The University Of Sheffield.

Structural studies into the mechanisms of pathogenicity of *Burkholderia*

By:

Hayley Owen

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

University of Sheffield

Department of Molecular Biology and Biotechnology

December 2016

Abstract

The aim of this thesis was to structurally characterise proteins potentially involved in the virulence mechanisms utilised by *B. pseudomallei*, the causative agent of the human disease melioidosis. The first project aims to continue investigations into an operon of four proteins of largely unknown structure and function, BPSS0211-BPSS0214. Structural studies on BPSS0211 provide evidence for conformational changes of its helical DUF1843 domain and a novel quaternary structure as part of its function. Structural studies of the DUF1842 domain of BPSS0212 reveal a lipocalin-like β-barrel fold, suggesting it may act as a ligand-binding domain. Biochemical studies indicated that BPSS0212 and BPSS0213 interact to form a complex of approximately 350kDa molecular weight, via the C-terminal DUF1843 domains. The structure of this complex was solved and revealed that eight subunits each of BPSS0212 and BPSS0213 associate into the presumed biologically active molecule.

The second project involves investigations into the type VI secretion system, which is known to be involved in virulence and is a complex composed of the products of at least 13 genes. Structural studies focussed on a putative predicted baseplate protein, TssA. Sequence comparisons have indicated across multiple species two distinct families of TssA, TssA-1 and TssA-2. Two domains, an N- and C-terminal domain, were identified from *B. cenocepacia* TssA-1. Constructs of the two TssA domains were produced, and the structures solved separately. The structure of the N-terminal domain has two bundles of helices forming an L-shaped molecule, and the combination of sequence and structure analysis shows that the TssA-2 N-terminal domain must be a tandem repeat of a structure resembling the first section of the N-terminal domain of TssA-1. The C-terminal domain forms a ring of 32 monomers with D16 symmetry, which has a distinctly different structure to the previously solved structure of the C-terminal domain of the TssA-2 family.

Acknowledgements

Firstly I would like to thank Prof. David Rice for his guidance, support and encouragement throughout my PhD. Secondly, thank you to Dr. Patrick Baker, whose undergraduate project and lectures sparked my interest into the world of crystallography, and who always has time to thoroughly answer my many questions. Thank you to Dr. John Rafferty, with whom I gained essential undergraduate laboratory experience and my first visit to the D11 crystallography lab, and who, along with Prof. Pete Artymiuk, offered advice and discussion during my project. Thank you to Dr. Matthew Day for showing me the ropes and starting me on the operon project. And thank you to Dr. Mark Thomas and Ruyue Sun for the opportunity to collaborate on the type VI secretion system project. I would also like to thank Dr. Svetlana Sedelnikova for her expertise in all things protein purification and preparation. Thank you to Dr. Svetomir Tzokov for his help with the EM data collection and analysis, as well as Jason Wilson for collecting the initial images. I would also like to thank Prof. Sheila Nathan and her students at UKM for providing the knockout mutant and performing the follow-up experiments.

Thank you to Fiona who has kept the lab running wonderfully. Thank you to Jason Wilson, Dr. Claudine Bisson, Dr. Sophie Bliss, Dr. George Mobbs, Sam Dix and Adli Aziz for the time spent together in and out of work, for the laughter and the crystallography in-jokes, as well as celebrating in the research successes and commiserating with me when it all didn't quite go to plan; for all the scientific discussions and for making it through the Diamond 24 hour trips together. And thank you to anyone I may have missed and to everyone in the crystallography lab who has created a happy and encouraging research environment.

I would like to thank all the other friends I have made during my time in Sheffield, and to those friends at home. And Gus for his random interjections of positivity during writing! To my wonderful friends Alice,

Lisa, Sarah, Rosa, Laura, Nicola and Victoria who provided comfort, smiles, food and advice when most needed, and Alice again for her hospitality during the final weeks.

Thanks to my Mum and Dad and to my sisters, for their love, support and encouragement, I couldn't have done it without you. Finally thank you to Chris, who makes me smile and is always there to keep me going.

Table of contents

Page

List of Figures

Page

Figure

List of Tables

Page

Table

List of abbreviations

Å – Angstrom

κ - kappa

ϕ – phi

ω, Ω - omega

A_{260} – absorbance at 260nm

A_{280} – absorbance at 280nm

A. hydrophila – Aeromonas hydrophila

AIDS – Acquired immunodeficiency syndrome

APS – ammonium persulfate

BLF1 – Burkholderia lethal factor 1

B. pseudomallei – Burkholderia pseudomallei

CC – correlation coefficient

CDD – conserved domain database

C. elegans - Caenorhabditis elegans

CNF1 – cytotoxic necrotising factor 1

CPS – Capsular polysaccharide

CT – computed tomographic

CTD – C-terminal domain

CV – column volume

Da – Dalton

DEAE – diethylaminoethanol

DMSO – dimethylsulfoxide

DNA – deoxyribonucleic acid

DTT - dithiothreitol

DUF – Domain of unknown function

E. coli – Escherichia coli

EDTA – ethylenediaminetetraacetic acid

EM – electron microscopy

eV – electron volt

FOM – figure of merit

G-CSF – granulocyte colony-stimulating factor

gDNA – genomic DNA

HEPES - 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid

HIV – Human Immunedeficiency Virus

IHA – indirect haemagglutination assay

IPTG – isopropyl β-D-1-thiogalactopyranoside

K_{av} – partition coefficient

kDa – kiloDalton

LB – Lysogeny Broth

LLG – log-likelihood gain

LPS - Lipopolysaccharide

MAD – multi-wavelength anomalous dispersion

mAU – milli absorbance unit

MBP – Maltose binding-protein

MES – 2-(*N*-morpholino)ethanesulfonic acid

MIX – marker of type six effectors

MNGC – multi-nucleated giant cells

MPD – 2-Methyl-2, 4-pentanediol

MW – molecular weight

NaCl – sodium chloride

NCS – non-crystallographic symmetry

NTD – N-terminal domain

OD – optical density

OMVs – outer membrane vesicles

P. aeruginosa – *Pseudomonas aeruginosa*

PAGE – polyacrylamide gel electrophoresis

PCR – polymerase chain reaction

PDB – protein data bank

PEG – polyethylene glycol

pI – isoelectric point

PRR – peptide recognition receptor

PVDF – Polyvinylidene fluoride

R_{factor} – R-factor

R_{free} free R-factor

R_{merge} – merging R-factor

R_{pim} – precision-indicating merging R-factor

RMSD – root-mean-square deviation

rpm – revolutions per minute

rRNA – ribosomal ribonucleic acid

SAD – single-wavelength anomalous dispersion

SAM – S-adenosyl-methionine

SDS – sodium dodecyl sulphate

S. marcescens – *Serratia marcescens*

SOC - super optimal broth with catabolite repression

SP – sulfopropyl

T2SS – type II secretion system

T3SS – type III secretion system

T6SS – type VI secretion system

TAE – tris-acetate-EDTA

TEMED – N,N,N',N'-tetramethylethylenediamine

T_m – melting temperature

TMP-SMX – trimethoprim-sulphamethoxazole

TPR – tetratricopeptide repeat

Tris – Tris(hydroxymethyl)methyl amine

TFZ – translation function Z-score

V – volts

V. cholerae – *Vibrio cholerae*

V_e – elution volume

V_o – void volume

V_t – total volume

Chapter 1. Introduction

1.1 Melioidosis

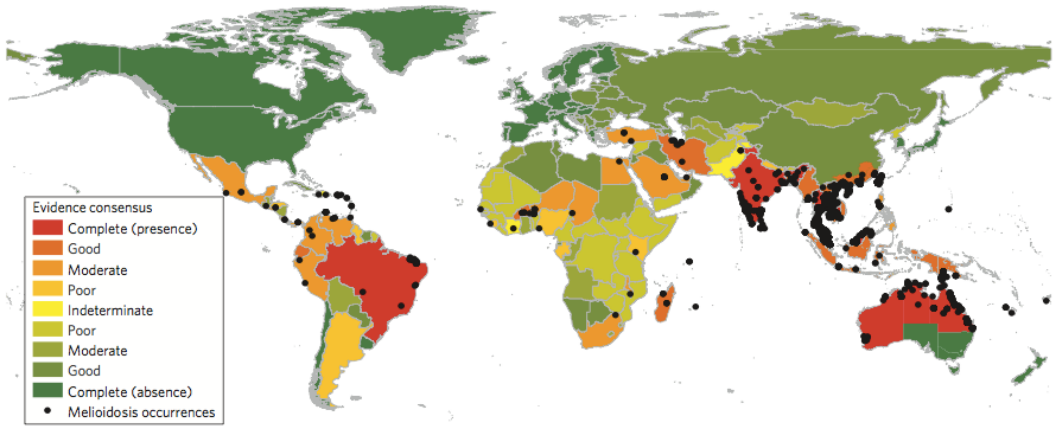
1.1.1 Disease

Melioidosis is a disease caused by the bacteria *B. pseudomallei* (*Burkholderia pseudomallei)* (Whitmore and Krishnaswami, 1912; reviewed in Wiersinga et al., 2012). It has been recognised as an opportunistic infection, as with early diagnosis and the essential treatment resources, the disease should not be fatal to a previously healthy individual (Currie et al., 2010; Zueter et al., 2016). The disease is endemic to South and South East Asia, Northern Australia, South America and West Sub-Saharan Africa (reviewed in Cheng and Currie, 2005; Limmathurotsakul et al., 2016). However, analysis of the worldwide prevalence of melioidosis has predicted the disease is underreported in 45 countries and present in a further 34, where it goes unreported (Figure 1.1) (Limmathurotsakul et al., 2016). During a study carried out in North East Thailand, following HIV/AIDS (human immunodeficiency virus/acquired immunodeficiency syndrome) and tuberculosis as the top two causes of death from infectious disease, melioidosis was found to be the third most common (Limmathurotsakul et al., 2010b). The mortality rate of melioidosis for 2015 has been estimated to be similar to that of measles, with estimated worldwide prevalence of 165,000 (68,000-412,000) (95% credible interval) cases of human melioidosis, resulting in 89,000 (36,000-227,000) (95% credible interval) deaths (Limmathurotsakul et al., 2016).

1.1.2 Infection

Melioidosis is contracted via inhalation, ingestion or percutaneous inoculation with soil or water infected with the bacteria *B. pseudomallei* (Currie and Jacups, 2003; Green and Tuffnell, 1968; Limmathurotsakul et al., 2013; Limmathurotsakul et al., 2014). Melioidosis as the cause of death has been identified in animals, predominantly goats and

A.



B.

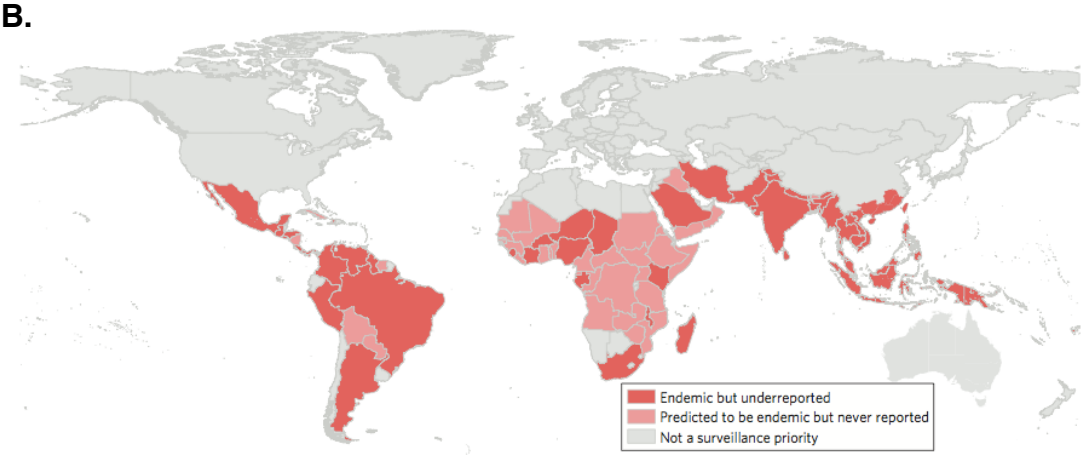


Figure 1.1. Global distribution of melioidosis

A. Map with evidence-based consensus on global distribution of melioidosis shown through the colour of each country. The evidence-based presence consensus of the disease ranges in a gradient from complete presence in red; through indeterminate in yellow, to complete absence in green. The location of cases of melioidosis or *B. pseudomallei* presence are shown as black dots - data from 1910-2014 (Limmathurotsakul et al., 2016). Image from Figure 1 in Limmathurotsakul et al., 2016. Reprinted by permission from Macmillan Publishers Ltd: Nature Microbiology (Limmathurotsakul et al., 2016), copyright (2016) https://www.nature.com/nmicrobiol/

B. Map with countries highlighted depending on whether melioidosis is endemic but goes underreported, if it has never been reported but is predicted to be endemic - these have been listed as priority countries for improvement of diagnostic and reporting facilities - or the country is not a surveillance priority (Limmathurotsakul et al., 2016). Image from Figure 3 in Limmathurotsakul et al., 2016. Reprinted by permission from Macmillan Publishers Ltd: Nature Microbiology (Limmathurotsakul et al., 2016), copyright (2016) https://www.nature.com/nmicrobiol/

transmission of the disease has been seen from person-person, in a mother to child via the placenta and breast milk (Abbink et al., 2001; Limmathurotsakul et al., 2006; Ralph et al., 2004). Furthermore there has been evidence of laboratory acquired melioidosis (Green and Tuffnell, 1968).

The risk of contracting melioidosis can be increased by certain factors, such as exposure to certain environmental conditions, activities and pre-existing medical conditions. Infection, by predicted inhalation due to aerosolisation of the bacteria, during periods of heavy rain and wind, such as monsoon season, has been shown to cause more severe manifestations of the disease (Currie and Jacups, 2003). ~80% of patients from a study in Northern Australia presented with symptoms during the rainy season (Currie et al., 2010). Growth of *B. pseudomallei* was shown to respond to an increase in soil water content and *B. pseudomallei* has been found in significantly more samples from wet season soil and groundwater than dry season soil in part of Northern Australia (Baker et al., 2011; Larsen et al., 2013). The bacteria can move from an initial site, aided by the local groundwater, into residential areas where groundwater has been known to be present and where cases of melioidosis have been reported (Baker et al., 2011). A number of tsunami survivors from Southern Thailand, presented with melioidosis, after either sustaining injury or undergoing inhalation during the tsunami (Chierakul et al., 2005). Similarly, inhalation of infected water in a near-drowning or choking event increases risk of melioidosis (Chierakul et al., 2005; Limmathrurotsakul et al., 2013). People who work or do other activities involving contact with soil or water are at a greater risk of infection by presumed inoculation through skin lesions (Limmathrurotsakul et al., 2013). The presence of open wounds also increases risk of infection (Chierakul et al., 2005; Limmathrurotsakul et al., 2013). Ingestion of untreated drinking water and eating food contaminated with *B. pseudomallei* has been shown to increase the risk of infection, with

melioidosis resulting from probable ingestion of drinking water containing *B. pseudomallei* (Limmathurotsakul et al., 2013; Limmathurotsakul et al., 2014; Thaipadungpanit et al., 2014). Pre-existing medical conditions such as diabetes mellitus, chronic kidney and lung disease, heart disease, thalassemia, pulmonary tuberculosis, cancer, taking steroids and being immunocompromised have been proposed to lead to an increased risk of acquiring melioidosis (Currie et al., 2004; Currie et al., 2010; Limmathurotsakul et al., 2013; Suputtamongkol et al., 1999; Zueter et al., 2016). Excessive consumption of alcohol and smoking have been linked to an increased risk of contracting melioidosis (Currie et al., 2004; Limmathurotsakul et al., 2013). Morbidity has been associated with age, a pre-existing disease or risk factor and septic shock (Currie et al., 2010; Zueter et al., 2016).

1.1.3 Clinical manifestations

Melioidosis can present as a number of manifestations and infection can be chronic, with symptoms lasting longer than 2 months (Currie et al., 2000a; Currie et al., 2010), or acute, with sudden aggressive onset (Currie et al., 2010; Zueter et al., 2016). The presentation of the disease during prolonged studies of clinical cases is more often acute, than chronic (Currie et al., 2010; Zueter et al., 2016) Pneumonia and other lung infections are the most common presentation, others include genitourinary infection, infections of the skin, liver and spleen, septic arthritis, neurological melioidosis and also bacteremia as a symptom alone (Currie et al., 2000b; Currie et al., 2010; Zueter et al., 2016). Abscesses are common on the skin – foot, neck and forearm, and organs - liver, spleen, kidney and prostate in men (Currie et al., 2000b; Currie et al., 2010; Zueter et al., 2016). Various other sites of infection have also been reported (reviewed in Limmathurotsakul and Peacock, 2011). Patients can also present with septic shock and/or bacteremia (Currie et al., 2010). The disease can also present with a secondary centre of infection, with and without bacteremia, often in the lungs, liver, bones,

joints and soft tissues (Currie et al., 2000b; Currie et al., 2010; Zueter et al., 2016). Some manifestations of the disease are found more commonly in specific countries than others, with supportive parotitis being more common in Thailand and brain inflammation – encephalomyelitis in Australia (Currie et al., 2000b; Lumbiganon and Viengnondha, 1995; McLeod et al., 2015) (Figure 1.2).

Death due to melioidosis has been reported due to sepsis, multiple organ failure, respiratory disease, septic shock and sudden and acute disease onset (Currie et al., 2010; Limmathurotsakul et al., 2010b; Zueter et al., 2016). Mortality rate has been related to age and presence of risk factors (Currie et al., 2010; Zueter et al., 2016), and septic shock presentation resulted in a higher mortality rate, compared to those who presented without (Currie et al., 2010). The mortality rate has decreased over the duration of extensive clinical case studies, from 30% to 9% (over the first 5 years to over the last 5 years, respectively), averaging 14% over 20 years in Northern Australia, and 49% to 40.5% over 10 years in Northeast Thailand (Currie et al., 2010; Limmathurotsakul et al., 2010b). In Malaysia the mortality rate was recorded at 32.9% over a 15-year clinical study (Zueter et al., 2016).

1.1.4 Incubation and latency

The period between initial infection with *B. pseudomallei* and presentation of symptoms can undergo an incubation of variable length. For acute infections, the period has been suggested to be between 1-21 days, (mean of 9 days) (Currie et al., 2000a; Currie et al., 2000c). The bacteria can remain latent within the body for a number of years before the disease begins to manifest, with the infection presenting as asymptomatic for long periods of time (Currie et al., 2010; Ngauy et al., 2005; Price et al., 2013). The longest recorded time of predicted latency is 62 years, in a World War II veteran (Ngauy et al., 2005). Re-activation of the disease is possibly as a result of a compromise of the immune system via an injury

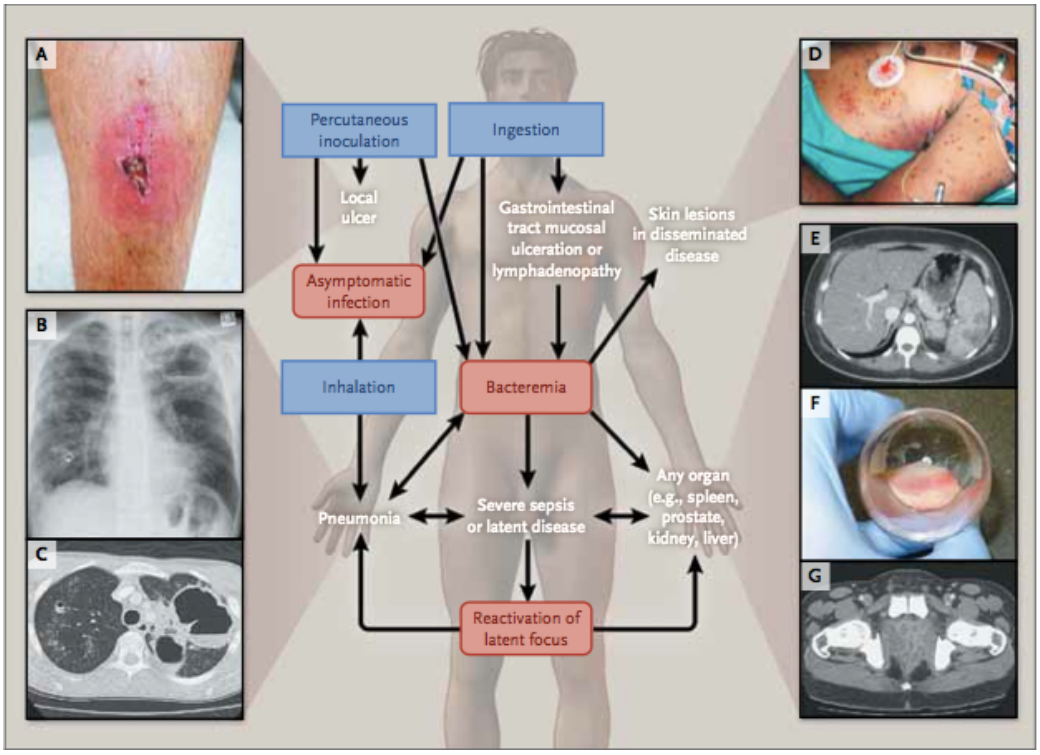


Figure 1.2. Route of infection and clinical manifestations of melioidosis. The routes of infection are shown in blue boxes, with progression of disease in red boxes and resulting clinical manifestations written in white. **A.** – cutaneous melioidosis infection, **B.** – lung abscesses, **C.** – CT (computed tomographic) scan of **B**, **D.** – skin manifestation, **E.** – CT scan showing abscesses on the spleen, **F.** – pus taken from the patient in **G**, **G.** – CT scan showing abscesses of the prostate and surrounding tissue (Wiersinga et al., 2012).
Image is from Figure 3 in Wiersinga et al., 2012.
Reproduced with permission from Wiersinga, W.J., Currie, B.J., and Peacock, S.J. (2012). Melioidosis. New England Journal of Medicine *367*, 1035-1044, Copyright Massachusetts Medical Society.

or infection (Ngauy et al., 2005; Suputtamongkol et al., 1994).

1.1.5 Diagnosis

Melioidosis can be misdiagnosed as it presents with a range of symptoms, some similar to other diseases including tuberculosis and influenza (Brent et al., 2007; Chan and Yip, 2015; Doker et al., 2014). The disease is therefore known as the 'great mimicker' (reviewed in Wiersinga et al., 2006). The absence of awareness of the diseases clinical manifestations can be problematic for accurate diagnosis (Barman et al., 2013; Hoffmaster et al., 2015; Koponen et al., 1991). The elucidation of the bacterial agent responsible for presented symptoms can require specialised equipment, which can be unavailable in certain areas (Hoffmaster et al., 2015; Suputtamongkol et al., 1994). Also the unfamiliarity with *B. pseudomallei* specifically may result in its misidentification as bacillus or a pseudomonad (Brent et al., 2007; Doker et al., 2014). The presence of *B. pseudomallei* in culture may seem unlikely or the misidentified version in cultures can be rejected as a contaminant (Brent et al., 2007; Doker et al., 2014; Koponen et al., 1991). The combination of clinical and event history, clinical diagnosis, laboratory identification and communication throughout are important in diagnosis (Koponen et al., 1991). The shorter time to diagnosis and application of treatment are proposed to relate to a decrease in mortality (Currie et al., 2010).

The culture of *B. pseudomallei* from samples taken from the patient is indicative of a positive diagnosis of melioidosis (Wuthiekanun et al., 1990; Wuthiekanun et al., 2001; Zueter et al., 2016), although this technique is not as sensitive as initially thought (Limmathurotsakul et al., 2010a). *B. pseudomallei* produce distinct colonies when grown on Ashdown agar and take ~2 days to grow (Wuthiekanun et al., 1990). The most popular serodiagnostic test is the IHA (indirect hemagglutination assay), used to identify the presence of *B. pseudomallei* by measuring the amount of

antibody response to *B. pseudomallei* antigens (Harris et al., 2009; Hoffmaster et al., 2015). However the IHA has been shown to be non-specific, lack sensitivity and results differ with disease presentation (Cheng et al., 2006; Harris et al., 2009). The IHA was carried out on a sample population of children in Thailand with no known melioidosis diagnosis (Wuthiekanun et al., 2006). It was shown that ~60-70% of children older than 4 years had a detectable antibody response, whilst 9-20% (increasing from 9-13/14 years) of children had an antibody titre the same as the amount indicated to represent the presence of melioidosis (Wuthiekanun et al., 2006). The seropositivity levels of healthy individuals in Northern Australia were found to be ~3%, with significant exposure of people to soil and water potentially containing *B. pseudomallei* (James et al., 2013). A number of other possible diagnostic techniques are being investigated (Hoffmaster et al., 2015).

1.1.6 Treatment

B. pseudomallei has shown resistance to a number of antibiotics including aminoglycosides (amikacin, gentamicin, netilmicin, tobramycin), ampicillin, carbenicillin, rifampicin, fosfomycin, amoxicillin, ticarcillin, pefloxacin, norfloxacin, erythromycin, clindamycin lincomycin, cefoxitin, cephalothin and cefsulodin, also penicillin had no effect on the *B. pseudomallei* infection in an animal model (Dance et al., 1989; Koponen et al., 1991; Miller et al., 1948b; Thibault et al., 2004).

The current treatment for melioidosis is a two-stage process. An intensive intravenous antibiotic stage using ceftazidime, meropenem or imipenem takes around 10-14 days (Currie et al., 2000b; Zueter et al., 2016). The initial step may be accompanied by the use of G-CSF (granulocyte colony-stimulating factor) to treat patients specifically with septic shock, this treatment is indicated to increase survival time, but does not influence patient mortality (Cheng et al., 2007; Currie et al., 2000b). The length of the initial treatment depends on the clinical symptoms presented, severity

of the disease and the response to the treatment (Limmathurotsakul et al., 2006; Zueter et al., 2016). The second eradication stage follows with a 3-5 month minimum oral antibiotic stage, using TMP-SMX (trimethoprim-sulphamethoxazole) or TMP-SMX with doxycycline (Chaowagul et al., 2005; Currie et al., 2000b; Limmathurotsakul et al., 2006; Meumann et al., 2012; Zueter et al., 2016). Both treatments are effective, however more side effects are attributed to the presence of the additional doxycycline treatment (Chusri et al., 2012). An alternative final treatment can be amoxicillin-clavulanate, which is used for pregnant women, children and for people with allergies or intolerance to the other treatments (Cheng et al., 2008; Zueter et al., 2016). Additional drainage of abscesses may be required (Currie et al., 2000b; Morse et al., 2009).

Resistance to TMP-SMX has been recorded in *B. pseudomallei* isolated from patients with melioidosis (Wuthiekanun et al., 2005). However, a study on a collection of isolates from patients with melioidosis found 99.7% were receptive to TMP-SMX (Saiprom et al., 2015). It has been shown that during treatment of the disease, ceftazidime resistance developed in the host, due to mutations in or near the *penA* gene, coding for a β-lactamase, such as a nucleotide mutation, G to A, in the promoter region and a mutation causing a change from Tyr to Cys in the PenA protein (Sarovich et al., 2012).

1.1.7 Relapse and reinfection

After an initial infection, some patients experience relapse or reinfection, and can have multiple episodes of recurrent infection (Currie et al., 2010; Limmathurotsakul et al., 2006; Maharjan et al., 2005). An unfinished course of treatment, diagnosis problems and a delay in hospital presentation has been related to recurrent infection (Zueter et al., 2016). Relapse is the return of the original infection, whereas reinfection occurs with a different strain of *B. pseudomallei* (Maharjan et al., 2005; Sarovich et al., 2014). It is possible for a patient to be infected with multiple

genotypes of *B. pseudomallei* at once (Pitt et al., 2007). Analysis of the *B. pseudomallei* genotypes in recurrent infections found ~25% were caused by reinfection and ~75% were relapse (Limmathurotsakul et al., 2006; Maharjan et al., 2005; Sarovich et al., 2014). Recurrence episodes associated with relapse occurred within a shorter time interval than those caused by reinfection (Maharjan et al., 2005; Sarovich et al., 2014).

Fewer cases of relapse have been recorded when an initial intravenous antibiotic stage of ~26 days was used, implemented due to poor adherence to the eradication treatment (Pitman et al., 2015). However, the type of oral antibiotics used in the second stage of treatment and the length of this stage have been suggested to be risk factors associated with relapse, such as amoxicillin-clavulanate as the oral antibiotic drug and treatment less than 8 weeks was deemed hazardous (Limmathurotsakul et al., 2006). Lack of adherence to the treatment regime has been linked to relapse infection (Currie et al., 2000a; Currie et al., 2010).

1.1.8 Vaccine development

There is currently no vaccine available against melioidosis, and with the emergence of possible resistance to the front-line antibiotics, the length of the current treatment and recurrence of the disease, a vaccine would be beneficial and cost effective (Currie et al., 2000a; Pitman et al., 2015; Wuthiekanun et al., 2005; Peacock et al., 2012). The production of a *B. pseudomallei* vaccine has to take into account the different routes of infection and types of infection (acute, chronic or asymptomatic), types of vaccine possible, standardisation of testing, host/bacteria interactions and the intracellular lifestyle of the bacteria, these and other issues including current vaccine candidates have been reviewed in Choh et al., 2013; Limmathurotsakul et al., 2015; Peacock et al., 2012 and Silva and Dow, 2013. A selection of current areas of vaccine research being carried out, are described below.

A vaccine development strategy to address the chronic stages of infection has been investigated. Specific proteins overexpressed in conditions resembling those seen during chronic infection were identified (Ooi et al., 2013). Four proteins were investigated, including one that produced a higher antibody response in patients with non-recurrent melioidosis infection (Champion et al., 2016; Suwannasaen et al., 2011). Proteins were combined and used to immunise mice, resulting in increased protection compared to control and previously identified protective antigens – LolC and CPS (capsular polysaccharide) (Champion et al., 2016).

Immunisation of mice with heat-killed whole cell *B. pseudomallei* resulted in significant protection from later challenge with intra-peritoneal application of bacterium, with 80-100% survival over 3 weeks, but less protection was noted for aerosol challenge, but still of significance (Sarkar-Tyson et al., 2009). The mice were also significantly protected from *B. pseudomallei* infection via aerosol when immunised with killed-whole cell *B. mallei* and *B. thailandensis* (Sarkar-Tyson et al., 2009).

Research has been carried out into possible immunisation with components of *B. pseudomallei* – CPS, LPS (lipopolysaccharide) and outer membrane proteins. Investigations into the affect of OmpW from *B. pseudomallei* as a possible vaccine revealed that administration of OmpW to two different mouse models provided increased protection from *B. pseudomallei* challenge (Casey et al., 2016). OMVs (outer-membrane vesicles) produced by *B. pseudomallei* have been found to contain CPS and LPS and a number of proteins associated with the periplasm and outer membrane (Nieves et al., 2011). Immunisation of mice with OMVs, both intranasally and subcutaneously, generates an antibody response, and significant protection is provided from aerosol challenge with *B. pseudomallei*, with subcutaneous immunisation (Nieves et al., 2011).

Investigations have been carried out into *B. pseudomallei* sensitive to a bacteriophage, as a possible treatment. Results indicate that pre-treatment of cells with bacteriophage prior to infection with *B. pseudomallei* increased the survival rate of epithelial cells, similarly mice that underwent bacteriophage pre-treatment before or after infection exhibited some protection from infection (Guang-Han et al., 2016).

1.1.9 Disease prevention

Another course of action against melioidosis is to promote the prevention and raise awareness of the disease. A set of guidelines has been proposed to reduce the risk of contracting melioidosis (Limmathurotsakul et al., 2013). These guidelines include reducing possible routes of infection and exposure to environmental sources of *B. pseudomallei* (Limmathurotsakul et al., 2013). Shoes, boots or rubber gloves should be worn if in contact with soil or water, and umbrellas should be used in the event of exposure to extreme weather, which is to be avoided in itself (Limmathurotsakul et al., 2013). Identification of *B. pseudomallei* in drinking water supplies led to the recommendation of chlorinating the water, boiling the water and using bottled water, which is in agreement with the guidelines (Limmathurotsakul et al., 2013; Thaipadungpanit et al., 2014). Washing post-exposure and of wounds should be carried out with clean water and soap (Limmathurotsakul et al., 2013). People should be made aware if they have any risk factors associated with contracting melioidosis or there are risk factors to avoid (Limmathurotsakul et al., 2013).

1.1.10 *B. pseudomallei* as a bioterror threat

Due to the virulence of *B. pseudomallei*, the availability and its stability in the soil, the small amount required for infection and potential unawareness of the disease, the bacteria has been designated as a tier 1 overlap select agent (CDC and HHS, 2012). *B. pseudomallei* has been subcatergorised as a Catergory B agent potential bioweapon on the

Centres of Disease Control list, on the basis of its mode of infection – via inhalation, its availability in the soil, its biosafety level requirements and the awareness of the disease (Rotz et al., 2002).

1.2 *Burkholderia pseudomallei*

1.2.1 Phylogeny

B. pseudomallei are Gram-negative, aerobic, rod-shaped, motile, non-spore forming β-proteobacteria, which are part of the larger genus *Burkholderia* (Miller et al., 1948a; Sessitsch et al., 2005; Wetmore and Gochenour, 1956; Whitmore, 1913; Yabuuchi et al., 1992). The *Burkholderia* genus contains ~100 species (Parte, 2014), and was initially described for seven species, previously of the *Pseudomonas* genus (Yabuuchi et al., 1992). The genus has been phylogenetically analysed using a number of methods: *recA* gene, genome comparisons, and the use of multilocus sequence analysis using a number of genes, and 16s rRNA (ribosomal ribonucleic acid), some of which have resulted in the separation of the species into two groups (Estrada-de los Santos et al., 2013; Payne et al., 2005; Sawana et al., 2014). One group contains species associated with plants and the environment and the other has pathogenic *Burkholderia* species (Estrada-de los Santos et al., 2013; Sawana et al., 2014). Further analysis of the phylogeny has suggested the formation of a further genus of *Paraburkholderia*, to accommodate the first group of environmental species, leaving the pathogenic species in the *Burkholderia* genus (Sawana et al., 2014). However three groups have been suggested based on 16S rRNA analysis, which splits the second group of pathogenic species into two - *B. cepacia* complex and *B. pseudomallei* group together, and the separate *B. glathei* group (reviewed in Depoorter et al., 2016).

The *Burkholderia genus* includes a number of subgroups. The *B. pseudomallei* group contains pathogenic *B. pseudomallei*, non-pathogenic *B. thailandensis* and *B. mallei* (the causative agent of the

equine disease, glanders) (Brett et al., 1998; Payne et al., 2005; reviewed in Whitlock et al., 2007). *B. oklahomensis,* which has been identified as *B. pseudomallei* previously, has been placed into the *B. pseudomallei* group in a specific type of phylogenetic analysis (reviewed in Depoorter et al., 2016; Glass et al., 2006). Investigations have shown that *B. mallei* likely evolved from a strain of *B. pseudomallei* (Losada et al., 2010). *B. thailandensis* was identified as a less virulent strain of *B. pseudomallei* in hamsters, and designated a new species based on 16S rRNA analysis (Brett et al., 1997; Brett et al., 1998). There is also the Bcc – *Burkholderia cepacia* complex, which includes a number species such as *B. cepacia*, *B. cenocepacia*, *B. diffusa, B. ambifaria, B. ubonensis* and *B. vietnamensis* (Payne et al., 2005; Vanlaere et al., 2008). The *Burkholderia* genus includes plant pathogens such as *B. glumae* and *B. gladioli* targeting rice plants and *B. glumae* affects various plants, causing bacterial wilt (Jeong et al., 2003; Nandakumar et al., 2009; Ura et al., 2006). *Burkholderia* species exist that are not pathogenic to plants, but are present within plants, plant roots and in the surrounding root soil, and are involved in nodule formation and nitrogen fixation (Caballero-Mellado et al., 2004; Estrada-De los Santos et al., 2001; Guentas et al., 2016; Moulin et al., 2001; Sessitsch et al., 2005). The *Burkholderia* genus also contains species associated with or involved in symbiosis with insects and fungi (Kikuchi et al., 2005; Lim et al., 2003; Santos et al., 2004).

1.2.2 Survival

B. pseudomallei is found in soil and water and can survive at a range of temperatures, 2-40°C and pH values, 3-8 (Chen et al., 2003; Robertson et al., 2010; Tong et al., 1996; Wuthiekanun et al., 1996). Bacteria were successfully grown in soils from paddy fields and farmland, with water content of preferably >15%, however *B. pseudomallei* has been shown to survive in dehydrated soil (Chen et al., 2003; Larsen et al., 2013). The optimal conditions for growth of *B. pseudomallei* in soil are ~37/42°C and pH 6.5-7.5 (Chen et al., 2003). *B. pseudomallei* can survive in distilled

water and rain water, with live bacteria found in distilled water after 16 years (Pumpuang et al., 2011; Robertson et al., 2010). Although *B. pseudomallei* can survive in low concentrations of salt, it is unable to survive in seawater or high salt concentrations, ~40% (Chen et al., 2003; Robertson et al., 2010).

1.2.3 Morphology

B. pseudomallei produces distinct colonies when grown on Ashdown agar (Chantratita et al., 2007). A total of seven different colony morphologies have been identified from clinical specimens, with morphotype I as the most commonly observed (Figure 1.3A) (Chantratita et al., 2007). In some of the clinical specimens analysed different morphologies were produced by the same bacterial culture (Chantratita et al., 2007). *B. pseudomallei* were shown to undergo reversible morphotype switching under starvation conditions (Chantratita et al., 2007). Morphotype switching was also seen in the presence of antibiotic, high temperature, lack of iron and in interaction with a mouse model (Chantratita et al., 2007). Starvation induces morphotype I to switch mainly to type III, II and V (Chantratita et al., 2007). *B. pseudomallei* cultured from distilled water after 16 years exhibited morphotype VII in 55% of colonies (Pumpuang et al., 2011).

Certain morphotypes showed significant association with phenotypes for expression of virulence factors such as biofilm formation, adherence, flagella, swarming and production of certain enzymes (Chantratita et al., 2007). However another study showed that certain morphotypes could not be associated with differences in biofilm formation (Koh et al., 2013). Morphotype I was able to replicate at a higher rate in macrophages and was more resilient to reactive oxygen species compared to types II and III (Tandhavanant et al., 2010). Morphotype II has been associated with increased biofilm production (Chantratita et al., 2007). Morphotype III was more resistant to antimicrobial peptide challenge, compared to types I

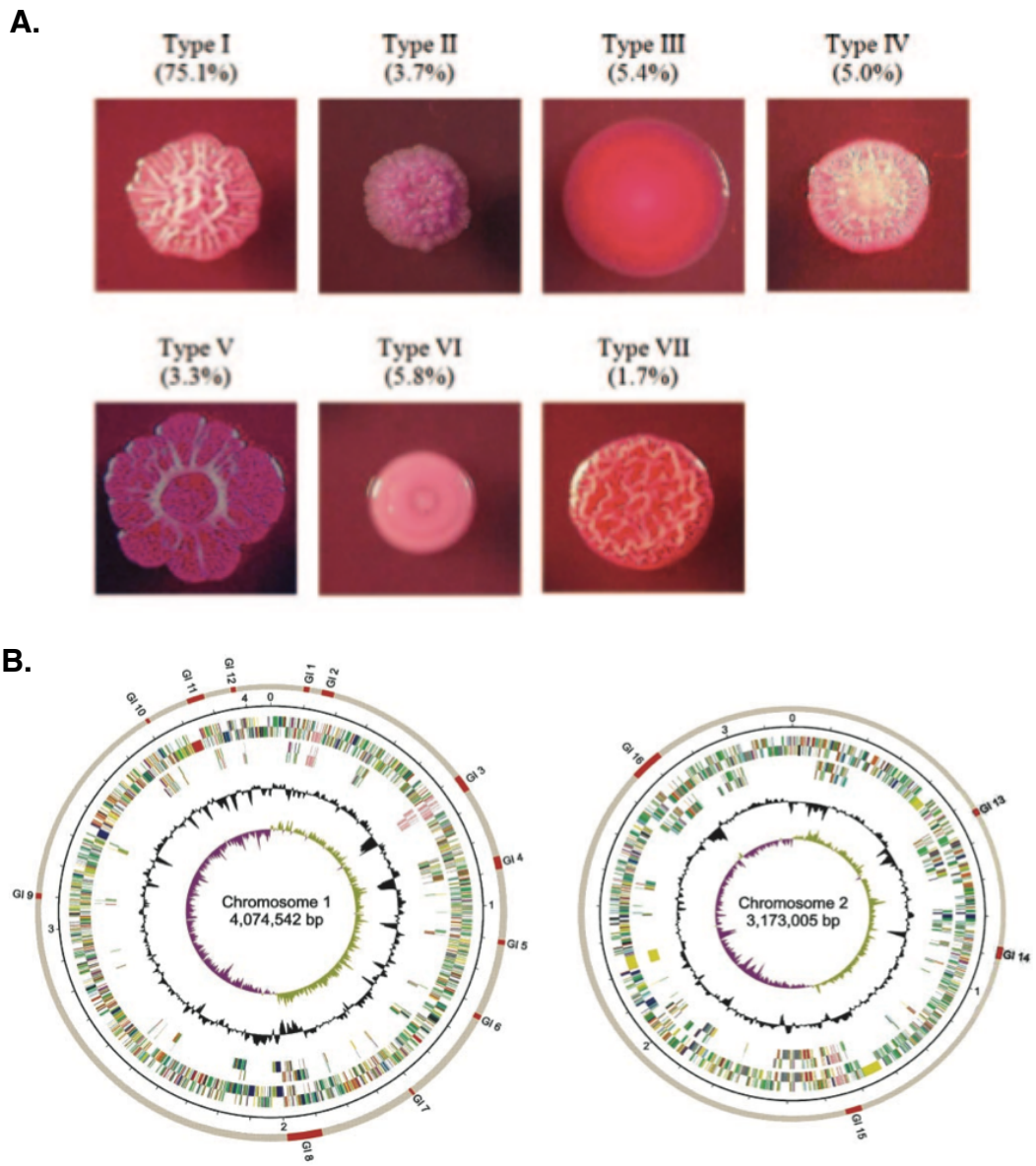


Figure 1.3. The seven colony morphotypes and genome of *B. pseudomallei*. A. The seven colony morphotypes of *B. pseudomallei* when grown on Ashdown's agar. Morphotypes are labelled type I to VII, with the percentage occurrence of each type in 212 *B. pseudomallei* samples from patients presenting with melioidosis (Chantratita et al., 2007). Image from Figure 1.A. in Chantratita et al., 2007, Biological relevance of colony morphology and phenotypic switching by *Burkholderia pseudomallei*. Journal of Bacteriology *189*, 807-817.
B. Schematic of the large and small chromosomes from *B. pseudomallei* K96243 (Holden et al., 2004). Image from Figure 1 in Holden, M.T.G., *et al.* (2004). Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. Proceedings of the National Academy of Sciences of the United States of America *101*, 14240-14245. Copyright (2004) National Academy of Sciences, U.S.A.

and II, and has also shown increased flagella production compared to type I (Chantratita et al., 2007; Tandhavanant et al., 2010). Investigations into the proteomic profiles of morphotypes I, II and III identified proteins involved in amino acid transport, metabolism, energy production and conversion, cell motility and chaperones were upregulated in both type II and III compared to type I (Chantratita et al., 2012). However down-regulated proteins included those with a role in metabolism, energy production and conversion, amino acid transport, secondary metabolite production and transport, and post-translational modification proteins (Chantratita et al., 2012).

1.2.4 Genome

The *B. pseudomallei* K96243 genome is formed from two circular chromosomes. The large chromosome (BPSL) is 4.07Mbp, whilst the small chromosome (BPSS) is 3.17Mbp (Holden et al., 2004) (Figure 1.3B). The larger chromosome encodes genes involved in core functions such as biosynthesis of amino acids, cofactors, nucleotides, fatty acids and macromolecules, metabolism, chemotaxis and the cell envelope and has 3460 coding sequences (Holden et al., 2004). Accessory genes encoding secondary metabolism, regulators, adaptation and two component systems are encoded on the smaller chromosome, which has 2395 coding sequences (Holden et al., 2004). Both chromosomes have core and accessory functions, but with a larger amount of core genes on the large chromosome and more accessory genes and genes for hypothetical proteins on the small chromosome (Holden et al., 2004). The genome has a G/C content of 67.7% and 68.5% for the large and small chromosomes, respectively (Holden et al., 2004).

1.2.5 Intracellular lifestyle

Once *B. pseudomallei* have entered the body they are able to infiltrate both phagocytic and non-phagocytic cells, through phagocytosis and invasion, respectively (Jones et al., 1996; Kespichayawattana et al.,

2000; Pruksachartvuthi et al., 1990). On internalisation inside the host cells *B. pseudomallei* are present in endocytic vesicles (Jones et al., 1996; Gauthier et al., 2001), are able to evade phagocytic killing (Egan and Gordon, 1996; Pruksachartvuthi et al., 1990) and have been shown to replicate intracellularly (Jones et al., 1996; Kespichayawattana et al., 2000). The bacteria proceed to escape from endosomes prior to fusion with the cell lysosome (Gauthier et al., 2001; Harley et al., 1994; Jones et al., 1996; Stevens et al., 2002), otherwise bacteria would be degraded (Hoppe et al., 1999). *B. pseudomallei* induces actin-rearrangement in the host cell cytoplasm, which results in the formation of actin tails at one end of the bacteria (Kespichayawattana et al., 2000). The bacteria use the actin tails to form protrusions that can contact neighbouring cells (Kespichayawattana et al., 2000; Stevens et al., 2002). The fusion of neighbouring cells, leads to the formation of MNGC (multi-nucleated giant cells) in both phagocytic and non-phagocytic cells, which produce plaques on cell death (French et al., 2011; Kespichayawattana et al., 2000). *B. pseudomallei* induces apoptosis of both phagocytic and non-phagocytic cells (Kespichayawattana et al., 2000) (Figure 1.4).

1.3 Investigations into the mechanisms of virulence

B. pseudomallei employs a range of virulence factors involved in nutrient acquisition, cellular invasion, toxins, adhesion and biofilm formation. These virulence factors have been reviewed extensively in the literature, Cheng and Currie, 2005; Galyov et al., 2010; Wiersinga et al., 2012 and Willcocks et al., 2016. A selection of virulence factors will be described below. Whilst a large number of virulence factors have been studied and characterised, there is still research taking place to identify more potential virulence factors (Moule et al., 2016). However there is still information regarding the biology and virulence mechanisms of *B. pseudomallei* which are unknown.

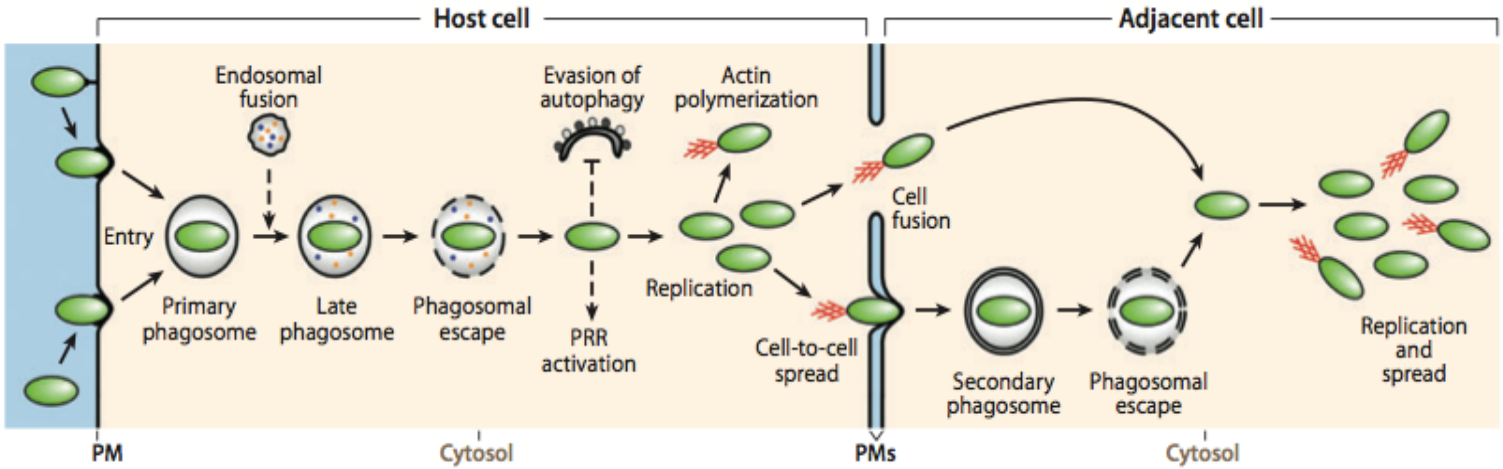


Figure 1.4. Intracellular lifestyle of *B. pseudomallei*.

B. pseudomallei invade or are phagocytosed into host cells, which results in the presence of the bacteria in the primary phagosome. The primary phagosome fuses with the endosome to produce the late phagosome. *B. pseudomallei* escape from the phagosome and trigger PRR (peptide recognition receptor) activation, but prevent autophagy. The bacteria then replicate, induce actin polymerisation and move to other cells (Galyov et al., 2010).

Figure from Galyov et al., 2010, Figure 2.

Reproduced with permission of Annual Review of Microbiology, Volume 64 © by Annual Reviews, http://www.annualreviews.org.

1.3.1 Sigma factors

RpoE is an alternative sigma factor present in an operon of four genes (Korbsrisate et al., 2005), and has been shown to be involved in the *B. pseudomallei* response to osmotic, oxidative and heat stress (Korbsrisate et al., 2005; Vanaporn et al., 2008). RpoE regulates the expression of other sigma factors and stress response proteins, as well as regulators, metabolic enzymes and cell wall synthesis pathway proteins (Korbsrisate et al., 2005; Thongboonkerd et al., 2007; Vanaporn et al., 2008). The operon has a role in regulating genes involved in the survival of *B. pseudomallei* in the mouse macrophage internal environment. Knockouts of the RpoE operon have decreased ability to form biofilms and are less virulent in mice (Korbsrisate et al., 2005; Thongboonkerd et al., 2007).

RpoS is another sigma factor involved in response to oxidative stress, carbon starvation and acidic growth conditions (Subsin et al., 2003). The transcription of RpoS occurs during exponential phase growth and is at its highest at the beginning of stationary phase (Subsin et al., 2003). In *B. pseudomallei* infection, RpoS regulates genes involved in cellular invasion, MNGC formation and induction of mouse cell macrophage apoptosis (Lengwehasatit et al., 2008; Utaisincharoen et al., 2006).

The RpoS regulon was identified and found to contain a range of genes involved in stress response also chaperones, lipid metabolism, energy metabolism and a number of hypothetical proteins (Osiriphun et al., 2009). The hypothetical protein BPSL1549 was found to be part of the RpoS regulon, with expression of the protein down-regulated in the absence of RpoS (Osiriphun et al., 2009). Investigation into BPSL1549 protein revealed it had structural similarity to CNF1 (cytotoxic necrotising factor 1) catalytic domain in *E. coli* (*Escherichia coli*) (Buetow et al., 2001; Cruz-Migoni et al., 2011). CNF1 is a glutamine deamidase that modifies Gln63-Glu63 in the protein Rho, a GTPase (Flatau et al., 1997). Rho is unable to hydrolyse GTP and is thus constitutively activated, leading to

cellular actin reorganisation (Flatau et al., 1997). BPSL1549 was found to act as a glutamine deamidase, targeting the protein eIF4A and thus inhibiting protein translation in the cell. The protein was lethal on administration to both mice and macrophages, and *B. pseudomallei* showed reduced virulence in mice on deletion of BPSL1549. BPSL1549 has been renamed as BLF1 – *Burkholderia* lethal factor 1 (Cruz-Migoni et al., 2011).

