sp", "protein_id", "rest"), sep =
"\\|") %>%
            separate(rest, into = c("protein_abbrev", "rest"), sep = "_",
extra = "merge") %>%
            separate(rest, into = c("species_id", "description"), sep = "
", extra = "merge") %>%
            separate(description, into = c("temp", "taxon_id"), sep =
"OX=", extra = "merge") %>%
            separate(taxon_id, into = c("taxon_id", "rest"), sep = " ",
extra = "merge") %>%
            separate(temp, into = c("protein_name", "species_name"), sep =
"OS=", extra = "merge") %>%
            distinct() %>%
            select(-sp)
}

# Function to parse the AAStringSet into tibble
AASS_2tibble <- function(data) {
      as_tibble(data.frame(sequence = as.character(data),
names=names(data), stringsAsFactors = FALSE))
}

# Loading each file into a list of files, without losing any information in
the name of each sequence
all_species <- map(.x = list.files(path = here("data"), full.names = TRUE,
pattern = "*.fasta"), .f = readAAStringSet)


# all_species is now a list of all the fasta sequences from each species
all_species


# This will now turn each species' file in our list into a list of proper
tibbles
all_species_tibble <- map_df(all_species, AASS_2tibble) %>%
      select(names, sequence)
all_species_tibble
```

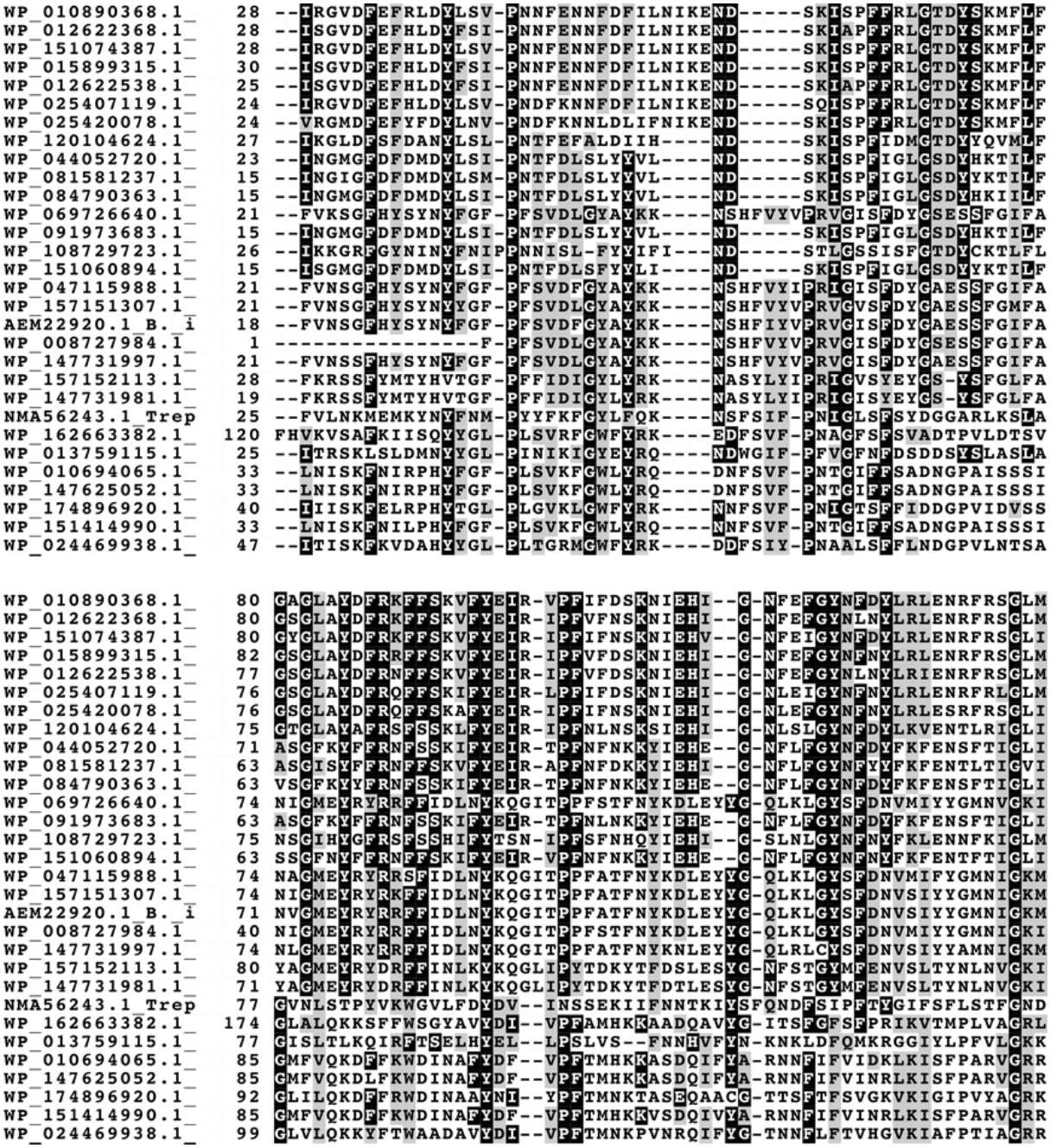
```
# This will parse and separate all the relevant columns in the giant tibble
all_species_tidy <- cleandata(all_species_tibble)
glimpse(all_species_tidy) # gives 5923 entries
all_species_tidy


# Some sanity checks
all_species_tidy %>%
      count(species_id) # 9
  all_species_tidy %>%
      count(taxon_id) # This gives 14
all_species_tidy %>% group_by(taxon_id) %>%
      count()


# Importing results from SpLip_results.txt
# Run SpLip, then simplify the text file using unix command sed -n '/>/p'
input.txt > SpLip_results.txt
results_splip <- read_tsv(here("data/SpLip_results.txt"), col_names =
FALSE) %>%
      separate(X1, into = c("stuff1", "protein_id", "rest"), sep = "\\|")
%>%
      separate(rest, into = c("protein_abbrev", "prediction_splip"), sep =
":", extra = "merge") %>%
      separate(protein_abbrev, into = c("protein_abbrev", "species_id"),
sep = "_") %>%
      mutate(prediction_splip = str_squish(prediction_splip)) %>%
      distinct() %>%
      select(protein_id, species_id, prediction_splip)

# https://services.healthtech.dtu.dk/service.php?LipoP-1.0
# Importing results from LipoP-1.0 LipoP_results.txt
results_lipop <- read_tsv(here("data/LipoP_results.txt"), col_names =
FALSE) %>%
      separate(X1, into = c("stuff1", "protein_id", "rest"), sep = "_",
extra = "merge") %>%
      separate(rest, into = c("rest2", "prediction_lipop"), sep = " ",
extra = "merge") %>%
      separate(prediction_lipop, into = c("prediction_lipop", "score"), sep
= " ", extra = "merge") %>%
      distinct() %>%
      select(protein_id, prediction_lipop)

#https://services.healthtech.dtu.dk/service.php?TMHMM-2.0
# Importing results from TMHMM_results.txt
results_tmhmm <- read_tsv(here("data/TMHMM_results.txt"), col_names =
FALSE) %>%
      separate(X1, into = c("stuff1", "protein_id", "rest"), sep = "\\|")
%>%
      separate(rest, into = c("protein_abbrev", "species_id"), sep = "_")
%>%
      separate(X5, into = c("rest", "prediction_tmhmm"), sep = "=") %>%
      distinct() %>%
      select(protein_id, species_id, prediction_tmhmm)

#http://phobius.sbc.su.se/
# Importing results from Phobius_results.txt
# When copying data from phobius web output, remove the header "SEQENCE ID
TM SP PREDICTION"
```

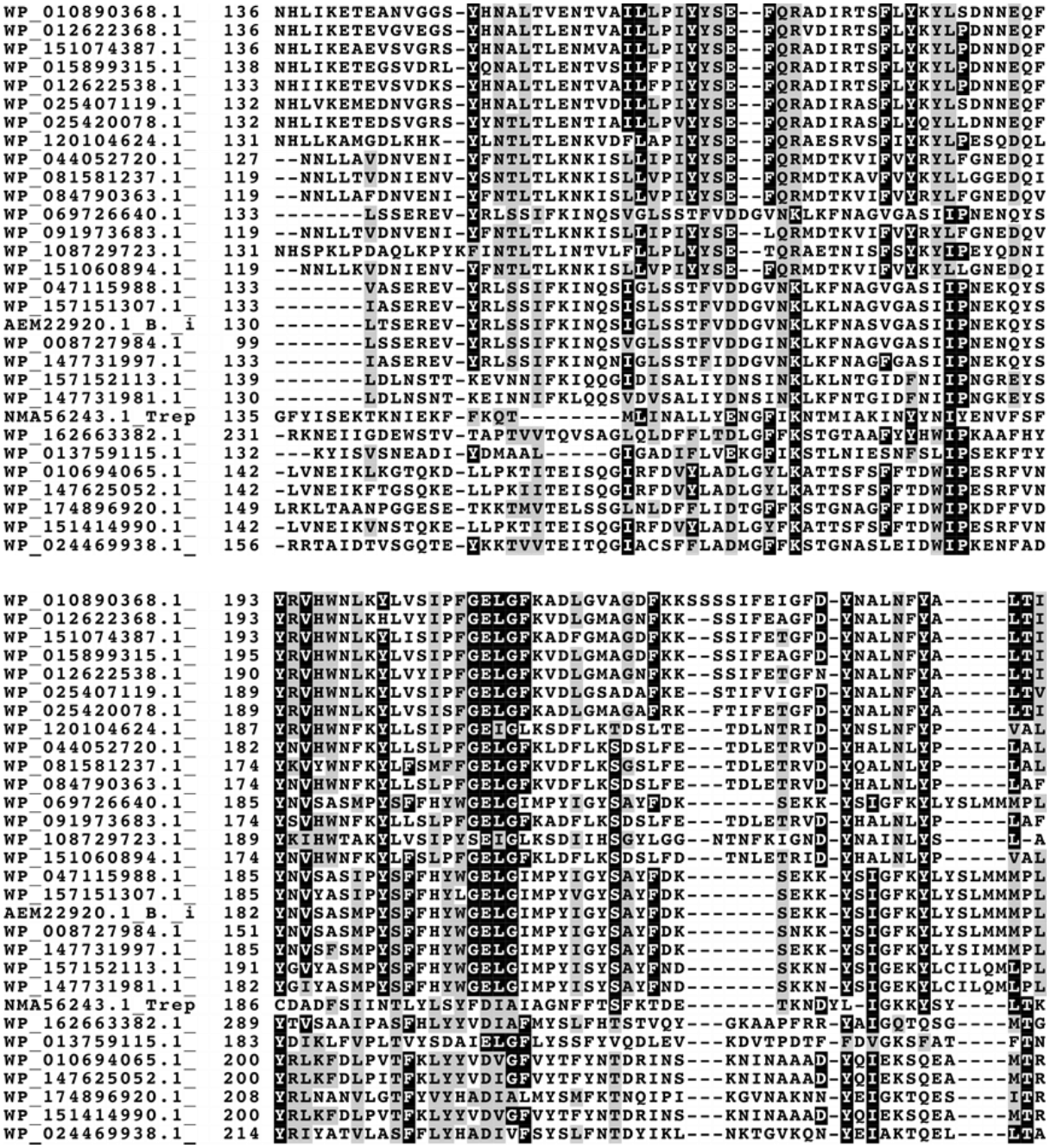
```
results_phobius <- read_tsv(here("data/Phobius_results.txt"), col_names =
FALSE) %>%
      separate(X1, into = c("stuff1", "protein_id", "rest"), sep = "\\|")
%>%
      separate(rest, into = c("protein_abbrev", "species_id"), sep = "_")
%>%
      separate(species_id, into = c("species_id", "rest"), sep = " ", extra
= "merge") %>%
      mutate(rest = str_squish(rest)) %>%
      separate(rest, into = c("phobius_TM", "phobius_SP",
"phobius_topology"), sep = " ", extra = "merge") %>%
      distinct() %>%
      select(protein_id, species_id, phobius_TM)