1.3.2 Target selection and the aims of project 1

A study comparing the stationary phase proteomes of *B. pseudomallei* and *B. thailandensis* identified fourteen proteins significantly differentially expressed between the two species (Wongtrakoongate et al., 2007) (Figure 1.5). Of these 14 proteins four were identified as hypothetical proteins, present or more highly overexpressed in *B. pseudomallei*, BPSL1549, BPSL1958, BPSS0212 and BPSS0683 (Wongtrakoongate et al., 2007). Another six of these proteins have been shown to be involved in virulence in other bacteria (Wongtrakoongate et al., 2007).

As described earlier, RpoS is involved in regulating genes for stress response and key stages in the intracellular lifestyle. The reporting of the *B. pseudomallei* RpoS regulon, through proteome analysis of *rpoS*+ and *rpoS*- strains, identified 13 hypothetical proteins down-regulated in the absence of RpoS (Osiriphun et al., 2009). Eight of these proteins were classified as under direct regulation by RpoS: BPSL0599, BPSL1549 (BLF1), BPSL3012, BPSS0212, BPSS0213, BPSS0683 BPSS1588 and BPSS2055 (Osiriphun et al., 2009). BPSS0212 and BPSS0213 were found to be part of an operon of four proteins – BPSS0211, BPSS0212, BPSS0213 and BPSS0214, with the RpoS dependent promoter in BPSS0211 (Osiriphun et al., 2009). The expression of the genes in the operon was found to be co-regulated (Osiriphun et al., 2009). Analysis of the upstream DNA (deoxyribonucleic acid) from the BPSS0211 coding region identified the RpoS promoter, CTAACCT described by Osiriphun et

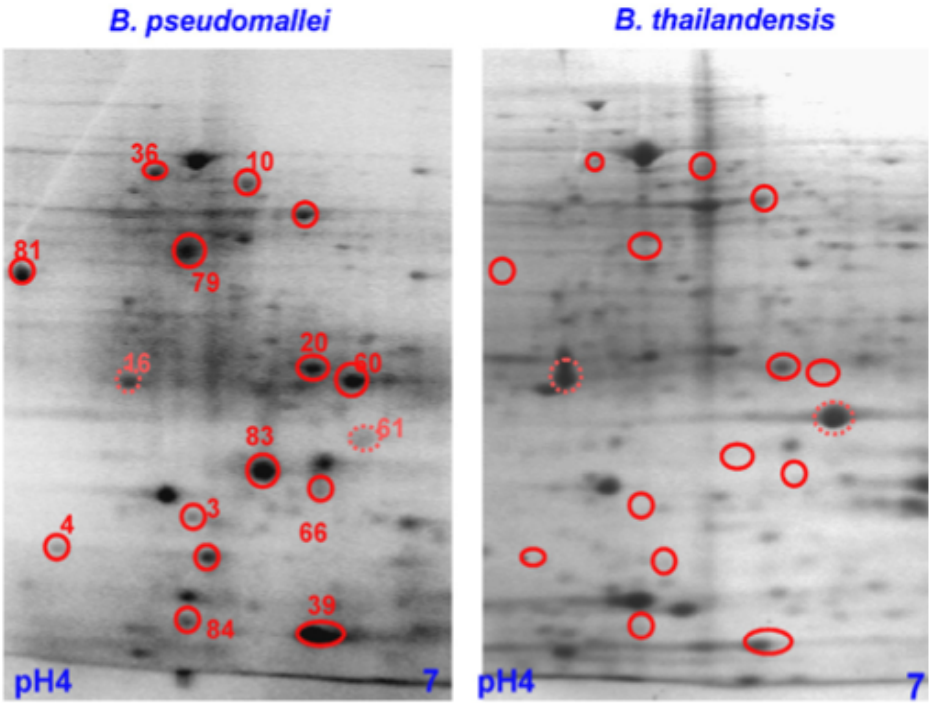


Figure 1.5. The two-dimensional gel of *B. pseudomallei* and *B. thailandensis* stationary phase proteomes. The numbers are associated with proteins present in *B. pseudomallei* stationary phase proteome. The circles represent protein spots with differential expression between the two species. All numbered spots, except spot 16 and 61 (dotted circles), were found to be present or have higher protein expression in *B. pseudomallei* than *B. thailandensis*. Spot 16 and 61 have higher expression in *B. thailandensis*. Proteins BPSL1549, BPSL1958, BPSS0212 and BPSS0683 are spots 79, 81, 83 and 84, respectively (Wongtrakoongate et al., 2007). Image from Figure 3 in Wongtrakoongate et al., 2007. Reprinted from Molecular and Cellular Probes, Volume 21, Wongtrakoongate, P., Mongkoldhumrongkul, N., Chaijan, S., Kamchonwongpaisan, S., and Tungpradabkul, S. Comparative proteomic profiles and the potential markers between *Burkholderia pseudomallei* and *Burkholderia thailandensis*. 81-91. Copyright (2007) with permission from Elsevier.

al., 2009 (Winsor et al., 2008).

The protein target selection for this project, stipulated previously by Dr. Matthew Day (Day, 2012), were that the proteins had to be differentially expressed in pathogenic species *B. pseudomallei* and in non-pathogenic species *B. thailandensis*, during stationary phase growth (Wongtrakoongate et al., 2007). Protein expression had to also be under the control of the alternative sigma factor RpoS (Osiriphun et al., 2009). The proteins present in both studies are BPSL1549 (BLF1), BPSS0212 and BPSS0683 (Osiriphun et al., 2009; Wongtrakoongate et al., 2007). BPSS0213 was identified in the stationary phase proteome of *B. pseudomallei*, but was not significantly differently expressed between the two species (Wongtrakoongate et al., 2007).

The overexpression of these hypothetical proteins in *B. pseudomallei* compared to *B. thailandensis* suggests that they may play a role in virulence mechanisms of *B. pseudomallei*, exhibited by BPSL1549 identified as BLF1. A student at University of Sheffield has carried out investigations on the protein BPSS0683 previously. Initial investigations were carried out, by Dr. Matthew Day, on the remaining nine target hypothetical proteins from both studies including BPSL0599, BPSL1958, BPSL3012, BPSS1588, BPSS2055 and the operon of four proteins BPSS0211-BPSS0214 (Day, 2012). Further work was carried out, by Dr. Matthew Day, on proteins BPSL0599, BPSL1958, BPSS0211, BPSS0212 and BPSS0213, this included the structure determination of the protein, BPSS0211 (Day, 2012). The project aims to continue the work carried out on the operon of four proteins BPSS0211, BPSS0212, BPSS0213 and BPSS0214 started by Dr. Matthew Day (Day, 2012), identified via the selection criteria described above.

1.3.3 Secretion systems

Gram-negative bacteria use a number of different secretion systems to transport proteins and effectors. Currently there are five identified secretion systems, which cross both cell membranes and one, which crosses the outer membrane, which all are at various stages of structural characterisation (reviewed in Costa et al., 2015) (Figure 1.6). *B. pseudomallei* encodes three T3SS (Type III Secretion System) and six T6SS (Type VI Secretion System) (Attree and Attree, 2001; Rainbow et al., 2002; Shalom et al., 2007; Winstanley et al., 1999).

1.3.4 Type III secretion system

The T3SS spans both Gram-negative bacterial membranes and delivers effectors directly into the target cell (reviewed in Costa et al., 2015). In *B. pseudomallei*, two T3SS are similar to that found in *Ralstonia solanacearum* (Rainbow et al., 2002; Winstanley et al., 1999) and the third T3SS-3 (Bsa T3SS) is similar to the SPI-1 cluster from the human pathogen *Salmonella typhimurium* (Attree and Attree, 2001; Stevens et al., 2002). *B. pseudomallei* strains with a mutation in a Bsa T3SS component were attenuated in both hamster and mouse models, indicating its role in virulence during infection (Stevens et al., 2004; Warawa and Woods, 2005). T3SS-3, as a virulence factor, has been linked to melioidosis infections in the lung and spread of disease to other areas of the body (Gutierrez et al., 2015). Virulence was reduced on tracheal administration of a T3SS-3 mutant in a mouse model and reduced liver and spleen colonisation by T3SS-3 mutants, on infection (Gutierrez et al., 2015). T3SS-1 and T3SS-2 may also contribute to the virulence of *B. pseudomallei* (Warawa and Woods, 2005). The Bsa-T3SS has been shown to be involved in the intracellular lifestyle of *B. pseudomallei* with roles in endocytic vesicle escape, intracellular replication and, actin rearrangement and protrusion formation (Stevens et al., 2002). There is conflicting evidence as to whether Bsa-T3SS has a role in non-phagocytic cell invasion (French et al., 2011; Stevens et al.,

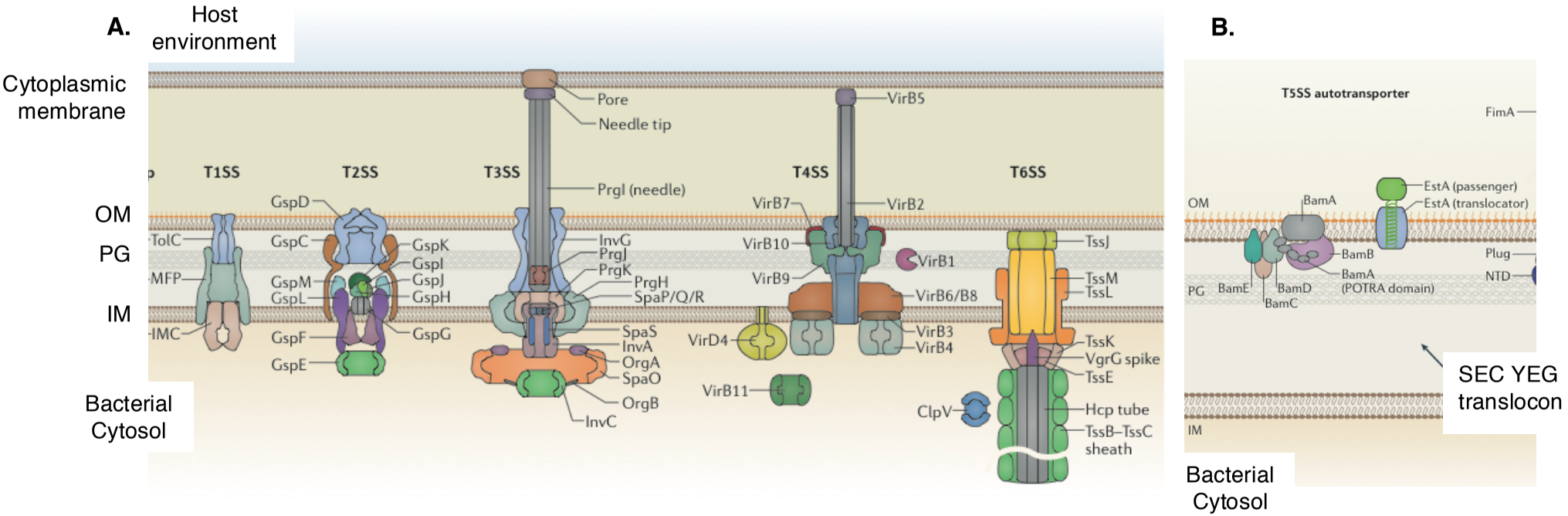


Figure 1.6. Gram-negative secretion systems

Schematic representation of the six secretion systems present in Gram-negative bacteria. **A.** T1SS, T2SS, T3SS, T4SS and T6SS cross both the inner and outer membrane. **B.** T5SS only crosses the outer membrane and uses the SEC translocation system to cross the inner membrane, along with other systems used to construct outer membrane components (not shown). A subtype of T5SS assembly into the OM is aided by the Bam complex. IM – inner membrane, PG – peptidoglycan, OM – outer membrane. (Costa et al., 2015).

Figure 1.6A and B, adapted from Costa et al., 2015 – Figures 1a and 2a, respectively.

Adapted by permission from Macmillan Publishers Ltd: Nature Reviews Microbiology (Costa et al., 2015), copyright (2015)
http://www.nature.com/nrmicro/index.html

2003).

The role of the Bsa-T3SS in the escape of *B. pseudomallei* from endocytic vesicles, which is impaired on mutation of the system, may contribute to the other effects seen downstream in the intracellular lifestyle of *B. pseudomallei* (Stevens et al., 2002; French et al., 2011). Investigations were carried out into the effect of a Bsa-T3SS mutation starting from the entry of cells into the cytoplasm, using a *B. thailandensis* model (French et al., 2011). The Bsa-T3SS was subsequently not required for downstream intracellular replication, actin rearrangement or cell-to-cell spread (French et al., 2011). However the combination of these investigations suggest that the T3SS is a key virulence factor in *B. pseudomallei* infection.

Another secretion system in *B. pseudomallei* is the T6SS. A large amount of research has taken place to establish the assembly and function of the T6SS and it's effectors, described in sections 1.4 and 1.5. However there is still some unknown information regarding the structure and role of certain components of the secretion system

1.4 Type VI secretion system

1.4.1 Introduction

The T6SS was identified from gene clusters in Gram-negative bacteria, including *V. cholerae* (*Vibrio cholerae)* and *P. aeruginosa* (*Pseudomonas aeruginosa)* (Das and Chaudhuri, 2003; Mougous et al., 2006; Pukatzki et al., 2006). The T6SS cluster has 13 core genes (Boyer et al., 2009; Shalom et al., 2007; Zheng and Leung, 2007) and additional non-essential genes or *tss*-associated genes have been found within these T6SS gene clusters that are potentially secreted proteins or may function within the T6SS (Shalom et al., 2007). The core genes are: TssA, TssB, TssC, TssD (Hcp), TssE, TssF, TssG, TssH (ClpV), TssI (VgrG), TssJ, TssK, TssL and TssM (Boyer et al., 2009; Shalom et al., 2007). The

T6SS has been shown to target proteins to cells in a cell-to-cell contact dependent manner (Hood et al., 2010; MacIntyre et al., 2010; Pukatzki et al., 2006).

1.4.2 Function

The T6SS has a role in virulence in a number of disease causing Gram-negative bacteria. In *B. pseudomallei* mutation of T6SS gene causes attenuation in hamster models (Burtnick et al., 2011). A T6SS from *P. aeruginosa* is predicted to be involved in chronic lung infections of cystic fibrosis patients (Mougous et al., 2006). A T6SS in *B. thailandensis* was found to be required for virulence in a mouse model (Schwarz et al., 2010). T6SS from *V. cholerae* is able to target and cause cross-linking of actin within macrophages (Pukatzki et al., 2007; Pukatzki et al., 2006). *A. hydrophila* (*Aeromonas hydrophila)* has a T6SS which secretes effectors into human cells and cause cell rounding and apoptosis (Suarez et al., 2010).

The T6SS can be used as a tool to allow specific microbes to outcompete other bacteria. For example a toxic effector protein from *P. aeruginosa* is targeted to other bacteria, but is not delivered to eukaryotic cells (Hood et al., 2010). *V. cholerae* T6SS was shown to be virulent towards other Gram-negative bacteria, but not other species of *V. cholerae* (MacIntyre et al., 2010). A T6SS from *B. thailandensis* shows contact and competes with other Gram-negative bacteria both in growth and biofilm formation, but only specific species (Schwarz et al., 2010). *S. marcescens* (*Serratia marcescens*) is able to compete with other Gram-negative bacteria and *S. marcescens* strains via the bacterial targeted killing effect of it's T6SS (Murdoch et al., 2011).

Therefore it has been predicted that a T6SS can target bacterial or eukaryotic cells (Hood et al., 2010). This has also been found to occur between multiple T6SS found within the same organism, with one T6SS

interacting with competing Gram-negative bacteria, whilst another is involved in virulence during infection of a eukaryotic model organism (Schwarz et al., 2010). However some T6SS may have multiple functions, with T6SS-dependent interaction with both eukaryotic and bacterial cells shown by *V. cholerae* (MacIntyre et al., 2010; Pukatzki et al., 2006).

1.4.3 Type VI secretion system proteins related to bacteriophage tail components

T3SS and T2SS (Type II Secretion System) have been shown to share structural homology with bacterial flagella and type IV pilus, respectively (reviewed in Diepold and Armitage, 2015; reviewed in Peabody et al., 2003). Analysis of the T6SS core gene products revealed that a number were similar to bacteriophage components. TssI protein N-terminal domains were shown to be a similar to two proteins present in the T4 bacteriophage – gp27 and gp5, in structure and sequence (Kanamaru et al., 2002; Leiman et al., 2009; Pukatzki et al., 2007). TssI N-terminus is similar to both gp27 and gp5, but lacking the lysozyme domain (Leiman et al., 2009; Pukatzki et al., 2007). In T4 bacteriophage, gp27 interacts with full-length gp5 to assemble into the tail spike of the bacteriophage injection machinery, with a stoichiometry of $(gp27)_3(gp5)_3$ (Kanamaru et al., 2002). TssD was shown to have a similar structure to T4 bacteriophage protein gp19 and bacteriophage λ protein gpV (Leiman et al., 2009; Mougous et al., 2006; Pell et al., 2009). TssE is similar to T4 bacteriophage baseplate protein gp25 (Leiman et al., 2009). It was also found that TssF, TssG and TssK were similar to part of gp6-N-terminus, gp53 and gp8, respectively (Brunet et al., 2015; Planamente et al., 2016). TssA has been predicted to be homologous to the C-terminus of gp6 (Planamente et al., 2016) and full-length TssA has been suggested to have a similar role to tail terminator proteins gp15 or gp3 (Zoued et al., 2016). Gp25, gp6, gp53 and gp8 are all part of the T4 bacteriophage baseplate, with gp27-gp5 as part of the central hub (Kanamaru et al., 2002; Kostyuchenko et al., 2003). The sheath proteins TssB/TssC form a

structure similar to the T4 phage tail (Kube et al., 2014).

The similarity in sequence and structure between the T6SS and phage components suggests that the T6SS may also assemble and function in a similar way (Leiman et al., 2009; Pell et al., 2009; Pukatzki et al., 2007). However the core components of the T6SS would assemble to form a cell-puncturing device, from inside of the cell to outside, as opposed to phage, which delivers DNA into the target cell from the outside (Basler et al., 2012; Kanamaru, 2009; Kanamaru et al., 2002; Leiman et al., 2009).

1.5 Type VI secretion system core components and mechanism

1.5.1 Membrane complex

The membrane complex, the base of which is present in the cytoplasm, spans the inner membrane, periplasm and it interacts with the outer membrane. A hole of around 15-20 Å (Angstrom) is present through the centre of the complex (Durand et al., 2015). It has 5–fold symmetry with 10 copies of each of the three proteins, TssJ, TssL and TssM (Durand et al., 2015). The complex is attached to the outer membrane via TssJ, an outer membrane anchored lipoprotein (Aschtgen et al., 2008; Durand et al., 2015). TssM extends through the inner membrane and periplasm, and is predicted to form part of the base, and TssL is present in the base and forms hook-like protrusions at the bottom of the base (Durand et al., 2015).

1.5.2 Sheath

TssB and TssC were found to assemble into a large molecular weight complex when co-purified (Bönemann et al., 2009). This complex forms tubules of varying lengths, an outer diameter of ~300 Å and an inner pore of ~100 Å (Bönemann et al., 2009; Kube et al., 2014). The sheath has been visualised in an extended and a contracted state (Basler et al., 2012). The extended state is denser, longer and thinner than the contracted state, which is wider and hollow, and around half the length

(Basler et al., 2012). The contracted sheath also exhibits helical ridges on its surface and has 12 cog-wheel protrusions visible in cross-section (Basler et al., 2012; Bönemann et al., 2009). The contracted sheath structure has tubules formed from stacks of six TssB and TssC heterodimers in a hexameric ring (Kube et al., 2014; Kudryashev et al., 2015). The stacked rings are rotated relative to each other resulting in the 12 surface protrusions and production of the helical ridges (Kube et al., 2014). An interaction is formed between three proteins from three separate heterodimers, which ultimately leads to the formation of the tubule like structure (Kudryashev et al., 2015). A model for the extended sheath structure has been predicted based on the T4 phage tail extended state (Kube et al., 2014). It has been suggested that the sheath may exist in another state just after contraction has occurred (Clemens et al., 2015).

1.5.3 TssH

TssH is an AAA+ protein (Mougous et al., 2006), involved in the disassembly of the tail sheath (Bönemann et al., 2009). In the contracted state, the TssH interaction site located at the N-terminal end of TssC is exposed on the outside of the sheath (Kube et al., 2014; Pietrosiuk et al., 2011). TssH N-terminal domain is then able to interact with TssC and disassemble the sheath into TssB/TssC heterodimers (Bönemann et al., 2009; Basler et al., 2012; Kube et al., 2014). It is predicted that when the sheath is in the extended state, the TssH interacting site on TssC is inaccessible to TssH (Kube et al., 2014).

1.5.4 Inner tube

The TssD protein assembles into rings, ~85 Å outer diameter and 40 Å inner diameter, each of six TssD subunits with 6-fold symmetry (Mougous et al., 2006). The inner tube is composed of multiple copies of the TssD ring stacked head to tail (Brunet et al., 2014), with the TssI trimer forming a spike at the end, and a PAAR protein further sharpens the spike (Leiman et al., 2009; Pukatzki et al., 2007; Shneider et al., 2013). This

inner tube, in a similar way to the phage tail tube, is predicted to act as the puncturing device (Pukatzki et al., 2007; Leiman et al., 2009). TssI and PAAR are predicted to be present in the baseplate prior to contraction of the sheath (Kanamaru et al., 2002; Leiman et al., 2009).

1.5.5 Baseplate

A potential baseplate structure has been visualised at one end of the sheath, with part of it present between the sheath and the membrane (Basler et al., 2012). A number of proteins have been shown to be similar to phage baseplate proteins – TssE – gp25, TssF – gp6 N-terminal, TssA – gp6 C-terminal, TssG – gp53, TssK – gp8, TssI – gp27-gp5, however TssG is also thought to be similar to gp7 (Brunet et al., 2015; reviewed in Leiman et al., 2010; Leiman et al., 2009; Planamente et al., 2016; Taylor et al., 2016). TssF, TssG and TssK localise to the inner membrane, and TssK is shown to interact with TssL, with both TssK and TssG interacting with TssM (Brunet et al., 2015; English et al., 2014; Zoued et al., 2013). TssK was found to form a multisubunit complex with TssF and TssG, with TssF and TssG forming a complex, prior to binding of TssK (English et al., 2014; Brunet et al., 2015). TssF and TssG have also been shown to interact with TssE and the TssF/TssG complex with TssI (Brunet et al., 2015). A complex of TssE, TssF, TssG and TssK was produced with a stoichiometry of 1:2:1:3, respectively (Taylor et al., 2016). Another complex of TssK, TssF and TssG has been produced with a proposed stoichiometry of 12:6:3, respectively (English et al., 2014). It is predicted that TssE, TssF, TssG, TssK, as well as TssI are part of the baseplate structure of the T6SS machinery (Brunet et al., 2015; Taylor et al., 2016).

1.5.6 TssA

Currently TssA from two different organisms, enteroaggregative *E. coli* (*Escherichia coli*) (referred to as *E. coli* from this point, unless otherwise specified) and *P. aeruginosa,* have been investigated. For *E. coli* TssA, constructs of two domains were produced, the N-terminal domain and C-

terminal domain (Zoued et al., 2016). The C-terminal is a dodecameric ring, with two layers of six subunits, creating a star-shaped lumen (Zoued et al., 2016). The N-terminal is predicted to surround the central C-terminal hexameric core with six long arms protruding out away from the core. The N-terminal forms a dimer in solution and the section of N-terminal crystallised – N-terminal section 2 – is also a dimer (Zoued et al., 2016). The structure of *P. aeruginosa* TssA C-terminal domain has been predicted to be homologous to the C-terminal domain of the phage baseplate protein gp6 (Planamente et al., 2016). *P. aeruginosa* TssA was seen to form rings with ~260 Å outer diameter and ~100 Å inner diameter. It is predicted to contain a dodecamer, or double-dodecamer, similar to gp6 protein, that forms dodecameric rings with 6-fold symmetry (Aksyuk et al., 2009; Planamente et al., 2016).

The TssA proteins from *E. coli* and *P. aeruginosa* have been located at an end of the B/C sheath protein, with *E. coli* TssA found to be associated with the cytoplasm and the membrane (Planamente et al., 2016, Zoued et al., 2016). However the evidence from both papers suggests these two TssA proteins have different functions (Planamente et al., 2016, Zoued et al., 2016).

E. coli TssA has been predicted to be involved in the recruitment of the baseplate in T6SS assembly via its interaction with the membrane complex and baseplate (Zoued et al., 2016). The *E. coli* TssA was seen to continue to associate with the sheath as it grows, moving away for the baseplate (Zoued et al., 2016). The C-terminal central core and N-terminal long arm extensions have been modeled onto the extended sheath containing the tail tube (Zoued et al., 2016). It has been hypothesised that the structure of *E. coli* TssA C-terminal domain undergoes a conformational change causing the star-shaped lumen to widen allowing the TssD tail tube components to enter, thus controlling the addition of tail tube and also sheath components (Zoued et al., 2016).

P. aeruginosa TssA has been suggested to form part of the baseplate, attach to the sheath and allow passage of the TssD tail tube through the TssA ring (Planamente et al., 2016).

1.5.7 Assembly and mechanism

The assembly and functional mechanism of the T6SS complex (Figure 1.7) has had a number of predicted models throughout the literature from research into the T6SS from different organisms. Described here is an amalgamation of the information from current literature, however different T6SS may undergo variations on the predicted core assembly and mechanism.

Assembly begins with TssJ, followed by TssL and TssM, which forms the membrane complex, but which does not fully cross the outer membrane (Durand et al., 2015). The predicted baseplate proteins, TssK, TssG, TssF, TssE, TssI (PAAR) then assemble onto the cytoplasmic side of the membrane complex (Zoued et al., 2013; English et al., 2014; Brunet et al., 2015). TssA, from *E. coli*, is associated with the growing tail tube and sheath (Zoued et al., 2016), and in *P. aeruginosa*, TssA is associated with one end of the sheath (Planamente et al., 2016). The tail tube, TssD, and sheath, TssB/TssC, then assemble onto the baseplate (Brunet et al., 2015, Basler et al., 2012). In order for the TssB/TssC sheath to assemble, TssD is required (Brunet et al., 2014). Two models have been suggested for the tail tube and sheath assembly – the TssD tube assembles followed by the TssB/TssC sheath or a TssD tube ring component assemblies into the a sheath ring component and then both are assembled (Basler et al., 2012, Zoued et al., 2016). The sheath is assembled in an extended state (Basler et al., 2012). TssM is now able to cross the outer membrane (Durand et al., 2015).

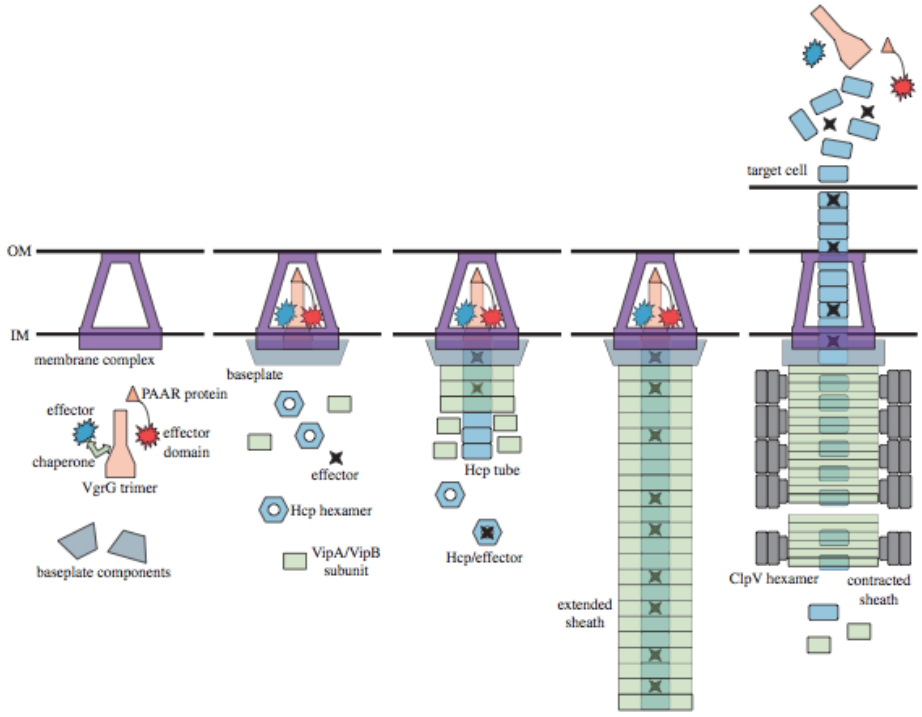


Figure 1.7. Type VI secretion system assembly and mechanism
The membrane complex (TssJ, TssL, TssM) assembles in the membrane first. The spike (TssI/VgrG, PAAR), with possible attached effector domains, along with TssE, TssF, TssG and TssK are predicted to form the baseplate, which then could attach to the membrane complex. The tail tube (TssD/Hcp) and sheath (TssB, TssC) assemble onto the baseplate. Effectors can also interact with TssD/Hcp tail tube components. Contraction of the sheath leads to the expulsion of the spike and inner tube into the target cell or outside of the cell and thus delivery of the effectors. TssH/ClpV then disassembles the contracted sheath and recycles the components (Basler, 2015).
Image from Figure 1 in Basler, 2015.
© 2015 Basler, M.
Basler, M. (2015). Type VI secretion system: secretion by a contractile nanomachine. Philosophical Transactions of the Royal Society B *370*, 20150021.
DOI: 10.1098/rstb.2015.0021
Link to article:
http://rstb.royalsocietypublishing.org/content/370/1679/20150021
Links to license: https://creativecommons.org/licenses/by/4.0/
https://creativecommons.org/licenses/by/4.0/legalcode

Once the T6SS machinery is assembled, the contraction of the tail sheath is predicted to cause the opening of the membrane complex to form a channel through which the tail tube can pass (Durand et al., 2015). In a similar mechanism to phage, a conformational change is predicted to take place within the baseplate that causes the tail sheath to contract (Leiman et al., 2004 and Kostyuchenko et al., 2005). Contraction of the tail sheath results in the expulsion of the tail tube from the cell (Basler et al., 2012; Brunet et al., 2013). Disassembly of the contracted sheath by TssH, allows recycling of the sheath protein components TssB and TssC into new sheath formations (Bönemann et al., 2009; Basler et al., 2012). The membrane complex remains and can be used again (Durand et al., 2015).

1.5.8 Effectors

T6SS effectors include structural core components of the tail tube – TssI and PAAR, as well as separate proteins (Pukatzki et al., 2007; Shneider et al., 2013). Inhibitor proteins have been identified in bacteria containing T6SS effectors, to protect them from the effects of their own virulence factors (Hood et al., 2010; Salomon et al., 2014).

1.5.9 Core component associated effectors

TssI and TssD have been identified in the supernatant of bacterial cultures, having been expelled from the cell by T6SS contraction (Pukatzki et al., 2006; Pukatzki et al., 2007; Basler et al., 2012). The PAAR protein, situated on the tip of the tail spike, has been identified as a domain on its own or within proteins containing other functional domains N- or C-terminal to the PAAR domain (Shneider et al., 2013). These functional domains include nucleases, transthyretin, peptidases, lipase, and a large amount of unknown domains (Shneider et al., 2013). TssI proteins, as mentioned previously, contain a conserved N-terminal region similar to the phage tail spike, but can also include a variable C-terminal region - these are referred to as ‘evolved’ TssI proteins (Pukatzki et al.,

2007; Leiman et al., 2009; Kanamaru et al., 2002). 'Evolved' TssI proteins have been found in *Vibrio, Pseudomonas, Yersinia, Burkholderia* and *Ralstonia* (Pukatzki et al., 2007). *V. cholerae*, has three TssI proteins, one C-terminal domain has an actin cross-linking domain, which has been shown to cause rounding of cells (Pukatzki et al., 2006; Sheahan et al., 2004). T6SS-5, from *B. thailandensis*, secretes a TssI which is a 'evolved' version with a C-terminal domain region of unidentified function, with a role in virulence and MNGC formation (Schwarz et al., 2014). Other TssI C-terminal domains identified have predicted functions including peptidoglycan-binding, mannose-binding, proteases and tropomyosin (Pukatzki et al., 2007).

The delivery of the effectors present in core component proteins is hypothesised to be via incorporation in the formation of the T6SS (Shneider et al., 2013). The PAAR tail tip protein can have effector domains attached, which from the orientation of the PAAR structure termini would protrude away from the tip (Shneider et al., 2013). Proteins attached to the PAAR tail tip have been secreted by the T6SS (Shneider et al., 2013). For the tail spike TssI, the 'evolved' proteins assemble into trimers via the N-terminal region (Pukatzki et al., 2007). Therefore the C-terminal effector domains of TssI are transported when the T6SS contracts and the tail spike and tube is expelled (Shneider et al., 2013; Pukatzki et al., 2007).

1.5.10 Cargo effectors

These cargo effectors are not domains within T6SS core components, but are transported by the T6SS (reviewed in Durand et al., 2014). Three exported proteins have been identified for H1-T6SS from *P. aeruginosa* – Tse1, Tse2, Tse3 (Hood et al., 2010). Tse2 was shown to be toxic against both prokaryotic and eukaryotic cells, however the T6SS-1 could only deliver Tse2 to prokaryotic cells (Hood et al., 2010). Tse2 interacts with the inside of the TssD ring, which is involved in the export of the

effector (Silverman et al., 2013). The EvpP protein from *Edwardsiella tarda* is part of the T6SS cluster. It is a secreted protein and is found to interact with the TssD equivalent EvpC (Zheng and Leung, 2007). A functional domain found to be associated with PAAR domains, is the transtherytin domain, which has been predicted to act as an adaptor domain (Shneider et al., 2013). Also TssI proteins without additional effector domains are involved in the secretion of non-component proteins (Hachani et al., 2011).

1.5.11 Type VI secretion system secretion signal

How proteins are chosen for secretion is yet to be fully determined, TssI and TssD do not contain identifiable signal sequences (Pukatzki et al., 2006; Williams et al., 1996). An N-terminal MIX (marker of type six effectors) motif has been identified in T6SS effectors from *Vibrio parahaemolyticus*, as well as other proteins including those previously known as T6SS effectors from different organisms, and within TssI and PAAR proteins (Salomon et al., 2014). MIX consists of three motifs; the conserved core motif is residues hRxGhhYhh (h – hydrophobic), and there are another two less conserved regions (Saloman et al., 2014).

1.5.12 Type VI secretion system in *Burkholderia*

There are six T6SS gene clusters in *B. pseudomallei* (Shalom et al., 2007). The location of these clusters in the *B. pseudomallei* K96243 genome is BPSL3097-BPSL3111 (T6SS-1), BPSS0095-BPSS0116 (T6SS-2), BPSS0185-BPSS0167 (T6SS-3), BPSS0515-BPSS0533 (T6SS-4), BPSS1493-BPSS1511 (T6SS-5) and BPSS2093-BPSS2109 (T6SS-6) (Shalom et al., 2007). The first cluster is located on the larger chromosome and the other five located on the smaller (Shalom et al., 2007). Schell et al., 2007, also annotated the T6SS clusters from *B. pseudomallei*, *B. thailandensis* and *B. mallei*. The annotation by Shalom et al., 2007, will be used here. *B. mallei*, and *B. thailandensis* both have five T6SS gene clusters (Schell et al., 2007; Angus et al., 2014). *B. mallei*

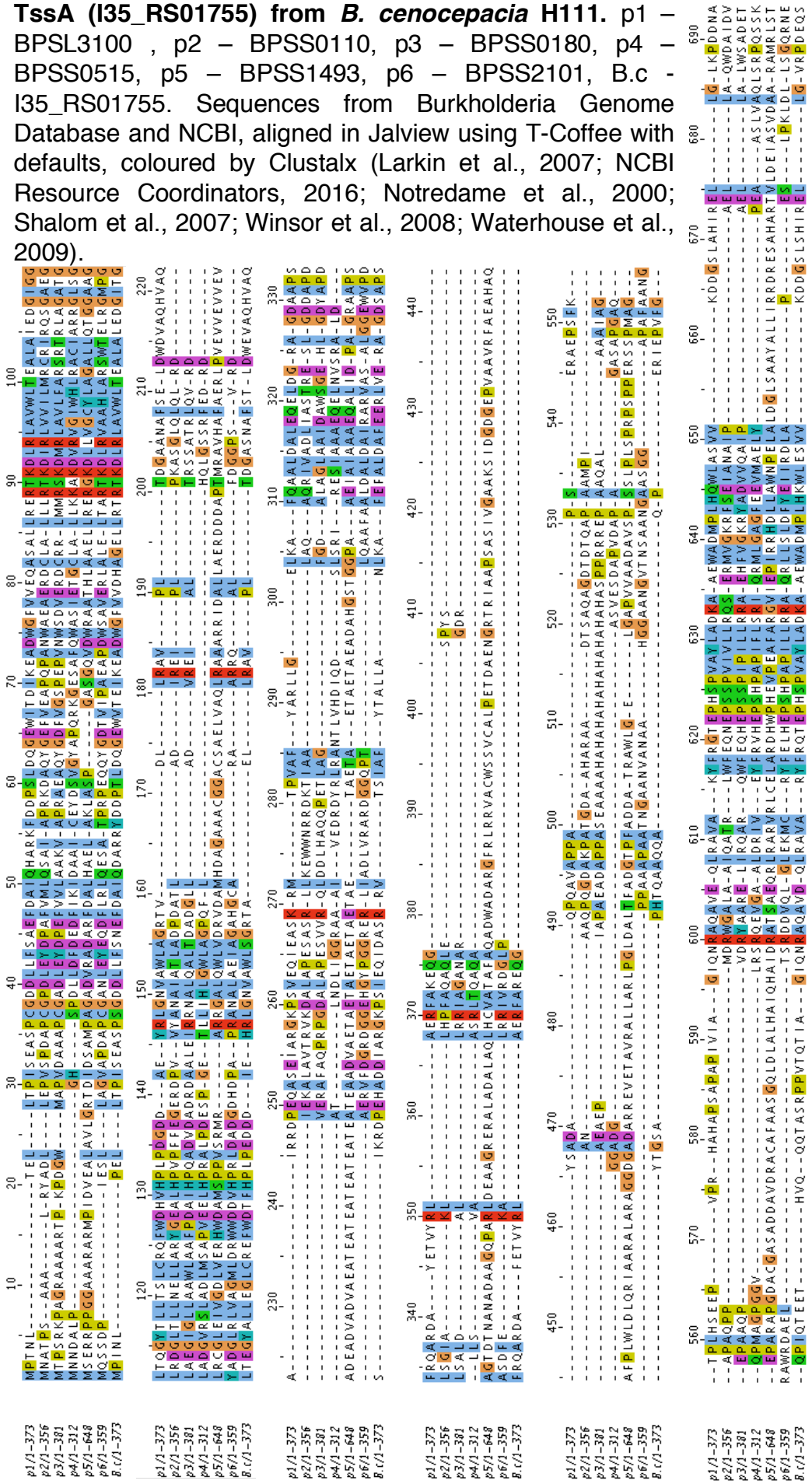
has equivalent to clusters 1, 3, 4, 5, and 6, and *B. thailandensis* has equivalent to clusters 1, 2, 4, 5 and 6, in *B. pseudomallei* (Schell et al., 2007; Schwarz et al., 2010). *B. cenocepacia* has one T6SS equivalent to *B. pseudomallei* T6SS-1 (Angus et al., 2014; Aubert et al., 2008).

The T6SS-5 has been shown to act as a key virulence factor in *B. pseudomallei* in hamster models, and is involved in intracellular growth of the bacteria (Burtnick et al., 2011; Chen et al., 2011). The secretion system is induced on the presence of *B. pseudomallei* within macrophages (Shalom et al., 2007; Chen et al., 2011), and when expressed in internalised bacteria it localises to one pole of the bacterium (Schwarz et al., 2014). Intracellular growth is postponed and occurs later than wildtype, and MNGC formation is absent with the mutation of a T6SS-5 core component (Burtnick et al., 2011). In *B. cenocepacia* T6SS-1 is involved in protrusion formation and actin rearrangement in macrophages (Aubert et al., 2008).

1.5.13 Aims of project 2

Investigations have been carried out into the role of the T6SS system in numerous Gram-negative bacteria, including *Burkholderia* species. However there is still unknown information regarding the structure and role of certain core components of the secretion system. The aim of the second project is to solve the structures of the remaining protein components from T6SS, and thus assign them a function. *B. cenocepacia* TssA is one of the proteins under investigation, an alignment was performed between the six TssA proteins from *B. pseudomallei* and the one TssA protein from *B. cenocepacia* (Figure 1.8).

Figure 1.8. Alignment of amino acid sequences of all six TssA proteins from *B. pseudomallei* K96243 with TssA (I35_RS01755) from *B. cenocepacia* H111. p1 – BPSL3100 , p2 – BPSS0110, p3 – BPSS0180, p4 – BPSS0515, p5 – BPSS1493, p6 – BPSS2101, B.c - I35_RS01755. Sequences from Burkholderia Genome Database and NCBI, aligned in Jalview using T-Coffee with defaults, coloured by Clustalx (Larkin et al., 2007; NCBI Resource Coordinators, 2016; Notredame et al., 2000; Shalom et al., 2007; Winsor et al., 2008; Waterhouse et al., 2009).



Chapter 2. Material and methods

The protocols outlined in this chapter cover the process of target gene cloning, transformation, protein overexpression, protein purification, the use of protein in biochemical investigations and crystallisation, and finally protein structure determination.

2.1 General recipes and methods

2.1.1 Antibiotics

	Concentration
Ampicillin	100 µg/mL
Carbenicillin	100 µg/mL

Solutions of antibiotics were prepared by dissolving the appropriate amount of antibiotic in MiliQ water.

2.1.2 Agarose gel

1% agarose gels were used for analysis of DNA products during the cloning process. 0.5 g of agarose was added to 50 mL TAE (Tris-acetate-EDTA(ethylenediaminetetraacetic acid)) buffer in a 100 mL Duran and dissolved using a microwave. 5 µL of GelRed (Biotium) (1 in 10000 dilution) was mixed into the dissolved agarose/TAE solution. Agarose gels were run with 5 µL of Hyperladder1 (Bioline) as a DNA marker. Agarose gels were set up with the TAE running buffer, loaded and run at 90-110 V (volts) for 40-50 minutes.

2.1.3 Miniprep, PCR purification and gel extraction protocols

Miniprep preparation of plasmid with or without insert was performed using the QIAprep Spin Miniprep Kit with microcentrifuge protocol (Qiagen) and products were eluted in autoclaved MiliQ water. PCR (polymerase chain reaction) purification of PCR products and post-digestion products was performed using the QIAquick PCR Purification Kit with microcentrifuge protocol (Qiagen) and products were eluted in

autoclaved MiliQ water. Gel extraction of PCR products and post-digestion products was performed using the QIAquick Gel Extraction Kit with microcentrifuge protocol (Qiagen) and products were eluted in autoclaved MiliQ water.

2.1.4 Transformation – cloning strains

For transformation of ligation reactions or empty plasmid for use in cloning, the *E. coli* strains used were: DH5α gold efficiency (Bioline), DH5α silver efficiency (Bioline) and DH5α (Invitrogen). 4-5 µL of plasmid, at 20-120 ng/uL, or ligation reaction solution was added to 50 µL pre-aliquoted *E. coli* cloning strain cells, (which had been thawed on ice), and incubated on ice for ~20-30 minutes. The solution was heat shocked for 30-45 seconds at 42°C, before being incubated on ice for 2 minutes. 300 µL of LB (Lysogeny Broth) (Bertani, 2004; Miller, 1972) (section 2.1.8) or pre-warmed SOC (super optimal broth with catabolite repression) was added to the cell solution, which was then incubated at 37°C, at 250 rpm (revolutions per minute) for 1-1.5 hours. Cell solution was plated onto LB agar (section 2.1.9), with added antibiotic, in two dilutions – 120-200 µL of culture, and 100-150 µL LB with 20-50 µL of culture. Plates were incubated at 37°C overnight.

2.1.5 Transformation – protein overexpression strains

For transformation of plasmid into overexpression strains, the following *E. coli* strains were used: BL21 DE3 (Agilent) and NEB Express (NEB). 2-3 µL of plasmid solution, at 20-120 ng/uL, was added to 50 µL pre-aliquoted *E. coli* overexpression strain cells, (which had been thawed on ice), and incubated on ice for ~20-30 minutes. The solution was heat shocked for 10-30 seconds at 42°C, before being incubated on ice for 2 minutes. 300 µL/950 µL of LB (section 2.1.8) or pre-warmed SOC was added to the cell solution, which was then incubated at 37°C, at 250 rpm for 1-1.5 hours. Cell solution was plated onto LB agar (section 2.1.9), with added antibiotic, in two dilutions – 120-200 µL of culture, and 100-150 µL

LB with 20-50 µL of culture. Plates were incubated at 37°C overnight.

2.1.6 Colony PCR

In order to confirm the presence of the insert in the ligated plasmid, colony PCR was carried out on plasmid isolated from ligation transformation colonies. A single transformation colony was added to 5 mL of LB (section 2.1.8) containing the appropriate concentration of antibiotic. Cultures were grown overnight at 37°C, 250 rpm. The overnight culture was pelleted at 5248 *g* for 20 minutes. Plasmid was isolated from the resulting cells pellets using the QIAprep Spin Miniprep Kit with microcentrifuge Protocol (Qiagen) and eluted in 50 µL autoclaved MiliQ water.

Colony PCR was carried out using the outlined protocol below:
Reaction:
4 µL BioMix Red (Bioline)
5 µL T7 Forward and Reverse Primers (10 pmol/µL)
T7 Forward – 3' – TAATACGACTCACTATAGG – 5'
T7 Reverse – 3' – GCTAGTTATTGCTCAGCGG – 5'
1 µL Plasmid

PCR protocol:
95°C for 1 minute
20 cycles of:
95°C for 30 seconds
54°C for 30 seconds
72°C for 1 minute
72°C for 5 minutes
The completed colony PCR reaction was run on a 1% agarose gel (section 2.1.2) to assess for the presence of the required insert band. Plasmids with suspected successful ligation were sent for DNA sequencing.

2.1.7 DNA Sequencing

In order to confirm the presence of the desired gene in the plasmid and that the gene has been cloned with the correct sequence, ligated plasmid was sent for DNA sequencing. Plasmid was prepared using the QIAprep Spin Miniprep Kit with microcentrifuge Protocol (Qiagen) and eluted in autoclaved MiliQ water. For sequencing carried out by the University of Sheffiled, Core Genomics Facility, 10-20 µL of DNA was aliquoted into an eppendorf tube and sent for sequencing. For sequencing carried out by GATC-Biotech, 5 µL of plasmid was combined with 5 µL of T7 forward or reverse primer at 5 pmol/µL, in an eppendorf tube, and then sent for sequencing.

2.1.8 LB media recipe

The components of LB media (Bertani, 2004; Miller, 1972) were dissolved in deionised water and autoclaved at 121°C for 15 minutes.

	Per litre
Tryptone	10 g
Yeast extract	5 g
NaCl (sodium chloride)	10 g

2.1.9 LB agar recipe and production of LB agar plates

The tryptone, yeast extract and NaCl components of LB agar (Bertani, 2004; Miller, 1972) were dissolved in deionised water, the agar was added last and the solution autoclaved at 121°C for 15 minutes.

	Per litre
Tryptone	10 g
Yeast extract	5 g
NaCl	10 g
Agar	15 g

Autoclaved LB agar was heated in the microwave until liquefied; it was then left to cool before the addition of antibiotics. 25 mL of liquid agar was aliquoted per plate.

2.1.10 SDS-PAGE - gel recipe and running protocol

Protein analysis during protein overexpression and protein purification was carried out using SDS-PAGE (sodium dodecyl sulphate-polyacrylamide gel electrophoresis) (Laemmli, 1970). All gels were prepared with 12% resolve and 6% stack, unless otherwise specified, the recipes for which are shown below. For a 10% solution of APS (ammonium persulfate), 0.1 g APS was dissolved in 1 mL of water, and the solution was preferably made fresh prior to use. The gel components were added in the order shown in the tables below (top to bottom), with the final two components, TEMED (N,N,N',N'-tetramethylethylenediamine) and 10% APS added last.

12% resolve	**per gel**
30% Acrylamide	2.5 mL
1 M Tris pH 8.8	2.35 mL
MiliQ water	1.28 mL
10% SDS	62.5 µL
TEMED	6.25 µL
10% APS	62.5 µL

15% resolve	**per gel**
30% Acrylamide	3.13 mL
1 M Tris pH 8.8	2.35 mL
MiliQ water	0.33 mL
10% SDS	62.5 µL
TEMED	6.25 µL
10% APS	62.5 µL

6% stack	**per gel**
30% Acrylamide	0.75 mL
1 M Tris pH 6.8	0.47 mL
MiliQ water	2.46 mL
10% SDS	37.5 µL
TEMED	3.75 µL
10% APS	37.5 µL

The resolve solution was poured between two glass sheets of desired separation. A 1:1 ratio of isopropanol:MiliQ water was layered over the setting resolve to prevent air bubbles setting into the gel. The isopropanol:MiliQ water layer was removed once the resolve was set. The stack was then poured on top and a well comb (for 9 or 15 wells) was added immediately. The set gel was set up in a gel running tank, with a buffer dam, if required. The gel running tank was filled with 1 x SDS-PAGE running buffer, the gel comb removed and the wells loaded with the prepared samples.

Gel samples were prepared by adding sample to NuPAGE LDS Sample Buffer (4x) (Thermo Fisher Scientific) and reducing agent (10x NuPAGE Sample Reducing Agent (Thermo Fisher Scientific) or 10x DTT (dithiothreitol)). Alternatively NuPAGE LDS Sample Buffer (4x) (Thermo Fisher Scientific) containing 2-Mercaptoethanol (20 µL 2-Mercaptoethanol in 300 µL NuPAGE LDS Sample Buffer (Thermo Fisher Scientific)) was used, to which the required sample volume was added. For protein overexpression analysis, and for protein purification ~20 µg of sample was loaded. The samples were then heated at ~100°C for ~5 minutes, followed by pulse centrifugation.

SDS-PAGE was run with 5 µL Mark12 (Thermo Fisher Scientific) as the protein molecular weight marker. Once loaded, the gel was run at 80 V for 10 minutes, followed by 200 V for 37-40 minutes. The gel was then removed from the gel tank and glass sheets, and stained using Coomassie blue gel stain, and then destained or, stained using Instant Blue (Expedeon).

2.1.11 SDS-PAGE - using pre-cast gels

Pre-cast gradient gels of 4-12% Bis-Tris gel (Novex) were also used for purifications involving the low molecular weight proteins – BPSS0211 and BPSS0211(1-123). Gel samples were prepared as in section 2.1.10. Pre-

cast gels were run with 5 µL Mark12 (Thermo Fisher Scientific) as the protein molecular weight marker and NuPAGE MES (2-(*N*-morpholino)ethanesulfonic acid) SDS running buffer (Thermo Fisher Scientific). Once loaded, the gel was run at 200 V for 35 minutes. The gel was then removed from the gel tank and plastic case, and stained using Coomassie blue gel stain, and then destained or, stained using Instant Blue (Expedeon).

2.1.12 Bradford assay

The Bradford assay was used to estimate protein concentration of samples during protein overexpression and protein purification (Bradford et al., 1976). The protocol is described below:

A standard was prepared containing 800 µL MiliQ water with 200 µL Bradford Reagent (Bio-Rad), and used as a blank. Samples for analysis were prepared as follows:

1-20 µL sample + 800 µL MiliQ water + 200 µL Bradford Reagent (Bio-Rad)

The absorbance of the sample was then measured at 595nm and the protein concentration in mg/mL calculated using the equation below.

OD(optical density)$_{595nm}$ x 15 / sample volume (µL) = protein concentration in mg/mL

The factor 15 is the factor, which relates the absorbance at 595nm to the protein concentration, calculated from the calibration graph of absorbance at 595nm (OD_{595nm}) vs protein concentration, for samples of known protein concentration.