#https://www.psort.org/psortb/
# Importing results from PSORTb_results.txt
# First line of PSORTb output has column names "SeqID Localization
      Score"
results_psortb <- read_tsv(here("data/PSORTb_results.txt"), col_names =
TRUE) %>%
      separate(SeqID, into = c("stuff1", "protein_id", "rest"), sep =
"\\|") %>%
      separate(rest, into = c("protein_abbrev", "rest"), sep = "_", extra =
"merge") %>%
      separate(rest, into = c("species_id", "protein_description"), sep = "
", extra = "merge") %>%
      distinct() %>%
      select(protein_id, species_id, Localization) %>%
      rename(Localization = "prediction_psortb")

results_psortb %>% distinct() %>%  group_by(prediction_psortb) %>% count()
%>% arrange(desc(n))

#https://services.healthtech.dtu.dk/service.php?SignalP-5.0
# Importing results from SignalP5_results2.txt
# remove timestamp line from start of file
results_signalP5 <- read_tsv(here("data/SignalP5_results.txt"), col_names =
TRUE) %>%
      separate("# ID", into = c("stuff1", "protein_id", "protein_abbrev",
"species_id", "rest"), sep = "_", extra = "merge") %>%
      select(protein_id, species_id, Prediction) %>%
      rename(Prediction = "prediction_signalP5") %>%
      distinct()

# Importing results from PRED_TMBB2_results.txt
# (using the dipeptide model)
results_pred_tmbb2 <-  read_tsv(here("data/PRED_TMBB2_results.txt"),
col_names = TRUE) %>% clean_names() %>%
      separate("protein_id", into = c("stuff1", "protein_id",
"protein_abbrev", "species_id", "rest"), sep = "_") %>%
      separate(beta_barrel_score_cut_off_is_0_43, into = c("X4",
"prediction_pred_tmbb2"), sep = "\\|") %>%
      select(protein_id, species_id, prediction_pred_tmbb2,
sequence_length) %>%
      distinct()

# The big merge
# https://stackoverflow.com/questions/8091303/simultaneously-merge-
multiple-data-frames-in-a-list
all_species_predictions <-
```

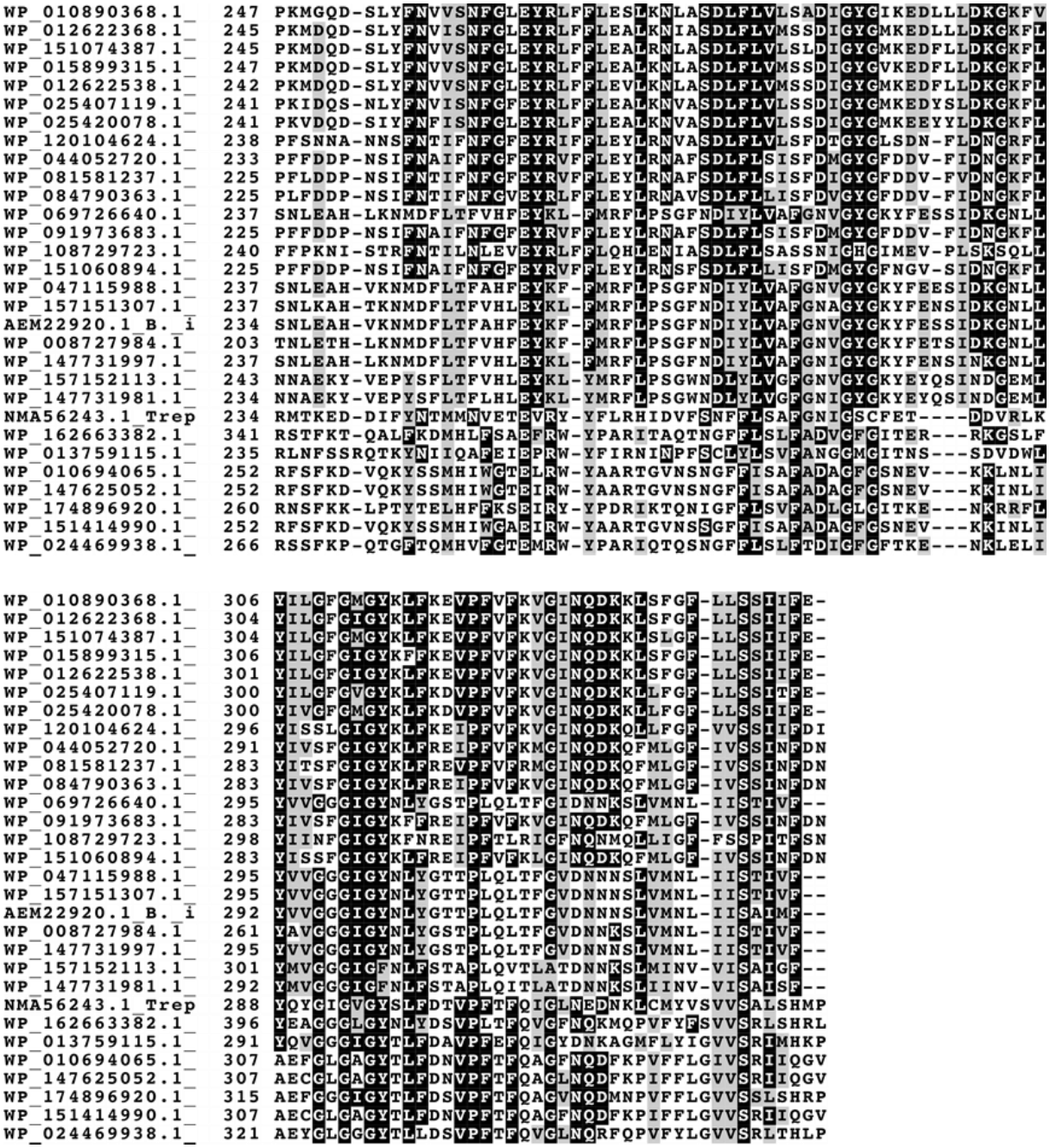
```
      purrr::reduce(.x = list(all_species_tidy, results_splip,
results_lipop, results_tmhmm, results_phobius, results_psortb,
results_signalP5, results_pred_tmbb2), .f = left_join, by = "protein_id")
%>%
      select(-starts_with("species_id."), -rest) %>%
      select(protein_id, protein_abbrev, sequence_length, protein_name,
species_id, taxon_id, sequence, starts_with("prediction")) %>%
      mutate(prediction_pred_tmbb2 = as.numeric(prediction_pred_tmbb2),
                          prediction_tmhmm = as.numeric(prediction_tmhmm))

glimpse(all_species_predictions)



# The big filtering
all_species_predictions %>% #
      filter(prediction_tmhmm == 0) %>% # 3465
      filter(!grepl("Cytoplasmic", prediction_psortb)) %>% # 1871
      filter(grepl("SPI)", prediction_signalP5)) %>% # 46
      filter(prediction_pred_tmbb2 > 0.43) # 4

filtered_predictions <-
      all_species_predictions %>%
#start 5923
  filter(prediction_splip == "NOT LIPOPROTEIN") %>%
#5128
      filter(prediction_lipop == "SpI" | prediction_lipop == "CYT" |
prediction_lipop == "TMH" ) %>%  #5005
      filter(prediction_tmhmm == 0)
%>%  #4301
      filter(!grepl("Cytoplasmic", prediction_psortb))
%>%  #2484
      filter(grepl("SPI)", prediction_signalP5))
                                                %>%  #63
      filter(prediction_pred_tmbb2 > 0.43)
#13


view(filtered_predictions)

output_table <- filtered_predictions %>% select(protein_id,
sequence_length, protein_name, taxon_id)

#create table
install.packages("formattable")
install.packages("data.table")
library(formattable)
library(data.table)
formattable(output_table)

formattable(output_table,
            align =c("l","c", "r"),
            list(`Indicator Name` = formatter(
              "span", style = ~ style(color = "grey",font.weight = "bold"))
```

Appendix 6- Multiple sequence alignment of BBJ25 protein sequences from *Borrelia*, *Brachyspira* and *Treponema* used for phylogenetic analysis

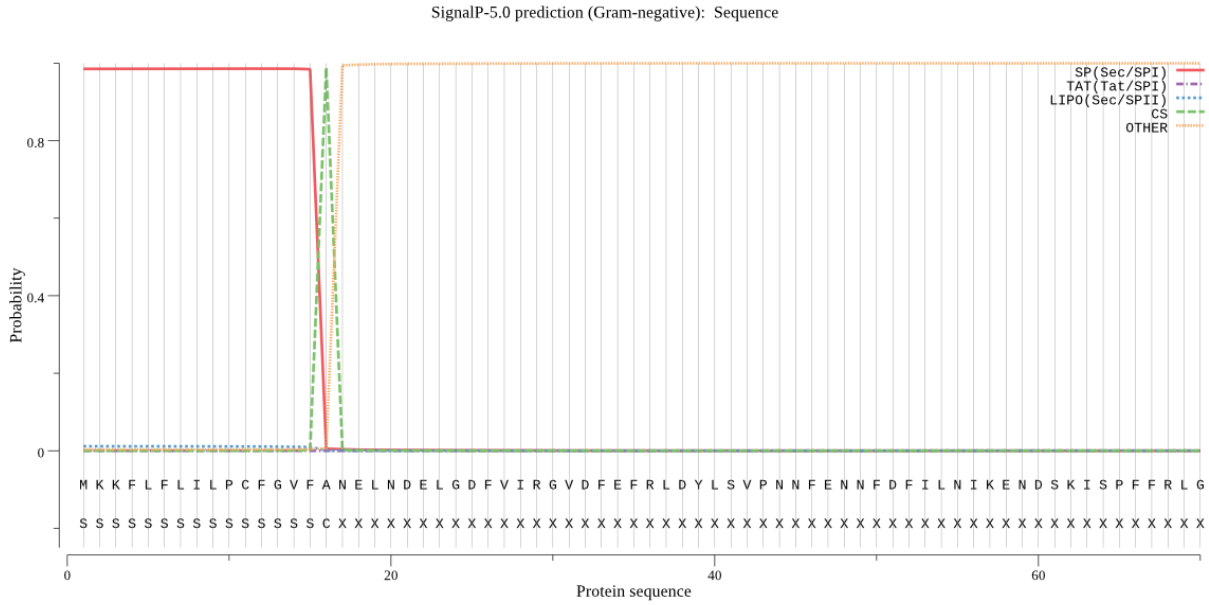



WP_010890368.1_ 28 --IRGVDFEFRLDYLSV-PNNFENNFDFILNIKENDSKISPFFRLGTDYSKMFLF
WP_012622368.1_ 28 --ISGVDFEFHLDYFSI-PNNFENNFDFILNIKENDSKIAPFFRLGTDYSKMFLF
WP_151074387.1_ 28 --IRGVDFEFHLDYLSV-PNNFENNFDFILNIKENDSKISPFFRLGTDYSKMFLF
WP_015899315.1_ 30 --ISGVDFEFHLDYFSI-PNNFENNFDFILNIKENDSKISPFFRLGTDYSKMFLF
WP_012622538.1_ 25 --ISGVDFEFHLDYFSI-PNNFENNFDFILNIKENDSKIAPFFRLGTDYSKMFLF
WP_025407119.1_ 24 --IRGVDFEFHLDYLSV-PNDFKNNFDFILNIKENDSQISPFFRLGTDYSKMFLF
WP_025420078.1_ 24 --VRGMDFEFYFDYLNV-PNDFKNNLDLIFNIKENDSKISPFFRLGTDYSKMFLF
WP_120104624.1_ 27 --IKGLDFSFDANYLSL-PNTFEFALDIIHNDSKISPFIDMGTDYYQVMLF
WP_044052720.1_ 23 --INGMGFDFDMDYLSI-PNTFDLSLYYVLNDSKISPFIGLGSDYHKTILF
WP_081581237.1_ 15 --INGIGFDFDMDYLSM-PNTFDLSLYYVLNDSKISPFIGLGSDYYKTILF
WP_084790363.1_ 15 --INGMGFDFDMDYLSI-PNTFDLSLYYVLNDSKISPFIGLGSDYHKIILF
WP_069726640.1_ 21 --FVKSGFHYSYNYFGF-PFSVDLGYAYKKNSHFVYVPRVGISFDYGSESSFGIFA
WP_091973683.1_ 15 --INGMGFDFDMDYLSI-PNTFDLSLYYVLNDSKISPFIGLGSDYHKTILF
WP_108729723.1_ 26 --IKKGRFGYNINYFNIPPNNISL-FYYIFINDSTLGSSISFGTDYCKTLFL
WP_151060894.1_ 15 --ISGMGFDFDMDYLSI-PNTFDLSFYYLINDSKISPFIGLGSDYYKTILF
WP_047115988.1_ 21 --FVNSGFHYSYNYFGF-PFSVDFGYAYKKNSHFVYIPRIGISFDYGAESSFGIFA
WP_157151307.1_ 21 --FVNSGFHYSYNYYGF-PFSVDLGYAYKKNSHFVYVPRVGVSFDYGAESSFGMFA
AEM22920.1_B._i 18 --FVNSGFHYSYNYFGF-PFSVDFGYAYKKNSHFIYVPRVGISFDYGAESSFGIFA
WP_008727984.1_ 1 ----------F-PFSVDLGYAYKKNSHFVYVPRVGISFDYGSESSFGIFA
WP_147731997.1_ 21 --FVNSSFHYSYNYFGF-PFSVDLGYAYKKNSHFVYVPRVGISFDYGAESSFGIFA
WP_157152113.1_ 28 --FKRSSFYMTYHVTGF-PFFIDIGYLYRKNASYLYIPRIGVSYEYGS-YSFGLFA
WP_147731981.1_ 19 --FKRSSFYMTYHVTGF-PFFIDIGYLYRKNASYLYIPRIGISYEYGS-YSFGLFA
NMA56243.1_Trep 25 --FVLNKMEMKYNYFNM-PYYFKFGYLFQKNSFSIF-PNIGLSFSYDGGARLKSLA
WP_162663382.1_ 120 FHVKVSAFKIISQYYGL-PLSVRFGWFYRKEDFSVF-PNAGFSFSVADTPVLDTSV
WP_013759115.1_ 25 --ITRSKLSLDMNYYGL-PINIKIGYEYRQNDWGIF-PFVGFNFDSDDSYSLASLA
WP_010694065.1_ 33 --LNISKFNIRPHYFGF-PLSVKFGWLYRQDNFSVF-PNTGIFFSADNGPAISSSI
WP_147625052.1_ 33 --LNISKFNIRPHYFGF-PLSVKFGWLYRQDNFSVF-PNTGIFFSADNGPAISSSI
WP_174896920.1_ 40 --IISKFELRPHYTGL-PLGVKLGWFYRKNNFSVF-PNIGTSFFIDDGPVIDVSS
WP_151414990.1_ 33 --LNISKFNILPHYFGF-PLSVKFGWLYRQNNFSVF-PNTGIFFSADNGPAISSSI
WP_024469938.1_ 47 --ITISKFKVDAHYYGL-PLTGRMGWFYRKDDFSIY-PNAALSFFLNDGPVLNTSA
WP_010890368.1_ 80 GAGLAYDFRKFFSKVFYEIR-VPFIFDSKNIEHI--G-NFEFGYNFDYLRLENRFRSGLM
WP_012622368.1_ 80 GSGLAYDFRKFFSKVFYEIR-IPFVFNSKNIEHI--G-NFEFGYNLNYLRLENRFRSGLM
WP_151074387.1_ 80 GYGLAYDFRKFFSKVFYEIR-IPFIFNSKNIEHV--G-NFEIGYNFDYLRLENRFRSGLI
WP_015899315.1_ 82 GSGLAYDFRRFFSKVFYEIR-IPFVFDSKNIEHI--G-NFEFGYNFNYLRLENRFRSGLM
WP_012622538.1_ 77 GSGLAYDFRNFFSKVFYEIR-IPFVFDSKNIEHI--G-NFEFGYNLNYLRIENRFRSGLM
WP_025407119.1_ 76 GSGLAYDFRQFFSKIFYEIR-LPFIFDSKNIEHI--G-NLEIGYNFNYLRLENRFRLGLM
WP_025420078.1_ 76 GSGLAYDFRQFFSKAFYEIR-IPFIFNSKNIEHI--G-NLEFGYNFNYLRLESRFRSGLI
WP_120104624.1_ 75 GTGLAYAFRSFSSKLFYEIR-IPFNLNSKSIEHI--G-NLSLGYNFDYLKVENTLRIGLI
WP_044052720.1_ 71 ASGFKYFFRNFSSKIFYEIR-TPFNFNKKYIEHE--G-NFLFGYNFDYFKFENSFTIGLI
WP_081581237.1_ 63 ASGISYFFRNFFSKVFYEIR-APFNFDKKYIEHI--G-NFLFGYNFNYFKFENTLTIGVI
WP_084790363.1_ 63 VSGFKYFFRNFSSKIFYEIR-TPFNFNKKYIEHE--G-NFLFGYNFDYFKFENSFTIGLI
WP_069726640.1_ 74 NIGMEYRYRRFFIDLNYKQGITPPFSTFNYKDLEYYG-QLKLGYSFDNVMIYYGMNVGKI
WP_091973683.1_ 63 ASGFKYFFRNFSSKIFYEIR-TPFNLNKKYIEHE--G-NFLFGYNFDYFKFENSFTIGLI
WP_108729723.1_ 75 NSGIHYGFRSFSSHIFYTSN-IPFSFNHQYIEHI--G-SLNLGYNFNYFKLENNFKIGLM
WP_151060894.1_ 63 SSGFNYFFRNFFSKIFYEIR-VPFNFNKKYIEHE--G-NFLFGYNFNYFKFENTFTIGLI
WP_047115988.1_ 74 NAGMEYRYRRSFIDLNYKQGITPPFATFNYKDLEYYG-QLKLGYSFDNVMIFYGMNIGKM
WP_157151307.1_ 74 NIGMEYRYRKFFIDLNYKQGITPPFATFNYKDLEYYG-QLKLGYSFDNVSIYYGMNIGKM
AEM22920.1_B._i 71 NVGMEYRYRRFFIDLNYKQGITPPFATFNYKDLEYYG-QLKLGYSFDNVSIYYGMNIGKM
WP_008727984.1_ 40 NIGMEYRYRRFFIDLNYKQGITPPFSTFNYKDLEYYG-QLKLGYSFDNVMIYYGMNIGKI
WP_147731997.1_ 74 NLGMEYRYRRFFIDLNYKQGITPPFATFNYKNLEYYG-QLRLCYSFDNVSIYYAMNIGKM
WP_157152113.1_ 80 YAGMEYRYDRFFINLKYKQGLIPYTDKYTFDSLESYG-NFSTGYMFENVSLTYNLNVGKI
WP_147731981.1_ 71 YAGMEYRYDRFFINLKYKQGLIPYTDKYTFDTLESYG-NFSTGYMFENVSLTYNLNVGKI
NMA56243.1_Trep 77 GVNLSTPYVKWGVLFDYDV--INSSEKIIFNNTKIYSFQNDFSIPFTYGIFSFLSTFGND
WP_162663382.1_ 174 GLALQKKSFFWSGYAVYDI--VPFAMHKKAADQAVYG-ITSFGFSFPRIKVTMPLVAGRL
WP_013759115.1_ 77 GISLTLKQIRFTSELHYEL--LPSLVS--FNNHVFYN-KNKLDFQMKRGGIYLPFVLGKK
WP_010694065.1_ 85 GMFVQKDFFKWDINAFYDF--VPFTMHKKASDQIFYA-RNNFIFVIDKLKISFPARVGRR
WP_147625052.1_ 85 GMFVQKDLFKWDINAFYDF--VPFTMHKKASDQIFYA-RNNFIFVINRLKISFPARVGRR
WP_174896920.1_ 92 GLILQKDFFRWDINAAYNI--YPFTMNKTASEQAACG-TTSFTFSVGKVKIGIPVYAGRK
WP_151414990.1_ 85 GMFVQKDFFKWDINAFYDF--VPFTMHKKVSDQIVYA-RNNFIFVINRLKISFPARVGRR
WP_024469938.1_ 99 GLVLQKKYFTWAADAVYDI--VPFTMNKPVNRQIFYG-TNNFLFTVHGVKIAFPTIAGRR

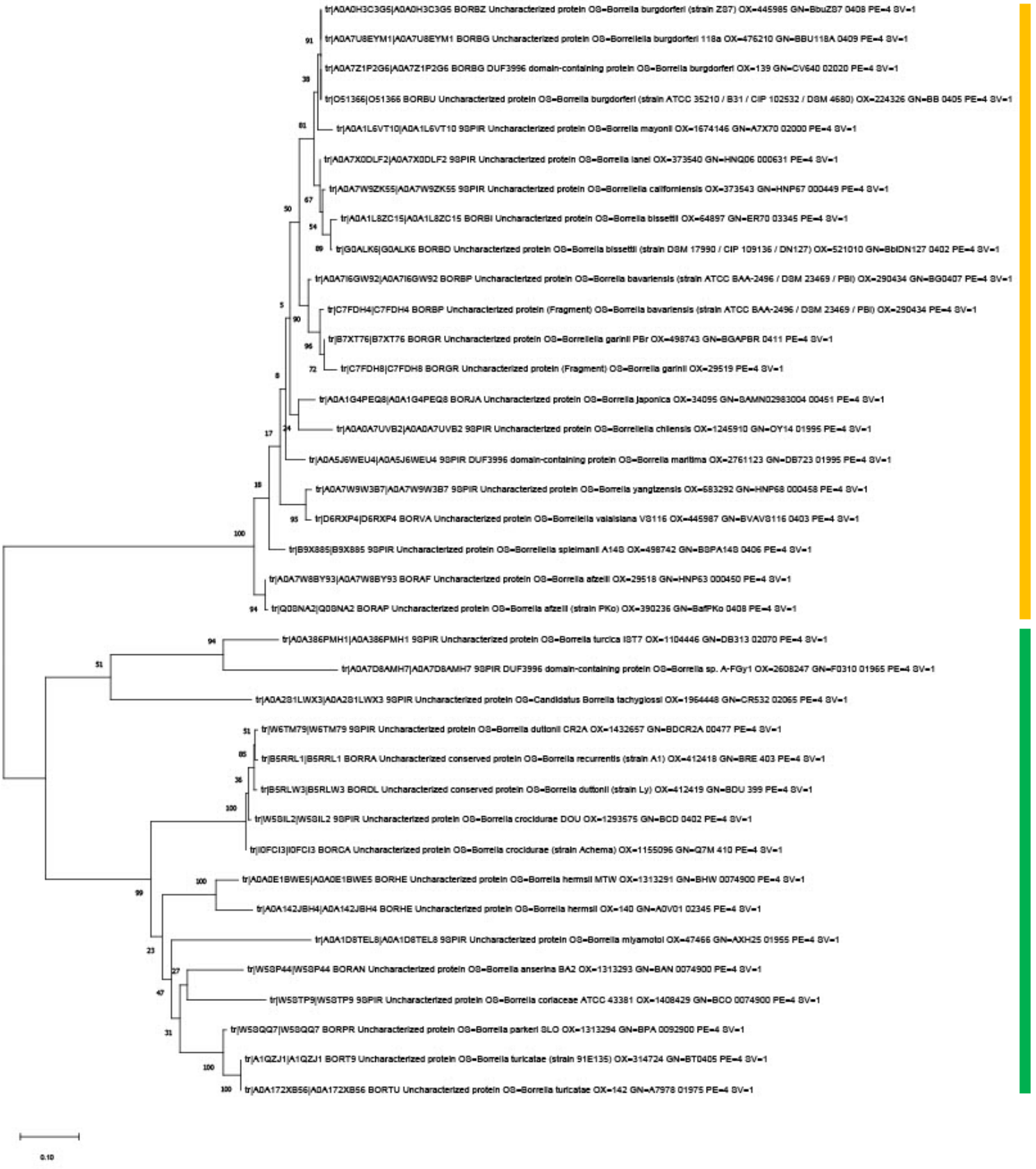

WP_010890368.1_ 136 NHLIKETEANVGGS-YHNALTVENTVAILLPIYYSE--FQRADIRTSFLYKYLSDNNEQF
WP_012622368.1_ 136 NHLIKETEVGVEGS-YHNALTLENTVAILLPIYYSE--FQRVDIRTSFLYKYLPDNNEQF
WP_151074387.1_ 136 NHLIKEAEVSVGRS-YHNALTLENMVAILLPIYYSE--FQRADIRTSFLYKYLPDNNDQF
WP_015899315.1_ 138 NHLIKETEGSVDRL-YQNALTLENTVSILFPIYYSE--FQRADIRTSFLYKYLPDNNEQF
WP_012622538.1_ 133 NHIIKETEVSVDKS-YHNALTLENTVAILFPIYYSE--FQRADIRTSFLYKYLPDNNEQF
WP_025407119.1_ 132 NHLVKEMEDNVGRS-YHNALTLENTVDILLPIYYSE--FQRADIRASFLYKYLSDNNEQF
WP_025420078.1_ 132 NHLIKETEDSVGRS-YYNTLTLENTIAILLPVYYSE--FQRADIRASFLYQYLLDNNEQF
WP_120104624.1_ 131 NHLLKAMGDLKHK--YLNTLTLENKVDFLAPIYYSE--FQRAESRVSFIYKYLPESQDQL
WP_044052720.1_ 127 --NNLLAVDNVENI-YFNTLTLKNKISLLIPIYYSE--FQRMDTKVIFVYRYLFGNEDQI
WP_081581237.1_ 119 --NNLLTVDNIENV-YSNTLTLKNKISLLVPIYYSE--FQRMDTKAVFVYKYLLGGEDQI
WP_084790363.1_ 119 --NNLLAFDNVENI-YFNTLTLKNKISLLVPIYYSE--FQRMDTKVIFVYRYLFGNEDQV
WP_069726640.1_ 133 -------LSSEREV-YRLSSIFKINQSVGLSSTFVDDGVNKLKFNAGVGASIIPNENQYS
WP_091973683.1_ 119 --NNLLTVDNVENI-YFNTLTLKNKISLLIPIYYSE--LQRMDTKVIFVYRYLFGNEDQV
WP_108729723.1_ 131 NHSPKLPDAQLKPYKFINTLTLINTVLFLLPLYYSE--TQRAETNISFSYKYIPEYQDNI
WP_151060894.1_ 119 --NNLLKVDNIENV-YFNTLTLKNKISLLVPIYYSE--FQRMDTKVIFVYKYLLGNEDQI
WP_047115988.1_ 133 -------VASEREV-YRLSSIFKINQSIGLSSTFVDDGVNKLKFNAGVGASIIPNEKQYS
WP_157151307.1_ 133 -------IASEREV-YRLSSIFKINQSISLSSTFVDDGVNKLKLNAGVGASIIPNEKQYS
AEM22920.1_B._i 130 -------LTSEREV-YRLSSIFKINQSIGLSSTFVDDGVNKLKFNASVGASIIPNEKQYS
WP_008727984.1_ 99 -------LSSEREV-YRLSSIFKINQSVGLSSTFVDDGINKLKFNAGVGASIIPNENQYS
WP_147731997.1_ 133 -------IASEREV-YRLSSIFKINQNIGLSSTFIDDGVNKLKFNAGFGASIIPNEKQYS
WP_157152113.1_ 139 -------LDLNSTT-KEVNNIFKIQQGIDISALIYDNSINKLKLNTGIDFNIIPNGREYS
WP_147731981.1_ 130 -------LDLNSNT-KEINNIFKLQQSVDVSALIYDNSINKLKFNTGIDFNIIPNGKEYS
NMA56243.1_Trep 135 GFYISEKTKNIEKF-FKQT--------MLINALLYENGFIKNTMIAKINYYNIYENVFSF
WP_162663382.1_ 231 -RKNEIIGDEWSTV-TAPTVVTQVSAGLQLDFFLTDLGFFKSTGTAAFYYHWIPKAAFHY
WP_013759115.1_ 132 ---KYISVSNEADI-YDMAAL-----GIGADIFLVEKGFIKSTLNIESNFSLIPSEKFTY
WP_010694065.1_ 142 -LVNEIKLKGTQKD-LLPKTITEISQGIRFDVYLADLGYLKATTSFSFFTDWIPESRFVN
WP_147625052.1_ 142 -LVNEIKFTGSQKE-LLPKIITEISQGIRFDVYLADLGYLKATTSFSFFTDWIPESRFVN