2.1.13 Absorbance assay to measure DNA and protein concentration

A NanoPhotometer (Implen) was used to estimate both DNA and protein concentration, using a lid of dilution factor 5x, 10x or 50x. DNA concentration was measured at A_{260nm} (absorbance at 260nm) and the Nanophotometer (Implen) was set for double stranded DNA giving the

DNA concentration in ng/µL. Protein concentration was measured at A_{280nm} (absorbance at 280nm) and calculated by:
A_{280nm} x lid dilution factor / extinction coefficient = protein concentration in mg/mL

The extinction coefficient for each protein was calculated by ProtParam (Gasteiger et al., 2005). For calculation of the BPSS0212/BPSS0213 complex concentration, the extinction coefficients of both proteins were averaged.

2.1.14 Gel filtration calibration graph - K_{av} vs LogMW

The calibration graph of K_{av} (partition coefficient) vs LogMW (molecular weight), used in section 5.5.2, was created by running a mix of proteins of known molecular weight on the Superdex200 pg gel filtration column, pre-equilibrated with 50 mM HEPES (4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid) pH 7.5, 100 mM NaCl, (Figure 2.1) and noting the elution volumes associated with each protein to calculate the K_{av} and compare it to the LogMW for each protein (Table 2.1). The proteins used were RibonucleaseA, Ovalbumin, Aldolase and Ferritin. The K_{av} was calculated using $(V_e-V_o)/(V_t-V_o)$, with V_e (elution volume) of the protein, V_o (void volume) of the column matrix (41 mL) and V_t (total volume) of the column matrix (116 mL). The LogMW was calculated from the known molecular weights of the four proteins.

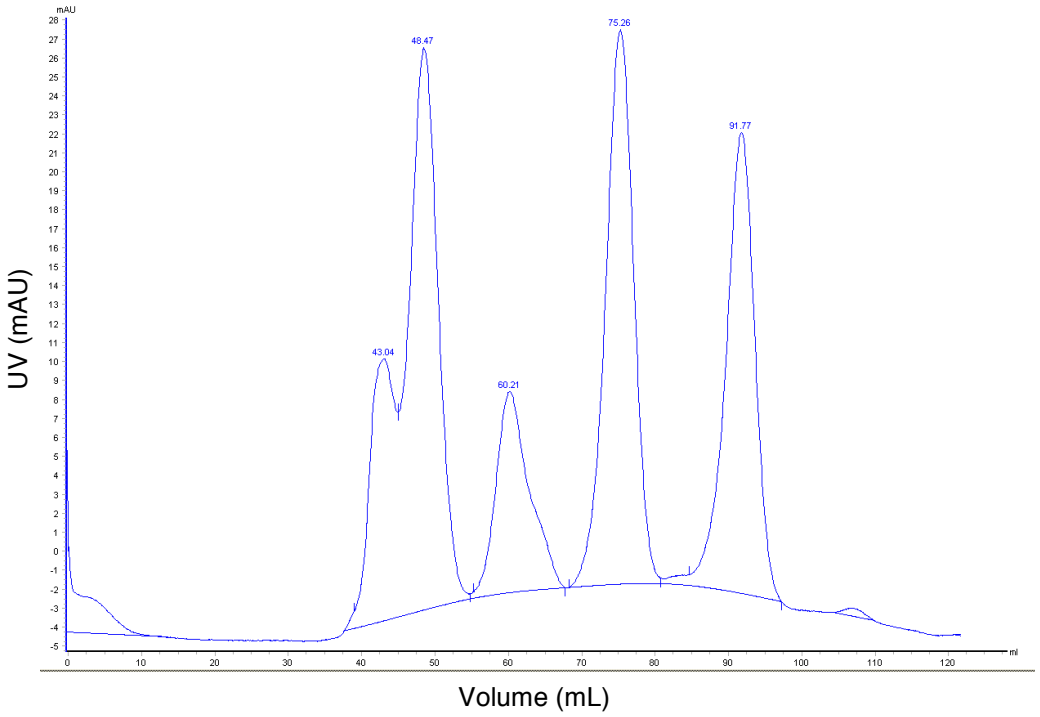


Figure 2.1. Calibration chromatogram for Superdex200 pg gel filtration column run with 50 mM HEPES pH 7.5, 100 mM NaCl. Chromatogram produced by running Ferritin, Aldolase, Ovalbumin and Ribonuclease A on the Superdex200 pg gel filtration column. mAU (mili absorbance unit) at A_{280} – blue. Elution volume values are shown for each elution peak.

Protein	**Elution volume (V_e)**	**K_{av} = (V_e-41)/75**	**MW (Da)**	**LogMW**
Ferritin	48.47	0.10	440000	5.64
Aldolase	60.21	0.26	158000	5.20
Ovalbumin	75.26	0.46	43000	4.63
RibonucleaseA	91.77	0.68	13700	4.14

Table 2.1 Calculation of the K_{av} value of each protein vs the LogMW. The K_{av} value is calculated as described in 2.1.14. The K_{av} value has been rounded to two significant figures. The molecular weight for each protein is given in Da (Daltons) (GE). The LogMW has been rounded to two significant figures. The K_{av} vs LogMW values were plotted to produce the graph in Figure 5.24.

2.2 Construct information

2.2.1 Primers

Primers for BPSS0211, BPSS0212+M, BPSS0213 and BPSS0214 were designed previously by Dr. Matthew Day, (Day, 2012). Primers for BPSS0211(1-123), BPSS0212 (Forward), BPSS0212ΔDUF(Domain of unknown function)1843 and BPSS0213ΔDUF1843 were designed using the *B. pseudomallei* K96243 genome sequence (Winsor et al., 2008, Holden et al., 2004). The restriction enzymes used were – NdeI (CATATG), which incorporated the first methionine ATG, BamHI (GGATCC) and XhoI (CTCGAG). Primers were ordered from Eurofins.

Primer	Primer sequence
0211(1-123) F	5'-TAAGTA**CATATG**GCTAATCACGGGCGGCGA-3'
0211(1-123) R	5'-ATATAT**GGATCC**CTACTTCCTCCGCTCGAGCTT-3'
0212+M F*	5'-ATATAT**CATATG**ATGTCCGAAGATCTTCG-3'
0212 F	5'-ATATAT**CATATG**TCCGAAGATCTTCGCGT-3'
0212 R*	5'-ATATAT**GGATCC**TCACTGCCGGCTTTCGA-3'
0212Δ DUF1843 F	5'-AGCTAT**CATATG**TCCGAAGATCTTCGCGTCGG-3'
0212Δ DUF1843 R	5'-ATATAT**CTCGAG**TCACGCGGCCGATCCGAT-3'
0213Δ DUF1843 F	5'-ATATAT**CATATG**GCAACTACCGGTCTCTTTCCCG-3'
0213Δ DUF1843 R	5'-ATATAT**CTCGAG**TCACGGCATCACGTGCTCGCTC-3'

F – Forward primer

R – Reverse primer

*0212+M F and 0212 R were designed by Dr. Matthew Day (Day, 2012).

2.2.2 Constructs cloned by Dr. Matthew Day

The amino acid sequences for the constructs produced previously by Dr. Matthew Day (Day, 2012), and used in the investigations described in this thesis.

BPSS0211

MSQAQGHPVTPYGVAIHQAIADGDLAQMKSLRTQAQALLAQQGNLATALELLEV
EIAKLERRK

BPSS0213

MATTGLFPVQLRVATPNLGAPVLWLNLLVNTVEKTASGFARITQTVYPPMHFRA
RVVGPFHQMRIDPHAPQSVTLTLSGSPTGPVAPQVVILELNALLNEGWQSGTAN
YRYFYESRWHSIEHAIVSKDNSRIPLDPPSEHVMPMYGVGLQEARASGDLSRMK
ALAQQAEQQLADHDVIAAELQKLEAEIARLEARR

2.2.3 Constructs cloned as part of this thesis

The amino acid sequences for the constructs cloned during the work carried out in this thesis.

BPSS0211(1-123)

MANHGRRAVRHEAPCRSMAASATFIDEIPRSCTNTTAGRTVPPRGRSQGCRPSS
TFSEANMSQAQGHPVTPYGVAIHQAIADGDLAQMKSLRTQAQALLAQQGNLATA
LELLEVEIAKLERRK

BPSS0212+M

MMSEDLRVGLFPVRYLVGTGLPGAPQLVLDLMVDTVDHSVVGRAAVSQAVSPPL
NFHADVWGSYVFRLGPPPRRDGSGAIVQISLQGNQGGPQSNSMITFYGELLLKG
DGKTGVASYRYYSNGSWHEVENVPVKADPELVPIEPGPVIGQSSMSAIGSAAMY
GVAIQSAAASGDLAHMRTLSAYARQQLESRDEIAAALSELKAEIAKLESRQ

BPSS0212

MSEDLRVGLFPVRYLVGTGLPGAPQLVLDLMVDTVDHSVVGRAAVSQAVSPPLN
FHADVWGSYVFRLGPPPRRDGSGAIVQISLQGNQGGPQSNSMITFYGELLLKGD

GKTGVASYRYYSNGSWHEVENVPVKADPELVPIEPGPVIGQSSMSAIGSAAMYG
VAIQSAAASGDLAHMRTLSAYARQQLESRDEIAAALSELKAEIAKLESRQ

BPSS0212ΔDUF1843

MSEDLRVGLFPVRYLVGTGLPGAPQLVLDLMVDTVDHSVVGRAAVSQAVSPPLN
FHADVWGSYVFRLGPPPRRDGSGAIVQISLQGNQGGPQSNSMITFYGELLLKGD
GKTGVASYRYYSNGSWHEVENVPVKADPELVPIEPGPVIGQSSMSAIGSAA

BPSS0213ΔDUF1843

MATTGLFPVQLRVATPNLGAPVLWLNLLVNTVEKTASGFARITQTVYPPMHFRA
RVVGPFHQMRIDPHAPQSVTLTLSGSPTGPVAPQVVILELNALLNEGWQSGTAN
YRYFYESRWHSIEHAIVSKDNSRIPLDPPSEHVMP

2.3 Cloning

All constructs were cloned into pET21a plasmid containing an ampicillin resistance marker.

2.3.1 Cloning template

Template DNA used for cloning was either gDNA (genomic DNA) from *B. pseudomallei* D286 or plasmid containing the template gene. *B. pseudomallei* D286 gDNA was isolated from a patient at Kuala Lumpur General Hospital (Lee et al., 2007) (supplied by Prof. Sheila Nathan, UKM), and diluted prior to use. Plasmid template was prepared using the QIAprep Spin Miniprep Kit with microcentrifuge Protocol (Qiagen) and eluted in autoclaved MiliQ water. DMSO (dimethylsulfoxide) was added to the PCR reaction due to the high GC content of the *B. pseudomallei* genome (Baskaran et al., 1996; Holden et al., 2004).

2.3.2 Primer preparation

Forward and reverse primers used in cloning were dissolved in the set volume stated by Eurofins to produce a 100 pmol/µL stock solution. 10 µL of 100 pmol/µL forward primer stock and 10 µL of 100 pmol/µL reverse primer stock was combined with 80 µL of autoclaved MiliQ water, to

produce the 10 pmol/µL solution of both primers used in cloning. The T_m (melting temperature) values for each primer were calculated using the OligoAnalzyer website (Integrated DNA Technologies).

2.3.3 Plasmid generation

E. coli DH5a cells (Invitrogen) were transformed with empty pET21a plasmid, as described in section 2.1.4. A single transformation colony was transferred into 5 mL of LB (section 2.1.8) containing 100 µg/mL of ampicillin, and grown overnight at 250 rpm and 37°C. Overnight cultures were pelleted at 5248 *g* for 20 minutes and the supernatant removed. Plasmid was isolated from cell pellets using the QIAprep Spin Miniprep Kit with microcentrifuge Protocol (Qiagen) and eluted in 50 µL autoclaved MiliQ water. Plasmid was digested using the same protocol as the corresponding insert or previously digested plasmid was used.

2.3.4 BPSS0211(1-123) cloning protocol

BPSS0211(1-123) was cloned from *B. pseudomallei* D286 gDNA. The PCR protocol reaction was prepared as follows:

20 µL Biomix Red (Bioline)
1 µL Forward and Reverse primer dilution (10 pmol/µL)
1 µL *B. pseudomallei* D286 gDNA (200 ng/µL)
0/1/2 µL DMSO (0/2.5/5% DMSO) (Sigma)
total volume up to 40 µL with autoclaved MiliQ water

The T_m values for the forward and reverse primers were 63.9°C and 63.8°C, respectively. The PCR reaction was performed using the following protocol:

95°C for 5 minutes
30 cycles of:
    95°C for 30 seconds
    60°C for 30 seconds
    72°C for 1 minute

72°C for 5 minutes

5 µL of each reaction (0, 2.5 and 5% DMSO) was loaded and run on a 1% TAE agarose gel (section 2.1.2). The 2.5% DMSO PCR experiment produced a DNA band of appropriate size. The other ~35 µL of solution was purified using QIAquick PCR Purification Kit with microcentrifuge protocol (Qiagen) and eluted in 50 µL autoclaved MiliQ water.

The purified insert was digested with NdeI restriction enzyme using the following recipe:
6 µL CutSmart buffer (NEB)
35 µL BPSS0211(1-123) purified insert
4 µL restriction enzyme NdeI (NEB)
up to 60 µL with autoclaved MiliQ water
The reaction was incubated at 37°C for 2 hours and the resulting reaction was purified using QIAquick PCR Purification Kit with microcentrifuge protocol (Qiagen) and eluted in 50 µL autoclaved MiliQ water.

The purified insert was then digested with the second restriction enzyme BamHI, using the following recipe:
6 µL CutSmart buffer (NEB)
35 µL BPSS0211(1-123) Nde1 digested purified insert
4 µL restriction enzyme BamHI (NEB)
up to 60 µL with autoclaved MiliQ water
The reaction was incubated at 37°C for 2 hours and the resulting reaction was purified using QIAquick PCR Purification Kit with microcentrifuge protocol (Qiagen) and eluted in 50 µL autoclaved MiliQ water.

The ligation reaction between BPSS0211(1-123) insert and pET21a plasmid was set up as follows:
8 µL pET21a plasmid (digested with NdeI and BamHI)
8 µL BPSS0211(1-123) insert (digested with NdeI and BamHI)

2 µL 10x T4 ligase buffer (NEB)

1 µL T4 ligase (NEB)

The reaction was incubated at 16°C for 2 hours. The ligated plasmid was transformed into *E. coli* cloning strain cells as outlined in section 2.1.4.

2.3.5 BPSS0212+M cloning protocol

BPSS0212+M was cloned from *B. pseudomallei* D286 gDNA. The PCR protocol reaction was prepared as follows:

20 µL Biomix Red (Bioline)

1 µL Forward and Reverse primer dilution (10 pmol/µL)

1 µL *B. pseudomallei* D286 gDNA (~1000 ng)

0/1/2 µL DMSO (0/2.5/5% DMSO) (Sigma)

up to 40 µL with autoclaved MiliQ water

The T_m values for the forward and reverse primers were 57.3°C and 62.4°C, respectively. The PCR reaction was performed using the following protocol:

95°C for 5 minutes

30 cycles of:

- 95°C for 30 seconds
- 58°C for 30 seconds
- 72°C for 1 minute

72°C for 5 minutes

Each reaction (0, 2.5 and 5% DMSO) was loaded and run on a 1% agarose TAE gel (section 2.1.2). The 2.5% DMSO PCR experiment produced a DNA band of appropriate size. The band was gel extracted using QIAquick Gel Extraction Kit with microcentrifuge protocol (Qiagen) and eluted in 50 µL autoclaved MiliQ water.

The gel extracted insert was double digested with restriction enzymes NdeI and BamHI using the following recipe:

6 µL 10x NEB Buffer 3 (NEB)
50 µL BPSS0212 gel extracted insert
2 µL restriction enzyme NdeI (NEB)
2 µL restriction enzyme BamHI (NEB)
The reaction was incubated at 37°C for 6 hours followed by 65°C for 30 minutes and the resulting reaction was purified using QIAquick PCR Purification Kit with microcentrifuge protocol (Qiagen) and eluted in 65 µL autoclaved MiliQ water.

The ligation reaction between BPSS0212+M insert and pET21a plasmid was set up as follows:
8 µL pET21a plasmid (digested with NdeI and BamHI)
8 µL BPSS0212+M insert (digested with NdeI and BamHI)
2 µL 10x T4 ligase buffer (NEB)
1 µL T4 ligase (NEB)
The reaction was incubated at 16°C for 1 hour. *E. coli* cloning strain cells were transformed with the ligated plasmid as outlined in section 2.1.4.

2.3.6 BPSS0212 cloning protocol

BPSS0212 was cloned from pET21a plasmid containing the BPSS0212+M gene. The PCR protocol reaction was prepared as follows:
20 µL Biomix Red (Bioline)
1 µL Forward and Reverse primer dilution (10 pmol/µL)
2 µL pET21a plasmid containing BPSS0212+M gene (27.3 ng/µL)
0/1/2 µL DMSO (0/2.5/5% DMSO) (Sigma)
up to 40 µL with autoclaved MiliQ water

The T_m values for the forward and reverse primers were 54.3°C and 62.4°C, respectively. The PCR reaction was performed using the following protocol:
95°C for 5 minutes
30 cycles of:

95°C for 30 seconds

55°C for 30 seconds

72°C for 1 minute

72°C for 5 minutes

5 µL of each reaction (0, 2.5 and 5% DMSO) was loaded and run on a 1% TAE agarose gel (section 2.1.2). The 2.5% DMSO PCR experiment produced a DNA band of appropriate size. The other ~35 µL of solution was purified using QIAquick PCR Purification Kit with microcentrifuge protocol (Qiagen) and eluted in 50 µL autoclaved MiliQ water.

The purified insert was digested with NdeI restriction enzyme using the following recipe:
6 µL CutSmart buffer (NEB)
40 µL BPSS0212 purified insert
4 µL restriction enzyme NdeI (NEB)
up to 60 µL with autoclaved MiliQ water
The reaction was incubated at 37°C for 2 hours, followed by 65°C for 30 minutes and the resulting reaction was analysed by 1% agarose TAE gel (section 2.1.2). The remainder of the reaction was purified using QIAquick PCR Purification Kit with microcentrifuge protocol (Qiagen) and eluted in 50 µL autoclaved MiliQ water.

The purified insert was then double digested with NdeI and a second restriction enzyme BamHI, using the following recipe:
6 µL Buffer NEB 3.1 (NEB)
40 µL BPSS0212 Nde1 digested purified insert
4 µL restriction enzyme BamHI (NEB)
2 µL restriction enzyme NdeI (NEB)
up to 60 µL with autoclaved MiliQ water
The reaction was incubated at 37°C for 2 hours, followed by 65°C for 30 minutes and the resulting reaction was analysed by 1% agarose TAE gel

(section 2.1.2). The remainder of the reaction was purified using QIAquick PCR Purification Kit with microcentrifuge protocol (Qiagen) and eluted in 50 µL autoclaved MiliQ water

The ligation reaction between BPSS0212 insert and pET21a plasmid was set up as follows:
8 µL pET21a plasmid (digested with NdeI and BamHI)
8 µL BPSS0212 insert (digested with NdeI and BamHI)
2 µL 10x T4 ligase buffer (NEB)
1 µL T4 ligase (NEB)
The reaction was incubated at 16°C for 2 hours. *E. coli* DH5α silver efficiency cells were transformed with the ligated plasmid as outlined in section 2.1.4.

2.3.7 BPSS0212ΔDUF1843 cloning protocol

BPSS0212ΔDUF1843 was cloned from pET21a plasmid containing the BPSS0212 gene. The PCR protocol reaction was prepared as follows:
20 µL Biomix Red (Bioline)
1 µL Forward and Reverse primer dilution (10 pmol/µL)
1 µL pET21a plasmid containing BPSS0212 (65 ng/µL)
0/1/2 µL DMSO (0/2.5/5% DMSO) (Sigma)
up to 40 µL with autoclaved MiliQ water

The T_m values for the forward and reverse primers were 63.5°C and 64.9°C, respectively. The PCR reaction was performed using the following protocol:
95°C for 5 minutes
30 cycles of:
    95°C for 30 seconds
    59°C for 30 seconds
    72°C for 1 minute
72°C for 5 minutes

5 µL of each reaction (0, 2.5 and 5% DMSO) was loaded and run on a 1% TAE agarose gel (section 2.1.2). The 0% DMSO PCR experiment produced a DNA band of appropriate size. The other ~35 µL of solution was purified using QIAquick PCR Purification Kit with microcentrifuge protocol (Qiagen) and eluted in 50 µL autoclaved MiliQ water.

The purified insert was digested with NdeI restriction enzyme using the following recipe:
6 µL CutSmart buffer (NEB)
40 µL BPSS0212ΔDUF1843 purified insert
4 µL restriction enzyme NdeI (NEB)
up to 60 µL with autoclaved MiliQ water
The reaction was incubated at 37°C for 2 hours, followed by 65°C for 30 minutes and the resulting reaction was purified using QIAquick PCR Purification Kit with microcentrifuge protocol (Qiagen) and eluted in 50 µL autoclaved MiliQ water.

The purified insert was then double digested with NdeI and a second restriction enzyme XhoI, using the following recipe:
6 µL CutSmart buffer (NEB)
40 µL BPSS0212ΔDUF1843 purified insert (digested with NdeI)
4 µL restriction enzyme XhoI (NEB)
2 µL restriction enzyme NdeI (NEB)
up to 60 µL with autoclaved MiliQ water
The reaction was incubated at 37°C for 2 hours, followed by 65°C for 30 minutes and the resulting reaction was purified using QIAquick PCR Purification Kit with microcentrifuge protocol (Qiagen) and eluted in 50 µL autoclaved MiliQ water.

The ligation reaction between BPSS0212ΔDUF1843 insert and pET21a plasmid was set up as follows:
1 µL pET21a plasmid (digested with NdeI and XhoI)

3 µL BPSS0212ΔDUF1843 purified insert (digested with NdeI and XhoI)
0.5 µL T4 ligase buffer (NEB)
1-2 µL T4 ligase (NEB)
The reaction was incubated at 16°C for 1.5 hours. *E. coli* cloning strain cells were transformed with the ligated plasmid as outlined in section 2.1.4.

2.3.8 BPSS0213ΔDUF1843 cloning protocol

BPSS0213ΔDUF1843 was cloned from pET21a plasmid containing the BPSS0213 gene. The PCR protocol reaction was prepared as follows:
20 µL Biomix Red (Bioline)
1 µL Forward and Reverse primer dilution (10 pmol/µL)
1 µL pET21a plasmid containing BPSS0213 (52 ng/µL)
0/1/2 µL DMSO (0/2.5/5% DMSO) (Sigma)
up to 40 µL with autoclaved MiliQ water

The T_m values for the forward and reverse primers were 61.6°C and 65.7°C, respectively. The PCR reaction was performed using the following protocol:
95°C for 5 minutes
30 cycles of:
    95°C for 30 seconds
    58°C for 30 seconds
    72°C for 1 minute
72°C for 5 minutes
5 µL of each reaction (0, 2.5 and 5% DMSO) was loaded and run on a 1% TAE agarose gel (section 2.1.2). The 0% DMSO PCR experiment produced a DNA band of appropriate size. The other ~35 µL of solution was purified using QIAquick PCR Purification Kit with microcentrifuge protocol (Qiagen) and eluted in 50 µL autoclaved MiliQ water.

The purified insert was digested with NdeI restriction enzyme using the following recipe:
6 µL CutSmart buffer (NEB)
40 µL BPSS0213ΔDUF1843 purified insert
4 µL restriction enzyme NdeI (NEB)
up to 60 µL with autoclaved MiliQ water
The reaction was incubated at 37°C for 2 hours, then 65°C for 30 minutes and the resulting reaction was purified using QIAquick PCR Purification Kit with microcentrifuge protocol (Qiagen) and eluted in 50 µL autoclaved MiliQ water.

The purified insert was then double digested with NdeI and a second restriction enzyme XhoI, using the following recipe:
6 µL CutSmart buffer (NEB)
35 µL BPSS0213ΔDUF1843 purified insert (digested with NdeI)
4 µL restriction enzyme XhoI (NEB)
2 µL restriction enzyme NdeI (NEB)
up to 60 µL with autoclaved MiliQ water
The reaction was incubated at 37°C for 2 hours, then 65°C for 30 minutes and the resulting reaction was purified using QIAquick PCR Purification Kit with microcentrifuge protocol (Qiagen) and eluted in 50 µL autoclaved MiliQ water.

The ligation reaction between BPSS0213ΔDUF1843 insert and pET21a plasmid was set up as follows:
1 µL pET21a plasmid (digested with NdeI and XhoI)
3 µL BPSS0213ΔDUF1843 purified insert (digested with NdeI and XhoI)
0.5 µL 10x T4 ligase buffer (NEB)
1-2 µL T4 ligase (NEB)
The reaction was incubated at 16°C for 1.5 hours. *E. coli* cloning strain cells were transformed with the ligated plasmid as outlined in section 2.1.4.

2.4 Protein overexpression

Small-scale overexpression tests were carried out for BPSS0211(1-123), BPSS0212ΔDUF1843 and BPSS0213ΔDUF1843. However, as overexpression conditions that produced soluble BPSS0211, BPSS0212 and BPSS0213 protein had been elucidated previously by Dr. Matthew Day (Day, 2012), these conditions continued to be used for large-scale overexpression of BPSS0211, BPSS0212+M, BPSS0212+M-SeMet, BPSS0212 and BPSS0213.

2.4.1 Small-scale protein overexpression tests

To identify the overexpression conditions that produce soluble protein, small-scale overexpression tests were carried out. *E. coli* BL21 DE3 cells (Agilent) were transformed with plasmid containing the desired gene (section 2.1.5). A single colony from the transformation plate was used to inoculate 5 mL of LB (section 2.1.8) containing the appropriate antibiotics, and the cells were grown overnight at 37°C, 250 rpm. The overnight starter culture was used to inoculate 50 mL of LB (section 2.1.8) in 250 mL conical flasks with a 2% inoculation (1 mL of culture). The cells were grown at 37°C until the OD_{600nm} reached between 0.6-0.8, using water as a reference. 2 x 1 mL pre-induction samples were taken, centrifuged at 13000 rpm/10-15500 *g* for 10 minutes, the supernatant was removed and pellets stored at -20°C. Protein overexpression was induced by the addition of 1 mM IPTG (isopropyl β-D-1-thiogalactopyranoside) (generon) and cells were incubated at a combination of temperatures from a selection of: 12°C, 18°C, 25°C, 30°C, 37°C, also 12°C with a 20-30 minute pre-induction ice incubation. 2 x 1 mL samples were taken at time points throughout the test overexpression, centrifuged at 13000 rpm/10-15500 *g* for 10 minutes, the supernatant was removed and pellets frozen at -20°C. Protein overexpression was analysed by SDS-PAGE of the pre-induction samples, time interval samples and post-induction samples.

2.4.2 Protein construct overexpression conditions

Construct	Temperature	Time	IPTG concentration
BPSS0211*	37°C	3 hours	1 mM
BPSS0211(1-123)	37°C	2 hours	1 mM
BPSS0212ΔDUF1843	37°C	2 hours	1 mM
BPSS0213ΔDUF1843	25°C	overnight	1 mM
$TssA_{CTD}$(303-373)-SeMet**	37°C	overnight	0.3 mM
Constructs below required a 20 minute ice incubation prior to induction of protein expression			
BPSS0212+M*	12°C	24 hours	1 mM
BPSS0212*	12°C	24 hours	1 mM
BPSS0212+M-SeMet*	12°C	48 hours	1 mM
BPSS0213*	12°C	24 hours	1 mM

*overexpression protocols from Dr. Matthew Day (Day, 2012). **overexpression protocol adapted from those carried out by Ruyue Sun.

2.4.3 Large-scale protein overexpression

E. coli BL21 DE3 cells (Agilent) were transformed with plasmid containing the appropriate gene (section 2.1.5). A single colony was used to inoculate 50 mL of LB (section 2.1.8), aliquoted in 250 mL conical flasks, containing the appropriate antibiotic. The starter culture was grown overnight at 37°C, 200-250 rpm. 2 L conical flasks containing 500 mL of LB (section 2.1.8), containing the appropriate antibiotic, were inoculated with a 2% inoculation, (10 mL) of the overnight starter culture. The cells were grown at 37°C until OD_{600nm} reached between 0.6-0.8, using water as a reference. 2 x 1 mL pre-induction samples were taken, centrifuged at 13000 rpm/10-15500 *g* for 10 minutes, the supernatant was removed and pellets frozen at -20°C. If required the flasks were incubated on ice for 20-30 minutes prior to induction. Protein overexpression was induced by the addition of IPTG (generon) and cells were incubated at the required

temperature and length of time (specified in section 2.4.2). 2 x 1 mL post-induction samples were taken, centrifuged at 13000 rpm/10-15500 *g* for 10 minutes, the supernatant was removed and pellets frozen at -20°C. Cells were harvested by centrifugation at ~4000 rpm in a Beckman-J6B centrifuge (Beckman Coulter) for 35 minutes or ~5000 rpm in an F10-6 x 500y rotor (Thermo Fisher Scientific) for 15 minutes. The resulting cell pellets were resuspended in ~100 mL supernatant and repelleted in 50 mL Falcon tubes at 5248 *g* for 20 minutes. The supernatant was removed and pellets were stored at -20°C. Protein overexpression was analysed by SDS-PAGE of the pre-induction and post-induction samples.

2.4.4 Selenomethionine-incorporated protein overexpression

Selenomethionine-incorporated protein overexpression was performed using selenomethionine minimal media, the recipe for which is outlined below:

Selenomethionine minimal medium	
	Per litre
K_2HPO_4	10.5 g
$(NH_4)_2SO_4$	1 g
KH_2PO_4	4.5 g
$Na_3C_6H_5O_7.2H_2O$	500 mg
Glycerol	5 g
Adenine	500 mg
Guanosine	500 mg
Thymine	500 mg
Uracil	500 mg
Glucose* (for $TssA_{CTD}$(303-373)-SeMet)	2 g

The components for the selenomethionine minimal media were dissolved in deionised water and then autoclaved.

Additives	Per litre
$MgSO_4.7H_2O$	1 mg/1 g
Thiamine	4 mg
L-lysine	100 mg
L-phenylalanine	100 mg
L-threonine	100 mg
L-isoleucine	50 mg
L-leucine	50 mg
L-valine	50 mg

Additives were produced as a 50x stock, dissolved in deionised water. The pH of the solution was adjusted to pH 12.0, then back to pH 8.0 in order for the amino acids to dissolve. The stock solution was stored at -20°C and thawed when required.

	Per litre
L-Selenomethionine	40 mg

Selenomethionine was added prior to inoculation.

BPSS0212+M-SeMet

Jason Tsang, an MBiolSci Masters project student, carried out the production of selenomethionine-incorporated BPSS0212+M protein, following the protocol outlined in Dr. Matthew Day's thesis (Day, 2012).

$TssA_{CTD}$(303-373)-SeMet

The selenomethionine incorporated protein overexpression protocol for $TssA_{CTD}$(303-373) uses information from the native overexpression protocol from Ruyue Sun, PhD student at Sheffield Medical School, and the pMAL expression protocol (NEB). Both selenomethionine minimal medium and rich medium were used in the protein overexpression. In addition 0.2% glucose, present in the rich media, was included in the selenomethionine minimal medium recipe* or added to the media prior to

inoculation as a stock solution. The glucose stock solution was produced by dissolving glucose into autoclaved MiliQ water, and then passing the solution through a filter. The recipe for pMAL rich media (NEB) is outlined below:

pMAL rich medium (NEB)	**Per litre**
Tryptone	10 g
Yeast extract	5 g
NaCl	5 g
Glucose	2 g

400 µL of NEB Express cells (NEB) (provided by Ruyue Sun) were transformed with 2 µL of pMAL plasmid containing $TssA_{CTD}$(303-373) (provided by Ruyue Sun) and plated onto LB agar (section 2.1.9) with 100 µg/mL ampicillin. Individual transformation colonies were used to inoculate each of 2 x 50 mL of rich medium containing 100 µg/mL ampicillin in 250 mL flasks. Cultures were grown overnight at 250 rpm and 37°C. 2 x 2 L conical flasks containing 500 mL rich medium and 100 µg/mL ampicillin were inoculated with 2% (10 mL) of the overnight starter culture. The cells were grown at 37°C, 250 rpm until OD_{600nm} reached ~0.6-0.7, using water as a reference. The cells were pelleted at 5353 *g* for 20 minutes. The cell pellets were then resuspended in 100 mL of selenomethionine minimal media (section 2.4.4). Resuspended cells were then split into 2 x 50 mL Falcon tubes and repelleted at 5000 *g* for 30 minutes. Each pellet was then resuspended in 10 mL of selenomethionine minimal media and each resuspended pellet added to 500 mL of selenomethionine minimal media (section 2.4.4) containing 100 µg/mL ampicillin, 10 mL of 50x additives (section 2.4.4), (additional 0.2% glucose – if not already present in medium) and 40 mg/L selenomethionine. The cells were then grown at 37°C and 250 rpm for 1 hour. 2 x 1 mL pre-induction samples were taken from each flask, centrifuged at 13000 rpm/10-15500 *g* for 10 minutes, the supernatant was removed and pellets frozen at -20°C. Protein overexpression was induced

by the addition of 0.3 mM IPTG (generon) and cells were incubated at 37°C, with protein overexpressed overnight. 2 x 1 mL post-induction samples were taken, centrifuged at 13000 rpm/10-15500 *g* for 10 minutes, the supernatant was removed and pellets frozen at -20°C. Cells were harvested by centrifugation at 5353 *g* for 20 minutes. Pellets were resuspended in ~100 mL supernatant prior to repelleting by centrifugation at 5248 *g* for 20 minutes. Supernatant was discarded and pellets stored at -20°C. Protein overexpression was analysed by SDS-PAGE of the pre-induction and post-induction samples.

2.4.5 Analysis of protein overexpression

To analyse the results of both small-scale and large-scale protein overexpression experiments, the pre- and post-induction cell samples (and any other time point samples) were run on an SDS-PAG (section 2.1.10). For small-scale protein overexpression tests the pre-induction, a number of time point samples and the post-induction samples were analysed. For large-scale protein overexpression the pre-induction and post-induction samples were analysed.

Cell pellets were thawed on ice and then resuspended in 50 µL of 50 mM Tris (Tris(hydroxymethyl)methyl amine) pH 8.0 or MiliQ. Resuspended cell pellets were sonicated at 16 microns for 3 x 5 seconds, incubated on ice between each cycle. The sonicated cells were centrifuged at 13000 rpm/10-15500 *g* for 10 minutes and the soluble fraction removed to a fresh eppendorf tube. The insoluble pellet was resuspended in 4% SDS and incubated at room temperature for 20-30 minutes. The insoluble samples were then centrifuged at 13000 rpm/10-15500 *g* for 10 minutes, and the sample for the gel was taken from the supernatant.

The protein concentration of each soluble sample was analysed using the Bradford Assay (section 2.1.12). The volume of soluble sample needed to produce a final concentration of 1 mg/mL or 20 µg in 15 µL, was

calculated. The gel samples were prepared as specified in section 2.1.10. For a specific sample the same volume of insoluble as soluble was loaded onto the gel. The samples were then heated at ~100°C for ~5 minutes, followed by pulse centrifugation. SDS-PAGE was performed as outlined in section 2.1.10.

2.5 Protein purification of individual protein constructs

The protein purification protocols for BPSS0211, BPSS0211(1-123) and BPSS0212+M, BPSS0212+M-SeMet and BPSS0212 are outlined in chapters 3 and 4, respectively.

2.5.1 Analysis of protein purification

To analyse the results of the protein purification process, samples from the different stages of purification were run on an SDS-PAG. For chromatography elution fraction samples, the same volume of sample was loaded onto the gel for each fraction. Samples were prepared and the gel run as outlined in section 2.1.10.

2.6 Protein purification for complex formation

The protein purification protocols for BPSS0211, BPSS0211(1-123), BPSS0212+M, BPSS0212, BPSS0213, BPSS0212ΔDUF1843 and BPSS0213ΔDUF1843 involved in complex formation experiments, as discussed in chapter 5, are outlined below. For purification in the presence of protease inhibitors and EDTA, One tablet of EDTA-free protease inhibitors (cOmplete - Roche) was added, per maximum 50 mL lysis buffer, prior to resuspension of cell pellet, and 2 mM EDTA was added to all purification buffers.

2.6.1 BPSS0211

E. coli BL21 DE3 cell pellet containing BPSS0211 was resuspended in 50 mM Tris pH 8.0 buffer and sonicated for 3 x 20 seconds at 16 microns, incubated on ice between each cycle, then centrifuged at 18000

rpm/39191 *g* in a JA.20 rotor (Beckman Coulter) for 15 minutes. The protein content of the cell extract was analysed using the Bradford Assay (section 2.1.12). The cell extract was separated into two batches and each loaded onto a HiTrap DEAE (diethylaminoethanol) FF 5 mL column (GE) pre-equilibrated with Buffer A (50 mM Tris pH8.0) in two separate runs. The column was washed with Buffer A for 15 mL (3 CV (column volume)) and 6 mL fractions were collected. The column was run with a gradient of 0-100% Buffer B (50 mM Tris pH 8.0, 1 M NaCl) in 0 mL (0 CV), followed by 100% Buffer B in 2CV and 6 mL fractions were collected. The column was run at 5 mL/minute. In both runs, BPSS0211 eluted in the flow-through fractions, all flow-through fractions were pooled. An equal amount of 4 M ammonium sulphate was added to the pooled flow-through fractions to produce a final concentration of 2 M ammonium sulphate. The mixture was left on ice for 20 minutes and then centrifuged at 18000 rpm/39191 *g* in a JA.20 rotor (Beckman Coulter) for 20 minutes. The resulting pellet was dissolved in 20 mL 50 mM Tris pH 8.0 and the ammonium sulphate concentration was analysed using a sugar refractometer and was adjusted to 1.5 M with 4 M ammonium sulphate. The solution was left on ice for 10 minutes and then centrifuged at 18000 rpm/39191 *g* in a JA.20 rotor (Beckman Coulter) for 20 minutes. The ammonium sulphate concentration of the supernatant was analysed using a sugar refractometer and was adjusted to 2 M using the addition of 4 M ammonium sulphate. The solution was incubated at room temperature for 5 minutes. The solution was centrifuged at 18000 rpm/39191 *g* in a JA.20 rotor (Beckman Coulter) for 10 minutes and the resulting pellet resuspended in 1 mL 50 mM Tris pH 8.0. The resuspended pellet was made up to 2 mL with 50 mM Tris pH 8.0, and run on a HiLoad 16/600 120 mL Superdex 200 pg gel filtration column (GE) pre-equilibrated with 50 mM Tris pH 8.0, 500 mM NaCl. The column was run at 1.5 mL/minute and 2.5 mL fractions were collected over 0.6 CV-0.9 CV.

2.6.2 BPSS0211(1-123)

The purification of BPSS0211(1-123) for complex formation was performed by Dr. Svetlana Sedelnikova (University of Sheffield). Following induction, *E. coli* BL21 DE3 cell pellet containing overexpressed BPSS0211(1-123) was resuspended in 50 mM HEPES pH 7.0 buffer, sonicated and centrifuged. The cell extract was loaded onto a HiTrap SP (sulfopropyl) HP 5 mL column (GE) pre-equilibrated with 50 mM HEPES pH 7.0. BPSS0211(1-123) was eluted over a gradient of 0-100% 50 mM HEPES pH 7.0, 1 M NaCl, in 50 mL (10 CV) at 5 mL/minute with 2.5 mL fractions collected.

2.6.3 BPSS0212(+M) – one-step purification

E. coli BL21 DE3 cell pellet containing overexpressed BPSS0212, grown as above, was resuspended in 10 mL/g of lysis buffer (50 mM Tris pH 9.0). Cells were sonicated at 16 microns for 3 x 20 seconds, incubated on ice between each cycle, and centrifuged at 18000 rpm/39191 *g* in a JA.20 rotor (Beckman Coulter) for 15-20 minutes. The protein content of the cell extract was analysed using the Bradford Assay (section 2.1.12). Cell extract was loaded onto a HiTrap DEAE FF 5 mL column (GE) pre-equilibrated with Buffer A (50 mM Tris pH 9.0), and 5 mL fractions were collected. The column was washed with Buffer A for 5 mL (1 CV). BPSS0212 was eluted over 0-25% Buffer B (50 mM Tris pH 9.0, 1 M NaCl) in 100 mL (20 CV) and 2.5 mL fractions were collected. The column was run at 5 mL/minute. For the purification of BPSS0212+M-SeMet, two DEAE columns were run sequentially due to the large amount of protein present in the cell extract.

For the initial complex formation experiment (complex 5.1), BPSS0212+M was purified using a HiTrap DEAE FF 5mL column (GE), and eluted over a gradient of 0-25% 50 mM Tris pH 9.0, 1 M NaCl, over 50 mL (10 CV).

2.6.4 BPSS0212 – two-step purification

E. coli BL21 DE3 cell pellet containing overexpressed BPSS0212, grown as above, was resuspended in 10 mL/g of lysis buffer (50 mM Tris pH 9.0). Cells were sonicated at 16 microns for 3 x 20 seconds, incubated on ice between each cycle, and centrifuged at 18000 rpm/39191 *g* in a JA.20 rotor (Beckman Coulter) for 15-20 minutes. The protein content of the cell extract was analysed using the Bradford Assay (section 2.1.12). Cell extract was loaded onto a HiTrap DEAE FF 5 mL column (GE) pre-equilibrated with Buffer A (50 mM Tris pH 9.0), and 5 mL fractions were collected. The column was washed with Buffer A for 5 mL (1 CV). BPSS0212 was eluted over 0-25% Buffer B (50 mM Tris pH 9.0, 1 M NaCl) in 100 mL (20 CV) and 2.5 mL fractions were collected. The column was run at 5 mL/minute. For the purification of BPSS0212 with protease inhibitors and EDTA, two DEAE columns were run together due to the large amount of protein present in the cell extract. Elution fractions were analysed by SDS-PAGE (section 2.1.10). Elution fractions containing BPSS0212 were pooled and concentrated using a Vivaspin concentrator (Sartorius Stedem), to ~2 mL or made up to 2 mL with gel filtration buffer (50 mM Tris pH 8.0, 500 mM NaCl), and loaded onto a HiLoad 16/600 Superdex200 pg gel filtration column (GE) pre-equilibrated with 50 mM Tris pH 8.0, 500 mM NaCl. The column was run at 1.1-1.5 mL/minute, with 2 mL fractions collected over 0.3 CV-1 CV.

2.6.5 BPSS0213 – one-step purification

E. coli BL21 DE3 cell pellet containing overexpressed BPSS0213, grown as above, was resuspended in 10 mL/g of lysis buffer (50 mM Tris pH 9.0). Cells were sonicated at 16 microns for 3 x 20 seconds, incubated on ice between each cycle, and centrifuged at 18000 rpm/39191 *g* in a JA.20 rotor (Beckman Coulter) for 15-20 minutes. The protein content of the cell extract was analysed using the Bradford Assay (section 2.1.12). Cell extract was loaded onto a HiPrep DEAE FF 16/10 20 mL column (GE) pre-equilibrated with Buffer A (50 mM Tris pH 9.0), and 5/6 mL fractions

were collected. The column was washed with Buffer A for 30 mL (1.5 CV). BPSS0213 was eluted over 0-15% Buffer B (50 mM Tris pH 9.0, 1 M NaCl) in 300 mL (15 CV) and 5 mL fractions were collected. Another elution gradient used for BPSS0213 was 0-12% Buffer B (50 mM Tris pH 9.0, 1 M NaCl) in 240 mL (12 CV) and 5 mL fractions were collected. The column was run at 5 mL/minute. Elution fractions were analysed by SDS-PAGE (section 2.1.10).

For the initial complex formation experiment (complex 5.1), BPSS0213 was purified using a HiTrap DEAE FF 5 mL column (GE), and eluted over a gradient of 0-25% 50 mM Tris pH 9.0, 1 M NaCl, over 50 mL (10 CV).

2.6.6 BPSS0213 – two-step purification

E. coli BL21 DE3 cell pellet containing overexpressed BPSS0213, grown as above, was resuspended in 10 mL/g of lysis buffer (50 mM Tris pH 9.0). Cells were sonicated at 16 microns for 3 x 20 seconds, incubated on ice between each cycle, and centrifuged at 18000 rpm/39191 *g* in a JA.20 rotor (Beckman Coulter) for 15-20 minutes. The protein content of the cell extract was analysed using the Bradford Assay (section 2.1.12). Cell extract was loaded onto a HiPrep DEAE FF 16/10 20 mL column (GE) pre-equilibrated with Buffer A (50 mM Tris pH 9.0), and 5/6 mL fractions were collected. The column was washed with Buffer A for 30 mL (1.5 CV). BPSS0213 was eluted over 0-15% Buffer B (50 mM Tris pH 9.0, 1 M NaCl) in 300 mL (15 CV) and 5 mL fractions were collected. Another elution gradient used for BPSS0213 was 0-12% Buffer B (50 mM Tris pH 9.0, 1 M NaCl) in 240 mL (12 CV) and 5 mL fractions were collected. The column was run at 5 mL/minute. Elution fractions containing BPSS0213 were pooled and concentrated using a Vivaspin concentrator (Sartorius Stedim), to ~2 mL or made up to 2 mL with gel filtration buffer (50 mM Tris pH 8.0, 500 mM NaCl). The concentrated protein was loaded onto a HiLoad 16/600 Superdex 200 pg gel filtration column (GE) pre-equilibrated with 50 mM Tris pH 8.0, 500 mM NaCl. The column was run

at 1.5 mL/minute and 2 mL fractions were collected over 0.3 CV-1 CV.

2.6.7 BPSS0212ΔDUF1843

E. coli BL21 DE3 cell pellet containing overexpressed BPSS0212ΔDUF1843 was resuspended in 50 mM Tris pH 8.0 buffer, sonicated for 3 x 20 seconds at 16 microns, incubated on ice between each cycle, and centrifuged at 18000 rpm/39191 *g* in a JA.20 rotor (Beckman Coulter) for 15 minutes. The cell extract was loaded onto a HiTrap DEAE FF 5 mL column (GE) pre-equilibrated with 50 mM Tris pH 8.0 (Buffer A), at 5 mL/minute, the flow-though was collected in 5 mL fractions. The column was washed with 10 mL (2 CV) of Buffer A, with 5 mL fractions collected. BPSS0212ΔDUF1843 was eluted over a gradient of 0-50% Buffer B (50 mM Tris pH 8.0, 1 M NaCl), in 50 mL (10 CV) followed by 50-100% Buffer B in 25 mL (5 CV). The column was run at 5 mL/minute with 2 mL fractions collected. Concentrated protein, pooled off the DEAE column, was loaded onto a HiLoad 16/600 Superdex 200 pg gel filtration column (GE) equilibrated with 50 mM HEPES pH 7.5, 100 mM NaCl, run at 1.5 mL/minute. 2 mL fractions were collected over 0.28 CV-1 CV.

2.6.8 BPSS0213ΔDUF1843

E. coli BL21 DE3 cell pellet containing overexpressed BPSS0213ΔDUF1843 was resuspended in 50 mM MES pH 6.0 buffer, sonicated for 3 x 20 seconds at 16 microns, incubated on ice between each cycle, and centrifuged at 18000 rpm/39191 *g* in a JA.20 rotor (Beckman Coulter) for 15 minutes. The cell extract was loaded onto a HiTrap SP HP 5mL column (GE) pre-equilibrated with 50 mM MES pH 6.0 (Buffer A), at 5 mL/minute, the flow-though was collected in 5 mL fractions. The column was washed with 10 mL (2CV) of Buffer A, with 5 mL fractions collected. BPSS0213ΔDUF1843 was eluted over a gradient of 0-50% Buffer B (50 mM MES pH 6.0, 1 M NaCl), in 50 mL (10 CV) followed by 50-100% Buffer B in 25 mL (5 CV). The column was run at 5

mL/minute with 2 mL fractions collected. The protein pooled off the SP HP column was concentrated. During concentration the protein began to precipitate. The protein sample was centrifuged at 13000 rpm/10-15500 *g* for 10 minutes and the 2 mL of supernatant taken and loaded onto a HiLoad 16/600 Superdex 200 pg (GE) gel filtration column equilibrated with 50 mM HEPES pH 7.5, 100 mM NaCl, run at 1.5 mL/minute. 2 mL fractions were collected over 0.28 CV-1 CV.

2.7 Complex formation

The purification protocol steps for each protein used in complex formation experiment are outlined in the table below and the complex formation protocols for each complex experiment are described.

2.7.1 Protein purification protocol steps used in complex formation experiments

Each of the complex experiments is described using the corresponding results section in the BPSS0212/BPSS0213 complex results chapter, chapter 5. Purification protocol steps are described as DEAE (DEAE anion-exchange chromatography), SP (SP cation-exchange chromatography), $AmSO_4$ (ammonium sulphate cut) and GF (gel filtration chromatography). Purification steps shown in brackets are prerequisite purification steps and the steps shown in bold are the final purification step of the protein prior to use in the complex formation experiment. High or low corresponds to the NaCl concentration of the gel filtration column buffer, 500 mM or 100 mM, respectively. DEAE alone refers to the one-step purification protocol and (DEAE) GF refers to the two-step purification protocol.

Complex	**BPSS0211**	**BPSS0212+M**	**BPSS0213**
Complex 5.1	(DEAE, $AmSO_4$) **GF**	**DEAE**	**DEAE**
		BPSS0212	
Complex 5.2	(DEAE, $AmSO_4$) **GF**	**DEAE**	**-**
Complex 5.2	(DEAE, $AmSO_4$) **GF**	**-**	**DEAE**
Complex 5.3	-	**DEAE**	(DEAE) **GF**
Complex 5.4	-	**DEAE**	(DEAE) **GF**
	BPSS0211(1-123)		
Complex 5.5	**SP**	**DEAE**	**DEAE**
Complex 5.6 – (SeMet) high	-	BPSS0212+M **DEAE**	**DEAE**
Complex 5.6 – (SeMet) low	-	BPSS0212+M **DEAE**	**DEAE**
Complex 5.7 – high	-	(DEAE) **GF**	(DEAE) **GF**
Complex 5.7 – low	Formed from Complex 5.7 - high		
		BPSS0212 ΔDUF1843	**BPSS0213 ΔDUF1843**
Complex 5.8	-	(DEAE) **GF**	(SP) **GF**

The gel filtration column used was the same for all complex formation experiments, however the void volume of the column changed after complex experiment 5.1. For complex experiment 5.1, the void volume was 45 mL – relating to calibration graph in Figure 5.5. For all other experiments the void volume was 41 mL – relating to calibration graph in Figure 5.10.

2.7.2 Complex 5.1 - BPSS0211, BPSS0212+M, BPSS0213

BPSS0211, BPSS0212+M and BPSS0213 were concentrated separately and then combined in a ~1:1:1 mg ratio and run down a HiLoad 16/600 Superdex200 pg 120 mL gel filtration column (GE) pre-equilibrated with 50 mM Tris pH 8.0, 500 mM NaCl. 2 mL fractions were collected over 0.2 CV-0.9 CV.

2.7.3 Complex 5.2 - BPSS0211/BPSS0212 and BPSS0211/BPSS0213

BPSS0211, BPSS0212 and BPSS0213 were purified separately. BPSS0211 was combined with BPSS0212, with a two times molar excess of BPSS0212. The proteins were concentrated using a Vivapsin concentrator (Sartorius Stedim) and loaded onto the HiLoad 16/600 Superdex200 pg gel filtration column (GE) equilibrated with 50 mM Tris pH 8.0, 500 mM NaCl. The column was run at 1.5 mL/minute and 2 mL fractions were collected from 0.2 CV-1 CV. Elution fractions were analysed by SDS-PAGE.