WP_174896920.1_ 149 LRKLTAANPGGESE-TKKTMVTELSSGLNLDFFLIDTGFFKSTGNAGFFIDWIPKDFFVD
WP_151414990.1_ 142 -LVNEIKVNSTQKE-LLPKTITEISQGIRFDVYLADLGYFKATTSFSFFTDWIPESRFVN
WP_024469938.1_ 156 -RRTAIDTVSGQTE-YKKTVVTEITQGIACSFFLADMGFFKSTGNASLEIDWIPKENFAD
WP_010890368.1_ 193 YRVHWNLKYLVSIPFGELGFKADLGVAGDFKKSSSSIFEIGFD-YNALNFYA-----LTI
WP_012622368.1_ 193 YRVHWNLKHLVYIPFGELGFKVDLGMAGNFKK--SSIFEAGFD-YNALNFYA-----LTI
WP_151074387.1_ 193 YRVHWNLKYLISIPFGELGFKADFGMAGDFKK--SSIFETGFD-YNALNFYA-----LTI
WP_015899315.1_ 195 YRVHWNLKYLVSIPFGELGFKVDLGMAGDFKK--SSIFEAGFD-YNALNFYA-----LTI
WP_012622538.1_ 190 YRVHWNLKYLVYIPFGELGFKVDLGMAGNFKK--SSIFETGFN-YNALNFYA-----LTI
WP_025407119.1_ 189 YRVHWNLKYLVSIPFGELGFKVDLGSADAFKE--STIFVIGFD-YNALNFYA-----LTV
WP_025420078.1_ 189 YRVHWNLKYLVSISFGELGFKADLGMAGAFRK--FTIFETGFD-YNALNFYA-----LTI
WP_120104624.1_ 187 YRVHWNFKYLLSIPFGEIGLKSDFLKTDSLTE---TDLNTRID-YNSLNFYP-----VAL
WP_044052720.1_ 182 YNVHWNFKYLLSLPFGELGFKLDFLKSDSLFE---TDLETRVD-YHALNLYP-----LAL
WP_081581237.1_ 174 YKVYWNFKYLFSMFFGELGFKVDFLKSGSLFE---TDLETRVD-YQALNLYP-----LAL
WP_084790363.1_ 174 YNVHWNFKYLLSLPFGELGFKVDFLKSDSLFE---TDLETRVD-YHALNLYP-----LAF
WP_069726640.1_ 185 YNVSASMPYSFFHYWGELGIMPYIGYSAYFDK-------SEKK-YSIGFKYLYSLMMMPL
WP_091973683.1_ 174 YSVHWNFKYLLSLPFGELGFKADFLKSDSLFE---TDLETRVD-YHALNLYP-----LAF
WP_108729723.1_ 189 YKIHWTAKYLVSIPYSEIGLKSDIIHSGYLGG--NTNFKIGND-YNAINLYS-----L-A
WP_151060894.1_ 174 YNVHWNFKYLFSLPFGELGFKLDFLKSDSLFD---TNLETRID-YHALNLYP-----VAL
WP_047115988.1_ 185 YNVSASIPYSFFHYWGELGIMPYIGYSAYFDK-------SEKK-YSIGFKYLYSLMMMPL
WP_157151307.1_ 185 YNVYASIPYSFFHYLGELGIMPYVGYSAYFDK-------SEKK-YSIGFKYLYSLMMMPL
AEM22920.1_B._i 182 YNVSASMPYSFFHYWGELGIMPYIGYSAYFDK-------SEKK-YSIGFKYLYSLMMMPL
WP_008727984.1_ 151 YNVSASMPYSFFHYWGELGIMPYIGYSAYFDK-------SNKK-YSIGFKYLYSLMMMPL
WP_147731997.1_ 185 YNVSFSMPYSFFHYWGELGIMPYIGYSAYFDK-------SEKK-YSIGFKYLYSIMMMPL
WP_157152113.1_ 191 YGVYASMPYSFFHYWGELGIMPYISYSAYFND-------SKKN-YSIGEKYLCILQMLPL
WP_147731981.1_ 182 YGIYASMPYSFFHYWGELGIMPYISYSAYFND-------SKKN-YSIGEKYLCILQMLPL
NMA56243.1_Trep 186 CDADFSIINTLYLSYFDIAIAGNFFTSFKTDE-------TKNDYL-IGKKYSY----LTK
WP_162663382.1_ 289 YTVSAAIPASFHLYYVDIAFMYSLFHTSTVQY---GKAAPFRR-YAIGQTQSG----MTG
WP_013759115.1_ 183 YDIKLFVPLTVYSDAIELGFLYSSFYVQDLEV---KDVTPDTF-FDVGKSFAT----FTN
WP_010694065.1_ 200 YRLKFDLPVTFKLYYVDVGFVYTFYNTDRINS---KNINAAAD-YQIEKSQEA----MTR
WP_147625052.1_ 200 YRLKFDLPITFKLYYVDIGFVYTFYNTDRINS---KNINAAAD-YQIEKSQEA----MTR
WP_174896920.1_ 208 YRLNANVLGTFYVYHADIALMYSMFKTNQIPI---KGVNAKNN-YEIGKTQES----ITR
WP_151414990.1_ 200 YRLKFDLPVTFKLYYVDVGFVYTFYNTDRINS---KNINAAAD-YQIEKSQEA----MTR
WP_024469938.1_ 214 YRIYATVLASFFLYHADIVFSYSLFNTDYIKL---NKTGVKQN-YEIAKTQEL----LTA


WP_010890368.1_ 247 PKMGQD-SLYFNVVSNFGLEYRLFFLESLKNLASDLFLVLSADIGYGIKEDLLLDKGKFV
WP_012622368.1_ 245 PKMDQD-SLYFNVISNFGLEYRLFFLEALKNIASDLFLVMSSDIGYGMKEDLLLDKGKFL
WP_151074387.1_ 245 PKMDQD-SLYFNVVSNFGLEYRLFFLEALRNLASDLFLVLSLDIGYGMKEDFLLDKGKFL
WP_015899315.1_ 247 PKMDQD-SLYFNVVSNFGLEYRLFFLEALKNLASDLFLVMSSDIGYGVKEDFLLDKGKFL
WP_012622538.1_ 242 PKMDQD-SLYFNVVSNFGLEYRLFFLEVLKNLASDLFLVMSSDIGYGMKEDFLLDKGKFL
WP_025407119.1_ 241 PKIDQS-NLYFNVISNFGFEYRLFFLEALKNVASDLFLVLSSDIGYGMKEDYSLDKGKFL
WP_025420078.1_ 241 PKVDQD-SIYFNFISNFGLEYRLFFLEALKNVASDLFLVLSSDIGYGMKEEYYLDKGKFL
WP_120104624.1_ 238 PFSNNA-NNSFNTIFNFGFEYRIFFLEYLRNVASDLFLVLSFDTGYGLSDN-FLDNGRFL
WP_044052720.1_ 233 PFFDDP-NSIFNAIFNFGFEYRVFFLEYLRNAFSDLFLSISFDMGYGFDDV-FIDNGKFL
WP_081581237.1_ 225 PFLDDP-NSIFNTIFNFGFEYRVFFLEYLRNAFSDLFLSISFDIGYGFDDV-FVDNGKFL
WP_084790363.1_ 225 PLFDDP-NSIFNTIFNFGVEYRLFFLEYLRNAVSDLFLLISFDVGYGFDDV-FIDNGKFL
WP_069726640.1_ 237 SNLEAH-LKNMDFLTFVHFEYKL-FMRFLPSGFNDIYLVAFGNVGYGKYFESSIDKGNLL
WP_091973683.1_ 225 PFFDDP-NSIFNAIFNFGFEYRVFFLEYLRNAFSDLFLSISFDMGYGFDDV-FIDNGKFL
WP_108729723.1_ 240 FFPKNI-STRFNTILNLEVEYRLFFLQHLENIASDLFLSASSNIGHGIMEV-PLSKSQLL
WP_151060894.1_ 225 PFFDDP-NSIFNAIFNFGFEYRVFFLEYLRNSFSDLFLLISFDMGYGFNGV-SIDNGKFL
WP_047115988.1_ 237 SNLEAH-VKNMDFLTFAHFEYKF-FMRFLPSGFNDIYLVAFGNVGYGKYFEESIDKGNLL
WP_157151307.1_ 237 SNLKAH-TKNMDFLTFVHLEYKL-FMRFLPSGFNDIYLVAFGNAGYGKYFENSIDKGNLL
AEM22920.1_B._i 234 SNLEAH-VKNMDFLTFAHFEYKF-FMRFLPSGFNDIYLVAFGNVGYGKYFESSIDKGNLL
WP_008727984.1_ 203 TNLETH-LKNMDFLTFVHFEYKF-FMRFLPSGFNDIYLVAFGNVGYGKYFETSIDKGNLL
WP_147731997.1_ 237 SNLEAH-LKNMDFLTFVHLEYKL-FMRFLPSGFNDIYLVAFGNIGYGKYFENSINKGNLL
WP_157152113.1_ 243 NNAEKY-VEPYSFLTFVHLEYKL-YMRFLPSGWNDLYLVGFGNVGYGKYEYQSINDGEML
WP_147731981.1_ 234 NNAEKY-VEPYSFLTFLHLEYKL-YMRFLPSGWNDLYLVGFGNIGYGKYEYQSINDGEML
NMA56243.1_Trep 234 RMTKED-DIFYNTMMNVETEVRY-YFLRHIDVFSNFFLSAFGNIGSCFET----DDVRLK
WP_162663382.1_ 341 RSTFKT-QALFKDMHLFSAEFRW-YPARITAQTNGFFLSLFADVGFGITER---RKGSLF
WP_013759115.1_ 235 RLNFSSRQTKYNIIQAFEIEPRW-YFIRNINPFSCLYLSVFANGGMGITNS---SDVDWL
WP_010694065.1_ 252 RFSFKD-VQKYSSMHIWGTELRW-YAARTGVNSNGFFISAFADAGFGSNEV---KKLNLI
WP_147625052.1_ 252 RFSFKD-VQKYSSMHIWGTEIRW-YAARTGVNSNGFFISAFADAGFGSNEV---KKINLI
WP_174896920.1_ 260 RNSFKK-LPTYTELHFFKSEIRY-YPDRIKTQNIGFFLSVFADLGLGITKE---NKRRFL
WP_151414990.1_ 252 RFSFKD-VQKYSSMHIWGAEIRW-YAARTGVNSSGFFISAFADAGFGSNEV---KKINLI
WP_024469938.1_ 266 RSSFKP-QTGFTQMHVFGTEMRW-YPARIQTQSNGFFLSLFTDIGFGFTKE---NKLELI
WP_010890368.1_ 306 YILGFGMGYKLFKEVPFVFKVGINQDKKLSFGF-LLSSIIFE-
WP_012622368.1_ 304 YILGFGIGYKLFKEVPFVFKVGINQDKKLSFGF-LLSSIIFE-
WP_151074387.1_ 304 YILGFGMGYKLFKEVPFVFKVGINQDKKLSLGF-LLSSIIFE-
WP_015899315.1_ 306 YILGFGIGYKFFKEVPFVFKVGINQDKKLSFGF-LLSSIIFE-
WP_012622538.1_ 301 YILGFGIGYKLFKEVPFVFKVGINQDKKLSFGF-LLSSIIFE-
WP_025407119.1_ 300 YILGFGVGYKLFKDVPFVFKVGINQDKKLLFGF-LLSSITFE-
WP_025420078.1_ 300 YIVGFGMGYKLFKDVPFVFKVGINQDKKLLFGF-LLSSIIFE-
WP_120104624.1_ 296 YISSLGIGYKLFKEIPFVFKVGINQDKQLLFGF-VVSSIIFDI
WP_044052720.1_ 291 YIVSFGIGYKLFREIPFVFKMGINQDKQFMLGF-IVSSINFDN
WP_081581237.1_ 283 YITSFGIGYKLFREVPFVFRMGINQDKQFMLGF-IVSSINFDN
WP_084790363.1_ 283 YIVSFGIGYKLFREIPFVFKVGINQDKQFMLGF-IVSSINFDN
WP_069726640.1_ 295 YVVGGGIGYNLYGSTPLQLTFGIDNNKSLVMNL-IISTIVF--
WP_091973683.1_ 283 YIVSFGIGYKFFREIPFVFKVGINQDKQFMLGF-IVSSINFDN
WP_108729723.1_ 298 YIINFGIGYKFNREIPFTLRIGFNQNMQLLIGF-FSSPITFSN
WP_151060894.1_ 283 YISSFGIGYKLFREIPFVFKLGINQDKQFMLGF-IVSSINFDN
WP_047115988.1_ 295 YVVGGGIGYNLYGTTPLQLTFGVDNNNSLVMNL-IISTIVF--
WP_157151307.1_ 295 YVVGGGIGYNLYGTTPLQLTFGVDNNNSLVMNL-IISTIVF--
AEM22920.1_B._i 292 YVVGGGIGYNLYGTTPLQLTFGVDNNNSLVMNL-IISAIMF--
WP_008727984.1_ 261 YAVGGGIGYNLYGSTPLQLTFGVDNNKSLVMNL-IISTIVF--
WP_147731997.1_ 295 YVVGGGIGYNLYGSTPLQLTFGVDNNNSLVMNL-IISTIVF--
WP_157152113.1_ 301 YMVGGGIGFNLFSTAPLQVTLATDNNKSLMINV-VISAIGF--
WP_147731981.1_ 292 YMVGGGIGFNLFSTAPLQITLATDNNKSLIINV-VISAISF--
NMA56243.1_Trep 288 YQYGIGVGYSLFDTVPFTFQIGLNEDNKLCMYVSVVSALSHMP
WP_162663382.1_ 396 YEAGGGLGYNLYDSVPLTFQVGFNQKMQPVFYFSVVSRLSHRL
WP_013759115.1_ 291 YQVGGGIGYTLFDAVPFEFQIGYDNKAGMFLYIGVVSRIMHKP
WP_010694065.1_ 307 AEFGLGAGYTLFDNVPFTFQAGFNQDFKPVFFLGIVSRIIQGV
WP_147625052.1_ 307 AECGLGAGYTLFDNVPFTFQAGLNQDFKPIFFLGVVSRIIQGV
WP_174896920.1_ 315 AEFGGGIGYTLFDSVPFTFQAGVNQDMNPVFFLGVVSSLSHRP
WP_151414990.1_ 307 AECGLGAGYTLFDNVPFTFQAGFNQDFKPIFFLGVVSRIIQGV
WP_024469938.1_ 321 AEYGLGGGYTLLDSVPFTFQVGLNQRFQPVFYLGVVSRLTHLP

Appendix 7- SignalP 5.0 analysis of BBJ25



Appendix 8- Phylogenetic analysis of BB405



Protein Expression and Purification 163 (2019) 105447

Contents lists available at ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

Expression, purification and metal utilization of recombinant SodA from *Borrelia burgdorferi*

G. Brown[a], A.H. Broxham[a], S.E. Cherrington[a], D.C. Thomas[a], A. Dyer[a,1], L. Stejskal[a,2], R.J. Bingham[a,*]

[a] Department of Biological and Geographical Sciences, School of Applied Sciences, University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, UK

ARTICLE INFO

Keywords:
Borrelia
Superoxide dismutase
Manganese
E. coli

ABSTRACT

Borrelia are microaerophilic spirochetes capable of causing multisystemic diseases such as Lyme disease and Relapsing Fever. The ubiquitous Fe/Mn-dependent superoxide dismutase (SOD) provides essential protection from oxidative damage by the superoxide anion. *Borrelia* possess a single SOD enzyme - SodA that is essential for virulence, providing protection against host-derived reactive oxygen species (ROS). Here we present a method for recombinant expression and purification of *Borrelia burgdorferi* SodA in *E. coli*. Metal exchange or insertion into the Fe/Mn-SOD is inhibited in the folded state. We therefore present a method whereby the recombinant *Borrelia* SodA binds to Mn under denaturing conditions and is subsequently refolded by a reduction in denaturant. SodA purified by metal affinity chromatography and size exclusion chromatography reveals a single band on SDS-PAGE. Protein folding is confirmed by circular dichroism. A coupled enzyme assay demonstrates SOD activity in the presence of Mn, but not Fe. The apparent molecular weight determined by size exclusion corresponds to a dimer of SodA; a homology model of dimeric SodA is presented revealing a surface Cys distal to the dimer interface. The method presented of acquiring a target metal under denaturing conditions may be applicable to the refolding of other metal-binding proteins.

1. Introduction

Superoxide ($O_2^{\bullet-}$), formed by the reduction of O_2, is a reactive free radical capable of generating numerous other free radicals and reactive oxygen species that may go on to damage cellular components [1]. Superoxide dismutase (SOD) is highly conserved throughout all domains of life and catalyzes the dismutation of $O_2^{\bullet-}$ to hydrogen peroxide (H_2O_2) [2]. The most ancient form, thought to have evolved over 2 Gyr ago, is the Fe/Mn-SOD [3]. Found in mitochondria (SOD2) and the cytoplasm of bacteria, the Fe/Mn-SOD exists as either a homodimer or homotetramer with one metal-binding site per chain [2].

Borrelia burgdorferi is a pathogenic spirochete with several characteristics that are dissimilar to other prokaryotes, including an unusual genome structure [4], extremely limited metabolic capacity [5], and the presence of polyunsaturated membrane lipids [6]. *Borrelia* is also highly unusual in its ability to grow without needing a source of Fe [7], instead its cellular metabolism has evolved to utilize Mn, which is imported to significant concentrations by the specific transporter BmtA [8]. No iron-specific enzymes have been identified in *Borrelia*, however Fe-binding proteins have been identified that play a role in Cu/Fe detoxification [9]. The lack of cellular Fe gives some protection against oxidative damage to DNA by Fenton chemistry. The major targets of host-derived reactive oxygen species are the polyunsaturated membrane lipids [6].

While a small number of SOD-enzymes have been identified that remain catalytically active with either Mn or Fe bound at the active site (cambialistic SODs), the majority have evolved to function only in the presence of their target metal ion and are named correspondingly as either Fe-SOD or Mn-SOD. Fe/Mn-SODs bind to either metal with the same trigonal-bipyramidal geometry, and have high sequence and structural homology [10]. Metal coordinating residues (His26, His81, Asp167 and His171 based on *E. coli* numbering) are identical between the Mn and Fe-SODS and mis-incorporation can occur. Metal selectivity is thought to be at least partly driven by bioavailability. Predicting metal specificity from protein sequence alone is challenging [11].

The genome of *Borrelia burgdorferi* contains a single SOD (*sodA*, bb0153), which was demonstrated to be essential for infectivity in a mouse model [12]. The inactivation of *sodA* resulted in a phenotype

* Corresponding author.
E-mail address: r.j.bingham@hud.ac.uk (R.J. Bingham).
[1] Present address: School of Immunology & Microbial Sciences, Faculty of Life Sciences & Medicine, King's College London, Guy's Hospital, London, SE1 9RT.UK.
[2] Present address: Institute of Structural and Molecular Biology, Birkbeck College, London, WC1E 7HX, UK.

https://doi.org/10.1016/j.pep.2019.105447
Received 2 May 2019; Received in revised form 13 June 2019; Accepted 30 June 2019
Available online 01 July 2019
1046-5928/

more vulnerable to challenge by activated macrophages and neutrophils, while *in vitro* growth was unaltered. Although early work suggested this might be an Fe-SOD [13], it was subsequently demonstrated to utilize Mn [14]. An elegant series of *in vivo* experiments demonstrated that *Borrelia* accumulates high levels of cellular Mn, and that this is required to activate the SodA enzyme. In contrast, *Borrelia* SodA expressed in the iron-rich environment of yeast mitochondria was inactive [14].

Previous studies of the metal utilization of *Borrelia* SodA have not used pure SodA, rather assays based on non-denaturing electrophoresis gels with *Borrelia* whole cell lysate [13,15] or purified cell extracts [14]. In addition, no studies to date have conclusively determined whether this enzyme is cambialistic, whereby it could use both manganese and iron as its co-factor. Therefore, the objective of this work was to produce and purify recombinant *Borrelia* SodA allowing *in vitro* investigation of enzyme activity in the presence of Mn and Fe. Here we report a method to produce sufficient quantity of SodA to allow structural and biochemical studies. We describe a protocol that utilizes addition of Mn and Fe salts to recombinant protein in the denatured state prior to refolding and purification, thereby preventing contamination of downstream processes with these reactive transition metal ions. A homology model of dimeric SodA is presented.

2. Materials and methods

2.1. Plasmid construction and gene expression

The *bb0153* gene was codon optimized for expression in *E. coli*, synthetically produced (Eurofins Genomics) and sub-cloned into pET-47b (+) (Novagen®). The resulting construct was transformed into T7 Express Competent *E. coli* (New England BioLabs). Transformed cells were grown in LB media at 37 °C in an orbital incubator and were induced at an OD_{600} of 0.6 by addition of isopropyl-β-D-1-thiogalactopyranoside (IPTG, Sigma-Aldrich) to a working concentration of 1 mM. After 4-h, induced cells were harvested by centrifugation and lysed on ice by pulsed sonication in 0.3 M NaCl, 50 mM Tris-HCl, pH 8.0.

2.2. Purification of inclusion bodies