BPSS0211 was combined with BPSS0213, with a two times molar excess of BPSS0213. The proteins were concentrated using a Vivapsin concentrator (Sartorius Stedim) to ~900 µL and made up to 2 mL with 50 mM Tris pH8.0. The two concentrated proteins were loaded onto the HiLoad 16/600 Superdex200 pg gel filtration column (GE) equilibrated with 50 mM Tris pH 8.0, 500 mM NaCl. The column was run at 1.5 mL/minute and 2.5 mL fractions were collected from 0.2 CV-1 CV. Elution fractions were analysed by SDS-PAGE.

2.7.4 Complex 5.3 - BPSS0212, BPSS0213

The complex of BPSS0212/BPSS0213 was produced by combining BPSS0213 with a two times molar excess of BPSS0212, then combined in a 1:1 volume ratio with MiliQ water, to dilute the salt concentration and reduce precipitation. The proteins were concentrated using a Vivapsin concentrator (Sartorius Stedim). During concentration slight precipitation of the protein occurred so the precipitation was pelleted and the soluble protein removed and concentration continued. The two concentrated proteins were centrifuged at 13000 rpm/10-15500 *g* prior to loading onto the gel filtration column to ensure no precipitated protein was present. The two concentrated proteins were loaded onto the HiLoad 16/600 Superdex200 pg gel filtration column (GE) equilibrated with 50 mM Tris pH 8.0, 500 mM NaCl. The column was run at 1.5 mL/minute and 2 mL

fractions were collected from 0.25 CV-0.95 CV. Elution fractions were analysed by SDS-PAGE and fractions 19-22 were pooled and diluted 1:5 with MiliQ water, to dilute the buffer to 10 mM Tris pH 8.0, 100 mM NaCl. The pooled protein was concentrated using a 10 kDa (kiloDalton) molecular weight cut-off Vivapsin (Sartorius Stedim) to ~10 mg/mL.

2.7.5 Complex 5.4 - BPSS0212, BPSS0213

The complex of BPSS0212/BPSS0213 was produced by combining BPSS0213, diluted 1:4 with MiliQ water to reduce the salt concentration, with a two times molar excess of BPSS0212. The proteins were concentrated in a 10 kDa molecular weight cut-off Vivaspin (Sartorius Stedim). The two concentrated proteins were centrifuged at 13000 rpm/10-15500 *g* prior to loading onto the gel filtration column to ensure no precipitated protein was present. 1.1 mL of concentrated protein was made up to 2 mL with gel filtration buffer (50 mM Tris pH 8.0, 500 mM NaCl) and then loaded onto the HiLoad 16/600 Superdex200 pg gel filtration column (GE) equilibrated with 50 mM Tris pH 8.0, 500 mM NaCl. The column was run at 1.5 mL/minute and 2 mL fractions were collected from 0.25 CV-0.95 CV. Elution fractions were analysed by SDS-PAGE and fractions 17-20 were pooled and diluted 1:5 with MiliQ water, to dilute the buffer to 10 mM Tris pH 8.0, 100 mM NaCl. The pooled protein was concentrated using a 10 kDa molecular weight cut-off Vivapsin (Sartorius Stedim) to ~10 mg/mL.

2.7.6 Complex 5.5 - BPSS0211(1-123), BPSS0212, BPSS0213

BPSS0211(1-123) was purified using an SP cation-exchange column, by Dr. Svetlana Sedelnikova (University of Sheffield). BPSS0212 and BPSS0213 were both purified using a DEAE anion-exchange column. BPSS0211(1-123) was combined with BPSS0212 and BPSS0213, in a 1:1:1 mg ratio and concentrated to ~2 mL. The concentrated protein solution was made up to 2 mL with 50 mM HEPES pH 7.5, 100 mM NaCl. The gel filtration load sample contained some precipitation. The sample

was centrifuged at 13000 rpm/10-15500 *g* for 10 minutes to pellet the precipitated protein. The supernatant sample was applied to a HiLoad 16/600 Superdex200 pg gel filtration column (GE), pre-equilibrated with 50 mM HEPES pH 7.5, 100 mM NaCl, run at 1.5 mL/minute. 2 mL fractions were collected over 0.28 CV-1 CV.

2.7.7 Complex 5.6 - BPSS0212+M-SeMet, BPSS0213

BPSS0212+M-SeMet and BPSS0213 were purified by DEAE anion-exchange chromatography and elution fractions were pooled separately. BPSS0212+M-SeMet and BPSS0213 were combined into two batches, each in a ~2:1 molar ratio, of BPSS0212+M-SeMet:BPSS0213. The complex formation experiment was carried out in both high salt and low salt conditions. Each batch was concentrated using a Vivaspin concentrator (Sartorius Stedim) and the total volume made up to 2 mL with either 50 mM Tris pH 8.0, 500 mM NaCl, for the high salt experiment or 50 mM HEPES pH 7.5, 100 mM NaCl, for the low salt experiment. For the high salt experiment the protein sample was loaded onto a HiLoad 16/600 Superdex 200 pg gel filtration column pre-equilibrated with 50 mM Tris pH 8.0, 500 mM NaCl. For the low salt experiment the protein sample was loaded onto a HiLoad 16/600 Superdex 200 pg gel filtration column (GE) pre-equilibrated with 50 mM HEPES pH 7.5, 100 mM NaCl. For both experiments the column was run at 1.5 mL/minute and 2 mL fractions were collected from 0.25 CV-1 CV.

2.7.8 Complex 5.7 - BPSS0212, BPSS0213 with protease inhibitors and EDTA

BPSS0212 and BPSS0213 were both purified separately in the presence of protease inhibitors and EDTA by DEAE anion-exchange chromatography followed by gel filtration chromatography and the elution fractions were pooled separately.

For the high salt gel filtration complex experiment, purified BPSS0212 and BPSS0213 were combined, in a ~1:1 mg ratio and concentrated to a volume ~1.3 mL, with the volume then made up to 2 mL with 50 mM Tris pH 8.0, 500 mM NaCl, 2 mM EDTA. The protein sample was then loaded onto a HiLoad 16/600 Superdex 200 pg gel filtration column (GE) pre-equilibrated with 50 mM Tris pH 8.0, 500 mM NaCl, 2 mM EDTA. The column was run at 1.5 mL/minute and 2 mL fractions were collected from 0.25 CV-1 CV.

BPSS0212/BPSS0213 complex fractions 15-20 from the first peak of the high salt gel filtration co-purification were pooled, with a total 12 mL, and 5 mL was concentrated to a volume ~1.2 mL, with the volume then made up to 2 mL with 50 mM HEPES pH 7.5, 100 mM NaCl, 2 mM EDTA. The 2 mL of protein sample was then loaded onto a HiLoad 16/600 Superdex 200 pg gel filtration column (GE) pre-equilibrated with 50 mM HEPES pH 7.5, 100 mM NaCl, 2 mM EDTA. The column was run at 1.5 mL/minute and 2 mL fractions were collected from 0.25 CV-1 CV.

2.7.9 Complex 5.8 - BPSS0212ΔDUF1843 and BPSS0213ΔDUF1843 complex formation

BPSS0212ΔDUF1843 and BPSS0213ΔDUF1843 were purified by DEAE anion-exchange chromatography and SP cation-exchange chromatography, respectively, each followed by gel filtration chromatography, and the elution fractions were pooled separately. BPSS0212ΔDUF1843 and BPSS0213ΔDUF1843 gel filtration pooled fractions were combined in a 2:1 molar ratio and concentrated in a 10 kDa molecular weight cut-off Vivaspin concentrator (Sartorius Stedim), to a volume of around 600 µL. This volume was made up to 2 mL with 50 mM HEPES pH 7.5, 100 mM NaCl. The concentrated protein sample was loaded onto a HiLoad 16/600 Superdex 200 pg gel filtration column (GE) pre-equilibrated with 50 mM HEPES pH 7.5, 100 mM NaCl. The column

was run at 1.5 mL/minute and 2 mL fractions were collected over 0.28 CV-1 CV.

2.8 Complex preparation for electron microscopy

A sample of gel filtration elution fraction from the complex 5.5 gel filtration chromatography run, containing the BPSS0212/BPSS0213 complex was taken and the concentration measured by absorbance at A_{280nm}. Samples at 0.1 mg/mL and 0.2 mg/mL were taken and applied to grids for analysis. Further sample dilutions were made as necessary.

2.9 Crystallography

An overview of the crystallographic methods carried out, as part of the research presented in this thesis will be described below. The principles underlying these techniques are described in outline in the text book - Crystallography made Crystal Clear – Gale Rhodes, and in detail in the text book - Biomolecular Crystallography – Bernard Rupp.

2.10 Crystallisation

Protein was concentrated and prepared in buffer containing a low concentration of buffering component and salt. The main type of crystallisation experiment used was vapour diffusion, using sitting-drops or hanging-drops. Sitting-drop experiments have a large well or reservoir containing the crystallisation solution and a smaller well containing a combination of concentrated purified protein and crystallisation solution in a specified ratio. Hanging-drop experiments have a large well or reservoir containing the crystallisation solution with a smaller drop, containing a combination of concentrated purified protein and crystallisation solution in a specified ratio suspended above on a siliconised coverslip.

2.10.1 Sitting-drop crystallisation trials

Initial sitting-drop experiments were set up with a large range of crystallisation conditions, using any number of 11 crystallisation screens,

each containing 96 different crystallisation conditions. The screens used were PACT, JCSG+, PEGs, pHClear, Classics, MPD, $AmSO_4$, Proplex, Morpheus, Midas and PGA (Qiagen, Molecular Dimensions). These crystallisation conditions suites seek to vary, within the conditions, the type of precipitant, the pH, ionic strength and the use of additives.

Standard trays were dispensed using a Matrix Hydra II Plus One (Thermo Scientific) with 50 µL of reservoir solution added to the large well and 200 nL:200 nL drops of protein:reservoir solution in the smaller well. Different drop ratios used are specified in the corresponding results chapters. Trays were centrifuged at 1200 rpm/200 *g* for 2 minutes to mix the protein and reservoir solution in the small well. Sitting-drop trays set up by hand contained 30-50 µL of reservoir solution added to the large well and 500 nL-1 µL:500 nL-1 µL drops of protein:reservoir solution in the small well. Pre-dispensed sitting-drop trays were also set down using a Formulatrix NT8 (Formulatrix), in trays containing one large well and three smaller wells, with 25-50 µL reservoir and 200 nL:200 nL or 200 nL:150 nL drops, protein:reservoir ratio.

2.10.2 Optimisation crystallisation trials

Optimisation experiments of the initial sitting-drop hit conditions were performed using the hanging-drop method and the sitting-drop method with microbridges. For the optimisation of a specific crystallisation condition, variations of the concentration of precipitant, buffer or salt and also the type of buffer, salt, and pH were used. The reservoir contained 500 µL or 1000 µL of crystallisation solution. 1-3 µL drops of precipitant were mixed with 1-3 µL protein on a siliconised coverslip, which was then suspended over the reservoir and sealed with immersion oil. Experiments were incubated at room temperature (~19°C) and/or 17°C and/or 7°C. The details of specific optimisation trials are outlined in the relevant results chapter.

To siliconise the coverslips, which were cleaned using detergent and washed with distilled water, dried, and then placed in a vacuum desiccator with a beaker of ~10-20 mL of dichlorodimethylsilane solution, in a fume cupboard. A vacuum was produced over ~15 minutes and maintained for ~45 minutes to allow siliconisation of the coverslips to take place. Siliconised coverslips were then removed from the vacuum desiccator and polished using lens tissue and an air stream.

2.10.3 Production of derivative crystals

Overexpressing protein in the presence of selenomethionine was used to produce selenomethionine-incorporated protein crystals. For the iodine incorporation, iodine crystals were produced using potassium iodide combined with ~1 mL concentrated sulphuric acid, in a fume cupboard. Once the reaction had completed, the resulting crystals were washed with water. The acid was then neutralised with bicarbonate and the iodine crystals washed again and dabbed onto filter paper. Crystals were stored in a glass vial. The iodine crystals were placed next to the crystallisation drop using wax and left for set periods of time (hours to overnight).

2.11 Soaking ligands in BPSS0212 crystals

Crystals of BPSS0212 were grown in sitting-drop trays in PACT and PEGs crystallisation conditions screens. For the soaking in of biotin, a 40 mM suspension of biotin was produced in 500 µL. 0.5 µL of the biotin suspension was mixed with 0.5 µL of reservoir solution. 0.4 µL of this mixture was taken and added to the crystal containing small well drop. This last step was repeated twice for one of the crystal drops tested. For the soaking of crystals with specific ligands, a small amount of specific ligand, sufficient to give a concentration in approximately the desired range was added as powder directly into the crystal containing drop and left for ~2 hours.

2.12 Cryoprotection

Crystals were cryoprotected with solutions containing the crystallisation solution, with ~2% increase in PEG (polyethylene glycol), if present, or double the concentration of ammonium sulphate, if present. Cryoprotectants of 25-30% ethylene glycol, glycerol, low molecular weight PEGs (PEG300/PEG400) or sucrose were added to the crystallisation solution components. An MPD (2-Methyl-2, 4-pentanediol) percentage of 40% or more in a crystallisation solution was used as a cryoprotectant. Cryoprotectants for specific crystals are described within the appropriate chapter. Crystals were looped from the drop using LithoLoops (Molecular Dimensions), passed through a drop of 1-2 µL of cryoprotectant if required and cooled in liquid nitrogen. The cooled looped crystals were loaded into pucks, each which holds 16 crystals, to be sent to the Diamond Light Source Synchrotron, Oxford.

2.13 Crystal screening and data collection

Screening and data collection of protein crystals was carried out at the Diamond Light Source Synchrotron, Oxford, on the MX beamlines. The Synchrotron X-ray source generally allows the wavelength of the X-rays to be tuned to a specific wavelength. Data collection was performed using the rotation method, and data were collected using a Pilatus detector (DECTRIS). Data collection strategies were designed based on the type of experiment being carried out – high resolution, molecular replacement or experimental phasing, MAD (multi-wavelength anomalous dispersion) or SAD (single-wavelength anomalous dispersion). The data collection strategies for specific crystals are outlined in the relevant results chapters. The crystals were stored in pucks, under liquid nitrogen, and then loaded onto the goniometer using a robot.

2.13.1 Testing

The crystals were tested for quality of diffraction by taking 4-5 images 45° apart with a phi oscillation of 0.2° or 0.5°. Images were then inspected for

quality of diffraction and to ensure the X-rays were hitting the crystal in all directions. Test images were autoindexed, if possible, to provide spacegroup and cell dimension predictions, along with possible data collection strategies, using Mosflm or EDNA (Battye et al., 2011; Incardona et al., 2009).

2.13.2 Data collection experiments

Once the crystals were tested and there was sufficient evidence for the data collection to go ahead, a data collection strategy was decided. For standard data collection a rotation range covering the region of reciprocal space required, which ultimately depends on the spacegroup, was collected generally using a 0.1-0.2° phi oscillation. The choice of transmission depended on the beamline used, the resolution of the data desired and amount of data required. Experimental phasing of structures was carried out using SAD and MAD. A fluorescence scan was carried out on the crystal at the appropriate element edge to check for incorporation of the element. For selenium the K edge was used. The *CHOOCH* program (Evans and Pettifer, 2001) provided the peak and inflection energies at which to collect the data, along with the corresponding f' and f'' values. For iodine soaked crystals the peak wavelength was set at 1.7 Å. For a MAD experiment, data were collected at peak, inflection and high-energy remote wavelengths, whereas for a SAD experiment data were collected only at peak wavelength. The crystals were tested at the wavelength to be used for data collection, usually the peak wavelength, at a range of transmission values to establish the lowest transmission value that produced test diffraction images that could be indexed. The data were then collected over a large enough rotation range to ensure a large value of multiplicity, again with 0.1-0.2° phi oscillation.

2.14 Data processing

Data processing involves a number of stages and can be carried out using individual programs or a pipeline of programs. For the projects described here data processing was carried out using the Fast DP or xia2 automatic pipelines (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Sauter et al., 2004; Winter, 2010; Winter and McAuley, 2011; Zhang et al., 2006). Data processing involves recording the location and intensity of the spots on all the diffraction images collected. The locations of the reflections determine the cell dimensions and spacegroup. Reflections measured from multiple images are merged and scaled together to produce a list of amplitudes for each reflection, along with a measure of error for the amplitude of each reflection. Fast DP data processing uses *XDS* for indexing and integration, with both *XDS* and *POINTLESS* used to determine the spacegroup (Evans, 2006: Kabsch, 2010; Winter and McAuley, 2011). Finally the data were scaled and merged in *XDS*, *POINTLESS* and aimless (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Winter and McAuley, 2011). The xia2 pipeline with settings 3d, 3da and 3daii, were used during the work described in this thesis. These settings describe the specific programs used for data processing. 3d, 3da and 3daii use *XDS* and *XSCALE*. 3d then uses Scala to merge, 3da and 3daii use Aimless to merge. *XDS* and *XSCALE* were used to index, integrate and scale the data, whilst Scala or Aimless were used to merge the data. With 3daii, XDS uses all the images to index the data, compared to 3d and 3da, which uses a selection of images to index the data (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Sauter et al., 2004; Winter, 2010; Zhang et al., 2006).

2.15 Phasing

The diffracted X-rays were recorded as a diffraction pattern, which is the reciprocal space representation of the crystal contents, produced by

Fourier transform, and contains information about the amplitude of each wave, but not the phase. The amplitude information was recorded from the data collection experiment and the information output via the data processing. In order to produce an electron density map and therefore visualise the protein structure within the crystal, another Fourier transform can be performed. However, phase information about each wave is required, but this information is not recorded during the diffraction experiment. Therefore no second Fourier transform can be carried out, as only the amplitudes from the diffraction pattern are available. The phase information must be determined either experimentally or from a model of the protein or homologous protein.

2.16 Structure determination

The phase problem can be solved using molecular replacement or experimental phasing. Molecular replacement was carried out using a model of the actual protein or a homologous protein, using the CCP4 suite (Collaborative Computational Project, Number 4, 1994) supported program PhaserMR (McCoy et al., 2007). Parameters to be determined were the orientation and position of the model molecule in the new cell. A successful result was gauged by the final TFZ (translation function Z) and LLG (log-likelihood gain) scores, and analysis of the output model (McCoy et al., 2007).

Structure solving by experimental phasing, using selenomethionine SAD or MAD and iodine SAD, was carried out using the *SHELX* program suite – *SHELXC*, *SHELXD* and *SHELXE* using the *HKL2MAP* interface (Pape and Shneider, 2004; Sheldrick, 2008) or Fast EP automatic pipeline (Diamond Light Source - automatic software pipeline). The *SHELX* program suite uses the three programs *SHELXC*, *SHELXD* and *SHELXE* (Sheldrick, 2008). *SHELXC* was used to analyse the amount of anomalous signal in the data (Sheldrick, 2008). The threshold used to show the presence of signal was <d''/sig> >0.8 (Sheldrick, 2008).

SHELXD identified and located the heavy atom sites, provided with the resolution limit, number of sites and the output from *SHELXC* (Sheldrick, 2008). Estimates for the number of heavy atoms sites expected were derived from an analysis of the number of relevant residues in the protein sequence (methionines for selenium sites and tyrosines for iodine). The number of observed sites depends on factors such as the incorporation, position mobility and clipping.

SHELXE was run with the predicted solvent content, heavy atoms sites produced by *SHELXD*, and rounds of autobuilding with the option of looking for α-helices (Sheldrick, 2008). *SHELXE* was performed in both the original and inverted hand. The outputs for both hands from *SHELXE* were compared, both visually in *Coot* and from analysis of the estimated mean FOM (figure of merit) and pseudo-free CC (correlation coefficient) values (Sheldrick, 2008). The .phs maps output from *SHELXE* were analysed in *Coot* for connectivity of the electron density and inclusion of the sites from *SHELXD* in electron density. The original and inverted estimated mean FOM and pseudo-free CC values were compared, with the higher value indicative of a successful result (Sheldrick, 2008). The Fast EP pipeline (Diamond Light Source - automatic software pipeline) uses output .mtz from Fast DP and the *SHELX* programs *SHELXC*, *SHELXD* and *SHELXE* to carry out automated experimental phasing (Sheldrick, 2008; Winter and McAuley, 2011).

2.17 Model building, refinement and validation

Models and maps output from programs were analysed and models built in *Coot* (Emsley et al., 2010). Additional density modification and phase combination was carried out as described in the relevant chapters. Refinement was carried using the CCP4 suite (Collaborative Computational Project, Number 4, 1994) supported program Refmac5 (Murshudov et al., 1997; Murshudov et al., 2011). To analyse the progress of refinement and to avoid accumulation of bias, the model and

map were analysed after each round of refinement. Successful refinement was measured as the reduction of the R_{free} (free R-factor) and R_{factor} (R-factor) together. Structures were validated using the validation options in *Coot* (Emsley et al., 2010), including analysis of the Ramachandran plot for each structure, rotamer analysis, difference map blobs and peptide omega analysis. MolProbity (Chen et al., 2011) was also used for validation statistics regarding possible Ramachandran outliers, rotamer outliers, bad bonds and bad angles, clash scores and low resolution geometry issues. For $TssA_{CTD}$ and $TssA_{NTD}$ additional wwPDB validation was also carried out (Berman et al., 2003).

Chapter 3. Investigations into the operon and the first protein, BPSS0211

Previous work carried out by Dr. Matthew Day on nine identified target proteins, including those encoded by the operon BPSS0211-BPSS0214, led to the elucidation of the structure of BPSS0211 (Day, 2012). The structure of BPSS0211 was solved in the presence of zinc, and consists of four helical monomers, which interact to form a dimer of dimers, with two zinc-binding sites shared across the dimer interface, and another four on the external protein surface (Day, 2012). Following the structure determination of BPSS0211, initial sequence and literature investigations were carried out on the BPSS0211-BPSS0214 operon to find information that may relate to its function and to continue the work carried out by Dr. Matthew Day.

3.1 Sequence analysis of operon proteins

3.1.1 Genome annotation

The protein sequences for the BPSS0211-BPSS0214 operon from the *B. pseudomallei* K96243 genome (Holden et al., 2004) were obtained from the Burkholderia Genome Database (Winsor et al., 2008), shown in Figure 3.1. The protein sequences for the BPSS0211-BPSS0214 operon from *B. pseudomallei* K96243 are the same as those from *B. pseudomallei* D286, used in cloning, except for a residue difference at position 174 in BPSS0214, which is Glu in K96243 and Asp in D286. In the Burkholderia Genome Database – genome *B. pseudomallei* K96243 – BPSS0211, BPSS0212 and BPSS0213 are all annotated as hypothetical proteins, with BPSS0214 as a heme/metallo cofactor biosynthesis-like protein (Winsor et al., 2008). The localisation of BPSS0211 and BPSS0212 is still unknown (Winsor et al., 2008), however BPSS0213 has been predicted to be extracellular, as a suggested member of the *B. pseudomallei* core secretome (Vander Broek et al., 2015). BPSS0214 is computationally predicted to be a cytoplasmic protein (Winsor et al., 2008; Yu et al., 2010).

BPSS0211

MSQAQGHPVTPYGVAIHQAIADGDLAQMKSLRTQAQALLAQQGNLATALELLEV
EIAKLERRK

Residues - 63

Molecular weight – 6793.82 Da

BPSS0212

MSEDLRVGLFPVRYLVGTGLPGAPQLVLDLMVDTVDHSVVGRAAVSQAVSPPLN
FHADVWGSYVFRLGPPPRRDGSGAIVQISLQGNQGGPQSNSMITFYGELLLKGD
GKTGVASYRYYSNGSWHEVENVPVKADPELVPIEPGPVIGQSSMSAIGSAAMYG
VAIQSAAASGDLAHMRTLSAYARQQLESRDEIAAALSELKAEIAKLESRQ

Residues - 212

Molecular weight – 22465.43 Da

BPSS0213

MATTGLFPVQLRVATPNLGAPVLWLNLLVNTVEKTASGFARITQTVYPPMHFRA
RVVGPFHQMRIDPHAPQSVTLTLSGSPTGPVAPQVVILELNALLNEGWQSGTAN
YRYFYESRWHSIEHAIVSKDNSRIPLDPPSEHVMPMYGVGLQEARASGDLSRMK
ALAQQAEQQLADHDVIAAELQKLEAEIARLEARR

Residues - 196

Molecular weight – 21667.80 Da

BPSS0214

MSTTDARPARYLFDSDYQRFVPVHAVWEITLACDLKCLHCGSRAGHRRTNELST
AECLEVIDALARLGTREVSLIGGEAYLRKDWTQLIRAIRSHGMYCAIQTGGRNL
TPKRLAQAVDAGLNGVGVSLDGLAPLHDKVRNVPGAFERALDTLRRARDAGIAV
SVNTQIGAQTM**E**DLPALMDTIIELGATHWQIQLTVAMGNAVDNDELLLQPYRLA
ELMPLLAKLYKDGVSRGLLMTVGNNIGYYGPYEHLWRGFGDERVHWSGCAAGQN
VIALEADGTVKGCPSLATVGFSGGNVRDMSLEDIWRTSEGIHFGRLRSVDDLWG
FCRTCYYADVCRGGCTWTSHSLLGKPGNNPYCHYRVLELQKQGLRERIAKVQDA
GPASFAVGRFDLVTERIADGEPVASVVRSGQVIELAWKNRGKRSPEVGRVPPKL
KMCRNCDGYVHAGEQTCPHCGGDIDAAARAHELDAQRRHALMNDLERLLGLPAS
TFGGG

Residues - 491

Molecular weight – 53737.21 Da

Figure 3.1. Protein sequences of all four proteins in the BPSS0211-BPSS0214 operon. Sequences from the Burkholderia Genome Database, *B. pseudomallei* K96243 (Winsor et al., 2008) and molecular weight calculated by ProtParam (Gasteiger et al., 2005). The residue difference between *B. pseudomallei* K96243 and *B. pseudomallei* D286 BPSS0214 is highlighted in bold.

3.1.2 *Burkholderia* homologs

The protein sequences of all four proteins were analysed using protein BLAST (Altschul et al., 1990) to identify any homologs in other organisms, any domain annotations and possible indications of function. BLAST analysis of BPSS0211, BPSS0212 and BPSS0213 within the *Burkholderia* species revealed homologs in *B. pseudomallei, B. mallei, B. thailandensis, B. oklahomensis, B. diffusa, B. ubonensis* and *B. ambifaria* (Altschul et al., 1990).

For BPSS0214, BLAST identified homologs with 81-100% identity and 97-100% query coverage present in *B. pseudomallei, B. mallei, B. thailandensis, B. oklahomensis, B. diffusa, B. ubonensis* and *B. ambifaria*. These homologs were annotated as GDL motif peptide-associated radical SAM(S-adenosyl-methionine)/SPASM maturase and radical SAM domain protein. Other BPSS0214 homologs annotated as PqqE, radical SAM domain and anaerobic sulfatase maturase were found with 23-51% identity over 52-80% coverage in *B. cepacia, B. cenocepacia, B. glumae, B. ubonsensis, B. pseudomultivorans, B. diffusa, B. multivorans, B. vietnamiensis, B. contaminans, B. lata, B. stabilis, B. ambifaria, B. territorii, B. arvi, B. hypogeia, B. turbans* and *B. catudaia* (Altschul et al., 1990).

Orthologs of BPSS0211, BPSS0212, BPSS0213 and BPSS0214 have been identified in operons containing six genes. The two possible extra genes in the operon are annotated as a methyl-accepting chemotaxis protein and a hypothetical protein (Winsor et al., 2008).

3.1.3 Homologs of operon proteins in other organisms

BLAST results for BPSS0211 produced both proteins annotated as hypothetical proteins and DUF1843 domain-containing proteins. BPSS0211 homologs were found in *Pseudomonas*, *Chromobacterium* and *Azospirillum* species, and truncated homologs in *Ralstonia*,

Pseudomonas, *Chromobacterium*, *Caulobacter*, *Nitrospirillum*, *Azospirillum*, *Nonomuraea*, *Tistrella*, and *Terriglobus* species (Altschul et al., 1990).

Protein BLAST analysis of BPSS0212 and BPSS0213 identified homologs corresponding to residues 1-212, (5-14)-(211-212), (1-8)-(207-212) and 77-212 for BPSS0212 and residues 1-196 and (1-6)-(172-196) for BPSS0213, in species of *Ralstonia*, *Chromobacterium*, *Pseudomonas* and *Sorangium* species. The homologs are annotated as hypothetical proteins, DUF1842 domain-containing proteins and proteins of unknown function (Altschul et al., 1990).

Closely related homologs of BPSS0214 were found in species of *Burkholderia, Ralstonia, Chromobacterium, Pseudomonas, Terriglobus, Caulobacter, Azospirillum, Tistrella, Sorangium* and *Nonomuraea*, with 57-99% identity over 76%-100% coverage. These homologs are annotated as GDL motif peptide-associated radical SAM/SPASM maturase, Y_X(10)_GDL-associated radical SAM protein, heme transporter CcmD and radical SAM protein (Altschul et al., 1990).

3.1.4 Domain analysis of operon proteins

The domain architecture of the proteins in the operon was analysed using the CDD (conserved domain database) (Marchler-Bauer et al., 2015). BPSS0211, BPSS0212 and BPSS0213 all have a DUF1843 domain present. In BPSS0211 the DUF1843 domain is the majority of the protein, whilst in BPSS0212 and BPSS0213 the DUF1843 domain is present at the C-terminus. BPSS0212 and BPSS0213 also both have a DUF1842 domain at the N-terminus (Altschul et al., 1990, Marchler-Bauer et al., 2015). Therefore BPSS0212 and BPSS0213 have the same domain architecture, with an N-terminal DUF1842 domain, followed by a C-terminal DUF1843 domain. BPSS0212 has a slightly longer interdomain linker than BPSS0213 (Figure 3.2).

BPSS0211

MSQAQGHPV**TPYGVAIHQAIADGDLAQMKSLRTQAQALLAQQGNLATALELLEV**
EIAKLERRK

BPSS0212

MSEDLR**VGLFPVRYLVGTGLPGAPQLVLDLMVDTVDHSVVGRAAVSQAVSPPLN**
FHADVWGSYVFRLGPPPRRDGSGAIVQISLQGNQGGPQSNSMITFYGELLLKGD
GKTGVASYRYYSNGSWHEVENVPVKADPELVPIEPGPVIGQSSMSAIGSA**AMYG**
VAIQSAAASGDLAHMRTLSAYARQQLESRDEIAAALSELKAEIAKLESRQ

BPSS0213

MAT**TGLFPVQLRVATPNLGAPVLWLNLLVNTVEKTASGFARITQTVYPPMHFRA**
RVVGPFHQMRIDPHAPQSVTLTLSGSPTGPVAPQVVILELNALLNEGWQSGTAN
YRYFYESRWHSIEHAIVSKDNSRIPLDPPSEHVM**PMYGVGLQEARASGDLSRMK**
ALAQQAEQQLADHDVIAAELQKLEAEIARLEARR

Figure 3.2. Domain architecture mapped onto the protein sequences of BPSS0211, BPSS0212 and BPSS0213. The DUF1842 domain is highlighted in blue and the DUF1843 domain is highlighted in red. The interdomain linker is underlined. Domain information from BLAST (Altschul et al., 1990) and CDD (Marchler-Bauer et al., 2015).

BPSS0212 and BPSS0213 share 38% identity over 97-98% sequence coverage. BPSS0211 shares 50% and 48% identity over 82% sequence coverage with BPSS0212 and BPSS0213, respectively (Altschul et al., 1990). BPSS0211, BPSS0212 and BPSS0213 amino acid sequences were aligned in Jalview (Waterhouse et al., 2009) to identify the sequence similarity between all three proteins (Figure 3.3).

BPSS0214 has a radical SAM domain, which binds a 4Fe-4S cluster, using a conserved CX_3CX_2C motif, where the three cysteine residues and SAM coordinate the 4Fe-4S cluster (Altschul et al., 1990; Marchler-Bauer et al., 2015). The C-terminus of BPSS0214 has an additional SPASM domain, which binds two 4Fe-4S clusters, and has the $CX_2CX_5CX_3C$ motif (Marchler-Bauer et al., 2015).

3.1.5 DUF1843 domain

The DUF1843 domain is found in ~304 proteins, in species of *Burkholderia, Ralstonia, Pseudomonas, Caulobacter, Terriglobus, Tistrella*, *Azospirillum*, *Bradyrhizobium*, *Hoeflea*, *Sorangium* and *Flavobacterium* (Pfam, PF08898, InterPro, IPR014994) (Finn et al., 2016; Mitchell et al., 2015). The DUF1843 domain is found within proteins as the DUF1843 domain alone or at the C-terminus of a protein, with the corresponding N-terminus containing a DUF1842 domain (Pfam, PF08898, InterPro, IPR014994). There is also a protein from *Terriglobus saanensis* - AciPR4_3596, which has two consecutive DUF1843 domains predicted (InterPro, IPR014994 - Mitchell et al., 2015).

3.1.6 DUF1842 domain

The DUF1842 domain is found in ~302 proteins, in species of *Burkholderia, Ralstonia, Pseudomonas, Chromobacterium, Caulobacter, Magnetospirillum, Microscilla, Flavobacterium, Pedobacter, Marinomonas, Pseudoaltermonas, Shewanella, Thiorhodococcus* and *Marichromatium* (Pfam, PF08896, InterPro, IPR014992) (Finn et al., 2016

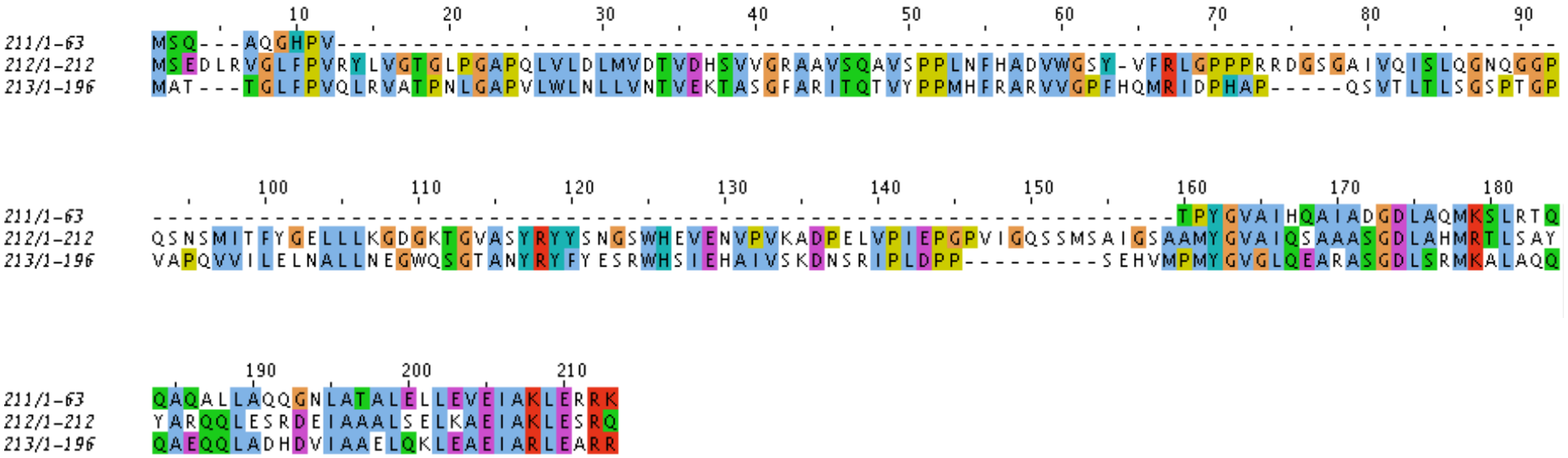


Figure 3.3. Alignment of BPSS0211, BPSS0212 and BPSS0213 amino acid sequences. Protein sequences of the three operon proteins BPSS0211 (residues 1-63), BPSS0212 (residues 1-212) and BPSS0213 (residues 1-196). The C-terminal of BPSS0211 is seen to align with BPSS0212 and BPSS0213. Alignment performed in Jalview using T-Coffee with defaults and coloured using Clustalx (Larkin et al., 2007; Notredame et al., 2000; Waterhouse et al., 2009). Sequences from *B. pseudomallei* K96243 from the Burkholderia Genome Database (Winsor et al., 2008).

; Mitchell et al., 2015). The DUF1842 domain is located at the N-terminus within proteins, alone and in proteins with a DUF1843 domain at the C-terminus (Pfam, PF08896, InterPro, IPR014992). There have also been identified two proteins from *Flavobacterium* species, with a DUF1842 domain and a superoxide dismutase, copper/zinc-binding domain within the same protein region (InterPro, IPR014992 - Mitchell et al., 2015).

There are currently no structures of the DUF1842 domain or DUF1843 domain in the PDB (Protein Data Bank) (Berman et al., 2003; Berman et al., 2000).

For BPSS0214 a number of domain hits were identified (Altschul et al., 1990; Marchler-Bauer et al., 2015), which are summarised in Figure 3.4. A BLAST run searching the PDB database (Altschul et al., 1990; Berman et al., 2003; Berman et al., 2000) found BPSS0214 shared 20% identity over 62% coverage with the structure of anSMEcpe (anaerobic sulfatase-maturing enzyme – *Clostridium perfringens*) PDB code 4K36 and 4K37 (Goldman et al., 2013).

3.1.7 Domain architecture

Another programme available via BLAST is CDART (Conserved Domain Architecture Retrieval Tool) (Geer et al., 2002), which identifies the domain of interest within the domain architecture of other proteins. The domain architectures identified for the DUF1842 domain are those described above – DUF1842 or DUF1842-DUF1843 (Geer et al., 2002).

The DUF1843 domain was found on its own or in combination with the DUF1842 domain (Geer et al., 2002). The presence of the DUF1843 domain with the DUF1842 domain together was known previously from BPSS0212 and BPSS0213. The DUF1843 domain was also found in phage tail tip fiber proteins from *Burkholderia* virus and in a host specificity protein or hypothetical protein from *B. pseudomallei*, *B.*

Accession	Name
cd01335	Radical SAM
TIGR04085	Radical SAM additional 4Fe-4S-binding SPASM domain
pfam13186	SPASM
pfam04055	Radical SAM
TIGR03913	Y_X(10)_GDL-associated radical SAM protein
COG0535	SkfB
PRK05301	PqqE
smart00729 family	Elp3, MiaB, MoaA etc.

MSTTDARPARYLFDSDYQRFVPVHAV**WEITLACDLKCLHC**GSRAGHRRTNELST
AECLEVIDALARLGTREVSLIGGEAYLRKDWTQLIRAIRSHGMYCAIQTGGRNL
TPKRLAQAVDAGLNGVGVSLDGLAPLHDKVRNVPGAFERALDTLRRARDAGIAV
SVNTQIGAQTMEDLPALMDTIIELGATHWQIQLTVAMGNAVDNDELLLQPYRLA
ELMPLLAKLYKDGVSRGLLMTVGNNIGYYGPYEHLWRGFGDERVHWSGCAAGQN
VIALEADGTVKGCPSLATVGFSGGNVRDMSLEDIWRTSEGIHFGRLRSVDDLWG
FCRTCYYADVCRGGCTWTSHSLLGKPGNNPYCHYRVLELQKQGLRERIAKVQDA
GPASFAVGRFDLVTERIADGEPVASVVRSGQVIELAWKNRGKRSPEVGRVPPKL
KMCRNCDGYVHAGEQTCPHCGGDIDAAARAHELDAQRRHALMNDLERLLGLPAS
TFGGG

Figure 3.4. BLAST conserved domains and annotation of the BPSS0214 amino acid sequence. The domain hits for BPSS0214 sequence from BLAST and CDD (Altschul et al., 1990; Marchler-Bauer et al., 2015). The section of BPSS0214 aligned to cd01335 conserved radical SAM motif is shown in bold and the conserved CX_3CX_2C motif is highlighted in blue. The section of BPSS0214 aligned to TIGR04085, is underlined and the motif of the radical SAM additional 4Fe-4S-binding SPASM domain, $CX_2CX_5CX_3C$, is highlighted in yellow. Domain analysis using CDD (Marchler-Bauer et al., 2015). BPSS0214 sequence from *B. pseudomallei* K96243 from the Burkholderia Genome Database (Winsor et al., 2008).

thailandensis and other *Burkholderia* species (Geer et al., 2002). Efflux transporters from *Bradyrhizobium* species and a transcriptional regulator from *Agrobacterium, Neorhizobium* and *Rhizobium* species were shown to contain the DUF1843 domain (Geer et al., 2002). Also the DUF1843 domain was found in a transposase from *Burkholderia* and *Paraburkholderia* species (Geer et al., 2002).

3.1.8 Proteomics investigation

All four proteins were analysed for the presence of signal sequences using SignalP (Petersen et al., 2011) and Predtat (Bagos et al., 2010), and no signal sequence was identified in any of the proteins. The BPSS0211 structure (Day, 2012) suggests the presence of a possible fold closely related to that of a leucine zipper motif. The 2ZIP-server (Bornberg-Bauer et al., 1998) predicted a potential leucine zipper in BPSS0211, but not its homologous domains in BPSS0212 or BPSS0213. None of the four proteins in the operon contain any helical transmembrane regions, shown using TMHMM (Sonnhammer et al., 1998).

The STRING database contains protein:protein interaction information (Szklarczyk et al., 2015). Predicted interactions are present between BPSS0211, BPSS0212 and BPSS0213, due to their location in the genome, with protein homology identified between BPSS0212 and BPSS0213, and BPSS0214 is predicted to interact with BPSS0212 and BPSS0213, but not BPSS0211, again due to gene location (Szklarczyk et al., 2015). BPSS0214 has the largest number of interactions, it is also predicted to interact with BPSL0961, BPSS1245, BPSL2256, BPSS2317, BPSS1331, BPSS0403, BPSL0355 and tar (Szklarczyk et al., 2015).

3.1.9 Phyre2 model

The Phyre2 server (Kelley et al., 2015) produced a number of convincing predicted protein models from the BPSS0214 sequence. The top model

hit is based on the template of an anaerobic sulfatase-maturing enzyme (anSME) (PDB:4K39) (Goldman et al., 2013), which shares 17% identity with BPSS0214 with 68% coverage (Kelley et al., 2015) (Figure 3.5). The model covers 66% of the sequence (residues 22-359) with 100% confidence (Kelley et al., 2015) (Figure 3.6). The model has the $(\beta/\alpha)_6$ of the radical-SAM fold, followed by another helix, then the $CX_{9-15}GX_4CX_nCX_2CX_5CX_3CX_nC$ consensus sequence found in anSME proteins (Benjdia et al., 2008; Goldman et al., 2013; Kelley et al., 2015). The other models generated with 100% confidence used templates including: radical SAM enzyme and BtrN, with 17-22% identity and 59-64% coverage (Kelley et al., 2015). Another 17 models were produced with 98.6-99.9% confidence and 11-22% identity (Kelley et al., 2015).

3.2 Literature investigations

Due to the sequencing of a number of *Burkholderia* strain genomes (Mukherjee et al., 2016; Winsor et al., 2008), the availability of the *B. pseudomallei* genome sequences has led to a range of large-scale transcriptomic, proteomic, metabolomic and structural genomic studies being carried out on *B. pseudomallei*. These studies have allowed the investigation into the role of many previously uninvestigated proteins. The current research was analysed for any mention of the proteins in the BPSS0211-BPSS0214 operon to further elucidate any function these proteins may have.

3.2.1 Essentiality of the BPSS0211-BPSS0214 operon

A study carried out by Moule et al., 2014 did not identify any of BPSS0211, BPSS0212, BPSS0213 and BPSS0214 proteins as essential genes in *B. pseudomallei*. The homolog of the BPSS0211-BPSS0214 operon in *B. thailandensis*, BTH_II2184, BTH_II2183, BTH_II2182 and BTH_II2181, was also not identified as essential in *B. thailandensis* (Baugh et al., 2013).

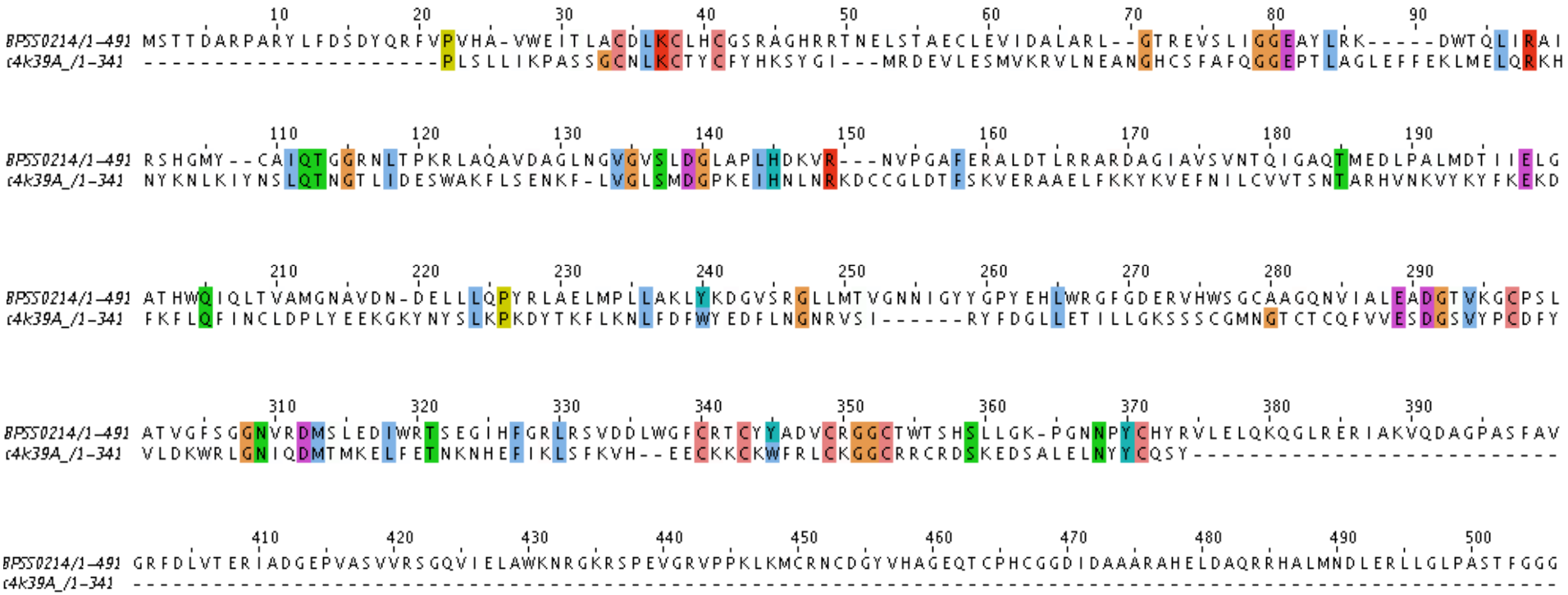


Figure 3.5. Phyre2 alignment of protein sequences for c4k39A template model and BPSS0214. The sequences of protein PDB:4K39 (anaerobic sulfatase-maturing enzyme - Goldman et al., 2013) residues 1-341, used as the template, and BPSS0214, residues 1-491, are aligned by Phyre2 (Kelley et al., 2015). Alignment visualised and coloured in Jalview using Clustalx colouring for 100% conservation (Larkin et al., 2007; Waterhouse et al., 2009).

MSTTDARPARYLFDSDYQRFV**PVHAVWEITLACDLKCLHCGSRAGHRRTNELST**
AECLEVIDALARLGTREVSLIGGEAYLRKDWTQLIRAIRSHGMYCAIQTGGRNL
TPKRLAQAVDAGLNGVGVSLDGLAPLHDKVRNVPGAFERALDTLRRARDAGIAV
SVNTQIGAQTMEDLPALMDTIIELGATHWQIQLTVAMGNAVDNDELLLQPYRLA
ELMPLLAKLYKDGVSRGLLMTVGNNIGYYGPYEHLWRGFGDERVHWSGCAAGQN
VIALEADGTVKGCPSLATVGFSGGNVRDMSLEDIWRTSEGIHFGRLRSVDDLWG
FCRTCYYADVCRGGCTWTSHSLLGKPGNNPYCHYRVLELQKQGLRERIAKVQDA
GPASFAVGRFDLVTERIADGEPVASVVRSGQVIELAWKNRGKRSPEVGRVPPKL
KMCRNCDGYVHAGEQTCPHCGGDIDAAARAHELDAQRRHALMNDLERLLGLPAS
TFGGG

Figure 3.6. Phyre2 model of residues 22-359 of BPSS0214
The model c4K39A (based on anaerobic sulfatase-maturing enzyme) (Goldman et al., 2013) is shown as cartoon, coloured by chainbows, the N-terminus is blue, leading to the C-terminus, in red. The sequence of BPSS0214 is shown with residues 22-359, covered by the model, highlighted in bold. Coordinates from Phyre2 (Kelley et al., 2015) and figure generated in PyMOL version 1.1r1 . BPSS0214 sequence from *B. pseudomallei* K96243 from the Burkholderia Genome Database (Winsor et al., 2008).

3.2.2 Expression of the BPSS0211-BPSS0214 operon

As mentioned previously, BPSS0212 and BPSS0213 were both identified in the stationary phase proteome of *B. pseudomallei*, with BPSS0212 more highly overexpressed in *B. pseudomallei*, than in *B. thailandensis* during stationary phase growth (Wongtrakoongate et al., 2007).

Comparative proteome analysis was carried out on morphotypes isolated from initial and relapse infections of *B. pseudomallei* (Velapatiño et al., 2012). The initial isolate was morphotype I, whereas the relapse isolate exhibited both morphotypes I and III (Velapatiño et al., 2012). BPSS0212 and BPSS0213 were quantified, using iTRAQ analysis, as having increased expression in the initial infection morphotype I compared to the relapse morphotype III (Velapatiño et al., 2012). BPSS0211 was also more highly overexpressed, but at lower levels than BPSS0212 and BPSS0213 (Velapatiño et al., 2012). BPSS0212 protein was also either absent or found to have lower expression, using 2D-PAGE, in morphotype III and morphotype I isolates of relapse *B. pseudomallei*, respectively, compared to *B. pseudomallei* isolated from the initial infection (Velapatiño et al., 2012). Therefore BPSS0211, BPSS0212 and BPSS0213 could have a possible role in establishing the initial infection of *B. pseudomallei*.

The expression of BPSS0211, BPSS0212, BPSS0213 and BPSS0214 is down regulated when *B. pseudomallei* is under oxidative stress by the addition of H_2O_2 (Jitprasutwit et al., 2014). Expression of the operon is also induced under low iron conditions (Tuanyok et al., 2005).

Ooi et al., 2013 carried out a study to investigate the effect specific environmental conditions had on the *B. pseudomallei* transcriptome. 82 different conditions were used. BPSS0211 was expressed in 51 conditions, whilst BPSS0212 and BPSS0213 were expressed in 54 conditions and BPSS0214 in 41 conditions. There are 41 conditions

where all four genes in the operon are expressed together, 10 conditions where only BPSS0211, BPSS0212 and BPSS0213 are expressed and three conditions where BPSS0212 and BPSS0213 are expressed. Genes that were co-expressed under similar conditions were grouped into clusters (Ooi et al., 2013). The operon BPSS0211-BPSS0214 was identified in cluster C095, along with proteins BPSL0350 (Fis family transcriptional regulator), BPSL0707 (cholesterol oxidase), BPSL1958 (hypothetical protein), BPSS0020 (hypothetical protein), BPSS0757 (porin protein), BPSS1657 (beta-glucosidase) and BPSS1858 (LysR family transcriptional regulator) (Ooi et al., 2013). The functional annotation of these genes is taken from the Burkholderia Genome Database (Winsor et al., 2008) and BLAST (Altschul et al., 1990).

3.2.3 Regulation of the BPSS0211-BPSS0214 operon

As mentioned previously, the expression of BPSS0211-BPSS0214 operon is under the control of the alternative sigma factor RpoS (Osiriphun et al., 2009). However, during chronic infection, RpoS was inactivated by a nonsense mutation in an isolate of *B. pseudomallei* 11.5 years post infection (Price et al., 2013). This suggests that the operon may not be required during the later stages of chronic infection. As well as their presence in the RpoS regulon, BPSS0212 and BPSS0213 expression has been found to be regulated by another sigma factor, RpoE (Thongboonkerd et al., 2007). RpoE negatively regulates BPSS0212 and BPSS0213 protein expression (Thongboonkerd et al., 2007).