The soluble fraction was removed by centrifugation (20,000 x g for 30 min at 4 °C), the insoluble material was resuspended in 0.3 M NaCl, 50 mM Tris-HCl, pH 7.0 and briefly pulse sonicated at an amplitude of 40% for a total length of 30 s. DNase1 was added at 1 μg/ml and the solution incubated at 4 °C for 4 h. Following incubation, samples were centrifuged (20,000 x g for 30 min at 4 °C) and weighed prior to resuspension in Wash Buffer 1 (Table 1) at 10 ml per gram of pellet. The suspension was briefly sonicated as previously described. Following this, the suspension was subjected to low speed centrifugation (~8000 x g) until a firm pellet was obtained. This process was repeated twice with inclusion Wash Buffer 1, twice with Wash Buffer 2, and twice with Wash Buffer 3 (Table 1). The inclusion bodies were then solubilized in Denaturing Buffer overnight at room temperature with shaking.

2.3. Purification of metal-free protein

Divalent metal ions present in the re-suspended bacterial pellet were removed by utilizing unchelated nitrilotriacetic acid (NTA). A 5 ml Ni-NTA column (His-Trap HP, GE Healthcare) was stripped of Ni^{2+} by addition of 100 mM EDTA. After equilibration of the column with Denaturing Buffer (8 M urea, 0.3 M NaCl, 50 mM Tris-HCl, pH 8.0), the resuspended pellet (volume 20 ml) was passed through the column twice to bind any free metal ions at a flow rate of 0.5 ml/min using an AKTA Prime FPLC system.

Table 1
Buffers for inclusion-body wash procedure.

Buffer name	Constituents
Wash Buffer 1	0.3 M NaCl, 50 mM Tris-HCl, 1 mM EDTA, 10 mM DTT, 5% Triton X-100, pH 8.0
Wash Buffer 2	0.3 M NaCl, 50 mM Tris-HCl, 1% Triton X-100, pH 8.0
Wash Buffer 3	0.3 M NaCl, 50 mM Tris-HCl, pH 8.0
Denaturing Buffer	8 M urea, 0.3 M NaCl, 50 mM Tris-HCl, pH 8.0

2.4. Addition of metal co-factors

Following the removal of metal ions by NTA the pH of the denatured protein solution was decreased to pH 4.0 by slow addition of dilute HCl. Acidification of the solution was required to both increase the solubility and minimize the oxidation of Fe^{2+}. $MnCl_2$ or $FeCl_2$ were added to samples of recombinant SodA to a final concentration of 1 mM. Recombinant protein samples were incubated at 4°C overnight before the pH was increased back to pH 8.0 by slow addition of dilute NaOH.

2.5. On-column refolding and purification

Recombinant protein samples (with addition of either $MnCl_2$ and $FeCl_2$ or control) were subjected to centrifugation at 16,000 g for 30 min prior to refolding/purification using a His-Trap HP column (GE Healthcare). The column was equilibrated with 10 column volumes (CV) of denaturing buffer (8 M urea, 50 mM Tris-HCl, 0.3 M NaCl, pH 8.0) and recombinant protein loaded at 0.5 ml/min. Protein folding was achieved by washing with a linear gradient from 8 M to 1 M urea over 2 h in 50 mM Tris-HCl, 0.3 M NaCl, pH 8.0 (flow rate 0.5 ml/min). The column was then washed with 10 column volumes of 1 M urea, 50 mM Tris-HCl, 0.3 M NaCl, 50 mM imidazole, pH 8.0. Recombinant protein was eluted in 50 mM Tris-HCl, 0.3 M NaCl, 0.3 M imidazole at pH 8.0.

Size Exclusion Chromatography was carried out using a Superdex 75 10/300 GL column (GE Healthcare) in 50 mM Tris-HCl, 0.3 M NaCl at pH 8.0. The exclusion volume/void volume was determined using blue dextran (molecular mass of ~2,000 kDa) which eluted at 8.7 ml. Apparent molecular weights were determined by comparison with a calibration curve prepared using known protein standards (Sigma-Aldrich).

Mn-SodA, Fe-SodA and apo-SodA were concentrated using a 10NMWL Amicon Ultra-15 centrifugal filter unit (Merck Millipore), centrifuged at 16,000 x g for 20 min at room temperature and protein concentrations ascertained by UV spectroscopy (extinction coefficient 38390 $M^{-1}cm^{-1}$) and Bradford analysis [16].

2.6. Circular dichroism

Circular dichroism experiments were carried out on a Jasco J-810 spectropolarimeter. Air and buffer blank background data were acquired and experimental measurements taken over a wavelength range of 185–260 nm. Experimental data was obtained at a protein concentration of 0.1 mg/ml over five individual scans and averaged. Data were analyzed using DichroWeb [17,18] using a mean residue weight of 111.8 Da, corresponding to the mean residue weight of SodA. Algorithms used for deconvolution included CDSSTR, CONTIN and SELCON3 with the SP175 reference set.

2.7. SOD enzyme activity assay

Superoxide dismutase activity was measured by an indirect enzyme assay (SOD Assay Kit, 19160, Sigma-Aldrich) and carried out according to the manufacturer's instructions with additional negative controls to account for possible interference from the utilized buffers. The assay was used to compare activity rates of recombinant Mn-SOD, Fe-SOD and apo-SOD and was run in quadruplicates with a protein concentration of 100 μg/ml in 96 well plates in a total well volume of 240 μl. Formazan dye production was determined by measuring absorbance for

20 min at 440 nm using a BMG Labtech SPECTROstar Nano microplate-reader. Statistical analysis by ANOVA and a multiple comparison of sample means by Tukey's test (TukeyHSD) was conducted in R [19].

2.8. Homology modelling

A homology model of *B. burgdorferi* SodA (UniProtKB accession code O30563) was generated using Phyre2 using intensive mode [20]. The highest ranked template was the Mn-dependent SOD from *Thermus thermophilus* with 50% sequence identity over 201 residues. The homodimer was generated by superimposition of the Phyre2 model onto the dimers of Mn-SOD from *E. coli* and *T. thermophilus* (PDB accession codes 1D5N and 3MDS). Molecular graphics and analyses were performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco [21].

3. Results and discussion

3.1. Cloning, expression and initial purification of SodA inclusion bodies

The coding sequence for gene bb0153 was synthesized and incorporated into the MCS of pET-47b (+) for expression of *B. burgdorferi* SodA with an N-terminal 6xHis-tag. Induction of T7 Express cells with IPTG resulted in high levels of SodA expression in inclusion bodies (Fig. 1A). No detectable recombinant protein was found in the soluble fraction despite attempts to optimize expression conditions such as: induction temperature, IPTG concentration and the addition of ethanol [22] (results not shown). A series of pellet-wash steps were employed (Methods section 2.2) to isolate inclusion bodies and remove background contamination (Fig. 1B).

3.2. Metal insertion

A range of factors contributed to our experimental approach to metal insertion under denaturing conditions. Buffers containing free transition metal ions may effectively catalyze the dismutation of oxygen radicals in the absence of enzyme. In addition, because of the neutral pH and oxidizing conditions, the addition of even small concentrations of Fe^{2+} metal salts resulted in insoluble oxidation products that may interfere with subsequent colorimetric SOD assays. Furthermore, *in vitro* studies using *E. coli* Mn-SOD showed that the direct reconstitution of metal ions by dialysis is not always possible [23]. *In vivo* studies have demonstrated that the folded form of apo-SOD2 (the mitochondrial homologue of *Borrelia* SodA with 40% sequence identity) could not acquire Mn and that protein unfolding was required during mitochondrial import for metal binding and activation [24]. For these reasons, it was decided that Fe and Mn ions would be added to *Borrelia* SodA prior to refolding. The subsequent purification steps employed (Ni-NTA and size exclusion chromatography) also served to remove unbound metal ions and prevent interference with subsequent colorimetric SOD assays.

To generate denatured apo-SodA, inclusion bodies were washed in EDTA buffer, solubilized in 8 M urea, and any residual metal ions were removed by passing the solution through stripped NTA resin. Metal salts (or water control) were then added to protein solutions in denaturing conditions to generate samples of apo-SodA, Fe-SodA and Mn-SodA. Centrifugation was required to remove insoluble oxidation products. All three samples (apo-SodA, Fe-SodA and Mn-SodA) were then refolded and purified independently.

3.3. Protein folding and final purification

Denatured protein samples were bound to Ni-NTA resin and allowed to refold by a slow reduction in urea concentration over 2 h. Protein was eluted by the addition of imidazole resulting a single band on SDS-PAGE (Fig. 1C). Size exclusion chromatography was utilized to remove misfolded or aggregated protein (Fig. 1D). Three peaks were detected, the first of which corresponded with the void volume (> 2000 kDa) and is most likely mis-folded/aggregated protein. The elution volumes of the two subsequent peaks correspond almost exactly to dimeric (47 kDa) and monomeric (23.5 kDa) SodA, and both peaks were confirmed to contain SOD activity by enzyme assay. The minor peak corresponding to monomeric SodA is most likely an artifact of recombinant expression and is unlikely to be physiologically relevant. Bacterial MnSod enzymes are most frequently dimeric [25], therefore the major peak detected from size exclusion that corresponds to dimeric SodA was selected for further study. The yield of refolded proteins were similar for Mn and Fe-SodA (8 mg purified protein per litre of *E. coli* cell culture). The yield of apo-SodA was particularly low at 0.9 mg/l.

Protein folding was confirmed by circular dichroism (CD) spectroscopy (Fig. 2, Table 2). Apo-SodA, Fe-SodA and Mn-SodA were confirmed to have CD spectra with high helix content (~60%) consistent with the long alpha-hairpin N-terminal domain and the Fe, Mn SOD C-terminal domain (alpha-beta (2)-alpha-beta-alpha (2)). Values are consistent with the Mn-Sod from *Thermus thermophilus* [26], the closest homologue to *Borrelia* SodA with 51% sequence identity over 200 residues.

3.4. Mn activates borrelia SodA and enhances protein folding

SOD activity was determined using an indirect colorimetric assay. Superoxide anions, produced by the oxidation of xanthine by xanthine oxidase, catalyze the reduction of WST-1 (2-(4-Iodophenyl)- 3-(4-nitrophenyl)-5-(2,4-disulfophenyl)-2H tetrazolium monosodium salt) to produce a formazan dye that absorbs light at 440 nm. SOD activity can be determined by measuring the inhibition of this reaction resulting from the removal of superoxide anions.

Apo-SodA and Fe-SodA failed to inhibit the enzyme activity colorimetric assay, giving average inhibition values close to zero (−1.1% ± 5.2 and 1.5% ± 6.3) and were not significantly different to each other ($P > 0.05$) (Fig. 3). This is in marked contrast to the Mn-SodA, which inhibited the reaction by 82.9% ± 5.8. These data confirm that *Borrelia* SodA is Mn-dependent and not cambialistic.

The results demonstrate that *Borrelia* SodA was able to chelate Mn in the presence of 8 M urea, and that this Mn then remained bound to the enzyme during the subsequent refolding and purification steps resulting in active SodA. It is possible that some residual native structure exists that allows metal binding, even in the presence of 8 M urea. The refolding of apo-SodA was particularly problematic and gave a very low yield (0.9 mg/L). Based on the improved yield of refolded protein in the presence of either Fe or Mn ions (8 mg/L), it appears that metal binding improved protein-folding. These results are in agreement with the literature. In live *Borrelia*, levels of SodA are significantly reduced in a *bmtA* mutant unable to import Mn [15]. When *Borrelia* SodA is heterogously expressed in the mitochondria or cytoplasm of yeast cells, immunoblots revealed that protein is only detected when cells were treated with 1 mM Mn [14]. This was demonstrated to be a stabilization of polypeptide and not a transcriptional effect. Therefore, we propose that Mn-binding by SodA promotes protein folding.

3.5. Homology model of dimeric Mn-SodA

The elution volume from size exclusion chromatography indicates that SodA primarily exists as a homodimer (Fig. 1D). A homology model of the SodA dimer was generated based on the tetrameric Mn-SOD from *T. thermophilus* and the dimeric Mn-SOD from *E. coli* (Fig. 4). The primary sequence of *Borrelia* SodA contains a single Cys residue (Cys136, Fig. 4). Because of the reducing environment in the cytoplasm, most intracellular Cys residues would not form a disulfide bond, however human SOD1 has been shown to contain an intrasubunit disulfide [29], therefore the possible involvement of a disulfide in the dimer interface in *Borrelia* SodA was important to consider. The homology model of the

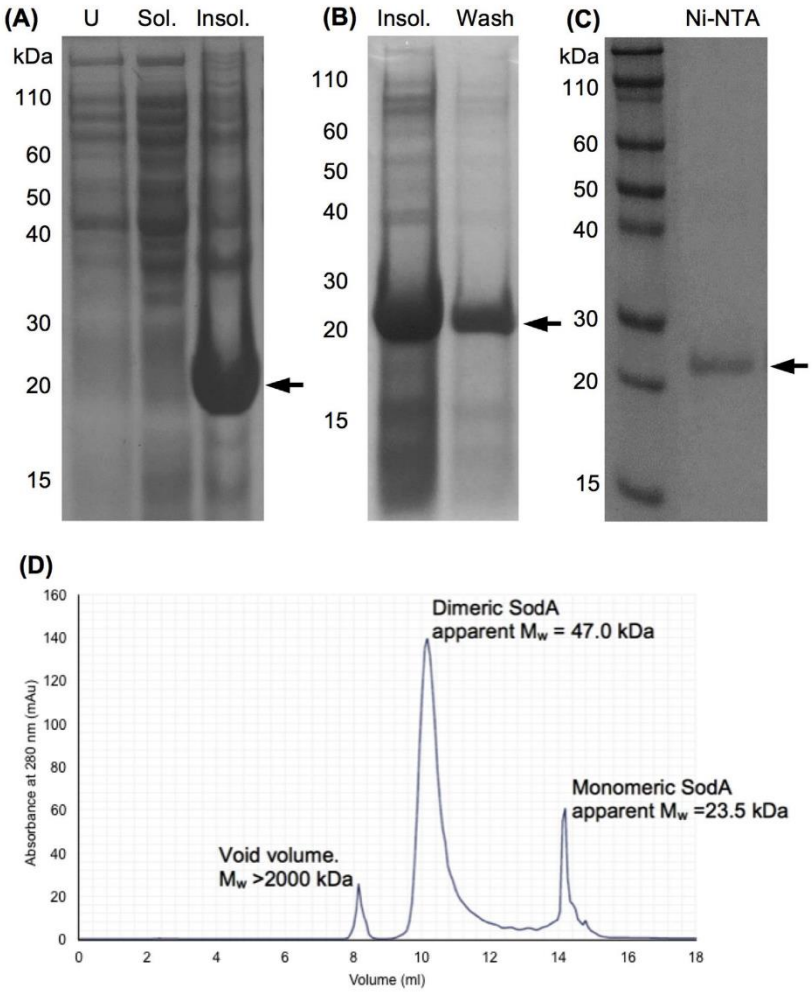


Fig. 1. Purification and refolding of recombinant *Borrelia* SodA. (A) SDS-PAGE analysis of uninduced whole cell lysate (U), followed by post-induction soluble (Sol.) and insoluble (Insol.) cell lysate fractions. Recombinant *Borrelia* SodA is indicated by black arrows. (B) The insoluble pellet (Insol.) was washed six times as described in Methods (2.2) to generate a sample of inclusion bodies (Wash). (C) The inclusion bodies were then solubilized in 8 M urea and purified by immobilized metal affinity chromatography (Ni-NTA). (D) Size exclusion chromatography of *Borrelia* Mn-SodA. Fractions from both peaks were confirmed as SodA by enzyme assay.

SodA dimer revealed that the two Cys136 residues are separated by ~43 Å and are therefore unlikely to be involved in an intrasubunit disulfide without considerable rearrangement of the dimer interface. This rearrangement is decidedly unlikely as residues at the dimer interface are highly conserved between the *E. coli, T. thermophilus* and *Borrelia* proteins with only a single amino acid difference observed, Tyr34/Phe34 (Supplementary Table 1).

4. Conclusion

Borrelia SodA recombinantly expressed in *E. coli* can be refolded in the presence of Mn. This therefore allows significant amounts of protein to be produced for structural biology and biochemical studies. This Mn facilitates protein folding and remains bound to the protein throughout the subsequent purification steps. As demonstrated by indirect colorimetric assay, and consistent with known literature, *Borrelia* SodA requires Mn for activity. Finally, by studying a homology model of *Borrelia* SodA a single Cys residue at the protein surface is revealed; this may stymie future attempts at crystallography unless it is chemically blocked, effectively reduced or removed by mutagenesis. .

Acknowledgements

Circular dichroism was collected at the University of York with the kind assistance of Dr. Andrew Leech. Funding was provided by Departmental PhD Studentships to GB, AHB and AD.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pep.2019.105447.