The BpsI quorum-sensing system is regulated by RpoS in *B. pseudomallei* (Wongtrakoongate et al., 2012). The elucidation of the BpsI quorum-sensing system regulon revealed the genes for BPSS0212 and BPSS0213 (Wongtrakoongate et al., 2012). Both genes were present in the regulon for RpoS and quorum sensing; therefore these two proteins are co-regulated by RpoS and quorum sensing (Wongtrakoongate et al.,

2012).

All of the quorum sensing system regulons for each of *B. pseudomallei*, *B. mallei* and *B. thailandensis* were identified and compared, to investigate whether any of the same genes were regulated by the three species (Majerczyk et al., 2014a). BPSS0211 is present in the regulon of system 3 (BpsI3-BpsR3) from *B. pseudomallei,* whereas BPSS0211, BPSS0212 and BPSS0213 are all present in the regulon of quorum sensing system 3 (BtaI3-BtaR3) in *B. thailandensis* (Majerczyk et al., 2014a; Majerczyk et al., 2014b). The BPSS0211-BPSS0214 operon is not present in any of the *B. mallei* quorum-sensing regulons (Majerczyk et al., 2014a).

3.2.4 BPSS0211-BPSS0214 operon investigation summary

The information provided by sequence analysis and literature searches has provided a range of information, however the function of the proteins within the *B. pseudomallei* BPSS0211-BPSS0214 operon remains difficult to establish. The structure elucidation of the operon proteins using protein X-ray crystallography may assist with the functional assignment of these proteins.

3.3 BPSS0211

BPSS0211 is the first protein encoded by the BPSS0211-BPSS0214 operon. Previous work carried out by Dr. Matthew Day, determined the structure of BPSS0211, in the presence of zinc, to 2.2 Å (Figure 3.7). The structure consists of four molecules of BPSS0211 forming a dimer of dimers, with two zinc ion-binding sites associated with each monomer (Day, 2012). One zinc ion site is shared across a 2-fold axis (Day, 2012). Another zinc ion is present on the external surface of the dimer of dimers and makes crystal contacts (Day, 2012). It was assumed that the crystallisation solution, JCSG+ E6 – which contains 0.2 M zinc acetate, provides the zinc present in the structure (Day, 2012). Dr. Matthew Day

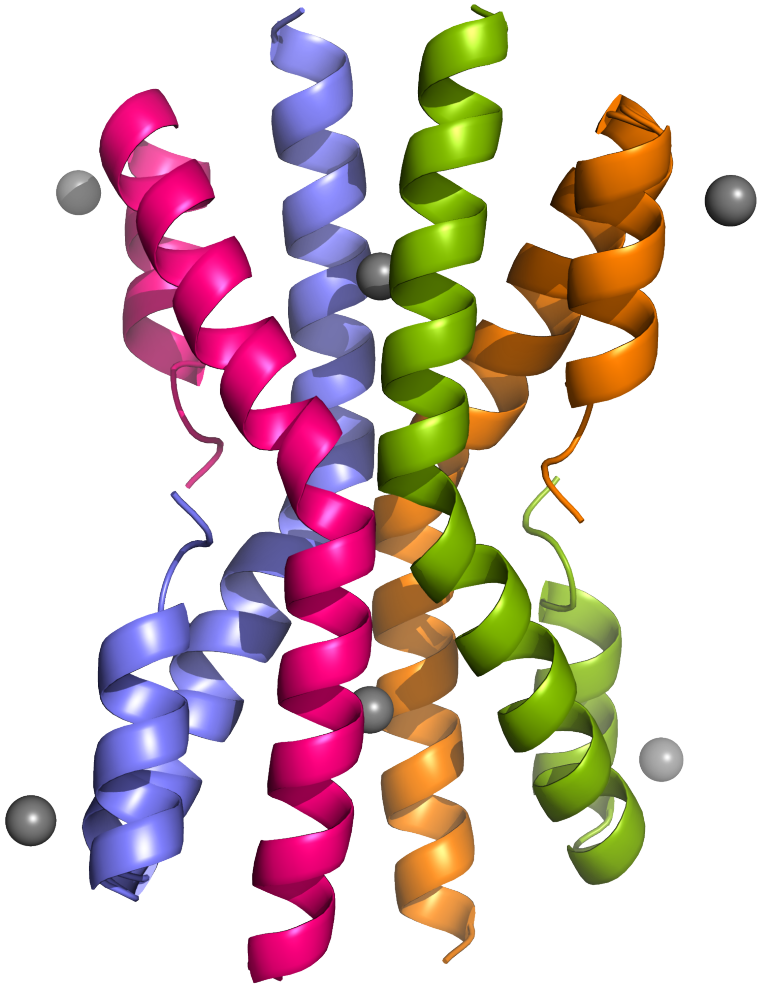


Figure 3.7. BPSS0211 dimer of dimers.
A dimer of dimers of BPSS0211 formed from four monomers. The two dimers are – pink/blue and orange/green. There are two zinc ion-binding sites per monomer, with one site shared across the 2-fold axis. The other site is present on the external surface of the dimer of dimers, the zinc atoms present in the structure are shown in grey. Image generated in PyMOL version 1.1r1 using BPSS0211 coordinates produced by Dr. Matthew Day (Day, 2012).

identified a different crystallisation condition, JCSG+ B6 that does not contain zinc, which also produces crystals of BPSS0211 (Day, 2012). The aim of this project is to continue the work carried out by Dr. Matthew Day in the crystallisation of BPSS0211 in zinc-free conditions and ultimately solve the structure. The BPSS0211 gene was cloned into plasmid pET21a and protein overexpressed in *E. coli* BL21 DE3, by Dr. Matthew Day previously (Day, 2012).

3.3.1 Native BPSS0211 protein purification

Native BPSS0211 was purified based on the protocol outlined in Dr. Matthew Day's Thesis (Day, 2012), by Dr. Matthew Day and Jason Tsang, an MBiolSci Masters project student. The purification was carried out as follows:

Following induction, *E. coli* cell pellet containing overexpressed BPSS0211 was resuspended in 50 mM Tris pH 9.0 buffer, sonicated for 3 x 20 seconds at 16 microns, and centrifuged at 70000 *g* for 15 minutes. The cell extract was loaded onto a DEAE-Sepharose column pre-equilibrated with 50 mM Tris pH 8.0 buffer. BPSS0211 was eluted in the flow-through fraction. An equal volume of 4 M ammonium sulphate was added to the flow-through, to produce a final concentration of 2 M ammonium sulphate. The mixture was stirred on ice for 20 minutes and then centrifuged at 70000 *g* for 20 minutes. The resulting pellet was dissolved in 20 mL 50 mM Tris pH 8.0 and the ammonium sulphate concentration was adjusted to 1.5 M with 4 M ammonium sulphate. The solution was centrifuged at 70000 *g* for 20 minutes. The ammonium sulphate concentration of the supernatant was adjusted to 2 M using the addition of 4 M ammonium sulphate and incubated at room temperature for 5 minutes. The solution was centrifuged at 70000 *g* for 20 minutes and the resulting pellet resuspended in 1 mL 50 mM Tris pH 8.0. The ammonium sulphate cut was followed by gel filtration chromatography, however BPSS0211 was not present, it was assumed the ammonium

sulphate cut was unsuccessful. The final supernatant ammonium sulphate concentration was increased to 2.4 M using 4 M ammonium sulphate and incubated overnight in the fridge. The solution was centrifuged at 70000 *g* for 20 minutes, the pellet resuspended in 1 mL 50 mM Tris pH 8.0 and loaded onto a Superose 6 column run with 50 mM Tris pH 8.0, 500 mM NaCl for 140 mL, with 2 mL fractions collected.

3.3.2 Crystallisation of native BPSS0211 protein

The concentrated BPSS0211 protein, ~6 mg/mL, was buffer exchanged into 10 mM Tris pH 8.0. Hanging-drop optimisation trials, designed by Dr. Matthew Day, were set down for a number of conditions that previously produced BPSS0211 crystals (Day, 2012). JCSG+ E6 – 0.2 M zinc acetate, 0.1 M imidazole pH 8.0, 20% (w/v) PEG3000 was optimised with 8-18% (w/v) PEG3000. Also JCSG+ B6 - 0.1 M phosphate-citrate buffer pH 4.2, 40% ethanol, 5% (w/v) PEG1000 was optimised with 0-9% (w/v) PEG1000, in increments of 3% and 30-45% (v/v) ethanol in increments of 3%. PACT B3 – 0.1 M MIB buffer pH 6.0, 25% (w/v) PEG1500 was optimised with 11-33% (w/v) PEG1500, in increments of 2% and MIB buffer at pH 5.5, pH 6.0 and pH 6.5. Crystallisation trials were carried out by Jason Tsang.

3.3.3 Data collection of native BPSS0211 zinc-free crystals

Crystals grown in the absence of zinc were cryoprotected in a range of solutions including – glycerol, ethylene glycol, PEG400 and sucrose, to test which was the best cryoprotectant. Cryoprotected crystals were cooled in liquid nitrogen and sent to the Diamond Light Source Synchrotron, Oxford, beamline I04-1. Data were collected on a crystal cryoprotected with 30% sucrose, grown in 0.1 M phosphate-citrate buffer pH 4.2, 36% (v/v) ethanol and 3% (w/v) PEG1000. 180° of data were collected with 0.4 second exposure over 0.2° phi oscillation, at a transmission of 43% (Figure 3.8). Data were processed using the xia2 3da pipeline to ~3 Å (Collaborative Computational Project, Number 4,

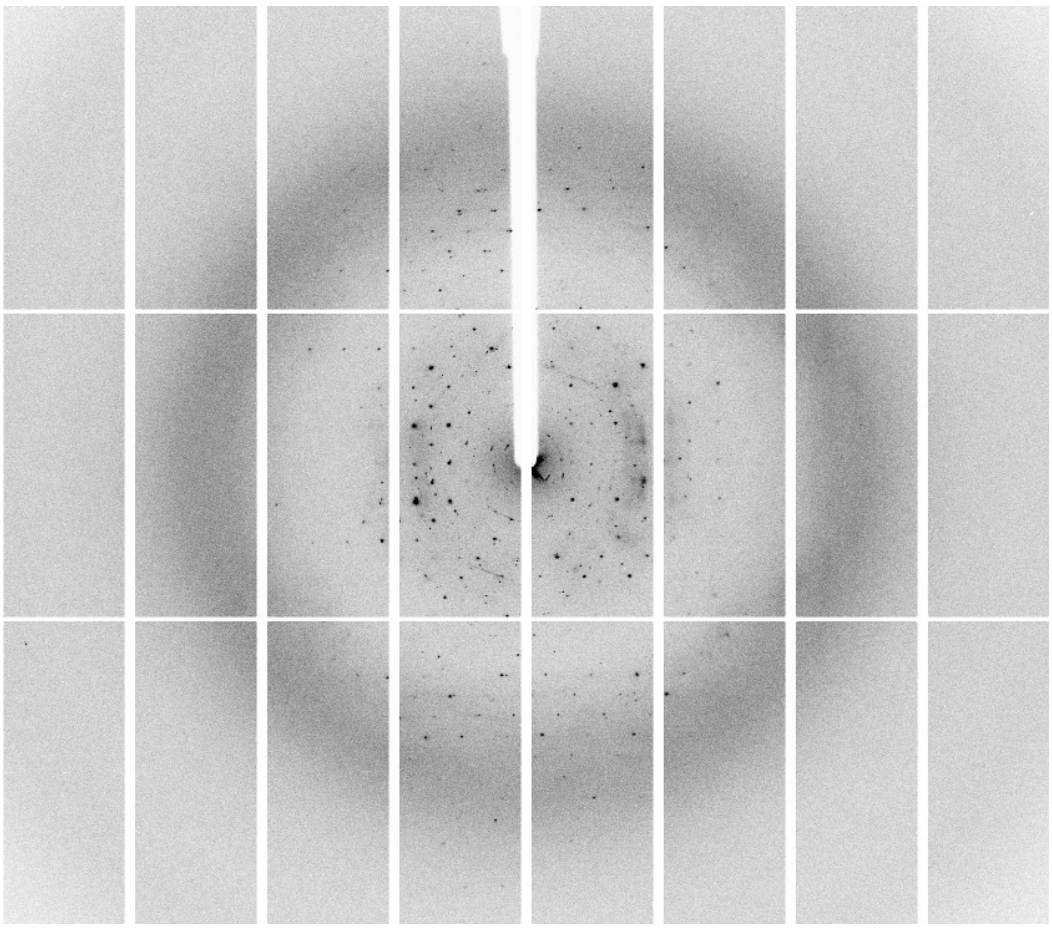


Figure 3.8. Diffraction data image of a crystal of native BPSS0211. Data collected at the Diamond Light Source Synchrotron, Oxford. Image generated from Diamond Light Source JPEGS.

1994; Evans, 2006; Kabsch, 2010; Sauter et al., 2004; Winter, 2010; Zhang et al., 2006) (Table 3.1).

3.3.4 Selenomethionine-incorporated BPSS0211 protein purification

A number of attempts were made to phase the native BPSS0211 data collected from crystals grown in zinc-free conditions using the previously solved structure of BPSS0211, via molecular replacement. Unfortunately this proved unsuccessful; therefore Jason Tsang, purified and set down initial crystallisation trials of selenomethionine-incorporated BPSS0211 protein, produced by Dr. Matthew Day.

E. coli cell pellet containing selenomethionine-incorporated BPSS0211 was resuspended in 50 mM Tris pH 9.0 buffer and sonicated for 3 x 20 seconds at 16 microns, then centrifuged at 43000 *g* for 15 minutes. The cell extract was loaded onto a DEAE-Sepharose column pre-equilibrated with 50 mM Tris pH 8.0. BPSS0211 elutes in the flow-through. An equal amount of 4 M ammonium sulphate was added to the flow-through fraction to produce a final concentration of 2 M ammonium sulphate. The mixture was stirred on ice for 20 minutes and then centrifuged at 43000 *g* for 25 minutes. The resulting pellet was dissolved in 20 mL 50 mM Tris pH 8.0 and the ammonium sulphate concentration was adjusted to 1.7 M with 4 M ammonium sulphate. The solution was centrifuged at 43000 *g* for 25 minutes. The ammonium sulphate concentration of the supernatant was adjusted to 2 M using the addition of 4 M ammonium sulphate and incubated with stirring on ice for 20 minutes. The solution was centrifuged at 43000 *g* for 25 minutes and the resulting pellet resuspended in 1 mL 50 mM Tris pH 8.0. The resuspended pellet was loaded onto a Superose6 column, run for 140 mL with 50 mM Tris pH 8.0, 500 mM NaCl buffer, with 2 mL fractions collected.

Data collection - BPSS0211	Native data	Selenomethionine data		
		Peak	Inflection	High Energy Remote
Wavelength (Å)	0.92000	0.97961	0.97980	0.96863
Resolution range (Å)	43.51-3.04 (3.12-3.04)	44.50-3.43 (3.52-3.43)	51.53-3.45 (3.54-3.45)	51.49-3.42 (3.51-3.42)
Space group	$P6_2$ 2 2	$P6_2$ 2 2	$P6_2$ 2 2	$P6_2$ 2 2
Unit cell parameters				
a (Å)	59.20	59.45	59.5	59.46
b (Å)	59.20	59.45	59.5	59.46
c (Å)	87.02	88.48	88.53	88.69
α (°)	90	90	90	90
β (°)	90	90	90	90
γ (°)	120	120	120	120
Total reflections	33207 (2770)	23832 (1839)	23480 (1819)	24283 (1849)
Unique reflections	1982 (142)	1450 (98)	1430 (98)	1470 (100)
Multiplicity	16.8 (19.5)	16.4 (18.8)	16.4 (18.6)	16.5 (18.5)
Completeness (%)	99.3 (99.2)	99.4 (98.9)	99.6 (98.8)	99.5 (98.9)
Mean I/σ	13.9 (4.3)	12.7 (3.2)	13.0 (3.0)	12.2 (3.6)
Wilson B factor ($Å^2$)	88.3	160*	169*	108.9
R_{merge}	0.095 (0.563)	0.130 (1.030)	0.128 (1.057)	0.142 (0.998)
R_{pim} (I)	0.025 (0.132)	0.038 (0.246)	0.036 (0.254)	0.039 (0.241)
Anomalous completeness	-	99.2 (98.7)	99.4 (98.6)	99.4 (98.7)
Anomalous multiplicity	-	10.2 (10.6)	10.1 (10.5)	10.2 (10.6)
Anomalous correlation	-	0.712 (0.099)	0.453 (0.129)	0.158 (0.020)
Anomalous slope	-	1.373	1.244	1.056

Table 3.1. Data collection processing statistics for BPSS0211 in the absence of zinc. xia2 data processing statistics for BPSS0211 native and selenomethionine MAD collected data, using the 3da pipeline (native and peak) (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Sauter et al., 2004; Winter, 2010; Zhang et al., 2006) and 3daii pipeline (inflection and high energy remote) (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Winter, 2010). * From Wilson (Winn et al., 2011). R_{merge} – merging R-factor, R_{pim} – precision-indicating merging R-factor.

3.3.5 Crystallisation of selenomethionine-incorporated BPSS0211 protein

Selenomethionine-incorporated BPSS0211 protein containing fractions were pooled and concentrated in a 5 kDa molecular weight cut-off Vivaspin concentrator (Sartorius Stedim). The concentration of the protein was analysed by the Bradford assay, and indicated the protein concentration was 6 mg/mL. However during crystallisation the protein began to precipitate, and A_{280nm} analysis revealed a concentration of 38 mg/mL, using the extinction coefficient calculated from the protein sequence by ProtParam (Gasteiger et al., 2005). The binding efficiency of the dye used in the Bradford assay can vary with different protein (Bradford, 1976). The 38 mg/mL protein preparation was set down in sitting-drop crystallisation trials, in a 1:1 ratio, using screens JCSG+, $AmSO_4$, PEGs, Classics and PACT. The protein preparation was diluted from 38 mg/mL to 8 mg/mL and set down in sitting-drop crystallisation trials using screens JCSG+, PACT, PEGs and Classics. Crystal hits were produced in a range of conditions, including JCSG+ B6, the zinc-free crystallisation condition (Figure 3.9). Hanging drop optimisation trials were set down for condition JCSG+ B6.

3.3.6 Data collection of selenomethionine-incorporated BPSS0211 zinc-free crystals

Crystals grown in optimised zinc-free crystallisation conditions, in cryoprotectant containing 0% PEG1000, 38% ethanol, 0.1 M phosphate-citrate buffer pH 4.2 and 30% glycerol derived from conditions JCSG+ B6 were cryoprotected, cooled in liquid nitrogen and sent to the Diamond Light Source Synchrotron, Oxford, for screening on beamline I02. A fluorescence scan was carried out at the selenium K edge to check for selenium incorporation in the crystal. The *CHOOCH* (Evans and Pettifer, 2001) output provided the peak and inflection wavelengths required for data collection – 12656.0 eV (electron volts) (f'' 8.91 e, f' -7.79 e) and 12654 eV (f'' 5.12 e, f' -11.39 e), respectively. The high-energy remote

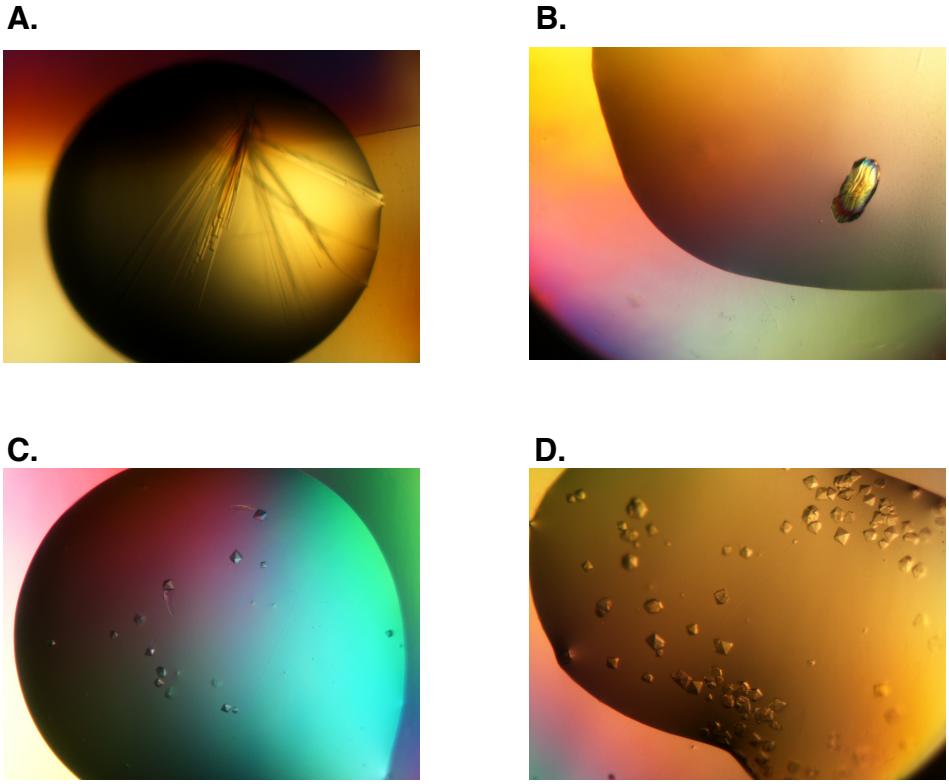


Figure 3.9. Crystal hits from sitting-drop crystallisation trials for selenomethionine-incorporated BPSS0211

Protein was set down at concentrations of 38 mg/mL and 8 mg/mL. **A.** PACT A2 (38 mg/mL) – 0.1 M SPG buffer pH 5.0, 25% (w/v) PEG1500. **B.** JCSG+ B6 (38 mg/mL) – 0.1 M Phosphate-citrate pH 4.2, 40% (v/v) ethanol, 5% (w/v) PEG1000. **C.** JCSG+ E6 (8 mg/mL) – 0.2 M zinc acetate, 0.1 M imidazole pH 8.0, 20% (w/v) PEG3000. **D.** JCSG+ E6 (38 mg/mL) – 0.2 M zinc acetate, 0.1 M imidazole pH 8.0, 20% (w/v) PEG3000.

wavelength was set at 12800 eV. Data for all three wavelengths were collected at 0.15° phi oscillation for 0.05 seconds over 180°, with a transmission of 20%. All three wavelengths of data were processed using the xia2 3da pipeline (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Sauter et al., 2004; Winter, 2010; Zhang et al., 2006) or 3daii pipeline (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Winter, 2010) (Table 3.1).

3.3.7 Experimental phasing of selenomethionine-incorporated BPSS0211 zinc-free crystal data

The autoSHARP pipeline (Vonrhein et al., 2007) was run with peak, inflection and high-energy remote datasets and a BPSS0211 sequence, which output a map, model and selenium sites. The final map produced by autoSHARP (Vonrhein et al., 2007) for the experimental phasing of BPSS0211 crystallised in the absence of zinc suggests the crystal contains an extended helix (Figure 3.10), unlike the bent helices seen in the BPSS0211 structure solved previously (Day, 2012) (Figure 3.7). One selenium site was identified and is thought to belong to Met28, as it is in the middle of the helical density, with the N-terminal methionine as the only other methionine in the sequence.

The location of Met28 was used to determine the register of the helix, with the direction of the helix having two possible orientations. The helical density contained a number of sidechain bulges, the orientation of which relative to the core was used to indicate the directionality of the helix. The model of the zinc-containing BPSS0211 was aligned in fragments into the density, to produce the initial model for zinc-free BPSS0211.

The model was refined against the high-energy remote dataset, at 3.42 Å, using Refmac5 (Murshudov et al., 1997; Murshudov et al., 2011). The phases output from autoSHARP (Vonrhein et al., 2007), providing the Hendrickson-Lattman coefficients was combined in CAD (Winn et al.,

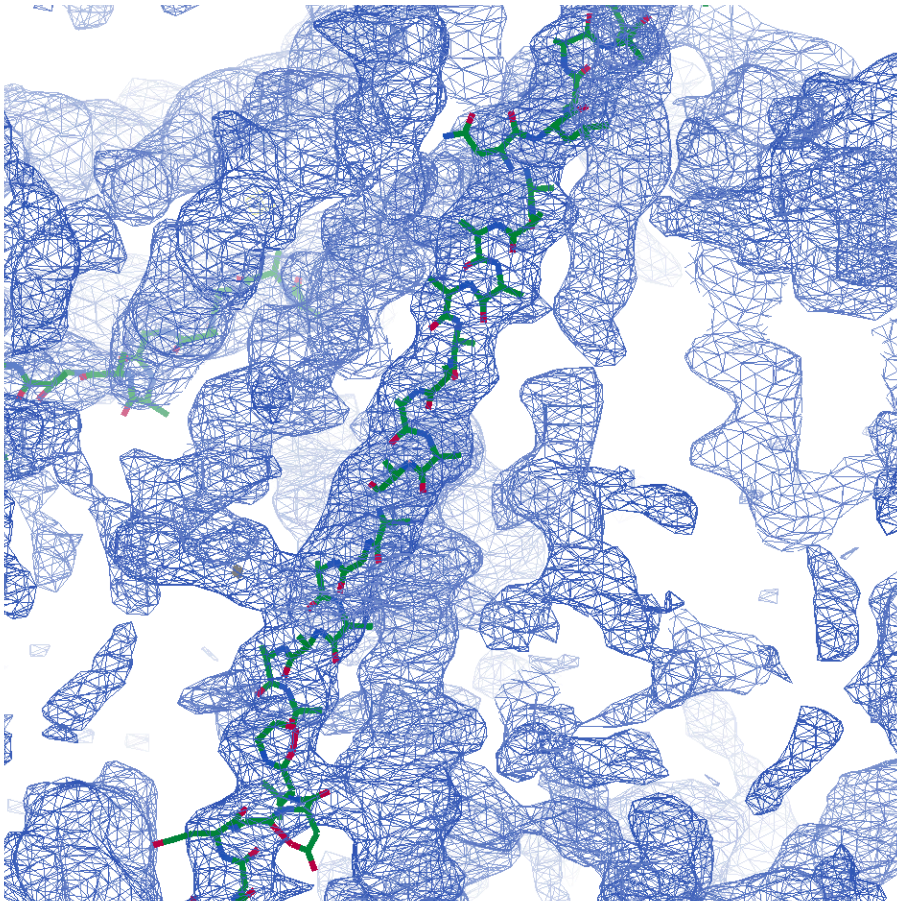


Figure 3.10. Selenium sites, map phases and model output from autoSHARP. The electron density output from autoSHARP (Vonrhein et al., 2007) is coloured blue and the selenium site is a grey pointer atom. The model is shown as atoms. Electron density for α-helices can be clearly seen. Map set to 0.2252e/A^3 = 0.98 rmsd (root mean square deviation). Figure generated in *Coot* (Emsley et al., 2010).

2011), with the calculated phases from the model. Phase combination was carried out in SigmaA (Winn et al., 2011). The model was then built to the map output from SigmaA (Winn et al., 2011) in *Coot* (Emsley et al., 2010). Cycles of building, refinement and phase combination were carried out against the high-energy remote data, to produce the current model of BPSS0211 in the absence of zinc. Higher resolution native data were collected previously to 3.04 Å (Table 3.1), however due to the change in c cell dimension from ~87 Å to ~88.5 Å, the model could not be refined against the native data. Therefore the BPSS0211 model in the absence of zinc produced from experimental phasing was used to molecular replace the native data. The output model then underwent cycles of refinement and validation using Refmac5 (Murshudov et al., 1997; Murshudov et al., 2011), against the native dataset. However there were some problems during refinement with the native data, and a satisfactory solution could not be produced, the route of this problem has yet to be elucidated. Therefore the model was refined to the high-energy remote dataset, at 3.42 Å, as previously mentioned.

3.3.8 Current zinc-free BPSS0211 structure

The current model of zinc-free BPSS0211 structure, built to the high-energy remote data, contains residues 12-58, of the expected 63 residues, with an R_{free} = 0.4442 and R_{factor} = 0.3596, with the current validation statistics outlined in Table 3.2. No waters or ligands were identified. A further feature of the crystal is unmodelled density seen between the layers of helices (Figure 3.11). Analysis of the separation between the helices revealed a distance of ~14 Å, this distance is too small for the placement of another helix, however a strand may be able to fit. There are five unmodelled residues at the C-terminus of BPSS0211, so one possibility is that these residues form a strand which protrudes up between the helices. The density however is too weak to interpret and any possible connection between the helical density and possible strand density cannot be seen at this resolution.

Refinement	**BPSS0211 (no zinc)**
R_{factor}	0.3596
R_{free}	0.4442
No. of non-H atoms	
Protein	348
Ligands/Metal ions	0
Water	0
Protein residues	47
RMSD (bonds) (Å)	0.0127
RMSD (angles) (°)	1.5711
Ramachandran favoured (%)	93.33
Ramachandran outliers (%)	0.00
Favoured rotamers (%)	73.58
Poor rotamers (%)	2.94
MolProbity score	1.86 (100^{th} percentile (N=611, 3.42Å ± 0.25Å)
Average B factors	
Main chain ($Å^2$)	114
Side chains ($Å^2$)	114

Table 3.2. Refinement and validation statistics for BPSS0211 in the absence of zinc. Statistics generated in Refmac5 (Murshudov et al., 1997, Murshudov et al., 2011), MolProbity (Chen et al., 2010) and Baverage (Winn et al., 2011).

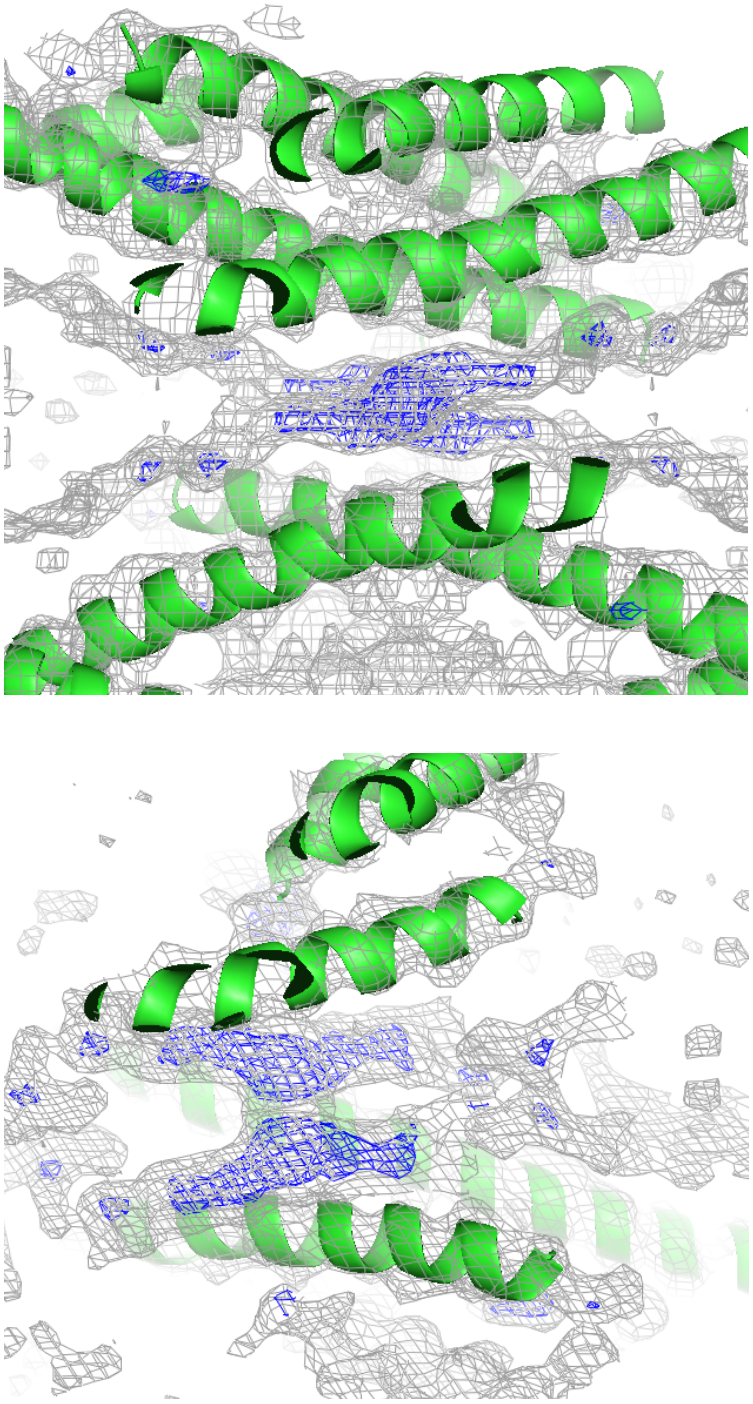


Figure 3.11. Unmodelled electron density seen between layers of BPSS0211 molecules in the crystal lattice. BPSS0211 structure shown as cartoon (green). The 2Fo-Fc (FWT, PHWT) map at 1.0 sigma (grey) and Fo-Fc (DELFWT, PHDELFWT) map at 3.0 sigma (blue). Map generated in FFT (Winn et al., 2011). Figures generated in PyMOL version 1.1r1.

The BPSS0211 structure, in the presence of zinc (Day, 2012), in which the long and short helices pack together, is what will be referred to as the closed conformation, in the absence of zinc, is in an extended or open conformation. The first helix in the BPSS0211 structure in the presence of zinc structure is flipped out and is part of a longer extended helix, or open conformation, in the absence of zinc (Figure 3.12). The difference in structure highlights the flexibility of the structure and that a conformational change takes place from open to closed. The change in conformation of the helices may be due to the pH of the crystallisation solution, from ~pH 8.0 to ~pH 4.2.

The zinc involved in crystal contacts in the zinc-containing structure is coordinated by the bent helical section of BPSS0211 in two molecules related by symmetry (Day, 2012), and therefore may play a role in stabilising the closed helix conformation.

3.3.9 Zinc-free BPSS0211 structure analysis

Multiple possible dimer interfaces were identified between helices in the crystal, of those seen; PDBePISA identified one as significant and potentially stable in solution (Krissinel and Henrick, 2007) (Figure 3.13). This dimer interface of BPSS0211, has an interface area of ~750 $Å^2$ out of a monomer surface area of 4457 $Å^2$ (Krissinel and Henrick, 2007), and the interacting region covers ~the last two thirds of both helices involved . The interacting residues in both monomers are M28, L31, R32, Q34, A35, L38, L39, Q41, Q42, L45, A46, A48, L49, L52, E53, E55, I56, A57 and K58 (Figure 3.14) (Krissinel and Hendrick, 2007).

The BPSS0211 extended conformation structure was built from the fitting of fragments of BPSS0211 closed conformation structure; therefore the overall fold of the protein should be maintained. This indicates that the way in which the monomeric chains of BPSS0211 interact and the interface used will also be retained. A number of extra residues are

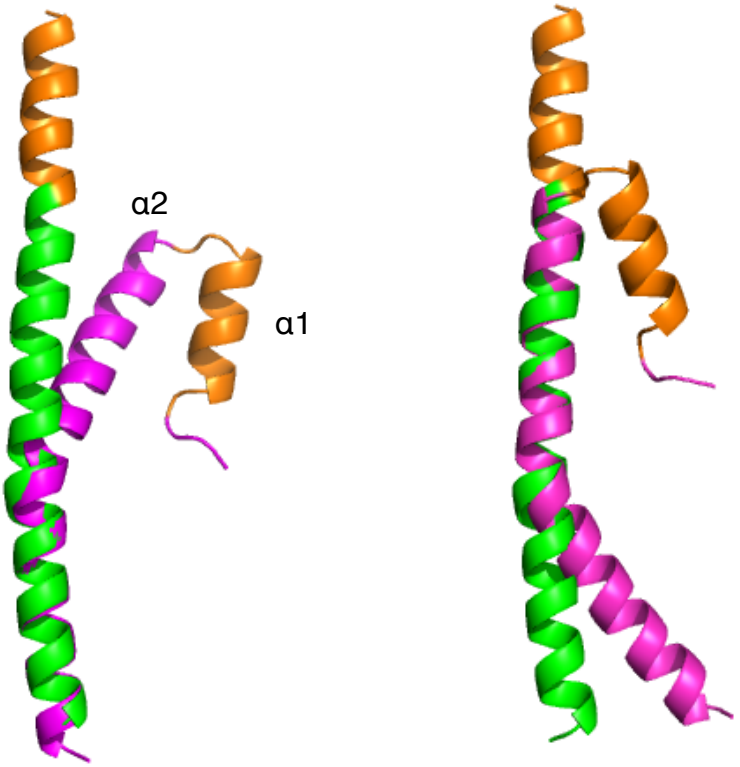


Figure 3.12. BPSS0211 open conformation structure compared to the BPSS0211 closed conformation structure. BPSS0211 in the absence of zinc is shown as the extended helix – open conformation (green/orange) and BPSS0211 in the presence of zinc is shown as two helices, a longer helix and a short helix – closed conformation (magenta/orange). The closed conformation can be aligned to two sections of the open conformation separately. The residues corresponding to the α1-α2 turn and α1 helix in the closed conformation are highlighted in the same colour (orange) in the open helix conformation. Images generated in PyMOL version 1.1r1, BPSS0211 closed conformation coordinates from Dr. Matthew Day (Day, 2012).

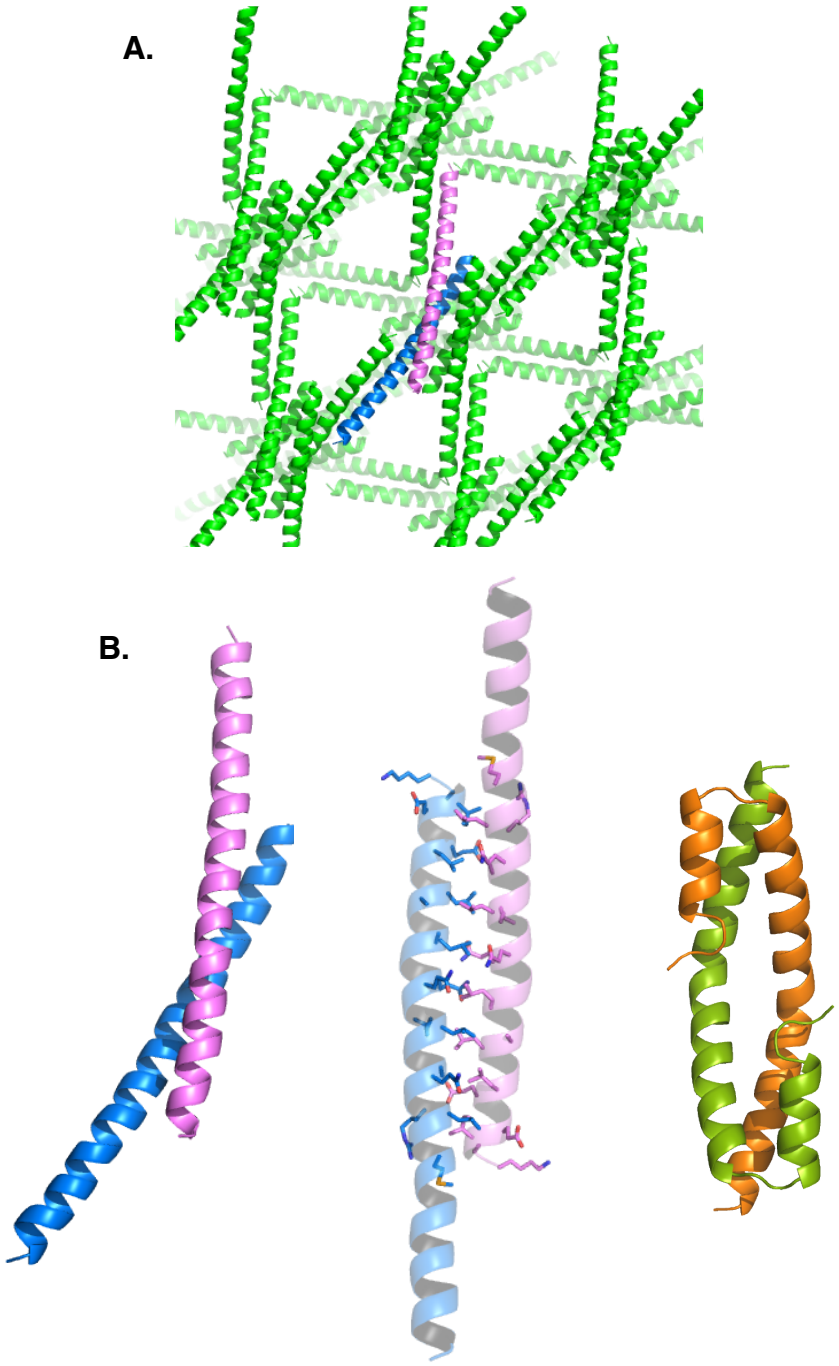
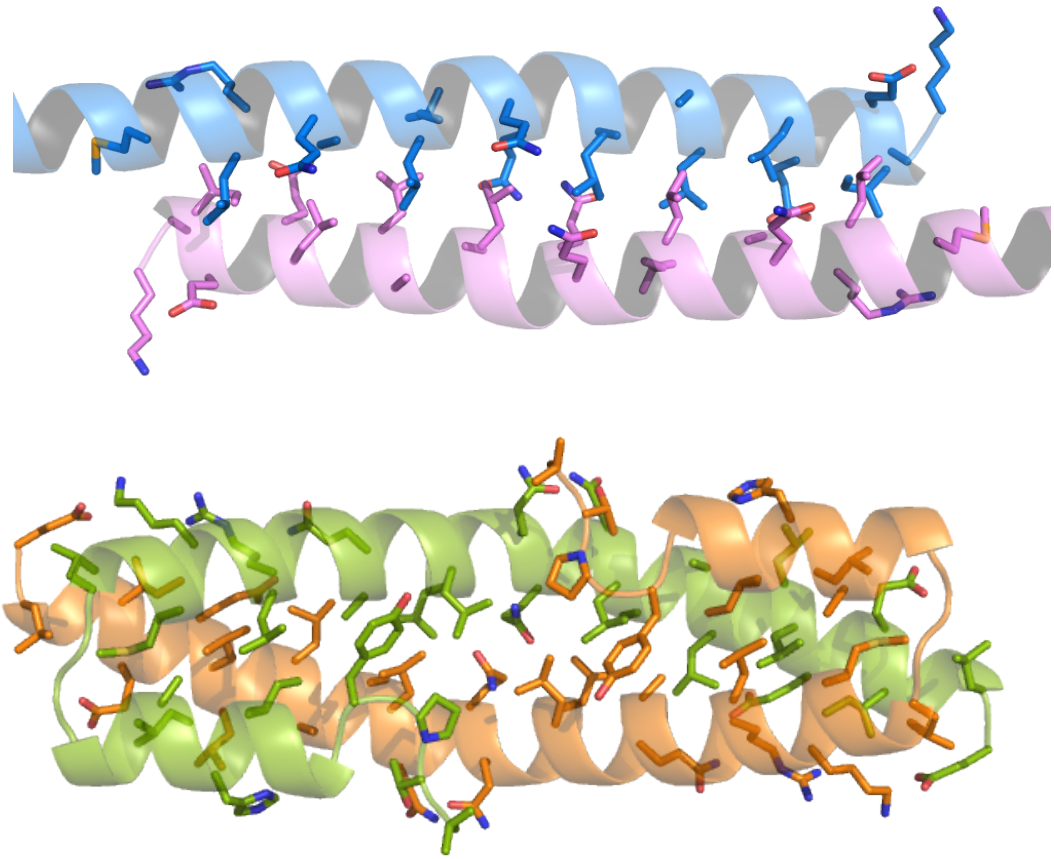


Figure 3.13. BPSS0211 dimer

A. The BPSS0211 open conformation crystal lattice is shown in green, with the dimer identified by PDBePISA (Krissinel and Hendrick, 2007) coloured blue and pink. **B.** The interacting residue sidechains are highlighted between the two open conformation BPSS0211 monomers (blue/pink). The interacting residues cover a similar region as the dimer interface between the α2 helices of the closed conformation of BPSS0211 dimer (orange/green). Images generated in PyMOL version 1.1r1, BPSS0211 closed conformation coordinates from Dr. Matthew Day (Day, 2012).



BPSS0211 (open)
~~MSQAQGHPVTP~~YGVAIHQAIADGDLAQMKSLRTQAQALLAQQGNLATALELLEV
EIAK~~LERRK~~

BPSS0211 (closed)
~~MSQAQGHP~~VTPYGVAIHQAIADGDLAQMKSLRTQAQALLAQQGNLATALELLEV
EIAKLE~~RRK~~

Figure 3.14. BPSS0211 dimer interface
The residues involved in the closed conformation BPSS0211 dimer interface (blue/pink) and the open helix conformation BPSS0211 dimer interface (orange/green) are mapped onto the structures, and highlighted in the corresponding BPSS0211 sequence. Crossed through residues indicate those absent from the structure. Interface and interface residues identified by PDBePISA (Krissinel and Hendrick, 2007) and images generated in PyMOL version 1.1r1, BPSS0211 closed conformation coordinates from Dr. Matthew Day (Day, 2012).

involved in the BPSS0211 interface in the presence of zinc due to the interaction between helices α1 and α2 within a monomer. Comparison of the dimer formation found the orientation of the helices to be similar, with the closed conformation helical dimer producing more curved helices than in the open conformation. A formation of helices similar to the dimer of dimers of BPSS0211 in the presence of zinc could be seen within the BPSS0211 open conformation crystal lattice, produced by the crystal symmetry (Figure 3.15). The low resolution of the data and its quality limits the precision of the structure. Nevertheless the current interpretation of the nature of the helical packing is consistent across both the zinc-bound (closed) and zinc-free (open) structures. Further work is required to produce higher resolution data.

3.4 BPSS0211(1-123)

3.4.1 BPSS0211(1-123) - a new protein annotation of BPSS0211

Analysis of the upstream DNA from the annotated BPSS0211 protein-coding region in *B. pseudomallei* K96243, using ExPasy translate (Gasteiger et al., 2003), identified a potential alternative start codon. The new N-terminal methionine is 60 residues upstream of the originally annotated N-terminal methionine, indicating the BPSS0211 protein may be nearly twice as long as first thought (Figure 3.16). An alignment between the BPSS0211(1-123) protein sequence from *B. pseudomallei* K96243 is compared to that from *B. pseudomallei* D286 in Figure 3.17. Two residue differences are identified at positions 39 (C39R) and 41 (A41V). The selection criteria for the investigation of this protein included the presence of the protein in the RpoS regulon. The RpoS site was identified as CTAACCT (Osiriphun et al., 2009) and is located upstream of the newly proposed BPSS0211(1-123) transcription site. The DNA upstream of the coding region for BPSS0211 homologs in other *Burkholderia* species was analysed to see if possible misannotation had taken place in other BPSS0211 homologous proteins. Interestingly the homolog of BPSS0211 in *B. thailandensis* E264, BTH_II2184, is

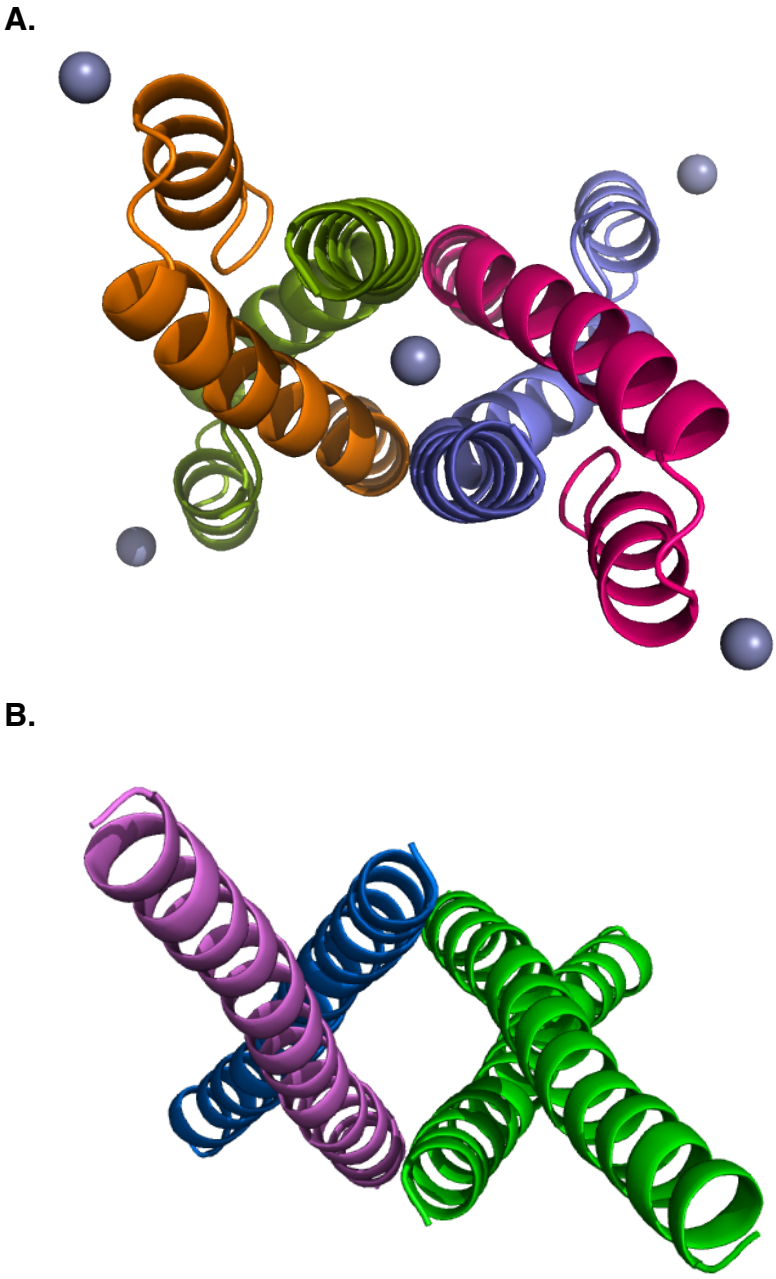


Figure 3.15. BPSS0211 helical bundles in crystal lattice
A. The dimer of dimers structure of the closed conformation BPSS0211 structure, with zinc ions shown as grey spheres. **B.** A four helical arrangement of BPSS0211 open conformation identified in the crystal lattice of the open conformation of BPSS0211. Images generated in PyMOL version 1.1r1, BPSS0211 closed conformation coordinates from Dr. Matthew Day (Day, 2012).

A.

CCGATCGGACATCGATCGTCGTCATCCGTTCCCTCCGGCGGGCGTGCGCCGCTT
CCCCGCCGCCGAGCGCGCGCCGGTTGCCGGCGACGCAACGATCCGGCCGGGCAT
TGTGCGATGGGCGCGCCGCAACGCGCGTCGCGCCGGTCGCCCGCCTGCGGGTAC
TACCAAATCTGGTAGATAGCGAAAGCGCGCAGGCGACCTAACCTTAGTACGCAG
ACCAATCGACAACATGGAAGCGCCCCGCCGCACCGGCGGGCGCACGGGGACAAC
CGCGTTACCGCCGTGGACCGCCGACGTGCCGGTTCAGAGGGATAACACGAATGG
CTAATCACGGGCGGCGAGCAGTGCGGCACGAGGCACCGTGTCGCAGCATGGCGG
CGAGCGCGACATTCATCGACGAAATTCCCCGATCCTGCACAAACACGACAGCCG
GATGCACTGCGCCGCCGCGGGGCAGATCGCAGGGCTGTCGGCCATCGTCCACTT
TCAGCGAGGCAAACATGAGTCAAGCACAGGGCCATCCGGTCACCCCTTATGGCG
TCGCCATCCATCAGGCGATCGCGGACGGCGATCTCGCGCAGATGAAATCGCTGC
GCACGCAGGCGCAGGCGCTGCTCGCCCAGCAGGGCAATCTCGCCACCGCGCTCG
AGCTTCTCGAAGTCGAGATCGCCAAGCTCGAGCGGAGGAAGTAG

B.