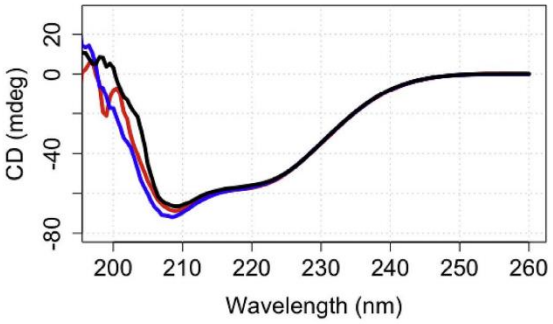


Fig. 2. Circular dichroism spectra of Apo-SodA (black), Fe-SodA (red) and Mn-SodA (blue) at 20°C. Experiments were carried out on a Jasco J-810 spectropolarimeter. Experimental data was obtained over five individual scans and averaged.

Table 2
Deconvolution of CD data for Apo-SodA, Fe-SodA and Mn-SodA. Proportion (%) of secondary structure types α-helix, β-strand, β-turn or unordered was determined using the CDSSTR method by DichroWeb [18,27] using the SP175 dataset [28].

	α-helix	β-strand	β-turn	unordered
Apo-SodA	58%	13%	13%	17%
Fe-SodA	62%	14%	9%	15%
Mn-SodA	59%	15%	9%	17%

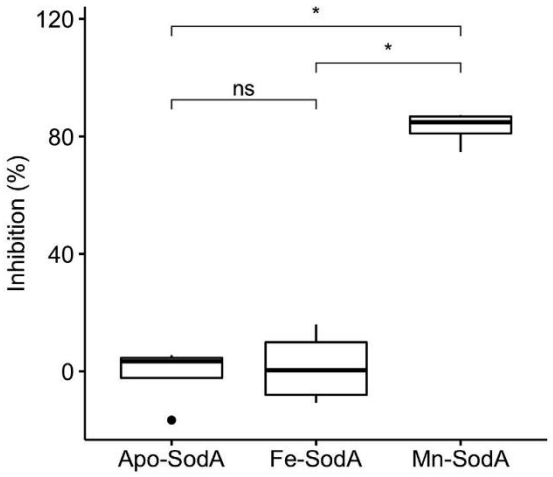


Fig. 3. Activity of *B. burgdorferi* SodA refolded from 8 M urea in the absence of metal ions (apo-SodA), or after addition of Fe/Mn. SOD activity was measured by an indirect enzyme assay based on the removal of superoxide anions from solution, preventing the reduction of WST-1 to formazan dye. Whiskers indicate the maximum and minimum of data, central bands indicate median values. Statistical analysis by ANOVA and a multiple comparison of sample means by Tukey's test revealed the enzyme activity of Mn-SodA to be significantly different to both Apo and Fe-SodA ($P < 0.05$).

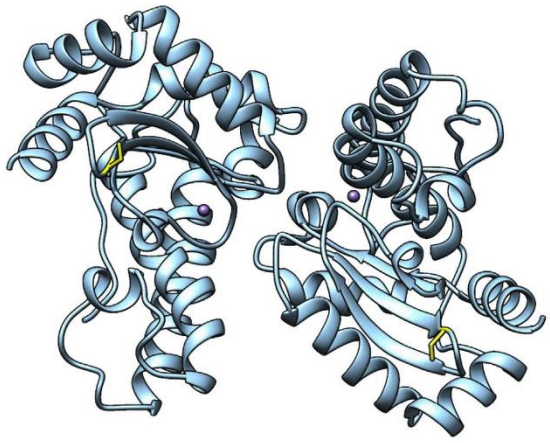


Fig. 4. Homology model of dimeric *Borrelia* SodA based on *T. thermophilus* and *E. coli* Mn-SODs generated using Phyre2. 203 residues were modelled at > 90% accuracy. Mn shown as purple spheres. Cys136 shown using yellow sticks representation. The distance between the two Cys residues is ~43 Å.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

GB, DCT, AHB, & SEC performed the research, analyzed the data and helped to draft the manuscript; AD, AHB, GB, LS and RJB conceived and designed the work and wrote the manuscript.

All authors contributed to and approved the final manuscript.

References

[1] L. Benov, How superoxide radical damages the cell, Protoplasma 217 (2001) 33–36.
[2] J.J. Perry, D.S. Shin, E.D. Getzoff, J.A. Tainer, The structural biochemistry of the superoxide dismutases, Biochim. Biophys. Acta 1804 (2010) 245–262.
[3] M.W. Smith, R.F. Doolittle, A comparison of evolutionary rates of the two major kinds of superoxide dismutase, J. Mol. Evol. 34 (1992) 175–184.
[4] P.E. Stewart, R. Byram, D. Grimm, K. Tilly, P.A. Rosa, The plasmids of *Borrelia* burgdorferi: essential genetic elements of a pathogen, Plasmid 53 (2005) 1–13.
[5] C.M. Fraser, S. Casjens, W.M. Huang, G.G. Sutton, R. Clayton, R. Lathigra, O. White, K.A. Ketchum, R. Dodson, E.K. Hickey, M. Gwinn, B. Dougherty, J.F. Tomb, R.D. Fleischmann, D. Richardson, J. Peterson, A.R. Kerlavage, J. Quackenbush, S. Salzberg, M. Hanson, R. van Vugt, N. Palmer, M.D. Adams, J. Gocayne, J. Weidman, T. Utterback, L. Watthey, L. McDonald, P. Artiach, C. Bowman, S. Garland, C. Fuji, M.D. Cotton, K. Horst, K. Roberts, B. Hatch, H.O. Smith, J.C. Venter, Genomic sequence of a Lyme disease spirochaete, *Borrelia* burgdorferi, Nature 390 (1997) 580–586.
[6] J.A. Boylan, K.A. Lawrence, J.S. Downey, F.C. Gherardini, *Borrelia* burgdorferi membranes are the primary targets of reactive oxygen species, Mol. Microbiol. 68 (2008) 786–799.
[7] J.E. Posey, F.C. Gherardini, Lack of a role for iron in the Lyme disease pathogen, Science 288 (2000) 1651–1653.
[8] Z. Ouyang, M. He, T. Oman, X.F. Yang, M.V. Norgard, A manganese transporter, BB0219 (BmtA), is required for virulence by the Lyme disease spirochete, *Borrelia* burgdorferi, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 3449–3454.
[9] P. Wang, A. Lutton, J. Olesik, H. Vali, X. Li, A novel iron- and copper-binding protein in the Lyme disease spirochaete, Mol. Microbiol. 86 (2012) 1441–1451.
[10] T.A. Jackson, C.T. Gutman, J. Maliekal, A.F. Miller, T.C. Brunold, Geometric and electronic structures of manganese-substituted iron superoxide dismutase, Inorg. Chem. 52 (2013) 3356–3367.
[11] R. Wintjens, D. Gilis, M. Rooman, Mn/Fe superoxide dismutase interaction fingerprints and prediction of oligomerization and metal cofactor from sequence, Proteins 70 (2008) 1564–1577.
[12] M.D. Esteve-Gassent, N.L. Elliott, J. Seshu, sodA is essential for virulence of *Borrelia* burgdorferi in the murine model of Lyme disease, Mol. Microbiol. 71 (2009) 594–612.
[13] C.A. Whitehouse, L.R. Williams, F.E. Austin, Identification of superoxide dismutase activity in *Borrelia* burgdorferi, Infect. Immun. 65 (1997) 4865–4868.
[14] J.D. Aguirre, H.M. Clark, M. McIlvin, C. Vazquez, S.L. Palmere, D.J. Grab, J. Seshu,

P.J. Hart, M. Saito, V.C. Culotta, A manganese-rich environment supports superoxide dismutase activity in a Lyme disease pathogen, *Borrelia* burgdorferi, J. Biol. Chem. 288 (2013) 8468–8478.
[15] B. Troxell, H. Xu, X.F. Yang, *Borrelia* burgdorferi, a pathogen that lacks iron, encodes manganese-dependent superoxide dismutase essential for resistance to streptonigrin, J. Biol. Chem. 287 (2012) 19284–19293.
[16] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. 72 (1976) 248–254.
[17] L. Whitmore, B.A. Wallace, DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data, Nucleic Acids Res. 32 (2004) W668–W673.
[18] L. Whitmore, B.A. Wallace, Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases, Biopolymers 89 (2008) 392–400.
[19] R.C. Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing Vienna, Austria, 2015http://wwwR-projectorg/.
[20] L.A. Kelley, S. Mezulis, C.M. Yates, M.N. Wass, M.J. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis, Nat. Protoc. 10 (2015) 845–858.
[21] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera–a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.
[22] G. Chhetri, P. Kalita, T. Tripathi, An efficient protocol to enhance recombinant protein expression using ethanol in Escherichia coli, MethodsX 2 (2015) 385–391.
[23] C.T. Privalle, W.F. Beyer Jr., I. Fridovich, Anaerobic induction of ProMn-superoxide dismutase in Escherichia coli, J. Biol. Chem. 264 (1989) 2758–2763.
[24] E. Luk, M. Yang, L.T. Jensen, Y. Bourbonnais, V.C. Culotta, Manganese activation of superoxide dismutase 2 in the mitochondria of Saccharomyces cerevisiae, J. Biol. Chem. 280 (2005) 22715–22720.
[25] R. Wintjens, C. Noel, A.C. May, D. Gerbod, F. Dufernez, M. Capron, E. Viscogliosi, M. Rooman, Specificity and phenetic relationships of iron- and manganese-containing superoxide dismutases on the basis of structure and sequence comparisons, J. Biol. Chem. 279 (2004) 9248–9254.
[26] S. Sato, K. Nakazawa, Purification and properties of superoxide dismutase from Thermus thermophilus HB8, J. Biochem. 83 (1978) 1165–1171.
[27] N. Sreerama, R.W. Woody, Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set, Anal. Biochem. 287 (2000) 252–260.
[28] J.G. Lees, A.J. Miles, F. Wien, B.A. Wallace, A reference database for circular dichroism spectroscopy covering fold and secondary structure space, Bioinformatics 22 (2006) 1955–1962.
[29] K. Sea, S.H. Sohn, A. Durazo, Y. Sheng, B.F. Shaw, X. Cao, A.B. Taylor, L.J. Whitson, S.P. Holloway, P.J. Hart, D.E. Cabelli, E.B. Gralla, J.S. Valentine, Insights into the role of the unusual disulfide bond in copper-zinc superoxide dismutase, J. Biol. Chem. 290 (2015) 2405–2418.