MANHGRRAVRHEAPCRSMAASATFIDEIPRSCTNTTAGCTAPPR
GRSQGCRPSSTFSEANMSQAQGHPVTPYGVAIHQAIADGDLAQM
KSLRTQAQALLAQQGNLATALELLEVEIAKLERRK

Figure 3.16. New protein annotation for BPSS0211 from *B. pseudomallei* K96243

A. The DNA sequence of the annotated BPSS0211 protein and the upstream DNA, the underlined sequence corresponds to the shorter construct of BPSS0211 and the red highlighted sequence corresponds to BPSS0211(1-123). **B.** The amino acid sequence from the transcription of the red highlighted DNA sequence. The shorter annotation of BPSS0211 is underlined. *B. pseudomallei* K96243 DNA sequence from the Burkholderia Genome Database (Winsor et al., 2008) and translated by ExPASy translate (Gasteiger et al., 2003).

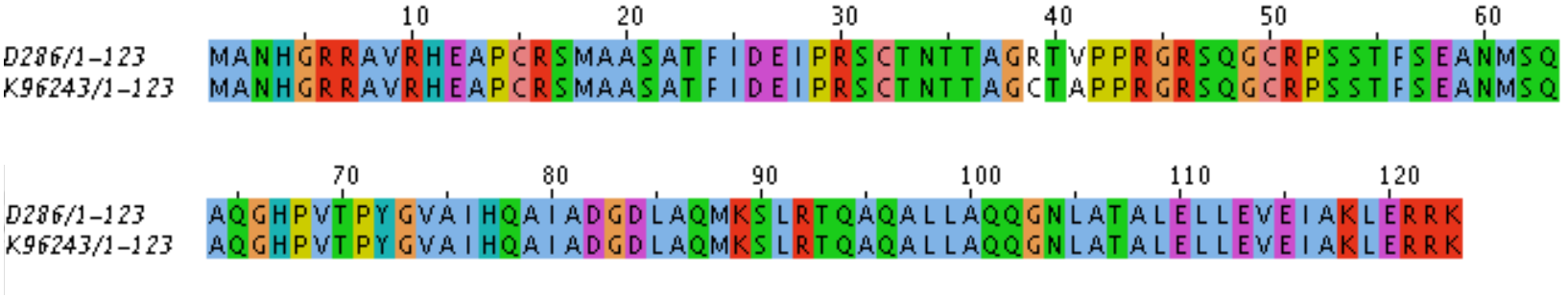


Figure 3.17. Comparison of the newly annotated BPSS0211(1-123) protein sequences from *B. pseudomallei* K96243 and *B. pseudomallei* D286. *B. pseudomallei* K96243 sequences from Burkholderia Genome Database and aligned in Jalview using ClustalOmega and coloured using Clustalx for 100% conservation (Larkin et al., 2007; Sievers et al., 2011; Winsor et al., 2008; Waterhouse et al., 2009)

annotated as a longer version of the protein at 131 residues. Also *B. mallei* ATCC 23344, BMAA1865 is annotated as a length in between the longer and shorter constructs at 106 residues. An alignment was carried out between the BPSS0211 homologs in other *Burkholderia* species (Winsor et al., 2008) (Figure 3.18). Interestingly another example of protein misannotation was found for the *bpss1516* gene in *B. pseudomallei* K96243 (Muangman et al., 2011). The new annotation of the gene is thought to have an extra 40 residues encoded upstream of the annotated start site (Muangman et al., 2011).

The BPSS0211(1-123) sequence was analysed using both the Phyre2 (Kelley et al., 2015) server and PSI-PRED (Buchan et al., 2013; Jones 1999) to produce a secondary structure prediction. Both programs predicted helical regions in the extended N-terminal section of BPSS0211(1-123). Analysis of the extra N-terminal residues using the signal sequence prediction programs Predtat and SignalP revealed no sign of a signal peptide (Bagos et al., 2010; Petersen et al., 2011).

3.4.2 Cloning and overexpression of BPSS0211(1-123)

Primers were designed for the new annotation of BPSS0211 and BPSS0211(1-123) was cloned into pET21a. The cloning protocol was carried out as specified in the materials and methods, section 2.3.4. Optimal soluble expression of BPSS0211(1-123) was produced under conditions of 37°C for 2 hours, with 1 mM IPTG induction (Figure 3.19). The molecular weight of this new construct is ~13236 Da, calculated using ProtParam (Gasteiger et al., 2005).

3.4.3 Initial purification of BPSS0211(1-123)

The pI (isoelectric point) of the BPSS0211(1-123) was predicted to be 9.73 using ExPasy ProtParam (Gasteiger et al., 2005). A purification protocol was designed, with the help of Dr. Svetlana Sedelnikova (University of Sheffield), using cation-exchange chromatography, with a

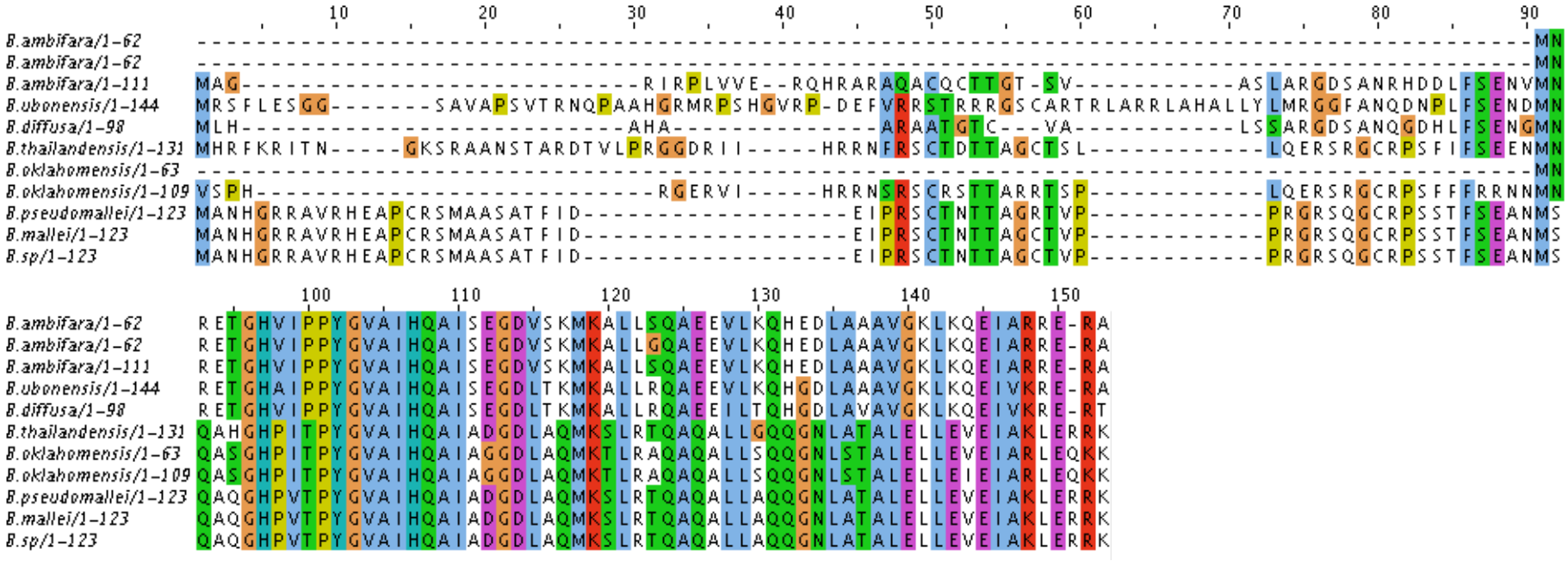


Figure 3.18. Alignment of BPSS0211 amino acid sequences from *Burkholderia* species

Burkholderia strains and protein used, in order, shown below: *B. ambifara* MC40-6 – BamMC406_5555, *B. ambifara* IOP40-10 – WP_006753565.1, *B. ambifara* MC40-6 - BamMC406_5555, *B. ubonensis* RF23-B916 – WI30-RS07190, *B. diffusa* - MSMB583WGS, *B. thailandensis* E264 – BTH_II2184 , *B. oklahomensis* C6786 – BG90_RS19035, *B. oklahomensis* E0147 – DM82_RS21300, *B. pseudomallei* D286 – BPSS0211 , *B. mallei* ATCC 23344 – BMAA1865, and *Burkholderia* species TSV202 – X994_RS05390. Sequences from Burkholderia Genome Database (Winsor et al., 2008) and *B. pseuodmallei* D286 genome, translation if required performed using ExPasy translate (Gasteiger et al., 2003) and aligned in Jalview using T-Coffee and coloured using Clustalx (Larkin et al., 2007; Notredame et al., 2000; Waterhouse et al., 2009).

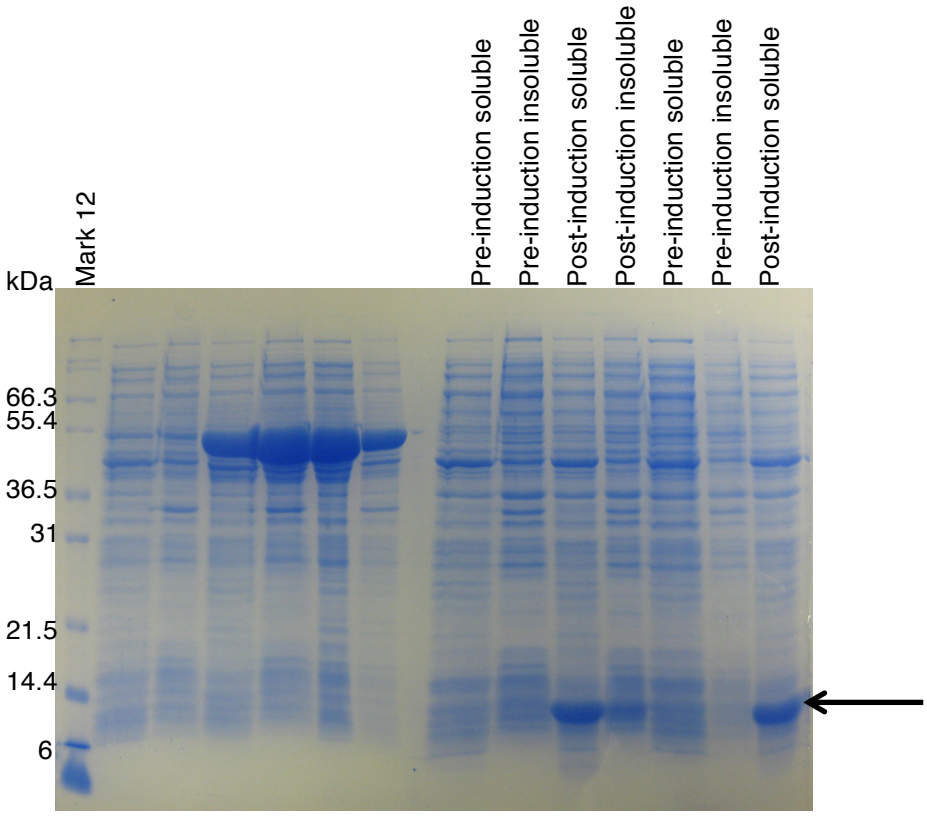


Figure 3.19. SDS-PAGE analysis of the BPSS0211(1-123) protein overexpression. Protein was overexpressed at 37°C for 2 hours. The first half of this gel is analysis of an unrelated protein. The molecular weight of BPSS0211(1-123) is ~13236 Da. The black arrow indicates the band of BPSS0211(1-123) present in the post-induction soluble fraction.

buffer pH of 7.0, followed by gel filtration chromatography.

Following induction, *E. coli* BL21 DE3 cell pellet containing overexpressed BPSS0211(1-123) was resuspended in 50 mM HEPES pH 7.0 buffer, sonicated for 3 x 20 seconds at 16 microns, and centrifuged at 18000 rpm/39191 *g* in a JA.20 rotor for 15 minutes. The cell extract was loaded onto a HiTrap SP HP 5 mL column (GE) pre-equilibrated with 50 mM HEPES pH 7.0 at 5 mL/minute. The column was washed with 2 CV at 5 mL/minute. BPSS0211(1-123) was eluted over a gradient of 0-100% 50 mM HEPES pH 7.0, 500 mM NaCl, in 10 CV at 5 mL/minute with 2 mL fractions collected (Figure 3.20 and 3.21). Pooled protein was concentrated in a 10 kDa molecular weight cut-off Vivaspin (Sartorius Stedim). The concentrated protein was made up to 2 mL with gel filtration buffer – 50 mM Tris pH 8.0, 500 mM NaCl. Unfortunately the protein solution precipitated prior to loading onto the gel filtration column. The 2 mL of load sample was centrifuged at 13000 rpm/10-15500 *g* for 10 minutes. The supernatant was removed from the pelleted protein and again the volume made up to 2 mL with 50 mM Tris pH 8.0, 500 mM NaCl and loaded onto a HiLoad 16/600 Superdex200 pg gel filtration column (GE) equilibrated with 50 mM Tris pH 8.0, 500 mM NaCl, run at 1.5 mL/minute. 2 mL fractions were collected over 0.2 CV-1 CV (Figure 3.22). The molecular weight was estimated as ~45 kDa using the calibration graph in Figure 5.10. This indicates the protein is running as a possible trimer, however there are uncertainties in predicting the molecular weight from the gel filtration elution volume, which is affected by shape (McGuiness, 1973).

3.4.4 Initial crystallisation of BPSS0211(1-123)

Gel filtration elution fractions were pooled and concentrated in a 10 kDa molecular weight cut-off Vivaspin (Sartorius Stedim), however during concentration, the BPSS0211(1-123) protein concentration, rather than increase it started to decrease. It was assumed the protein was adhering

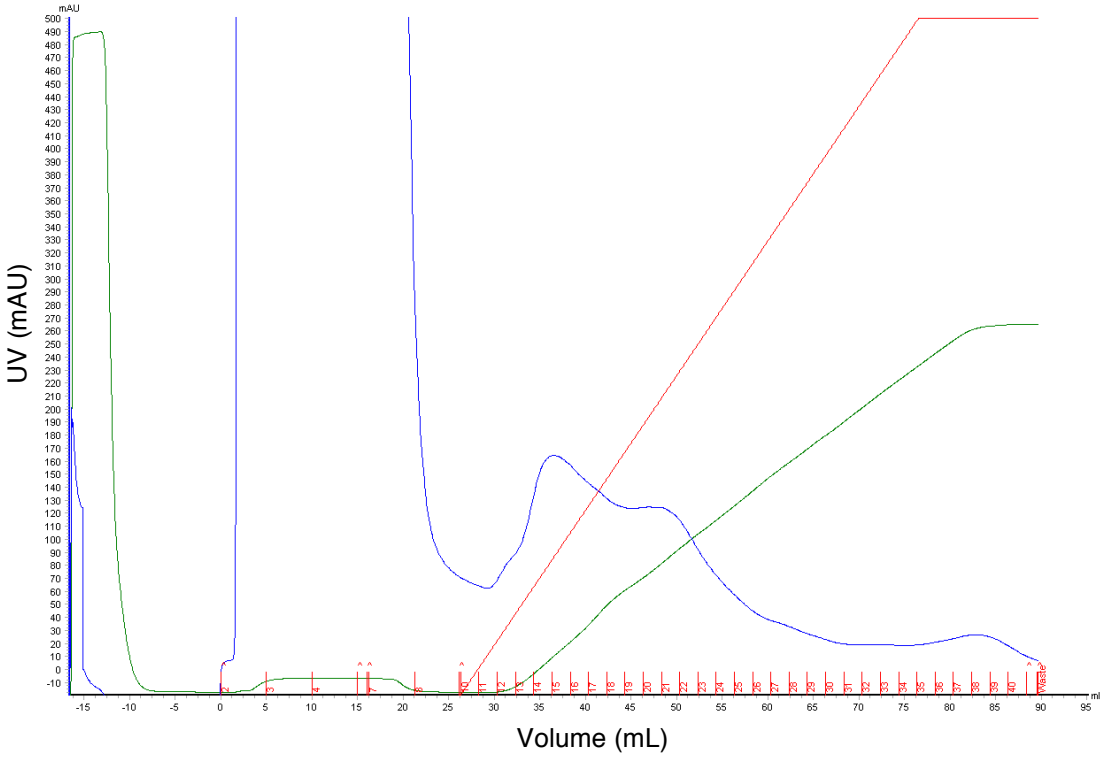


Figure 3.20. Chromatogram of BPSS0211(1-123) run on a SP HP cation-exchange column. mAU at A_{280} – blue, conductivity – green, elution fractions – red numbers, elution gradient – red.

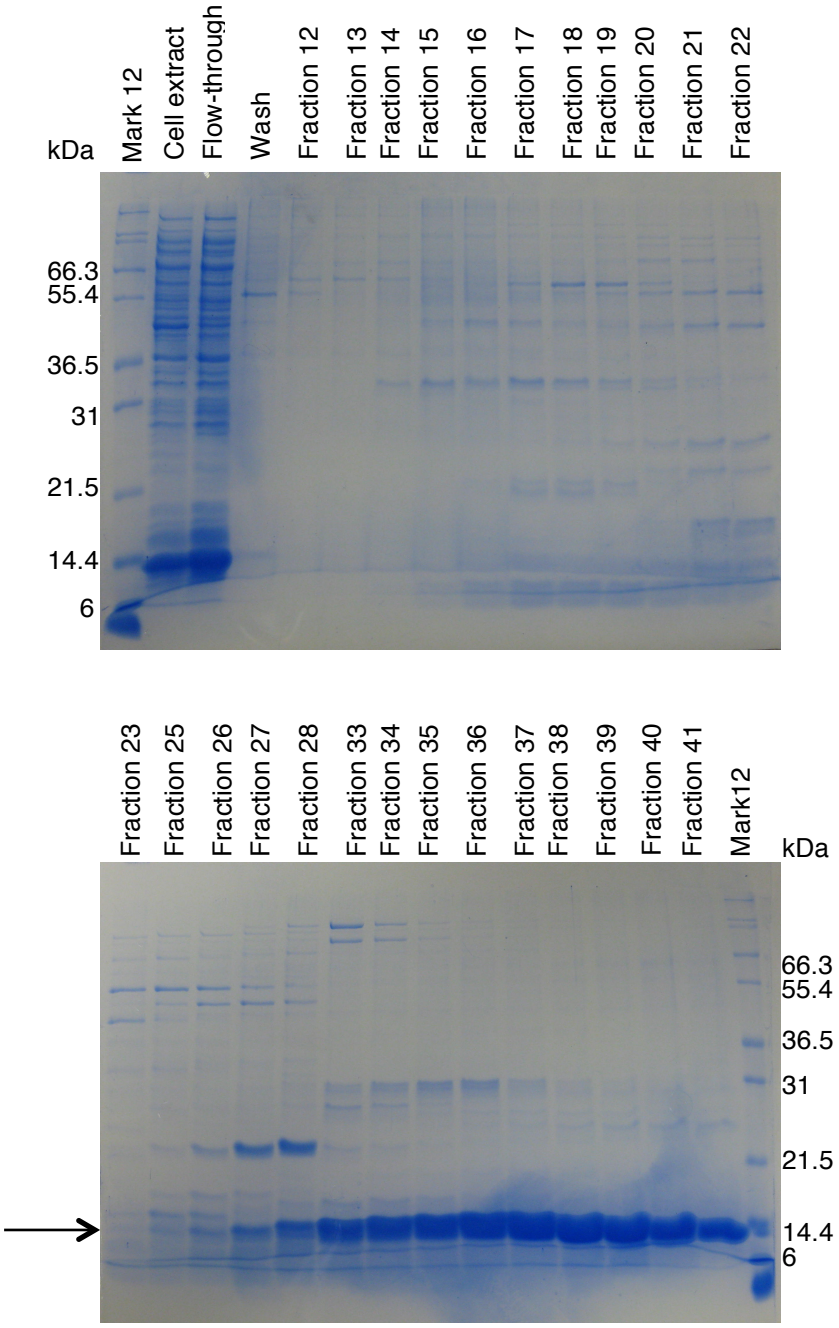


Figure 3.21. SDS-PAGE analysis of the BPSS0211(1-123) protein SP HP cation exchange purification. The molecular weight of BPSS0211(1-123) is ~13236 Da, and the corresponding protein band is indicated by a black arrow.

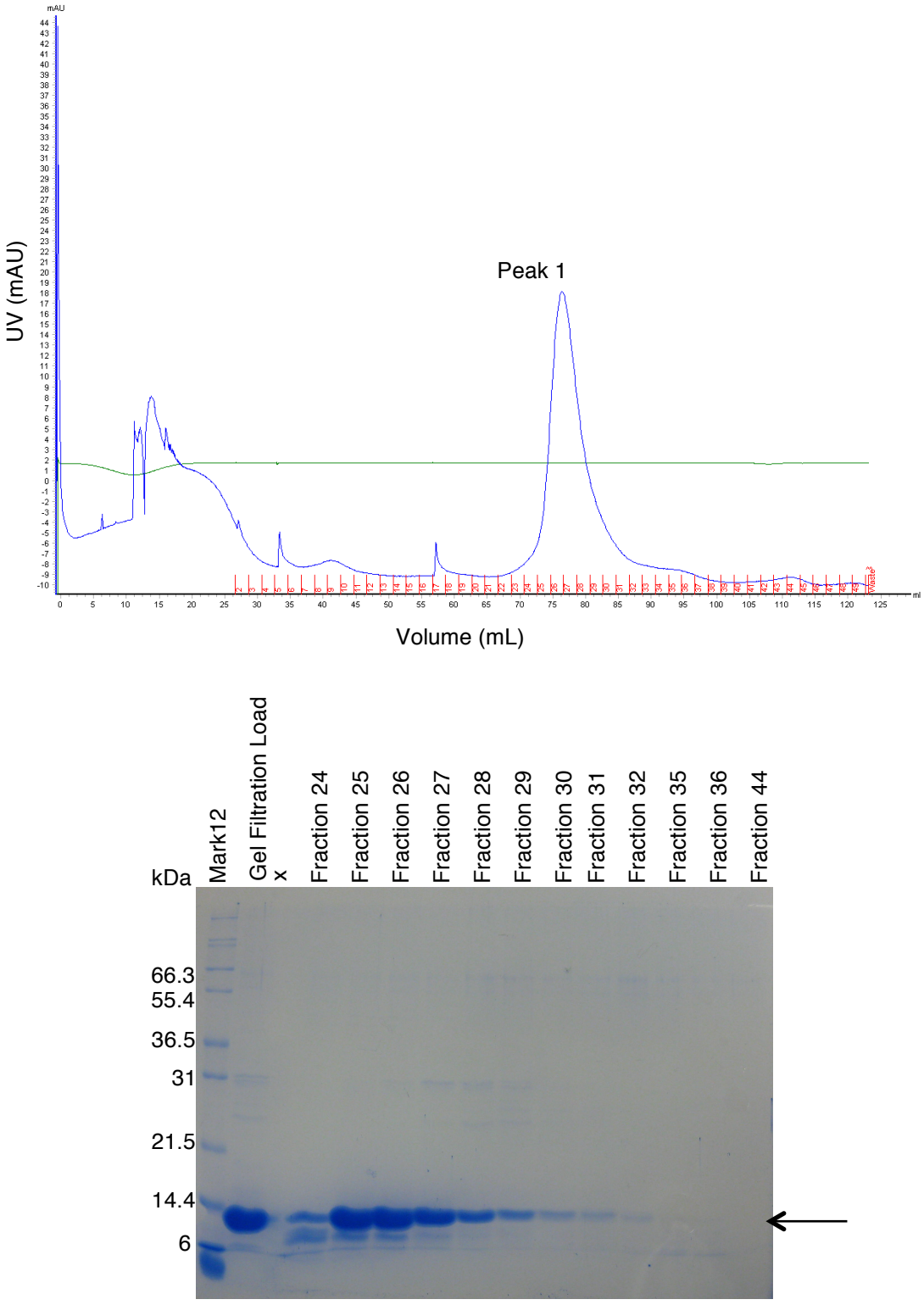


Figure 3.22. Chromatogram and SDS-PAGE analysis of the BPSS0211(1-123) protein gel filtration purification. Peak 1 – 76.42 mL. The molecular weight of BPSS0211(1-123) is ~13236 Da, and the corresponding protein band is indicated by the black arrow.

to the concentrator filter. Protein was concentrated to ~5.5 mg/mL and remained in buffer 50 mM Tris pH 8.0, 500 mM NaCl. Sitting-drop crystallisation robot trays were set down with crystallisation suites JCSG+, PACT, Classic, pHClear, Proplex, MPD and $AmSO_4$. A number of crystal hits were produced predominantly in the MPD screen and other tray conditions containing MPD, however some of the crystals were produced in unmixed drops. The crystals produced were small and rectangular (Figure 3.23).

Hanging-drop optimisation crystallisation screens were set up for conditions MPD C2 (0.2 M sodium malonate, 40% (v/v) MPD), MPD H7 (35% (v/v) MPD), JCSG+ H4 (0.2 M calcium chloride, 0.1 M Bis-Tris pH 5.5, 45% (v/v) MPD), MPD D5 (0.2 M sodium bromide, 40% (v/v) MPD), MPD B9 (0.2 M sodium formate, 40% (v/v) MPD), pHClear H4 (0.1 M HEPES pH 7.0, 40% (v/v) MPD) and MPD B5 (0.2 M sodium nitrate, 40% (v/v) MPD). For all optimisation trays only the percentage MPD was altered. The MPD concentration was varied from 35-45% (v/v) in 5% increments for conditions MPD D5, MPD B9, MPD B5, MPD C2 and pHClear H4. The MPD H7 condition was optimised using 30-40% (v/v) MPD in increments of 5% and for JCSG+ H4 the MPD concentration was varied from 40-50% (v/v) MPD in 5% increments. The crystals produced by the hanging-drop optimisation trials were no larger than those produced from sitting-drop optimisation trials, and were too small to be sent for data collection.

3.4.5 BPSS0211(1-123) crystals sent for screening

Crystals from the sitting-drop crystallisation trials of BPSS0211(1-123) in conditions MPD D7, MPD A9, MPD D11, MPD F12, MPD B1 and MPD B12 were cooled in liquid nitrogen and sent to beamline I04 at the Diamond Light Source Synchrotron, Oxford for screening. Crystals from conditions MPD B1, MPD B12, and MPD A9 produced low-resolution protein diffraction (Figure 3.24 and Figure 3.25). The other crystals sent

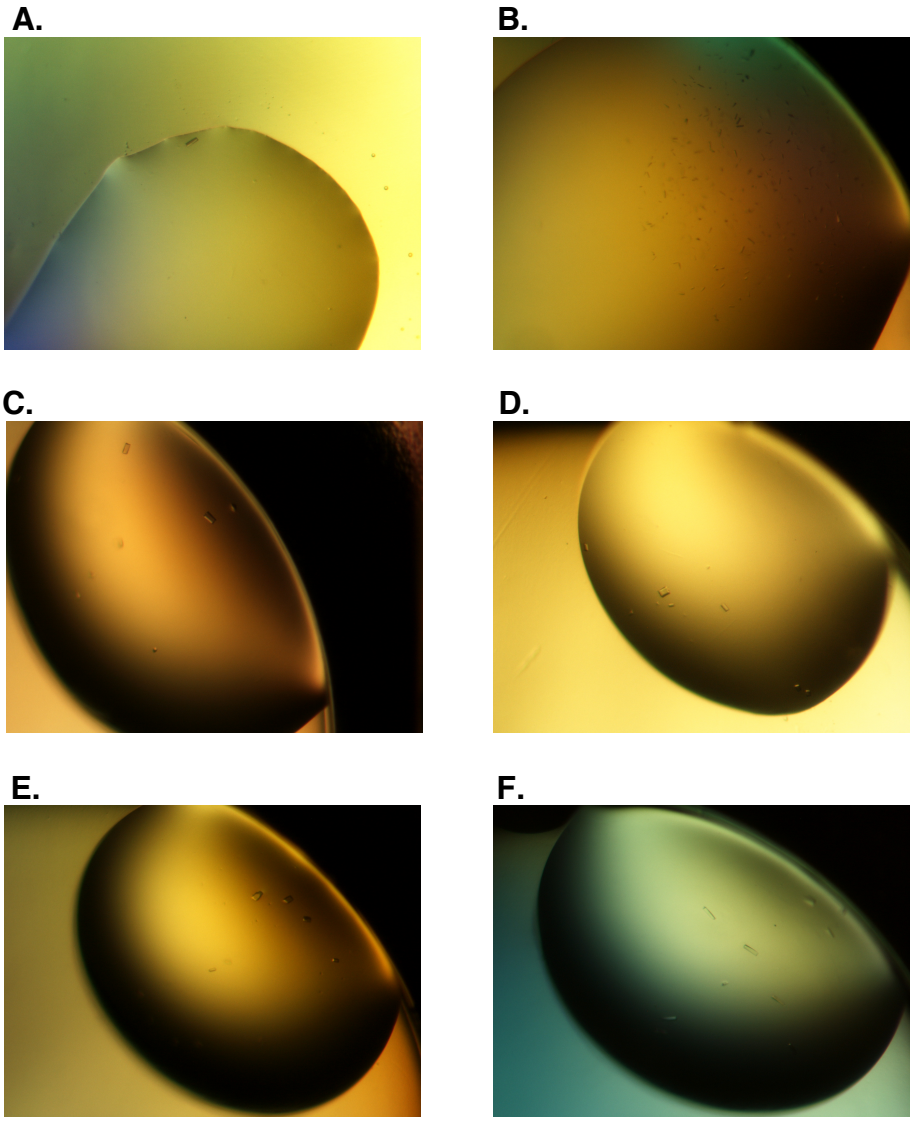


Figure 3.23. Crystal hits from sitting-drop crystallisation trials of BPSS0211(1-123). Protein was set down at 5.5 mg/mL in 50 mM Tris pH 8.0, 500 mM NaCl. **A.** pHClear H4 - 0.1 M HEPES, 40% (v/v) MPD, pH 7.0. **B.** MPD H7 – 35% (v/v) MPD. **C.** MPD D7 – 0.2 M ammonium phosphate, 40% (v/v) MPD. **D.** MPD F12 - 0.1 M BICINE pH 9.0, 65% (v/v) MPD. **E.** MPD B1 – 0.2 M sodium thiocyanate, 40% (v/v) MPD. **F.** MPD B12 – 0.2 M lithium acetate, 40% (v/v) MPD.

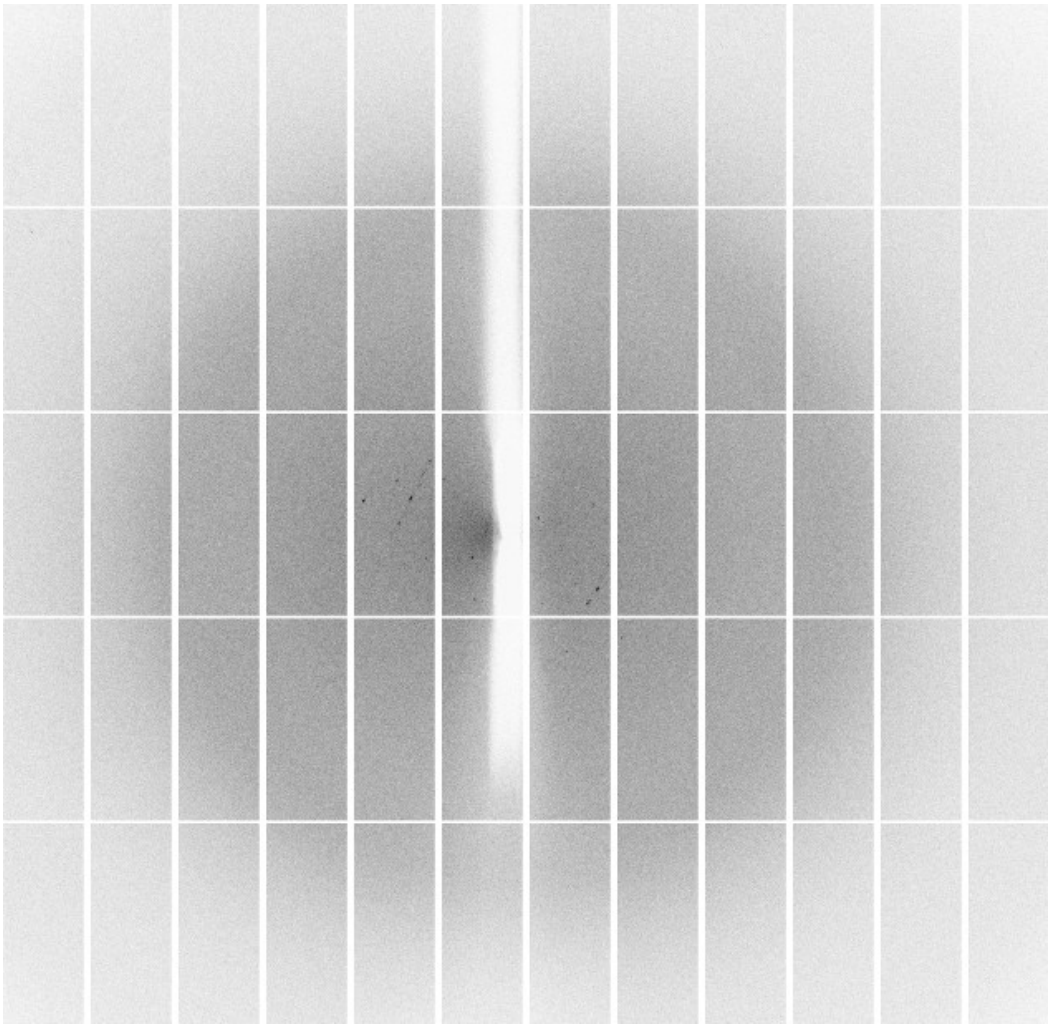


Figure 3.24. Diffraction test image of a crystal of BPSS0211(1-123) from condition MPD B1. Diffraction test image taken for 0.5 seconds over 0.5° oscillation, with the edge of the detector plate set to 2.5 Å, collected at the Diamond Light Source Synchrotron, Oxford. Image from Diamond Light Source JPEGS.

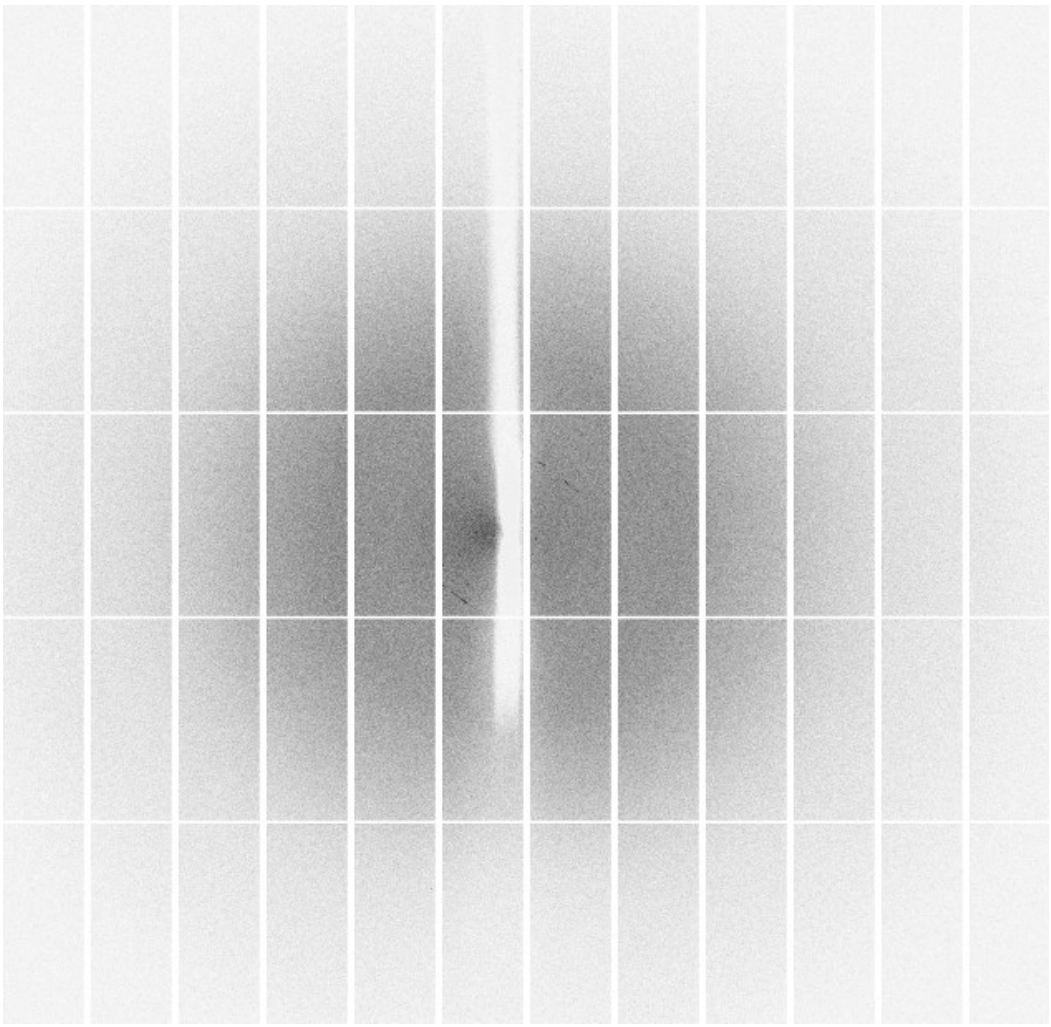


Figure 3.25. Diffraction test image of a crystal of BPSS0211(1-123) from condition MPD B12. Diffraction test image taken for 0.5 seconds over 0.5° oscillation, with the edge of the detector plate set to 2.0 Å, collected at the Diamond Light Source Synchrotron, Oxford. Image from Diamond Light Source JPEGS.

did not produce diffraction, possibly due to the small size of the crystal.

3.4.6 Further crystallisation optimisation of BPSS0211(1-123) crystals

BPSS0211(1-123) protein was purified using the above protocol for further crystallisation trials to be set down at a higher protein concentration and additional hanging-drop optimisations. However, protein could only be concentrated to ~6 mg/mL. Repeat sitting-drop crystallisation robot screens were set down for screens PACT, JCSG+, Proplex and a new screen, Morpheus, was used.

Hanging-drop optimisation crystallisation trials were set down for condition MPD B1 and MPD B12, with 30-50% (v/v) MPD. For MPD B1, the concentration of sodium thiocyanate was varied from 0.1 M-0.3 M in 0.05 M increments. For MPD B12 the concentration of lithium acetate was varied from 0.05 M-0.25 M in 0.05 M increments. Trays were set down with 1000 µL reservoir and two drops per well of 2:2 and 2:1 ratio of protein:reservoir.

3.4.7 Optimisation of the BPSS0211(1-123) purification protocol

During previous preparations of the BPSS0211(1-123) protein, precipitation occurred prior to loading onto the final gel filtration column and the protein could not be concentrated past ~5.5 mg/mL, as it was assumed to stick to the concentrator as the concentration increased. Therefore the purification protocol of BPSS0211(1-123) was optimised with a variety of different buffers and pH, carried out by Dr. Svetlana Sedelnikova (University of Sheffield), who also prepared the protein for crystallisation.

The cell extract containing BPSS0211(1-123) was loaded onto a 5 mL HiTrap SP HP column (GE) pre-equilibrated with 50 mM HEPES pH 7.5. BPSS0211(1-123) was eluted over a gradient of 15-100% 50 mM HEPES

pH 7.5, 1 M NaCl, in 10 CV at 5 mL/minute with 2.5 mL fractions collected. The purified protein, concentration 4.5 mg/mL, was buffer exchanged into three separate buffers, 5 mM Tris pH 8.0 50 mM NaCl, 50 mM sodium citrate pH 5.5 and 10 mM MES pH 6.5 100 mM NaCl. The final two samples were carried through, ~4 mg/mL in 10 mM MES pH 6.5, 100 mM NaCl and 5 mg/mL in 50 mM sodium citrate pH 5.5. A further sample of purified protein was added to an equal volume of 0.2 M sodium citrate pH 5.5 and concentrated further using a Vivaspin concentrator (Sartorius Stedim). The protein was then buffer exchanged into 50 mM sodium citrate pH 5.5, with a final concentration of ~10 mg/mL.

Gel filtration chromatography was carried out on purified protein samples from the SP HP column, with two different buffers – 50 mM Tris pH 8.0, 500 mM NaCl and 50 mM sodium citrate pH 5.5. The peak fractions from the gel filtration run with 50 mM sodium citrate pH 5.5, were pooled and concentrated to 9.5 mg/mL.

3.4.8 Crystallisation of new BPSS0211(1-123) protein preparations

The three protein preparations of BPSS0211(1-123), ~4 mg/mL (in 10 mM MES pH 6.5, 100 mM NaCl), 5 mg/mL (50 mM sodium citrate pH 5.5) and 10 mg/mL (50 mM sodium citrate pH 5.5) were crystallised in sitting-drop crystallisation trays with screens PACT, JCSG+ and Proplex with 200 nL:200 nL drops and 50 µL reservoir. The BPSS0211(1-123) sample produced in 50 mM sodium citrate pH 5.5 at 9.5 mg/mL was set down in sitting-drop crystallisation trials with screens PACT, JCSG+, Morpheus, Proplex and PGA, with 200 nL:200 nL drops and 25 µL reservoir. The crystallisation trials did not produce any crystals.

3.5 Conclusion and further work on BPSS0211 and BPSS0211(1-123)

3.5.1 BPSS0211

The structure of BPSS0211 has been produced in the absence of zinc, and has been shown to have a related structure to BPSS0211 in the presence of zinc. The function of BPSS0211 has yet to be elucidated, however the conformational change demonstrated between the structures may play a role. Further work would involve obtaining higher resolution data for the zinc-free structure in order to confirm the current model produced here. The unmodelled N- and C-termini residues present in both BPSS0211 structures could be removed and a new construct designed. The shorter construct may have improved crystal packing and thus produce crystals of higher diffraction quality.

3.5.2 BPSS0211(1-123)

BPSS0211(1-123) protein has been overexpressed, purified and crystallised. Crystals produced in conditions MPD B1, MPD B12 and MPD A9 produced protein diffraction, however the crystals were not of a high enough quality for data to be collected. Therefore these crystallisation conditions could be further optimised to try and produce crystals of a high enough quality for data collection. Additionally further crystallisation with protein purified using the optimised protocol may produce crystals in other crystallisation conditions from different untested screens. The structure of the longer BPSS0211 construct could produce further information regarding the function.

Chapter 4. BPSS0212

BPSS0212 is the second protein encoded by the BPSS0211-BPSS0214 operon under investigation. The gene was re-cloned, overexpressed, purified, crystallised and the structure solved using selenomethionine MAD.

4.1 Purification of BPSS0212

The initial purification of BPSS0212 was based on the protocol used in previous work carried out by Dr. Matthew Day (Day, 2012) and was performed with the aid of Dr. Svetlana Sedelnikova (University of Sheffield).

4.1.1 Initial purification protocol

Following induction, *E. coli* BL21 DE3 cell pellet containing overexpressed BPSS0212, was resuspended in 10 mL/g of 50 mM Tris pH 8.0 and sonicated at 16 microns for 3 x 20 seconds. The cell lysate was then centrifuged at 20000 rpm/~48400 *g* in a JA.20 rotor (Beckman Coulter) for 15 minutes. The cell extract was loaded onto a DEAE-Sepharose anion-exchange column pre-equilibrated with 50 mM Tris pH 8.0. BPSS0212 was eluted and collected in 5 mL fractions over a gradient of 0-80% 150 mM NaCl over ~175 mL, followed by 80-100% 250 mM NaCl over 25 mL, remaining at 100% 250 mM NaCl for a further 40 mL. Fractions were analysed for the presence of protein using the Bradford assay and SDS-PAGE. The fractions containing BPSS0212 were pooled, diluted, to lower the salt concentration, and loaded onto a 6 mL ResourceQ column (GE) pre-equilibrated with 50 mM MES pH 6.5, and eluted over a gradient of 0-20% 50 mM MES pH 6.5, 1 M NaCl over 20 CV. Fractions containing BPSS0212 were pooled and concentrated using a 30 kDa molecular weight cut-off Vivaspin concentrator (Sartorius Stedim) at 4870 *g*. During concentration a number of problems were encountered, including precipitation of the protein, which were overcome by experimentation involving altering the salt concentration of the sample

and using a new concentrator during protein concentration. BPSS0212 was finally concentrated to ~3 mL and loaded on a gel filtration column pre-equilibrated with 50 mM Tris pH 9.0, 500 mM NaCl. The column was run at 1 mL/minute over 140 mL. 2 mL fractions were collected. The gel filtration fractions were analysed by SDS-PAGE.

4.1.2 Degradation of BPSS0212 as a result of purification

The elution fractions from each column in the initial purification were analysed by SDS-PAGE (Figures 4.1 and 4.2). As described previously by Dr. Matthew Day (Day, 2012), BPSS0212 undergoes degradation during purification. The degradation of the protein results in two dominant protein bands. A band is present at ~28 kDa, which is thought to correspond to the full-length protein, albeit that it appears to run at an anomalously high molecular weight, and a truncated version of the protein at around ~23 kDa.

The DEAE-Sepharose elution fractions show that BPSS0212 undergoes a small amount of degradation, during this stage of purification. However significant degradation can be seen to have occurred during the ResourceQ column. When BPSS0212 was to be loaded onto the gel filtration column, having been eluted from the ResourceQ column, the protein precipitated. The ResourceQ column was run at pH 6.5, close to the pI of the BPSS0212, which is calculated to be 5.34 (ProtParam – Gasteiger et al., 2005), which may have caused the instability of the protein and thus resulted in its precipitation. As protease inhibitors and EDTA were not present during the purification, the increased degradation may have also been due to contaminating proteases active at pH 6.5.

4.1.3 BPSS0212 construct contains an extra N-terminal methionine

The final protein product was analysed by mass spectrometry and the results gave a molecular weight of 22596 Da, which corresponds to the calculated molecular weight of 22465.4 Da and an extra 131.6 Da

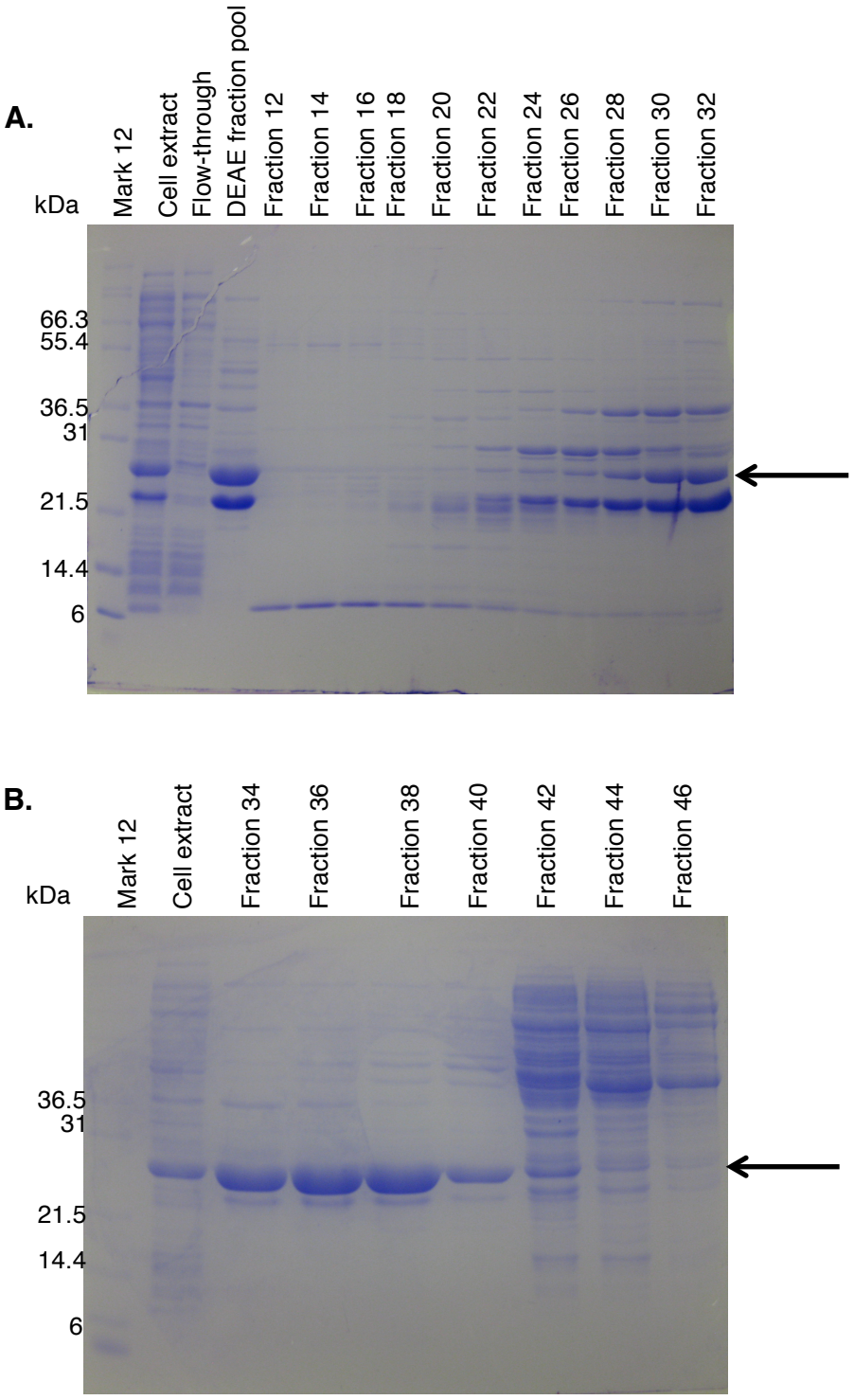


Figure 4.1. SDS-PAGE analysis of BPSS0212 DEAE anion-exchange chromatography purification. The black arrow indicates full-length BPSS0212. BPSS0212 molecular weight is 22465.4 Da.

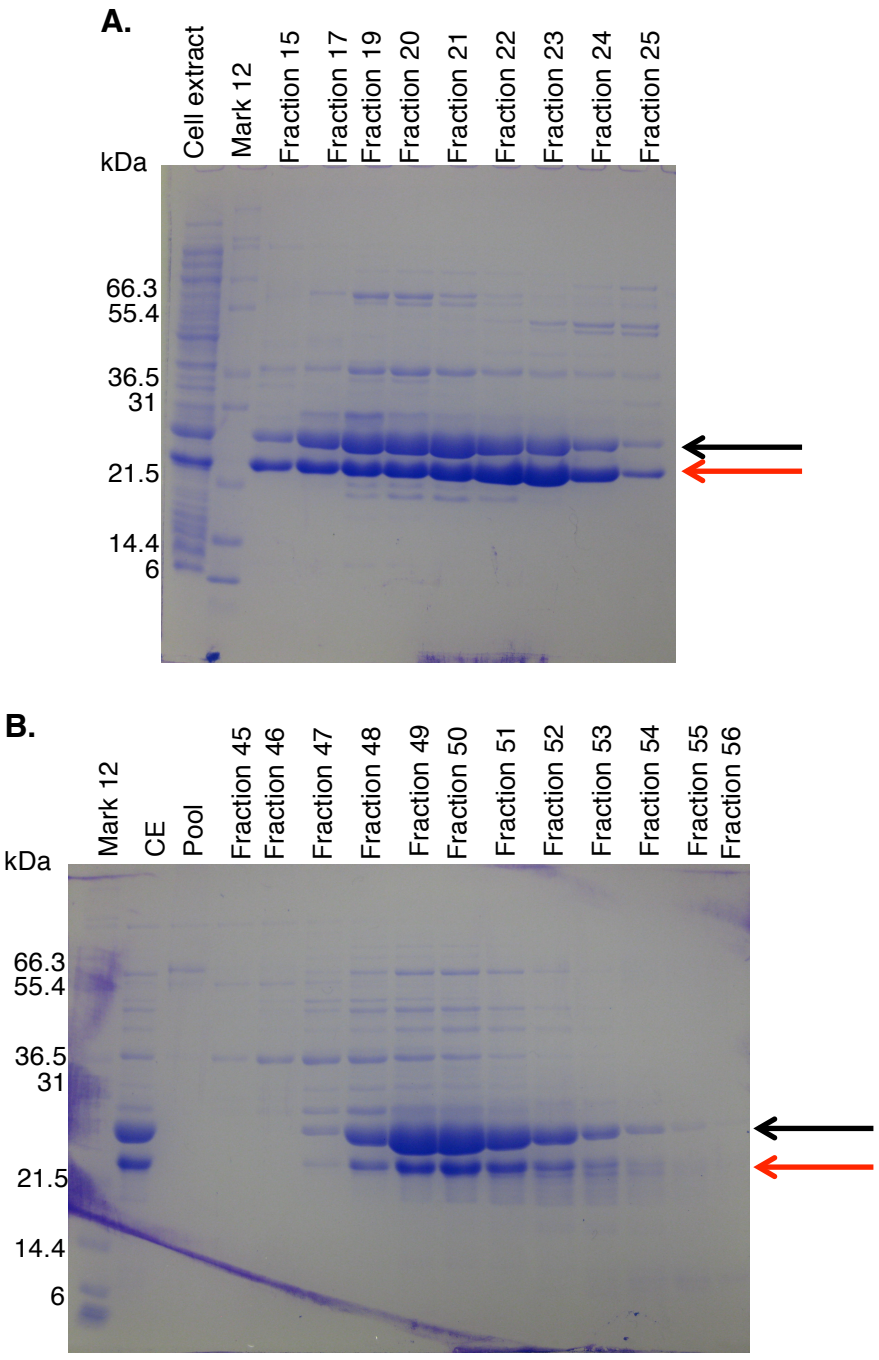


Figure 4.2. SDS-PAGE analysis of BPSS0212 ResourceQ anion-exchange chromatography and gel filtration chromatography purification. The black arrow indicates full-length BPSS0212 and the red arrow indicates BPSS0212 degradation products. **A.** BPSS0212 is degraded in the fractions eluted from the ResourceQ column. **B.** Gel filtration fractions also show degradation of BPSS0212.

(Figure 4.3). The 131.6 Da suggests that an extra methionine is present in the protein (Carr et al., 1991). Analysis of the PCR primers for BPSS0212 revealed an extra N-terminal methionine had been inserted via the forward primer (Day, 2012) during PCR. This construct is subsequently referred to as BPSS0212+M. Elsewhere the term BPSS0212 refers to the full-length wild-type protein.

4.1.4 Protocol optimisation to prevent degradation of BPSS0212+M

The purification protocol of BPSS0212+M was optimised to try to reduce the degradation and improve the stability of the protein. The ResourceQ column purification step was removed from the protocol, and the pH of the DEAE-anion exchange buffer and gel filtration buffer was increased from pH 8.0 to pH 9.0 to improve the stability of the protein, as pH 9.0 is further away from the pI of BPSS0212+M at 5.34. However, when BPSS0212+M was purified with the optimised two-step purification – DEAE anion-exchange chromatography followed by gel filtration chromatography, and with both purification buffers at pH 9.0, the protein was found to degrade when run on the gel filtration column. The analysis of SDS-PAGs from the optimised purification shows that BPSS0212+M was degrading to similar molecular weight bands as seen previously (Figure 4.4). Although BPSS0212+M still degrades the stability of the protein was increased, as on concentration of the protein before gel filtration, the protein did not precipitate. All following purifications of BPSS0212+M were carried out via DEAE anion-exchange chromatography and gel filtration chromatography, both with buffers at pH 9.0, using the protocol outlined below in section 4.1.6.

4.1.5 Selenomethionine-incorporated BPSS0212+M protein purification

Selenomethionine-incorporated BPSS0212+M protein, which will be referred to as BPSS0212+M-SeMet, was purified, as below, for the production of selenomethionine-incorporated crystals of

Figure 4.3. Mass spectrometry analysis of BPSS0212 from initial purification. The calculated molecular weight for BPSS0212 is 22465.4 Da. The molecular weight identified in the mass spectrometry analysis is 22597.0 Da, an extra 131.6 Da more than the expected molecular weight. Mass spectrometry was performed by Simon Thorpe, University of Sheffield.

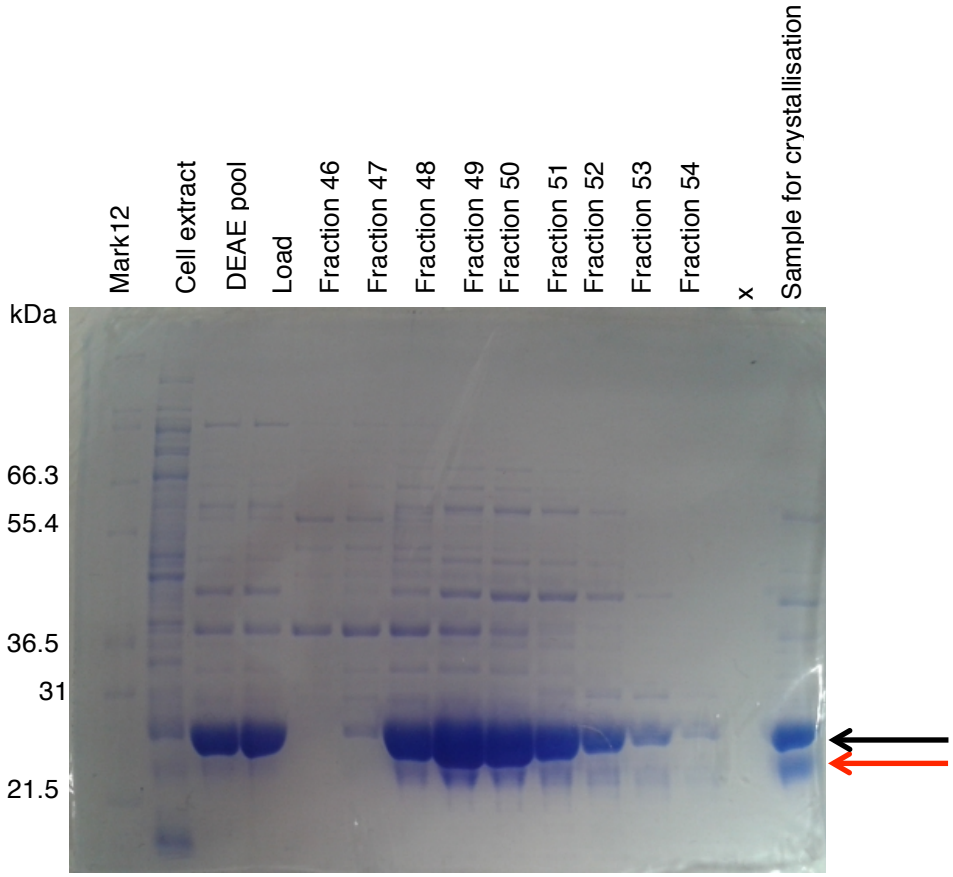


Figure 4.4. SDS-PAGE analysis of BPSS0212+M gel filtration chromatography step of the optimised purification protocol. The black arrow indicates full-length BPSS0212+M and the red arrow indicates degraded BPSS0212+M. The gel filtration elution fraction analysis by SDS-PAGE shows that degradation of BPSS0212+M still takes place with the removal of the ResourceQ purification step and the DEAE-anion exchange buffer at pH 9.0. x indicates an empty lane.

BPSS0212+M to solve the structure via experimental phasing.

E. coli BL21 DE3 cell pellet containing overexpressed BPSS0212+M-SeMet protein, was resuspended in 10 mL/g of 50 mM Tris pH 9.0. The resuspended cells were sonicated at 16 microns for 3 x 20 seconds. The cell lysate was then centrifuged at 24500 rpm/~72600 *g* in a JA25.50 rotor (Beckman Coulter) for 15 minutes. The cell extract was loaded onto a DEAE-anion exchange column pre-equilibrated with 50 mM Tris pH 9.0; the flow-through was collected. BPSS0212+M-SeMet was eluted at 4 mL/minute over a gradient of 0-100% 50 mM Tris pH 9.0, 250 mM NaCl in 300 mL, remaining at 100% for 40 mL and 8 mL fractions were collected. Fractions were analysed for the presence of protein using SDS-PAGE. Fractions containing BPSS0212+M-SeMet were pooled and concentrated using a 10 kDa molecular weight cut-off Vivaspin concentrator (Sartorius Stedim). BPSS0212+M-SeMet was concentrated to a volume ~1.5 mL and run on a gel filtration column pre-equilibrated with 50 mM Tris pH 9.0, 500 mM NaCl. The column was run at 1 mL/minute for 140 mL and 2 mL fractions were collected. Purification gels are shown in Figure 4.5.

4.1.6 Final purification

The purification of BPSS0212, a recloned construct to remove the extra N-terminal methionine, was carried out as below.

E. coli BL21 DE3 cell pellet containing overexpressed BPSS0212, was resuspended in 50 mM Tris pH 9.0. Cells were sonicated at 16 microns for 3 x 20 seconds and centrifuged at 18000 rpm/39191 *g* in a JA.20 rotor (Beckman Coulter) for 15 minutes. Cell extract was loaded onto a HiTrap DEAE FF 5 mL column (GE) pre-equilibrated with 50 mM Tris pH 9.0, at 5 mL/minute and 5 mL fractions were collected. BPSS0212 was eluted at 5 mL/minute over 0-25% 1 M NaCl in 100 mL (20 CV) and 5 mL fractions were collected. Elution fractions containing BPSS0212 were pooled and concentrated in a 10 kDa molecular weight cut-off Vivaspin (Sartorius

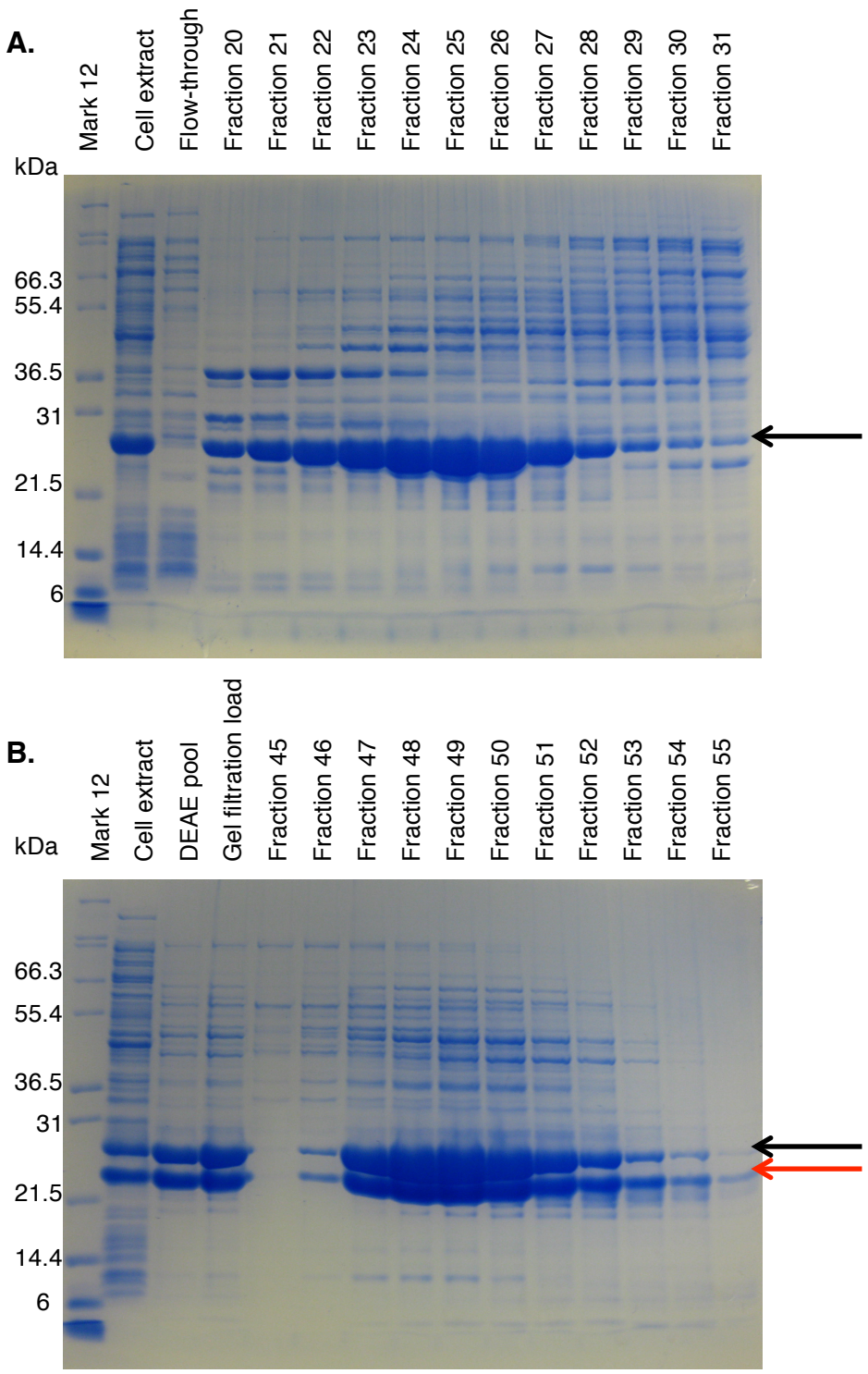


Figure 4.5. SDS-PAGE analysis of selenomethionine-incorporated BPSS0212+M DEAE anion-exchange and gel filtration chromatography purification. The black arrow indicates full-length BPSS0212+M and the red arrow indicates degraded BPSS0212+M. BPSS0212+M molecular weight is 22597 Da. **A.** DEAE anion-exchange chromatography gel - fractions 24-27 were pooled. **B.** Gel filtration chromatography gel - fractions 47-48 were pooled.

Stedim) and loaded onto a 120 mL HiLoad 16/600 Superdex200 pg gel filtration column (GE) pre-equilibrated with 50 mM Tris pH 9.0, 500 mM NaCl. The column was run at 1.5 mL/minute, with 2 mL fractions collected over 0.3 CV-1 CV.

4.2. Crystallisation

4.2.1 Crystallisation of BPSS0212+M

Despite evidence of continued degradation the initial purified preparation of BPSS0212+M protein was put into 10 mM Tris pH 8.0, 100 mM NaCl and concentrated to 6 mg/mL and 12 mg/mL. Robot sitting-drop crystallisation trials were set down using the PACT and PEGs screens for both concentrations of protein. Each small well drop contained, as standard, a 1:1 ratio of 200 nL protein to 200 nL reservoir solution. Protein was also set down in the ratio 2:2 and 5:2 (protein:reservoir), for both protein concentrations. Crystal hits were found in both PACT and PEGs screens, at both protein concentrations. The crystals produced a number of different forms (Figure 4.6). The production of protein crystals suggests that there is enough pure full-length BPSS0212+M protein present to form crystals, or the degradation of BPSS0212+M produces a fragment or fragments that are able to crystallise.

4.2.2 BPSS0212+M crystallisation optimisation

Fresh protein was produced, using the two-step purification protocol, with the ResourceQ column step removed, DEAE anion-exchange chromatography, with buffer at pH 8.0, and gel filtration chromatography, with buffer at pH 9.0. Hanging drop optimisation trials of conditions from PACT rows E and G and from PEGs row D were set up with 500 µL reservoir and 2 µL:2 µL (protein:reservoir) hanging drops. The optimisation conditions for the PACT screen rows E and G used 15-30% (w/v) PEG3350 in increments of 3% and either 0.2 M sodium bromide, 0.2 M potassium thiocyanate, 0.2 M sodium nitrate or 0.2 M sodium acetate. The conditions for the PEGs screen D7-D12 were optimised using 0.1 M

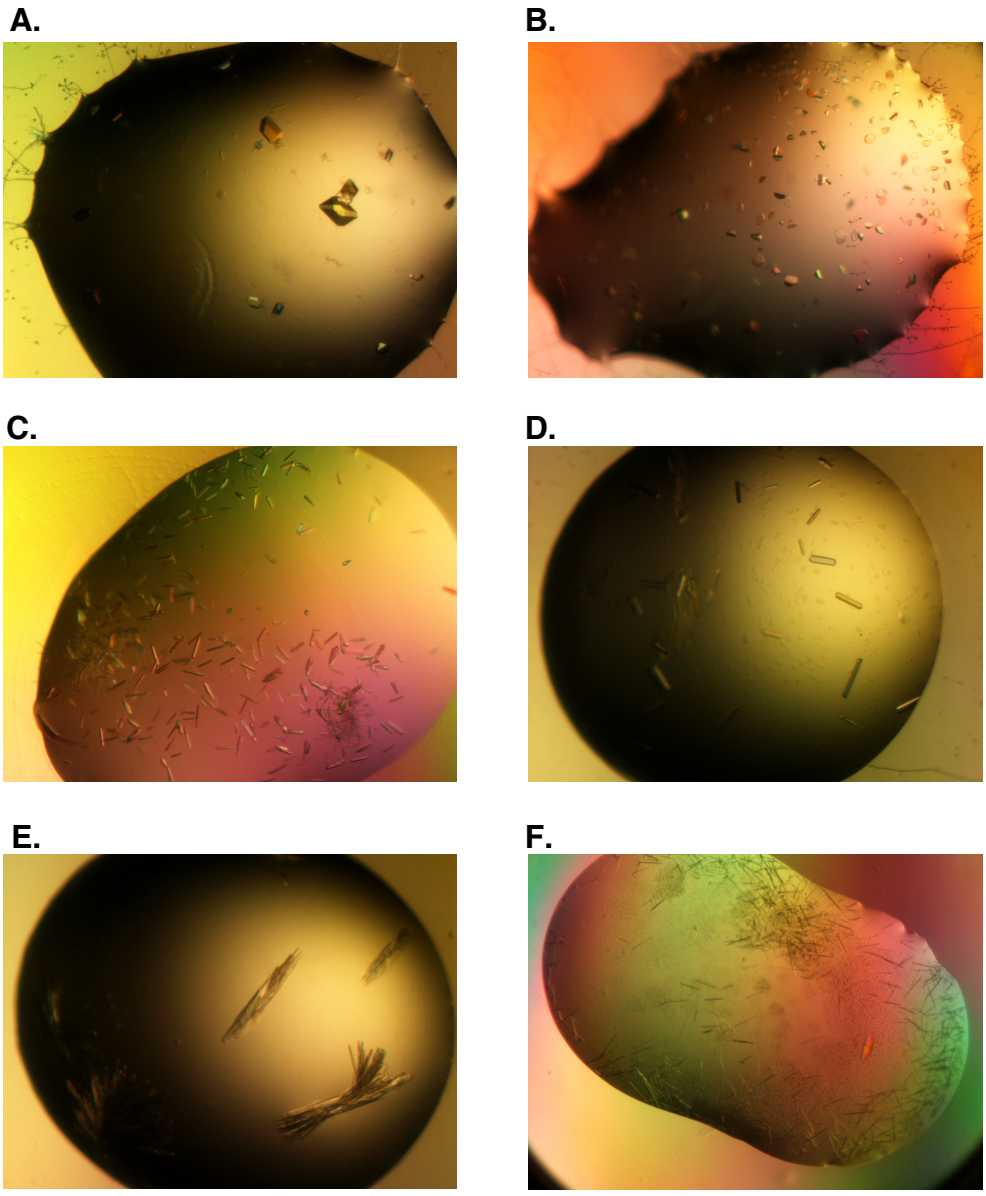


Figure 4.6. Crystal hits from sitting-drop crystallisation trials of native BPSS0212+M protein.

Protein was set down at concentrations of 6 mg/mL and 12 mg/mL. The ratio of protein:precipitant for each crystal condition is shown in brackets. **A.** PACT E4 – 0.2 M potassium thiocyanate, 20% (w/v) PEG3350 (6 mg/mL, 2:2). **B.** PACT E3 – 0.2 M sodium iodide, 20% (w/v) PEG3350 (6 mg/mL, 2:2). **C.** PACT H3 – 0.2 M sodium iodide, 0.1M Bis-Tris propane pH 8.5, 20% (w/v) PEG3350 (6 mg/mL, 2:2). **D.** PEGs F2 – 0.2 M potassium thiocyanate, 20% (w/v) PEG3350 (6 mg/mL, 2:2). **E.** PEGs E11 – 0.2 M potassium iodide, 20% (w/v) PEG3350 (6 mg/mL 5:2). **F.** PEGs E9 – 0.2 M ammonium chloride, 20% (w/v) PEG3350 (6 mg/mL 2:2).

Tris.HCl pH 8.5 and 15-30% (w/v) PEG – PEG3000, 4000, 6000, 8000, 10000 or 20000, in increments of 3%. However only needles or precipitate were produced in the drops (Figure 4.7).

4.2.3 Repeat crystallisation and further optimisation trials of BPSS0212+M

The formation of BPSS0212+M protein crystals may be affected by the difference in degradation between protein preparations. This was analysed by preparing more BPSS0212+M protein and repeating the crystallisation trials, to see if crystals could be produced in the same or similar crystallisation conditions. BPSS0212+M was purified by DEAE anion-exchange chromatography and gel filtration chromatography, with both steps run with buffers at pH 9.0. Protein was concentrated to ~10 mg/mL. Robot sitting-drop crystallisation trials were set down using a 1:1 ratio of protein:reservoir in the crystallisation screens PACT, JCSG+, PEGs, pHClear, Classic and MPD. Crystal hits were found of the same morphology and in similar crystallisation conditions as previous crystallisation trials, containing PEG3350 and salt, which shows that the preparation of BPSS0212+M protein is consistent enough to reproduce crystals (Figure 4.8).

4.2.4 Further crystallisaiton trials with BPSS0212

Further sitting-drop crystallisation trials were set down with the new Proplex and Morpheus suites, using BPSS0212 prepared using the final purification protocol in section 4.1.6. A number of conditions produced needle crystals, not large enough to be sent for data collection, and no larger crystals of the morphology seen previously were produced.

4.2.5 Selenomethionine-incorporated BPSS0212+M protein crystallisation

Selenomethionine-incorporated BPSS0212+M protein was produced and the protein was purified using the final optimised purification two-step

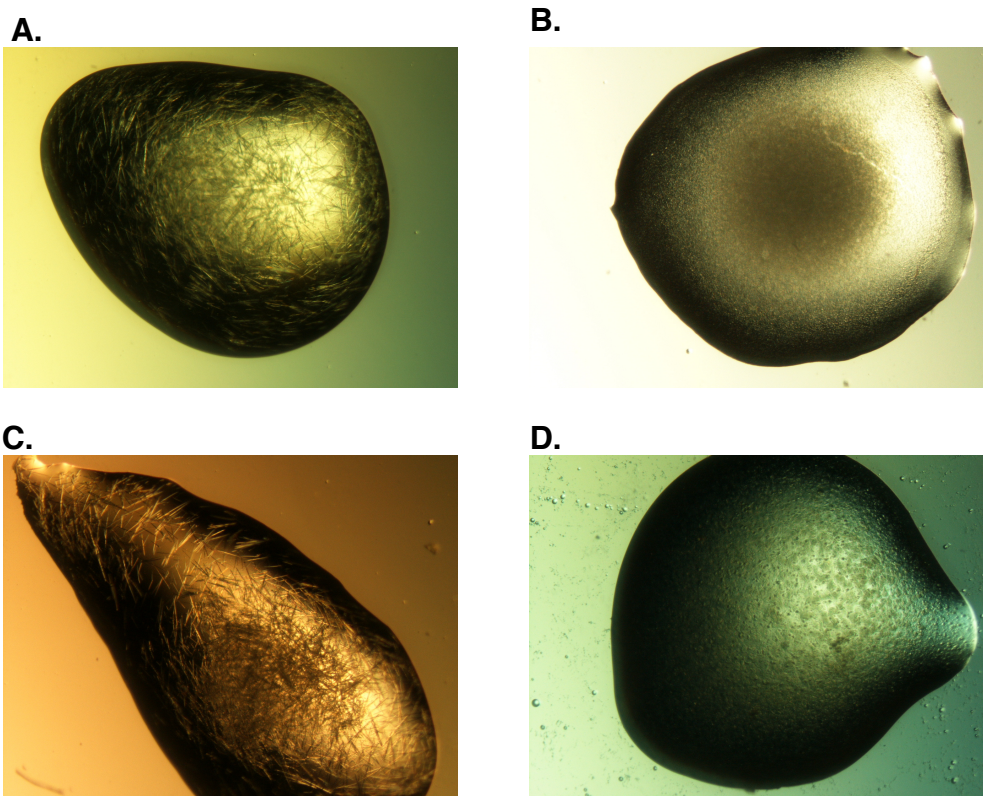


Figure 4.7. Crystal hits from hanging-drop optimisation trials of native BPSS0212+M protein.
A-C. PACT optimisation, using 15-30% (w/v) PEG3350 and a variety of salts at 0.2 M concentration. **D.** PEGs D7-D12 optimisation, using 0.1 M Tris.HCl pH 8.5 and 15-30% (w/v) PEG – PEG3000, 4000, 6000, 8000, 10000 or 20000.

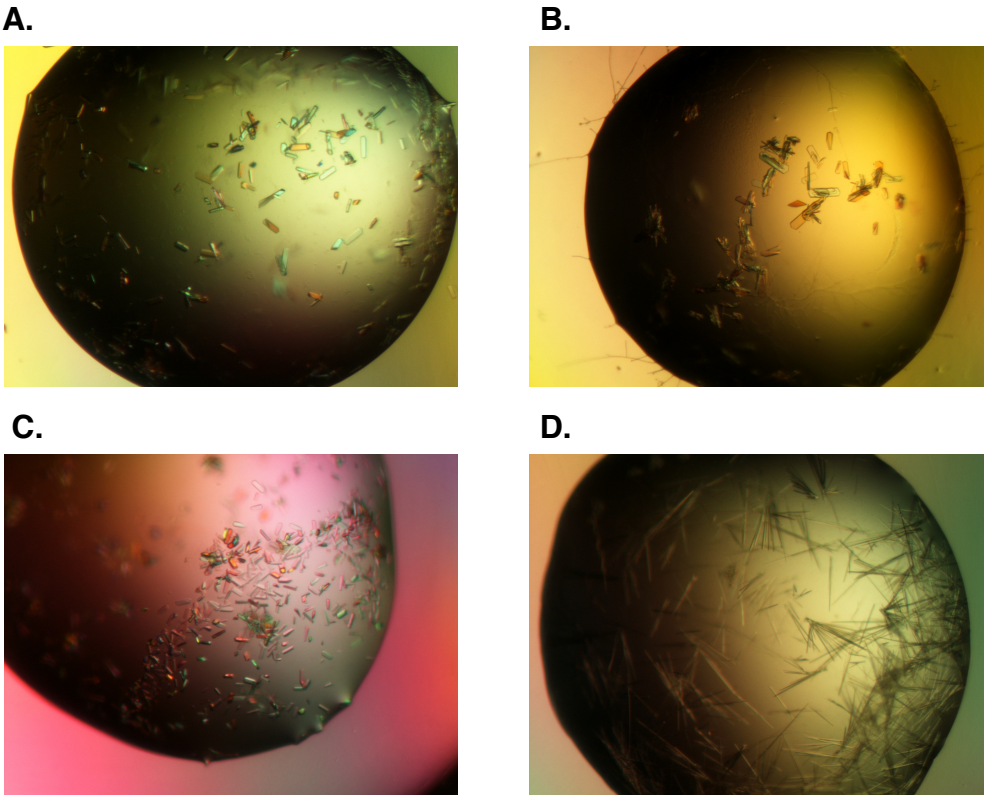


Figure 4.8. Crystal hits from sitting-drop crystallisation trials of native BPSS0212+M protein.
Protein was set down at ~10 mg/mL. **A.** PEGs F10 – 0.2 M potassium formate, 20% (w/v) PEG3350. **B.** PACT E6 – 0.2 M sodium formate, 20% (w/v) PEG3350. **C.** PEGs H6 – 0.2 M di-potassium phosphate, 20% (w/v) PEG3350. **D.** PACT G5 – 0.2 M sodium nitrate, 0.1 M Bis-Tris propane pH 7.5, 20% (w/v) PEG3350.

protocol of DEAE anion-exchange chromatography, followed by gel filtration chromatography, both at pH 9.0, outlined in section 4.1.5. The protein was concentrated to 6 mg/mL and 11 mg/mL, and set down in robot sitting-drop crystallisation trials. Five crystallisation suites, PACT, JCSG+, PEGs, Classic and Midas, were used. PACT and PEGs had drops set down in three different ratios, 1:1, 2:2 and 5:2 (protein:reservoir), for both protein concentrations. The JCSG+ and Classic suites had both protein concentrations set down in a 1:1 ratio, whilst the Midas screen only had 6 mg/mL set down in a 1:1 ratio. All screens were stored at 17°C. The PACT screen was also set down with protein at 11 mg/mL in a 1:1 ratio and stored at room temperature, ~19°C. Crystal hits were found in a number of conditions, at both protein concentrations (Figure 4.9).

Optimisation trays were set down based on conditions that successfully produced crystals. Hanging-drop optimisation trials were set down for PACT E5 (0.2 M sodium nitrate, 20% (w/v) PEG3350) and PEGs E1 (0.2 M sodium fluoride, 20% (w/v) PEG3350). For condition PACT E5 optimisation, PEG3350 percentage was varied from 18-23% in increments of 1%, and 0.2 M sodium nitrate. Protein:reservoir ratios of 1:1, 2:2, 5:2, 3:2/3:3 were used. For the PEGs E1 optimisation the sodium fluoride concentration was varied from 0.05 M-0.6 M in increments of 0.05 M, with 20% (w/v) PEG3350. As these hanging-drop crystallisation trials did not optimise the crystal forms grown previously, it was hypothesised that the crystals grew better in a sitting-drop experiment. Large scale sitting-drop optimisations were set up for conditions PEGs E1 and PEGs E11 using microbridges with 18-24% (w/v) PEG3350 and the appropriate salt. Trays were set down with 500 µL reservoir and 2 µL:2 µL drops for 6 mg/mL protein with condition PEGs E1, 1 µL:1 µL drops for 11 mg/mL protein with conditions PEGs E1 and E11. Unfortunately the crystals did not optimise using the microbridges.

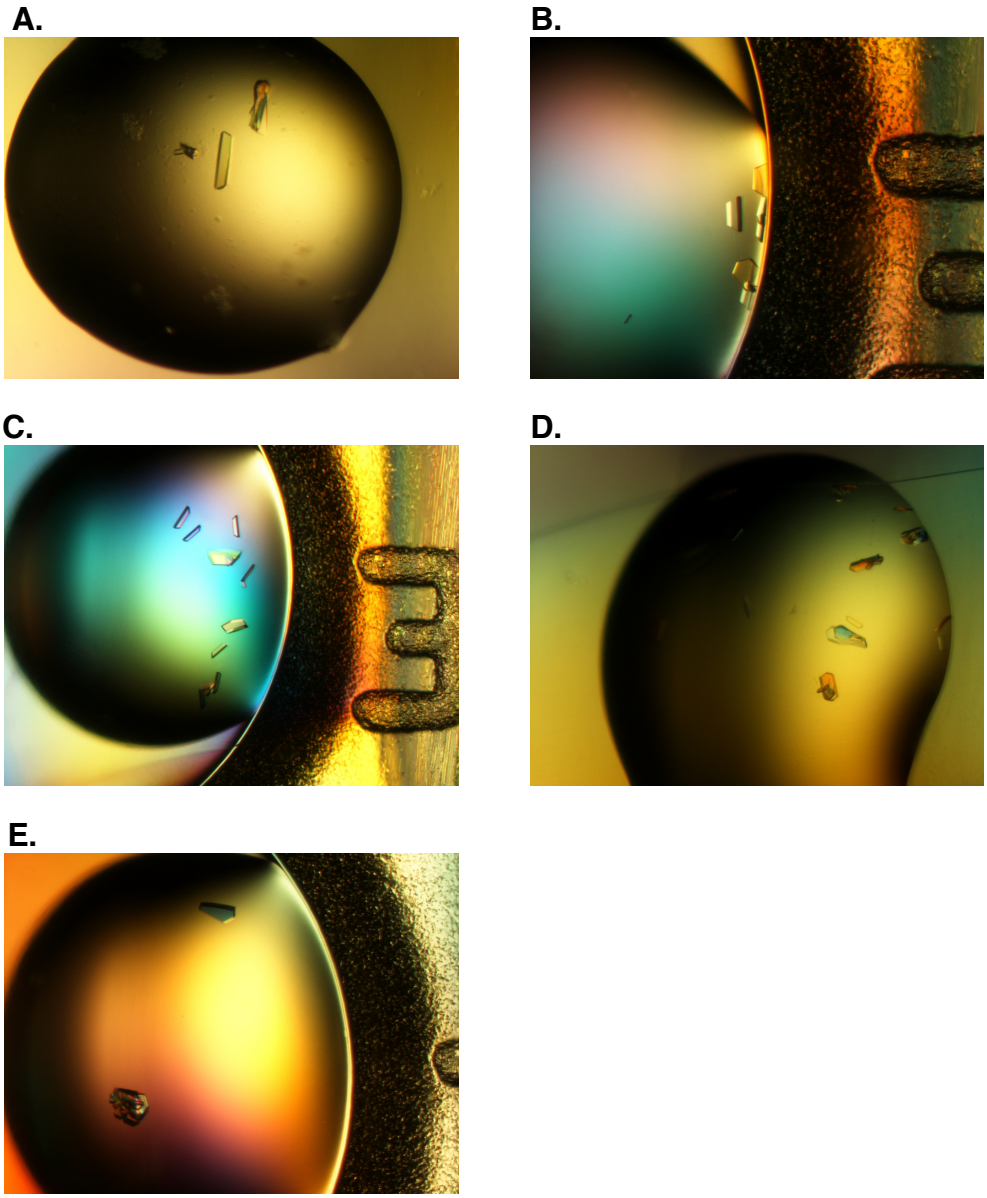


Figure 4.9. Crystal hits from sitting-drop crystallisation trials of selenomethionine-incorporated BPSS0212+M.
Protein was set down at concentrations of 6 mg/mL and 11 mg/mL. The ratio of protein:precipitant for each crystal condition is shown in brackets. **A.** PEGs E11 – 0.2 M potassium iodide, 20% (w/v) PEG3350 (6 mg/mL, 1:1). **B.** PEGs E1 – 0.2 M sodium fluoride, 20% (w/v) PEG3350 (6 mg/mL, 2:2). **C.** PEGs E1 – 0.2 M sodium fluoride, 20% (w/v) PEG3350 (11 mg/mL, 1:1). **D.** PACT A3 – 0.1 M SPG buffer pH 6.0, 25% (w/v) PEG1500 (11 mg/mL, 1:1, Room Temp ~19°C). **E.** PEGs E1 - 0.2 M sodium fluoride, 20% (w/v) PEG3350 (11 mg/mL, 2:2).

4.3 BPSS0212+M protein degradation and crystallisation

4.3.1 Crystallisation of degraded BPSS0212+M protein

The degradation of BPSS0212+M protein following purification was analysed by SDS-PAGE. A selection of elution fractions from the purification of BPSS0212+M protein used in the crystallisation trials outlined in section 4.2.3, as well as the final purified protein sample were stored at ~4°C for ~21 days and then run on an SDS-PAG (Figure 4.10). The results show that the protein appears to continue to degrade to a defined band on SDS-PAG, representing a possible stable protein degradation product. To assess whether this degradation product of BPSS0212+M could produce crystals, a sitting-drop robot crystallisation trial was conducted ~22 days after the initial purification, with a 1:1 ratio of protein:reservoir using the PEGs crystallisation condition suite. Crystals were produced of a similar morphology to those seen previously (Figure 4.11). This suggests that BPSS0212+M purified protein produces crystals when it has degraded to a homogeneous, stable fragment.

4.3.2 N-terminal sequencing of BPSS0212+M

In order to identify which part of the protein was being removed during degradation, crystals of BPSS0212+M were N-terminal sequenced (Figure. 4.12). A number of crystals were washed in MiliQ, and then dissolved in MiliQ and run on an SDS-PAG, which was used in a western blot experiment. The required protein bands were cut out of the resulting blot and sequenced using the Edman degradation sequencing method (western blot and N-terminal sequencing was carried out by Dr. Arthur Moir, University of Sheffield). The sequencing results have the first residue numbered in the protein sequence specified. The upper protein band from the sample containing two hanging drops (red arrow in Figure 4.13) was sequenced, the results show M1-MSEDL as the major sequence, and minor sequences of S3-EDLR and D5-LRVG. The lower band from the same sample (yellow arrow in Figure 4.13) has a number of sequences including M1-MSEDL, V8-GLFP, S3-EDLR and D5-LRVG.

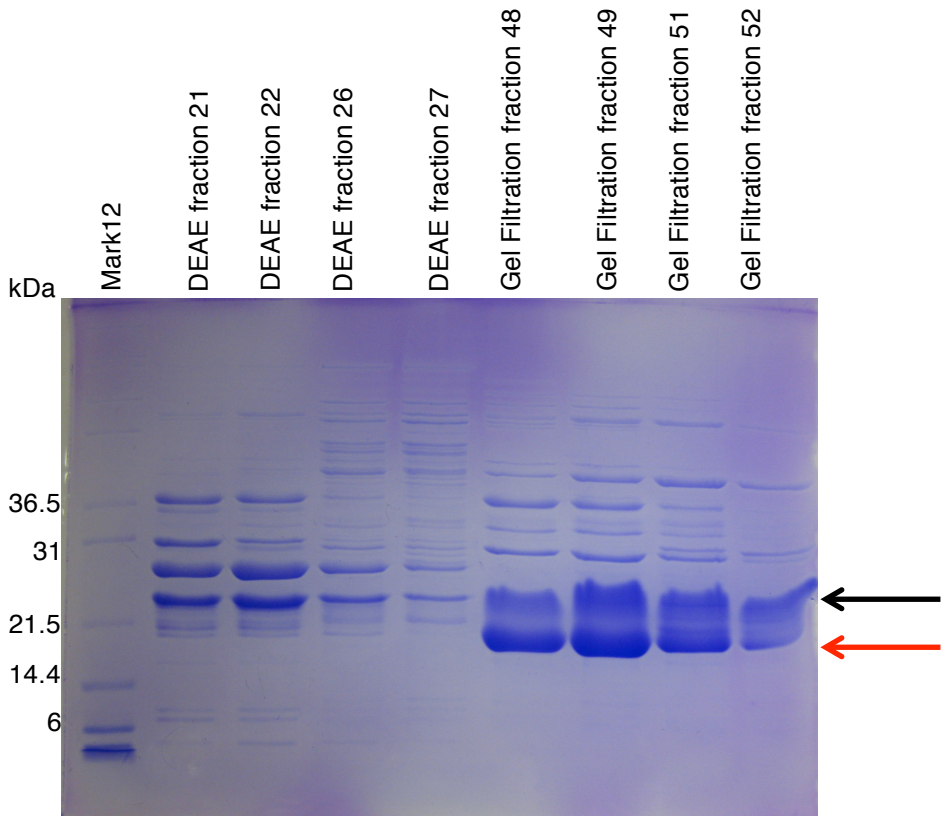


Figure 4.10. SDS-PAGE analysis of BPSS0212+M purification samples 21 days post-purification. BPSS0212+M full-length protein (black arrow) in the gel filtration fractions can be seen to degrade to a specific molecular weight (red arrow).

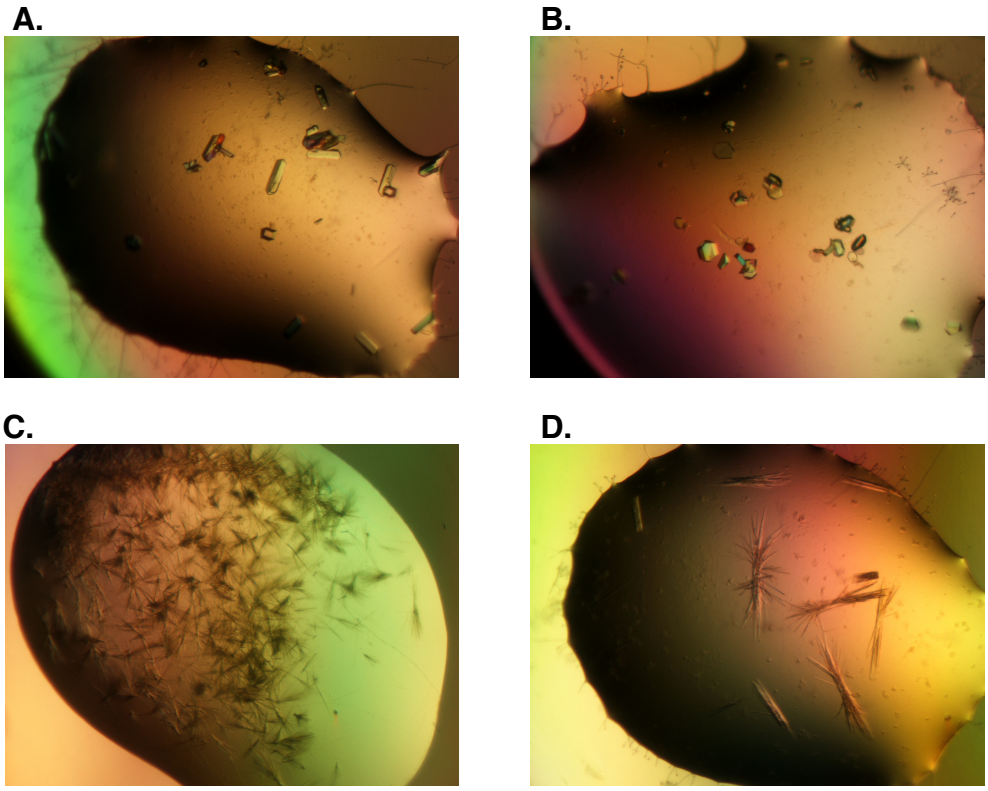


Figure 4.11. BPSS0212+M protein crystals, from protein crystallised 22 days post-purification.

Protein was set down at ~7 mg/mL. **A.** PEGs G10 – 0.2 M potassium sulphate, 20% (w/v) PEG3350. **B.** PEGs H6 – 0.2 M di-potassium phosphate, 20% (w/v) PEG3350. **C.** PEGs D9 – 0.1 M Tris.HCl pH 8.5, 25% (w/v) PEG6000. **D.** PEGs E8 – 0.2 M potassium chloride, 20% (w/v) PEG3350.

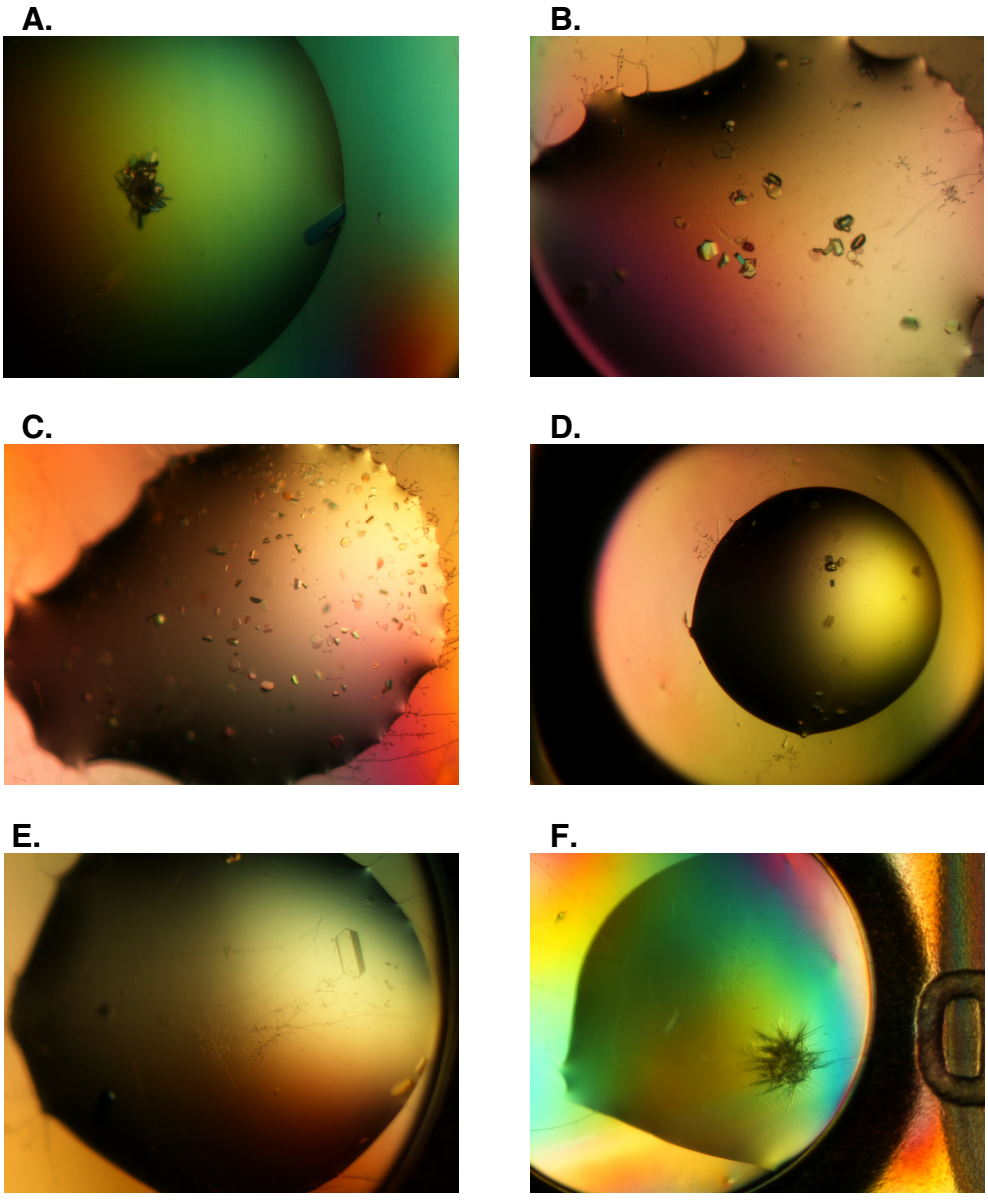
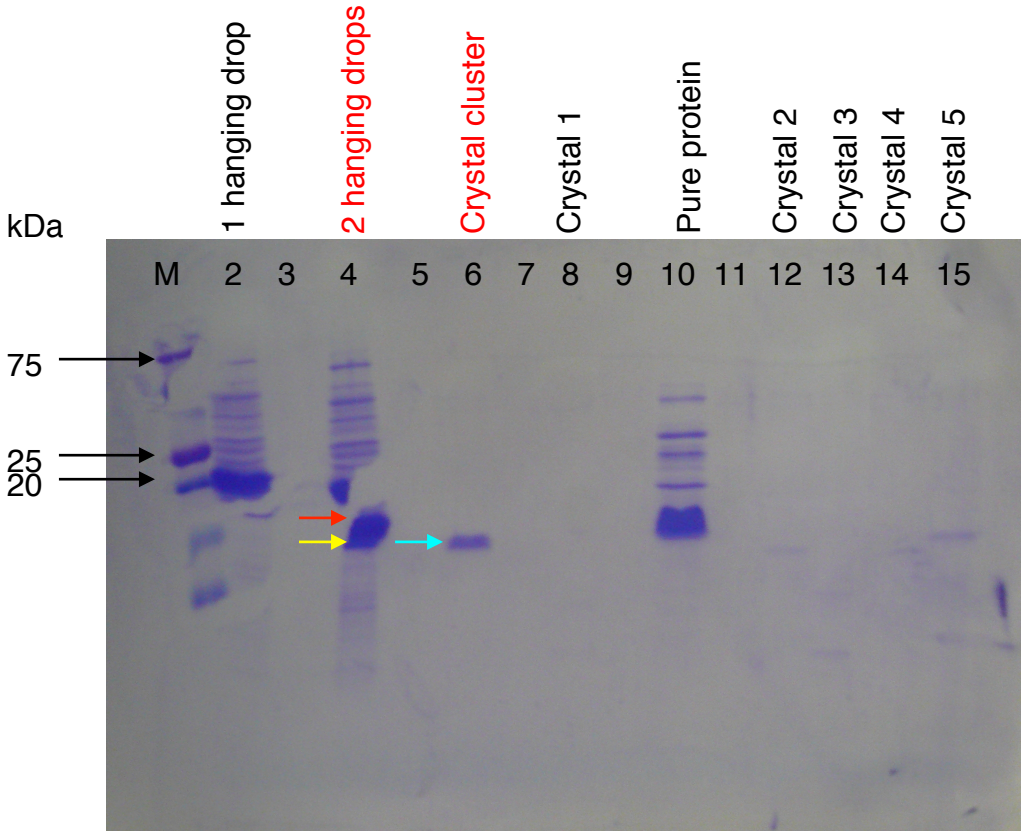


Figure 4.12. BPSS0212+M crystals chosen to run on an SDS-PAG for N-terminal sequencing.

A. PACT E5 – 0.2 M sodium nitrate, 20% (w/v) PEG3350. **B.** PEGs H6 – 0.2 M di-potassium phosphate, 20% (w/v) PEG3350. **C.** PACT E3 – 0.2 M sodium iodide, 20% (w/v) PEG3350. **D.** PACT G4 – 0.2 M potassium thiocyanate, 0.1 M Bis-Tris propane pH 7.5, 20% (w/v) PEG3350. **E.** JCSG+ B2 – 0.2 M sodium thiocyanate, 20% (w/v) PEG3350. **F.** PEGs D1 – 0.1 M sodium HEPES pH 7.5, 25% (w/v) PEG3000.



MMSEDLRVGLFPVRYLVGTGLPGAPQLVLDLMVDTVDHSVVGRAAVSQAVSPPL
NFHADVWGSYVFRLGPPPRRDGSGAIVQISLQGNQGGPQSNSMITFYGELLLKG
DGKTGVASYRYYSNGSWHEVENVPVKADPELVPIEPGPVIGQSSMSAIGSAAMY
GVAIQSAAASGDLAHMRTLSAYARQQLESRDEIAAALSELKAEIAKLESRQ

Figure 4.13. SDS-PAG blotted onto PVDF membrane - used for N-terminal sequencing. M – prestained marker – 5 µL, Lanes 3, 5, 7, 9 and 11 are empty. Lane 2 and 4 contain 1 and 2 hanging drops, respectively. Lane 6 contains the cluster of crystals seen in Figure 4.12A. Lane 8 – crystals in Figure 4.12C, lane 12 – crystals in Figure 4.12B, lane 13 – crystals in Figure 4.12D, lane 14 – crystals in Figure 4.12E and lane 15 – crystals in Figure 4.12F. Lane 10 contains a sample of BPSS0212+M protein purified ~45 days previously. The lanes highlighted in red and the coloured arrows indicate the protein bands that were N-terminal sequenced. The amino acid sequence of BPSS0212+M is also shown. PVDF - Polyvinylidene fluoride. (Western blot carried out by Dr. Arthur Moir) Sequence from Burkholderia Genome Database (Winsor et al., 2008).

This suggests the upper band contains majority full-length protein and the lower band contains full-length protein but also cleavage of up to seven N-terminal amino acids. A cluster of crystals from condition PACT E5 (Figure 4.12A) produced a strong single band on the gel, which was removed and sequenced. These particular crystals were of BPSS0212+M and the results of the Edman degradation indicated the sequence starts at G9-LFPV, however the results were quite weak. The results for the N-terminal sequencing suggests clipping of up to eight residues, ~900 Da, which does not account for the difference in molecular weight seen on the SDS-PAG, therefore this suggests some clipping is occurring at the C-terminus of the protein as well.

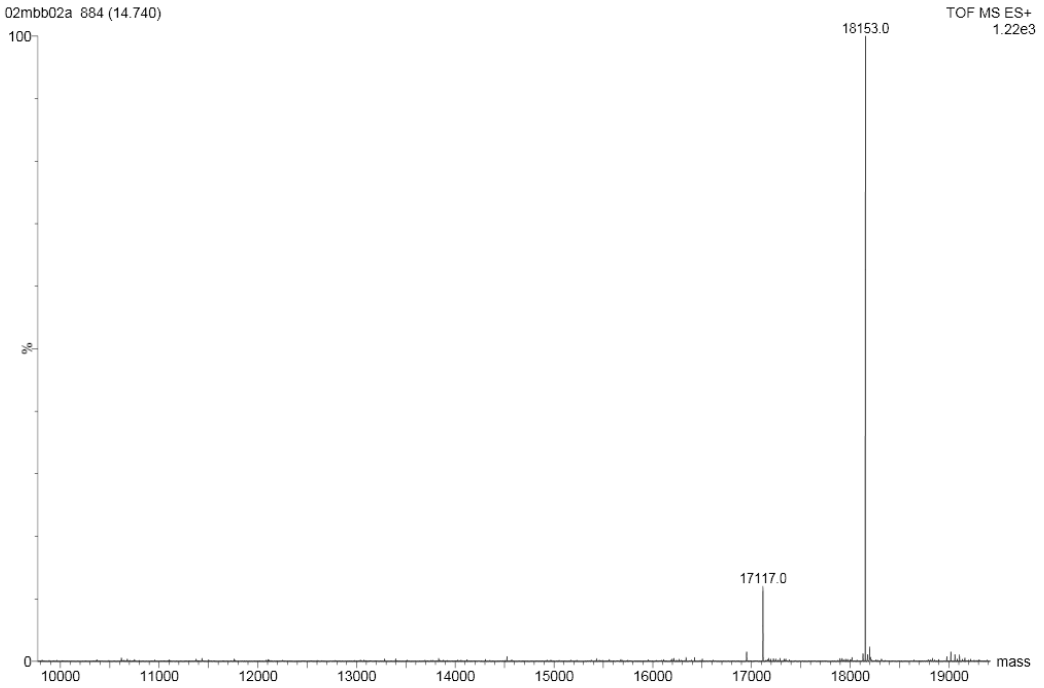
4.3.3 Mass spectrometry of degraded BPSS0212+M protein

A sample of the purified BPSS0212+M protein stored at ~4°C for ~21 days, as described in section 4.3.1, was sent for mass spectrometry analysis. The results produced molecular weights of 17117 Da, 18153 Da and higher, that were too large to be BPSS0212+M (22597 Da) (Figure 4.14). Analysis of these masses to try to identify the corresponding amino acid sequence found possible matches to masses of ~18153 Da, but not to 17117 Da, using the FindPept program (Gattiker et al., 2002; Gasteiger et al., 2005). The results produced two matches for G20-R192 and G9-S182. Compared to the N-terminal sequencing results this suggests that the G9-S182 is most likely. (Mass Spectrometry conducted by Simon Thorpe, University of Sheffield.)

4.4. Structure determination of BPSS0212+M

4.4.1 Native data collection

Native protein crystals grew in condition PACT E4 from the initial robot trays (section 4.2.1), using protein:reservoir ratio 2:2 and protein concentration of 6 mg/mL. Crystals were cryoprotected with 30% ethylene glycol, cooled in liquid nitrogen and sent to beamline I03 at the Diamond Light Source Synchrotron, Oxford. Data were collected at 0.97630 Å



MMSEDLRVGLFPVRYLVGTGLPGAPQLVLDLMVDTVDHSVVGRAAVSQAVSPPL
NFHADVWGSYVFRLGPPPRRDGSGAIVQISLQGNQGGPQSNSMITFYGELLLKG
DGKTGVASYRYYSNGSWHEVENVPVKADPELVPIEPGPVIGQSSMSAIGSAAMY
GVAIQSAAASGDLAHMRTLSAYARQQLESRDEIAAALSELKAEIAKLESRQ

Figure 4.14. Mass spectrometry results for BPSS0212+M protein stored for ~21 days. The mass spectrometry results for a sample of BPSS0212+M protein stored at ~4°C for ~21 days. The region predicted, by FindPept (Gattiker et al., 2002; Gasteiger et al., 2005) to correspond to the molecular weight shown of 18153 Da is highlighted in red on the amino acid sequence of BPSS0212+M. Mass spectrometry conducted by by Simon Thorpe, University of Sheffield. Sequence from Burkholderia Genome Database (Winsor et al., 2008).

wavelength, with 0.2° phi oscillation over 180°, with 0.2 second exposure and 50% transmission (Figure 4.15). Data were processed using the xia2 pipeline (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Sauter et al., 2004; Winter, 2010; Zhang et al., 2006). The data were processed into spacegroup C2, with cell dimensions of a=85.6 Å, b=40.4 Å, c=41.4 Å, α=γ=90°, β=118.8°, to a resolution of 1.48 Å. The full processing statistics are shown in Table 4.1.

4.4.2 Initial phasing attempt using BPSS0211

As BPSS0211 shares 50% identity over 24% coverage of BPSS0212 (Altschul et al., 1990), an initial phasing attempt was made to solve the BPSS0212+M structure, using molecular replacement, with the previously solved model of BPSS0211 (Day, 2012). However this proved to be unsuccessful, therefore selenomethionine-incorporated crystals of BPSS0212+M were produced, the purification of which is described in section 4.1.5.

4.4.3 Selenomethionine data collection

Selenomethionine-incorporated BPSS0212+M protein crystals were cryoprotected using 28-30% ethylene glycol and cooled in liquid nitrogen, before being sent to beamline I02 at the Diamond Light Source Synchrotron, Oxford. A fluorescence scan was performed at the selenium K edge, to confirm the incorporation of selenium into the protein and to identify the wavelengths at which to collect the data. The selenium absorbance spectrum was output in the programme *CHOOCH* and the wavelengths desired identified (Evans and Pettifer, 2001). The peak wavelength was identified as 0.9792 Å, with f'' 7.18 e and f' -8.3 e and the inflection wavelength as 0.97996 Å, with f'' 4.04 e and f' -11.2 e. The high-energy remote wavelength was set at 0.96863 Å. Peak data were collected with 0.15° phi oscillation over 360°, with 0.05 second exposure and transmission of 20%. This data collection strategy was repeated at the inflection and high-energy remote wavelengths. The three

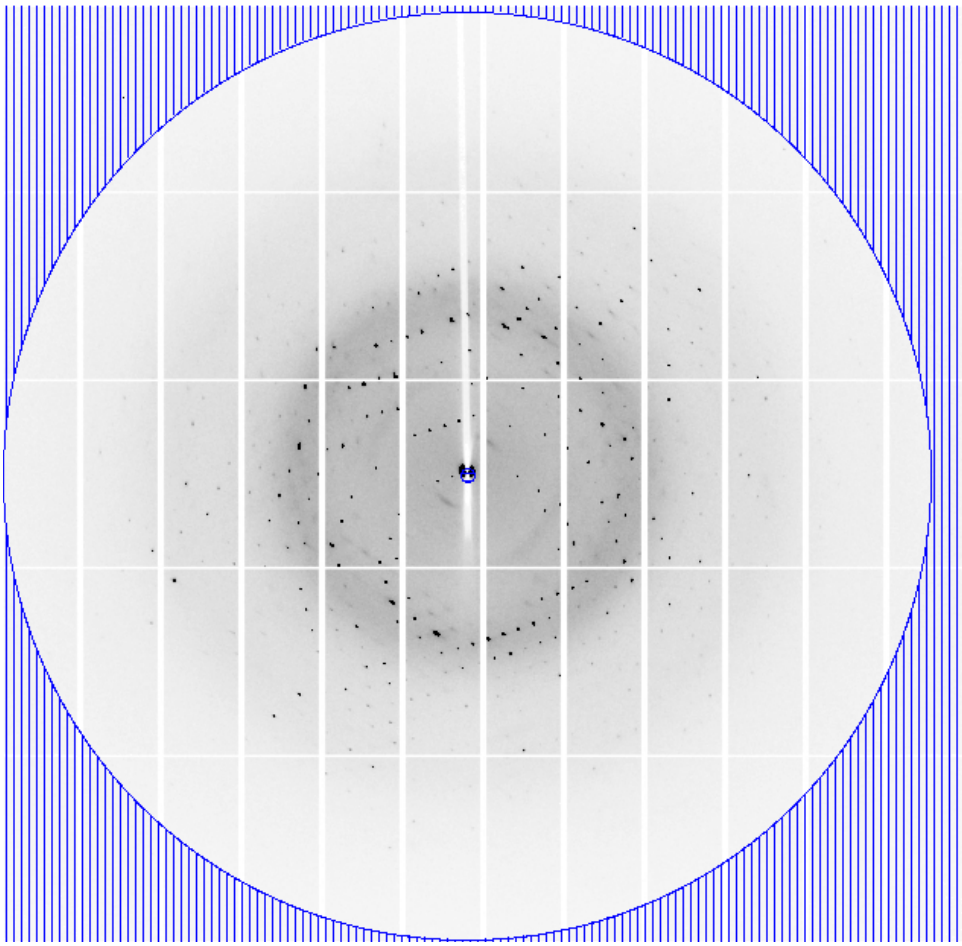


Figure 4.15. Diffraction test image of a crystal of BPSS0212+M from condition PACT E4. Diffraction test image taken, at the Diamond Light Source Synchrotron, Oxford, prior to data collection, with the resolution ring (blue) set to 1.5 Å. Image generated in iMosflm (Battye et al., 2011).

Data collection - BPSS0212+M	Native data	Selenomethionine data		
		Peak	Inflection	High Energy Remote
Wavelength (Å)	0.97630	0.97972	0.97996	0.96863
Resolution range (Å)	36.18-1.48 (1.52-1.48)	23.10-1.62 (1.66-1.62)	23.12-1.59 (1.64-1.59)	14.59-1.65 (1.70-1.65)
Space group	C 1 2 1	C 1 2 1	C 1 2 1	C 1 2 1
Unit cell parameters				
a (Å)	85.58	85.22	85.29	85.44
b (Å)	40.84	40.80	40.83	40.89
c (Å)	41.27	41.50	41.53	41.64
α (°)	90	90	90	90
β (°)	118.76	118.33	118.32	118.29
γ (°)	90	90	90	90
Total reflections	67718 (4792)	105073 (7389)	110539 (8044)	99388 (7139)
Unique reflections	20755 (1505)	15983 (1152)	16778 (1207)	15212 (1127)
Multiplicity	3.3 (3.2)	6.6 (6.4)	6.6 (6.7)	6.5 (6.3)
Completeness (%)	99.1 (99.5)	99.4 (98.3)	99.1 (96.5)	99.3 (99.3)
Mean I/σ	15.3 (2.3)	16.9 (3.4)	15.1 (2.7)	15.4 (2.8)
Wilson B factor ($Å^2$)	22	20	21	23
R_{merge}	0.029 (0.406)	0.093 (1.575)	0.108 (2.288)	0.077 (1.190)
R_{pim} (I)	0.022 (0.312)	0.050 (0.695)	0.051 (0.979)	0.039 (0.530)
Anomalous completeness	-	98.8 (98.3)	98.5 (96.5)	98.7 (99.0)
Anomalous multiplicity	-	3.4 (3.3)	3.4 (3.4)	3.3 (3.2)
Anomalous correlation	-	0.652 (0.086)	0.363 (0.008)	0.395 (-0.008)
Anomalous slope	-	1.759	1.266	1.247

Table 4.1. Data collection processing statistics for BPSS0212+M
xia2 3d data processing statistics for BPSS0212+M native and selenomethionine MAD collected data at peak, inflection and high energy remote (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Sauter et al., 2004; Winter, 2010; Zhang et al., 2006). The high-resolution shell information is shown in brackets.

wavelengths of data were processed using the xia2 pipeline (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Sauter et al., 2004; Winter, 2010; Zhang et al., 2006), to an average resolution of ~1.62 Å, in spacegroup C2 with average cell dimensions of a=85.3 Å, b=40.8 Å, c=41.5 Å, α=γ=90°, β=118.3°, processing statistics are shown in Table 4.1.

4.4.4 Experimental phasing of BPSS0212+M data

The structure of BPSS0212+M was solved using MAD experimental phasing, using the *SHELX* program and the *HKL2MAP* interface (Sheldrick, 2008; Pape and Shneider, 2004). The peak, inflection and high-energy remote data sets were analysed using *SHELXC* (Sheldrick, 2008) for the presence of anomalous signal. *SHELXD* (Sheldrick, 2008) was then used to identify the selenium atom sites. *SHELXE* (Sheldrick, 2008) was run in the original and inverted hand and a map and initial model was produced for each. The heavy atom sites and both maps output by *SHELXE* (Sheldrick, 2008) were viewed in *Coot* (Emsley et al., 2010).

4.4.5 Building of BPSS0212+M structure

The chosen map and model output from *SHELXE* (Sheldrick, 2008) were from the inverted hand (Figure 4.16). The structure of BPSS0212+M was built using *Coot* (Emsley et al., 2010) and refined to the native data to 1.48 Å resolution, using Refmac5 (Murshudov et al., 1997, Murshudov et al., 2011). The final R_{factor} is 0.1691 and R_{free} is 0.2159. The final model consists of 128 protein residues from a single molecule and 35 water molecules. The structure of BPSS0212+M was validated using MolProbity (Chen et al., 2010) and statistics are shown in Table 4.2. There is one bad rotamer present at Leu27, the electron density and surrounding core hydrophobic packing confirms the conformation of this residue.

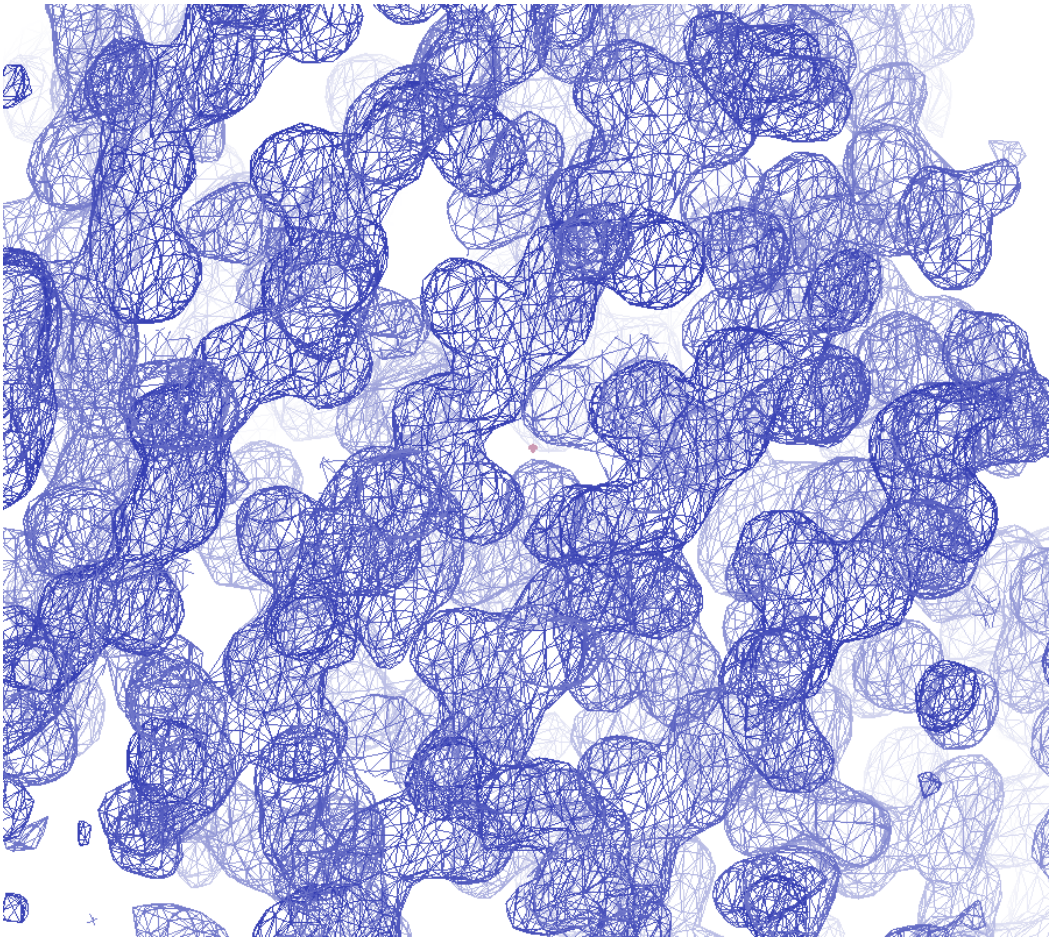


Figure 4.16. Final inverted hand map produced by *SHELXE* for selenomethionine MAD experimental phasing of BPSS0212+M data. The electron density output from *SHELX* (.phs) is coloured blue. Electron density for β-strands can be clearly seen. Image generated in *Coot* (Emsley et al., 2010), map produced by *SHELXE* (Sheldrick, 2008).

Refinement	BPSS0212+M-FormA
R_{factor}	0.1691
R_{free}	0.2159
No. of non-H atoms	
Protein	907
Ligands/Metal ions	0
Water	35
Protein residues	128
RMSD (bonds) (Å)	0.0127
RMSD (angles) (°)	1.6690
Ramachandran favoured (%)	98.28
Ramachandran outliers (%)	0.00
Favoured rotamers (%)	96.94
Poor rotamers (%)	1.02 (1 residue, Leu 27)
MolProbity score	0.92 (100^{th} percentile (N=4443, 1.48Å ± 0.25Å)
Average B factors	
Main chain ($Å^2$)	32
Side chains ($Å^2$)	44
Water ($Å^2$)	39

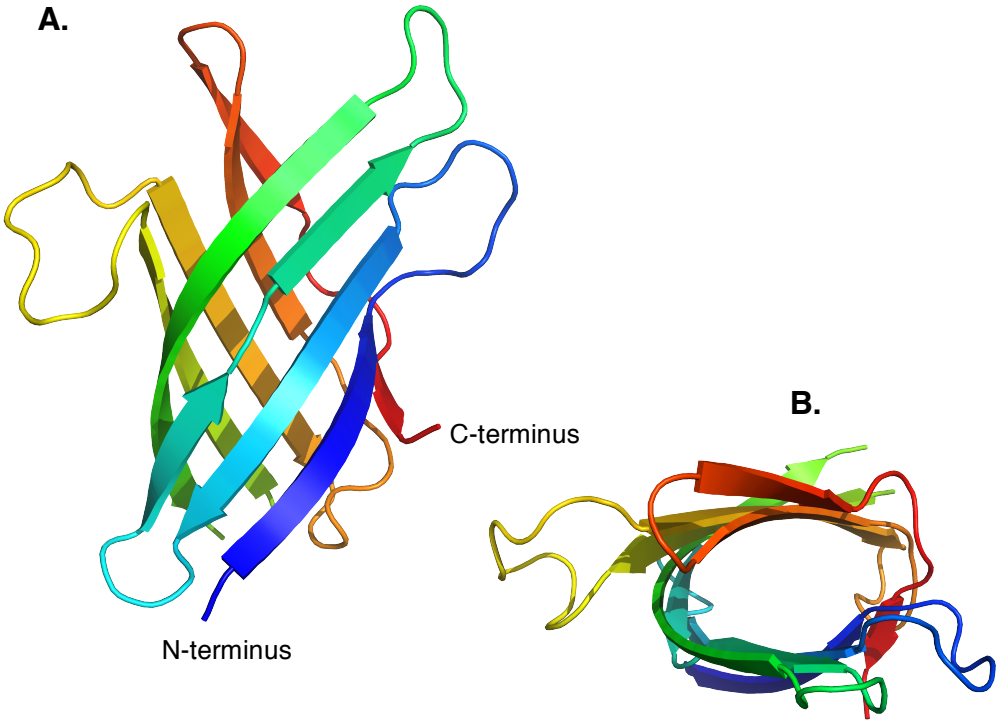
Table 4.2. Refinement and validation statistics for BPSS0212+M-FormA. Statistics generated in Refmac5 (Murshudov et al., 1997, Murshudov et al., 2011), MolProbity (Chen et al., 2010) and Baverage (Winn et al., 2011).

4.5 BPSS0212+M structural analysis

4.5.1. Analysis of the structure of BPSS0212+M-FormA

The model built into the electron density consists of residues 9-136 of BPSS0212+M, corresponding to the DUF1842 domain and part of the interdomain linker between the DUF1842 domain and the DUF1843 domain. The residues not present in the protein model due to the lack of representative electron density are the first seven residues, residues 68-76 present in a loop and the last 77 residues. The absence of a significant portion of the protein is not surprising considering the degradation seen during purification. The last 77 residues of the C-terminal region of BPSS0212+M represent the DUF1843 domain and the other part of the interdomain linker. The previously solved structure of the DUF1843 domain from BPSS0211 shows the DUF1843 domain is helical (Day, 2012). The absence of the last 77 residues conflicts with the number of residues predicted to be present by mass spectrometry, which predicted a molecular weight corresponding to residues 9-182. On analysis of the crystal lattice, the C-terminal end of the BPSS0212+M-FormA structure is adjacent to a cavity that is not large enough to accommodate the missing residues. This suggests further cleavage of the protein had taken place.

The structure of BPSS0212+M-FormA consists of β-strands, connected by loops, which form a squashed 8-stranded antiparallel β-barrel - β8 and β9 form a split strand. The β-barrel is formed from two β-sheets. The first sheet consists of β1, β2, β3 (a strand in two parts containing a β-bulge), β4, and β9. β4 forms the beginning of the second sheet, followed by β5, β6, β7 and β8. β9 is preceded by a large loop region, which allows the β9 strand to reach round and form part of the first β-sheet (Figure 4.17). Therefore the two β-sheets are attached to each other, forming a closed β-barrel. The location of the strands are mapped onto the sequence shown in Figure 4.17D. There is an alternative rotamer at Arg43, which has been modelled in two positions, each at 50% occupancy. The



C.

MMSEDLRVGLFPVRYLVGTGLPGAPQLVLDLMVDTVDHSVVGRAAVSQAVSPPL
NFHADVWGSYVFRL~~GPPPRRDGS~~GAIVQISLQGNQGGPQSNSMITFYGELLLKG
DGKTGVASYRYYSNGSWHEVENVPVKAD~~PELVPIEPGPVIGQSSMSAIGSAAMY~~
~~GVAIQSAAASGDLAHMRTLSAYARQQLESRDEIAAALSELKAEIAKLESRQ~~

D.

β1 β2 β3
~~MMSEDLRV~~GLFPVRYLVGTGLPGAPQLVLDLMVDTVDHSVVGRAAVSQAVSPPL
β4 β5 β6
NFHADVWGSYVFRL~~GPPPRRDGS~~GAIVQISLQGNQGGPQSNSMITFYGELLLKG
β7 β8 β9
DGKTGVASYRYYSNGSWHEVENVPVKAD~~PELVPIEPGPVIGQSSMSAIGSAAMY~~

~~GVAIQSAAASGDLAHMRTLSAYARQQLESRDEIAAALSELKAEIAKLESRQ~~

Figure 4.17. BPSS0212+M-FormA structure and sequence annotation. BPSS0212+M-FormA shown as cartoon, coloured in chainbows, blue (N-terminus) to red (C-terminus). The arrows represent β-strands. **A.** View perpendicular to the β-barrel. **B.** View parallel to the β-barrel. **A.** and **B.** images generated in PyMOL version 1.1r1. **C.** BPSS0212+M protein sequence, the highlighted region is present in the structure - the yellow highlighted sequence is the DUF1842 domain (Pfam - Finn et al., 2016), and the green sequence is the interdomain region between the DUF1842 and DUF1843 domains. **D.** The locations of the β-strands are mapped onto the BPSS0212+M sequence. The crossed out protein regions are unmodelled residues.

packing of the BPSS0212+M-FormA molecules in the crystal lattice, analysed by PDBePISA (Krissinel and Henrick, 2007) does not suggest any significant interactions or assemblies between molecules.

4.6 Analysis of other BPSS0212+M structures

Crystals from a solution of a putative BPSS0212+M/BPSS0213 complex were sent to Diamond Light Source Synchrotron, Oxford, to investigate whether the crystals contained the complex of BPSS0212+M/BPSS0213, or just BPSS0212+M or BPSS0213 alone. A number of the crystals sent contained BPSS0212+M alone and the data collected processed to different cell dimensions to the previously solved structure. Therefore these crystals may have contained some protein regions not seen in the previously solved structure or even the full-length BPSS0212+M. The structures were solved by molecular replacement using the previously solved BPSS0212+M-FormA model, in the program PhaserMR (McCoy et al., 2007). A selection of these structures exhibiting different spacegroups and structural elements will be described below.

4.6.1 BPSS0212+M-FormB

FormB crystallised in conditions PACT D10 (0.2 M magnesium chloride, 0.1 M Tris pH 8.0, 20% (w/v) PEG6000), in spacegroup C2 with cell dimensions of a=83.8 Å, b=40.7 Å, c=83.2 Å, α=γ=90°, β=107.2° and diffracted to a resolution of 2.34 Å. The crystal form had the same spacegroup as BPSS0212+M-FormA but with a longer cell dimension in the c-axis. The asymmetric unit contains two molecules of BPSS0212+M-FormB, chain A and chain B, each forming a eight-stranded antiparallel β-barrel. A significant interface was identified between the bottom (where the N- and C-termini reside) of two molecules of chain B, as assessed by PDBePISA (Krissinel and Henrick, 2007). Compared to BPSS0212+M-FormA both BPSS0212+M-FormB molecules have an extra three residues at the N-terminus, starting at residue Leu6, modelled as an Ala, however both protein chains still terminate at residue 136. Therefore the

structure does not cover the DUF1843 domain, only the extra N-terminal residues of BPSS0212+M, corresponding to the short amino acid sequence before the start of the DUF1842 domain, the DUF1842 domain and part of the DUF1842-DUF1843 interdomain linker (Figure 4.18A and B).

4.6.2 BPSS0212+M-FormC

FormC crystallised in conditions PACT A1 (0.1 M SPG buffer pH 4.0, 25% (w/v) PEG1500), in spacegroup $P2_12_12_1$ with cell dimensions of a=40.5 Å, b=49.6 Å, c=59.0 Å, α=β=γ=90°, and diffracted to a resolution of 1.80 Å. One molecule of BPSS0212+M-FormC is present in the asymmetric unit, forming a eight-stranded antiparallel β-barrel, with the same number of residues compared to BPSS0212+M-FormA, ending at residue 136. The β5-β6 loop, residues 89-98, is in a different conformation compared to BPSS0212+M-FormA (Figure 4.18C and D). This suggests the loop is flexible and may be involved in protein:protein interactions. Two glycine molecules were identified in the solvent regions, assumed to come from the SPG buffer component of the crystallisation solution.

4.6.3 BPSS0212+M-FormD

FormD crystallised in conditions PACT G2 (0.2 M sodium bromide, 0.1 M Bis-Tris propane pH 7.5, 20% (w/v) PEG3350), in spacegroup $P2_12_12_1$ with cell dimensions – a=40.3 Å, b=51.1 Å, c=58.3 Å, α=β=γ=90°, diffracting to a resolution of 1.69 Å. One molecule of BPSS0212+M-FormD is present in the asymmetric unit, forming a eight-stranded antiparallel β-barrel, with an extra eight C-terminal residues ending at residue 144, compared to BPSS0212+M-FormA. These residues protrude out from the structure forming an unstructured region. The electron density surrounding the C-terminus in one molecule and loop residues 48-55 in a symmetry related molecule is difficult to interpret, resulting in a section of this loop, residues 51-53, remaining unbuilt. A chloride ion was modelled into electron density surrounding the protein, assumed to be

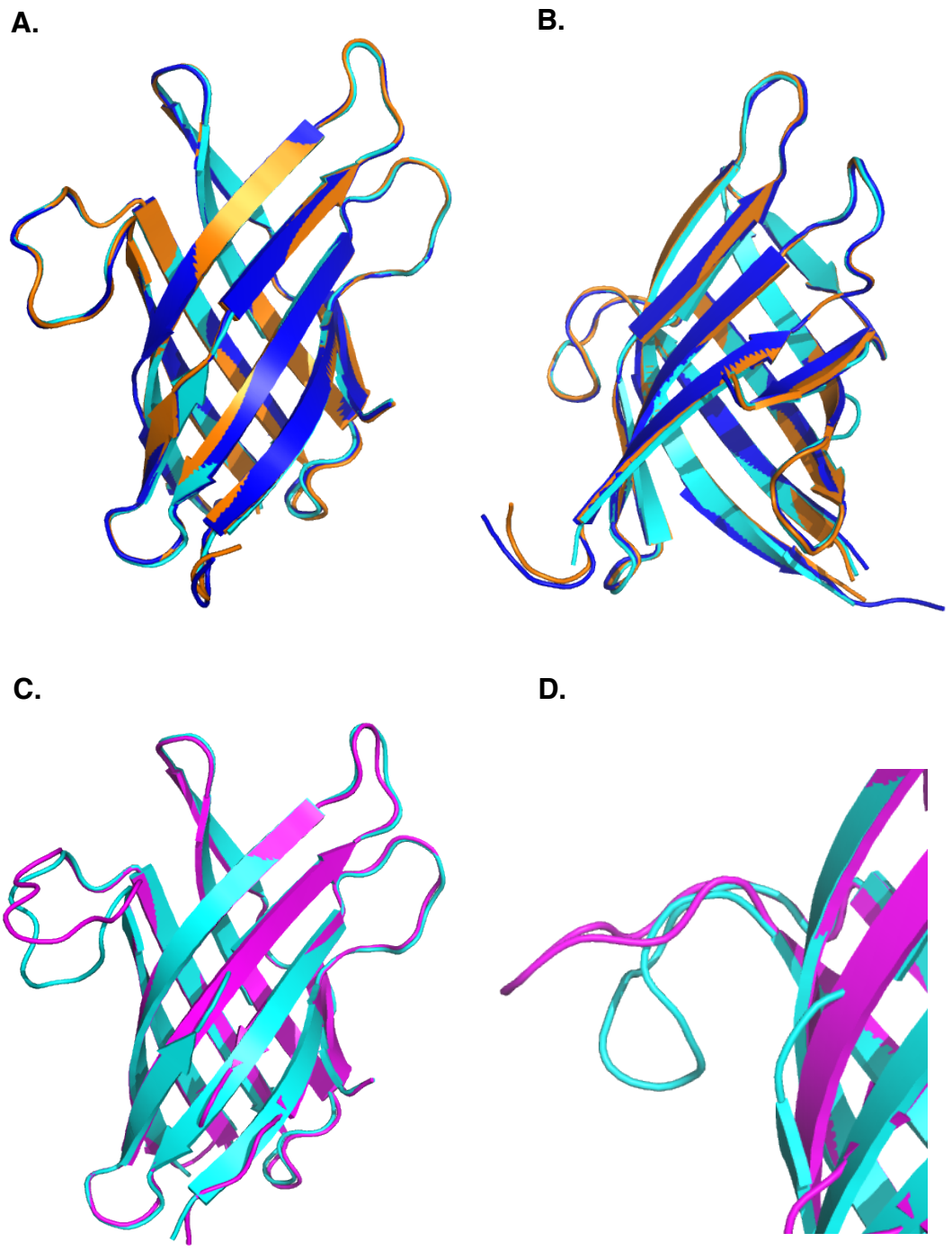


Figure 4.18. Comparison of BPSS0212+M structures BPSS0212+M-FormB and BPSS0212+M-FormC to BPSS0212+M-FormA. All structures are shown as cartoon representation, the arrows representing β-strands. BPSS0212+M-FormA is shown in blue. **A.** and **B.** - The two molecules in the asymmetric unit of BPSS0212+M-FormB – chain A (orange) and chain B (dark blue). **C.** and **D.** - BPSS0212+M-FormC is magenta. A conformational change is seen in loop residues 89-98 between the FormA and FormC structures. Structures aligned in SSM – *Coot* (Emsley et al., 2010; Krissinel and Hendrick, 2004) and images generated in PyMOL version 1.1r1.

from the purification buffer. In a similar way to FormC, the loop between strands β5 and β6, residues 89-98, is in a different conformation compared to BPSS0212+M-FormA (Figure 4.19A). The β5-β6 loop tucks into one open end of the barrel, the site of the N- and C-termini, in a symmetry related molecule (Figure 4.19B). Analysis of the BPSS0212+M-FormD structure showed a potential metal site. Currently it has been assigned as a zinc ion on the basis of the high electron density peak, the coordination ligands and the distance to the ligands. It is involved in crystal contacts between two molecules of BPSS0212+M-FormD and is coordinated by His38 from one molecule, and His57 and Asp59 from a symmetry related molecule, along with one water molecule (Figure 4.19C and D). The crystallisation solution does not contain any zinc, suggesting the metal is present due to some cross contamination that has taken place with another crystallisation condition, however where the zinc ultimately comes from is unclear.

4.6.4 Multiple BPSS0212+M structures - conclusion

The multiple structures of BPSS0212+M have modelled residues 6-144, which correspond to the DUF1842 domain, with an N-terminal extension and part of the interdomain linker between the DUF1842 and DUF1843 domains. The interdomain linker forms most of β8 and all of β9, followed by an unstructured region that extends out from the protein. The DUF1843 domain has not been found in any of the BPSS0212+M structures solved thus far. The structures have shown that the β5-β6 loop is flexible, as it has been found in two different conformations. The β5-β6 loop and N-terminus of the protein have both been shown to interact with one open end of the barrel, the site of the N- and C-termini, suggesting a possible binding site. The presence of a metal ion in the structure may not suggest any biological relevance, as it is involved in crystal contacts on the exterior of the protein.

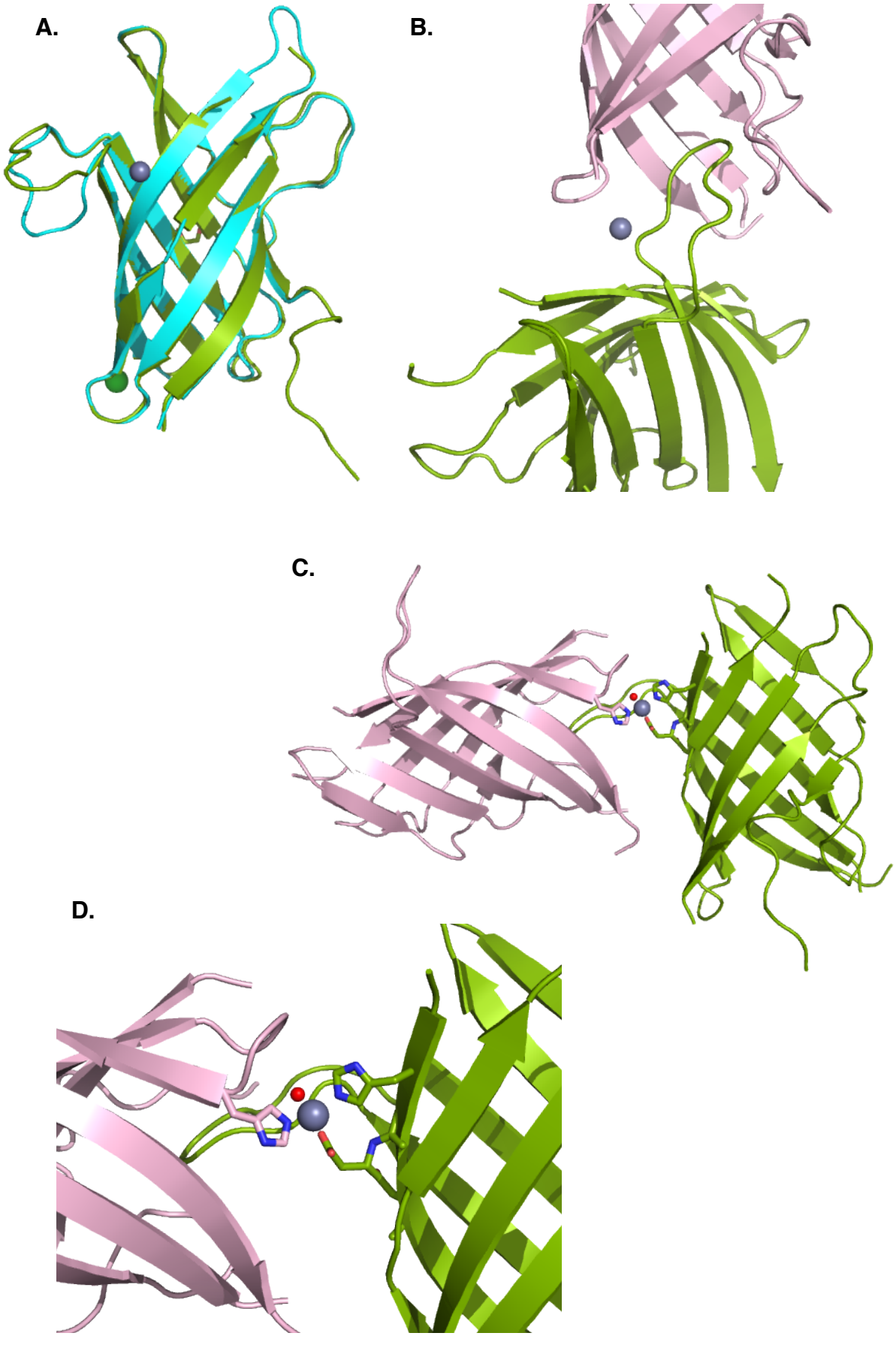

A.
B.
C.
D.

Figure 4.19. Comparison of BPSS0212+M-FormA to BPSS0212+M-FormD. All structures are cartoon representation, the arrows representing β-strands. **A.** BPSS0212+M-FormA is shown in blue. BPSS0212+M-FormD is green. The C-terminus of FormD is extended by eight residues. **B.** BPSS0212+M-FormD (green) and symmetry related molecule (pink). The green loop residues 89-98 pack into the bottom of the pink β-barrel. **C.** BPSS0212+M-FormD (green) and symmetry related molecule (pink). A predicted zinc ion (grey) is coordinated by His57 and Asp59 of the green molecule and His38 from the symmetry related pink molecule and a water molecule (red). **D.** Zoomed in on the zinc binding site. Structures aligned in SSM – *Coot* (Emsley et al., 2010; Krissinel and Hendrick, 2004) and images generated in PyMOL version 1.1r1.

4.7 Conserved residues

4.7.1 BPSS0212 vs BPSS0213

BPSS0212 and BPSS0213 share 38% identity with 97-98% coverage (Altschul et al., 1990). The residues present in both BPSS0212 and BPSS0213 were mapped onto the structure of BPSS0212+M-FormA. The location of these residues is spread throughout the structure (Figure 4.20). The majority of conserved residues are in the core of the protein and at either open end of the β-barrel. The conserved core residues suggest they are required to maintain the overall structure or a possible binding site. There is a cluster of residues at the open end of the barrel, where the N- and C-termini are located (Figure 4.21). The cluster consists of G9, L10, F11, P12, V13, V33, T35, V36, V81 and L106 as seen in BPSS0212+M. Another larger cluster of conserved residues is present at the other open end of the barrel. The residues involved are, as found in BPSS0212+M, G23, A24, P25, Q48, V50, P52, R118, Y119, W125, H126 and E129. The conservation of residues at either end of the barrel suggests a possible ligand-binding site or protein interaction site. One of the largest differences in sequence is that BPSS0212 has a four residue insertion at residues 72-75, which is not found in BPSS0213, this insertion is located in the β4-β5 loop, not modelled in BPSS0212+M. There is also a nine residue insertion present in the DUF1842-DUF1843 domain linker, not found in BPSS0213. The structure of BPSS0213 DUF1842 domain is predicted to be very similar to the BPSS0212 DUF1842 domain based on the sequence similarity and pattern of conserved residues.

4.7.2 DUF1842 domain residue conservation

The protein sequences of members of the DUF1842 domain (PF08896) family were aligned by Pfam (Finn et al., 2016), and viewed in Jalview (Waterhouse et al., 2009) (Figure 4.22). The consensus sequence for the family has been mapped onto the structure of BPSS0212+M-FormA to highlight the location of the conserved residues. The conserved residues

Figure 4.20. BPSS0212 vs BPSS0213 amino acid sequence comparison with conserved residues mapped onto BPSS0212+M-FormA structure. A. Amino acid sequences of BPSS0212 and BPSS0213 aligned using the TCoffee defaults in Jalview and coloured using Clustalx – highlighting the 100% conserved residues (Larkin et al., 2007; Notredame et al., 2000; Waterhouse et al., 2009). **B.** 100% conserved residues between BPSS0212 and BPSS0213, present in the structure determined sequence region (BPSS0212+M-FormA residues 9-136), are highlighted on the BPSS0212+M-FormA model (blue) as pink sticks, coloured by element. Images generated in PyMOL version 1.1r1 .

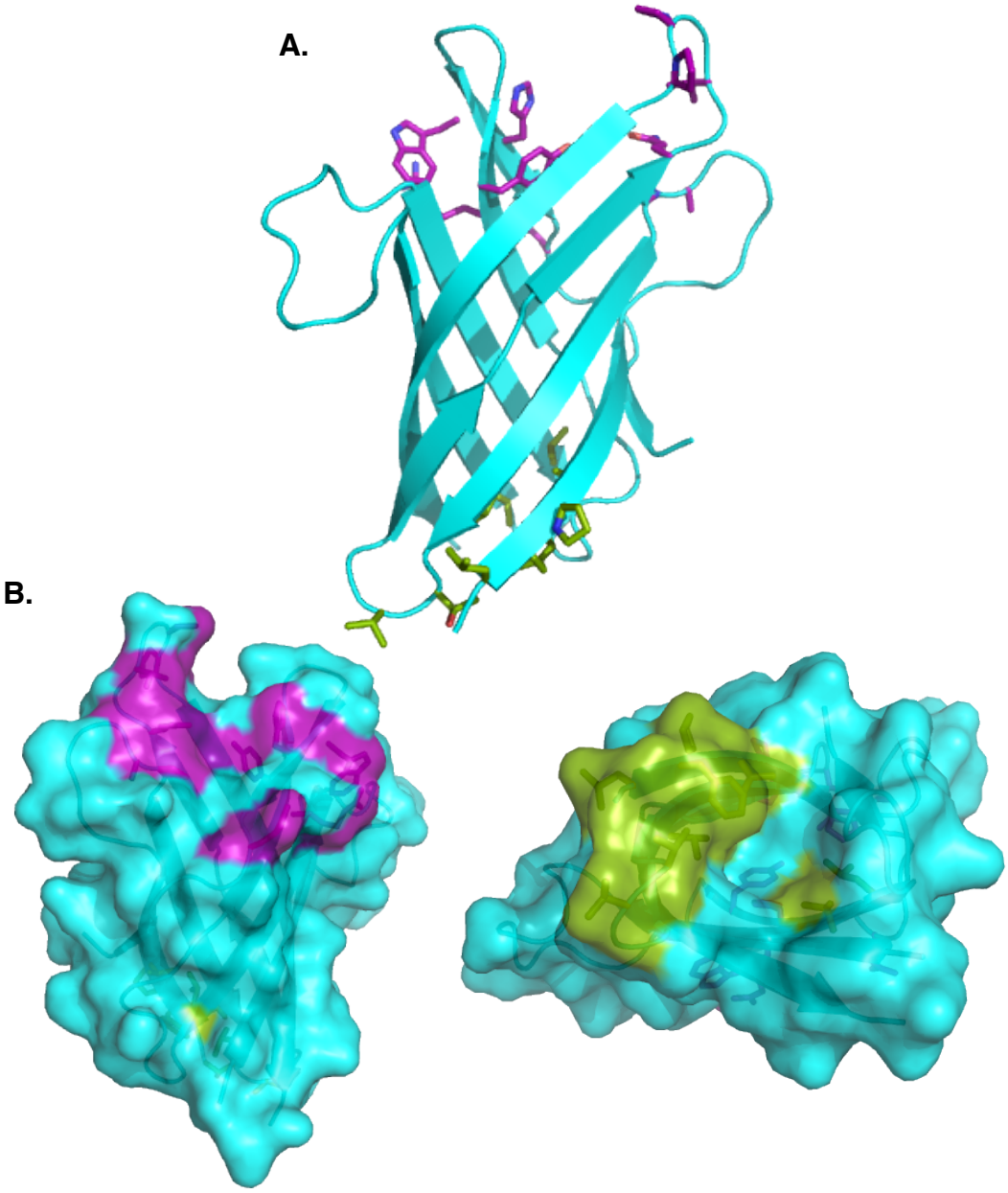


Figure 4.21. Conserved residue clusters between BPSS0212 vs BPSS0213 mapped onto BPSS0212+M-FormA structure. A. Two clusters (purple and green) of conserved residues, shown as sidechain sticks, were found at either end of the BPSS0212+M-FormA β-barrel. **B.** Patches (purple and green) highlighted on the BPSS0212+M-FormA structure, shown as cartoon with transparent surface. Residues associated with the clusters - purple – G23, A24, P25, Q48, V50, P52, P53, R118, Y119, W125, H126, E129 and green – G9, L10, F11, P12, V13, V33, T35, V36, V81, L106. Images generated in PyMOL version 1.1r1.

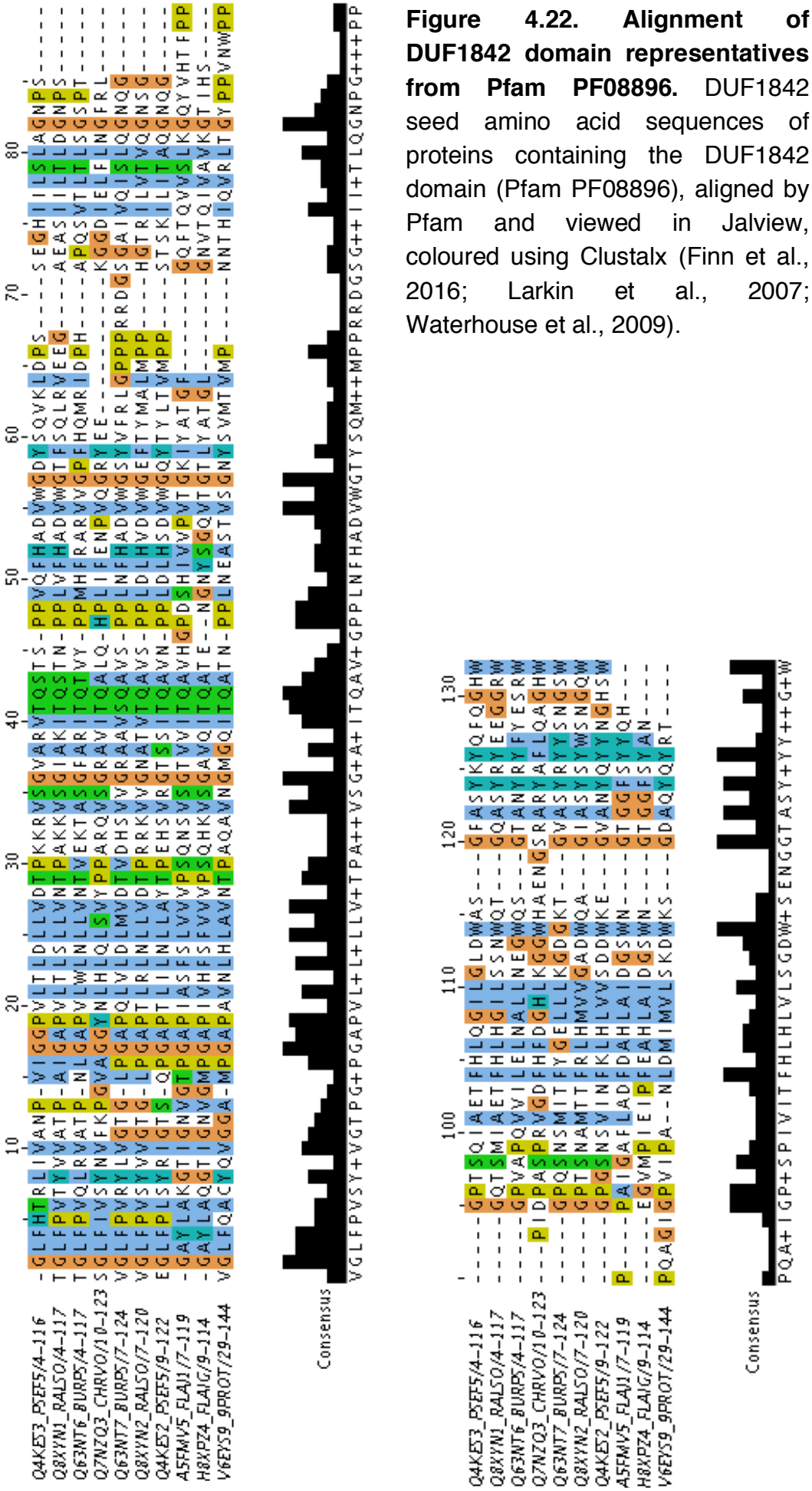


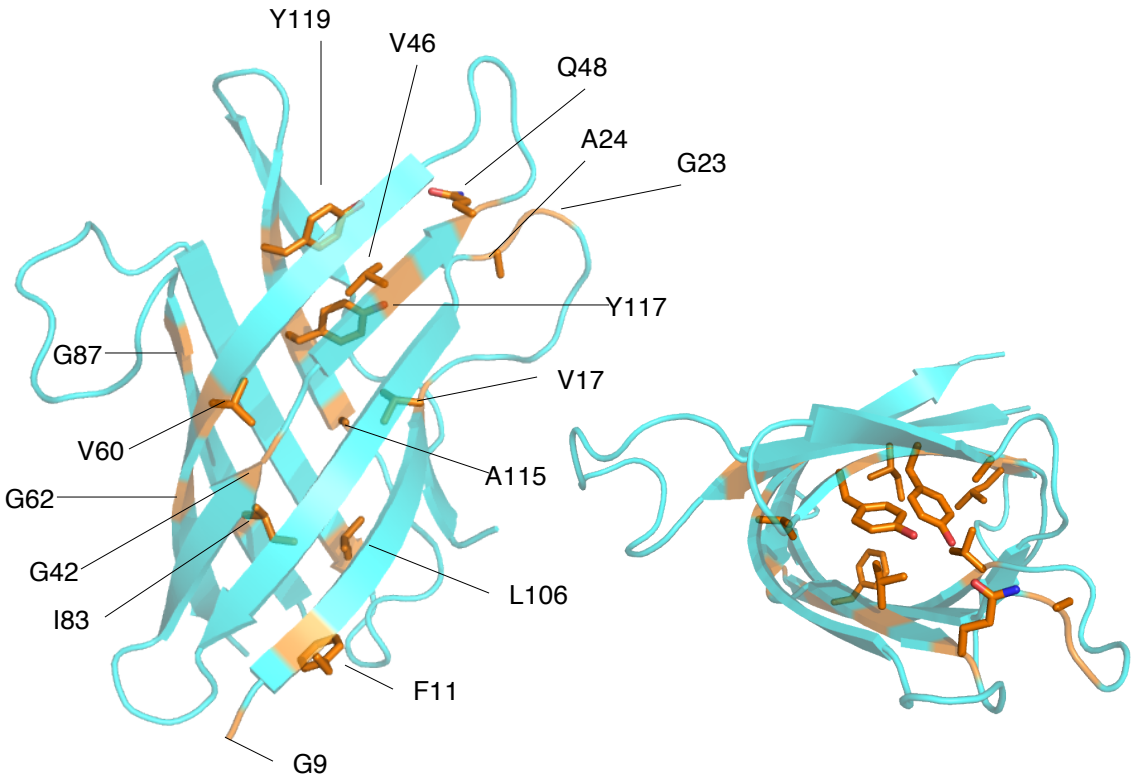
Figure 4.22. Alignment of DUF1842 domain representatives from Pfam PF08896. DUF1842 seed amino acid sequences of proteins containing the DUF1842 domain (Pfam PF08896), aligned by Pfam and viewed in Jalview, coloured using Clustalx (Finn et al., 2016; Larkin et al., 2007; Waterhouse et al., 2009).

are defined as minimum conservation of 75%. The conserved residues are G9, F11, V17, G23, A24, G42, V46, Q48, V60, G62, I83, G87, L106, A115, Y117 and Y119 (BPSS0212+M numbering) and are located predominantly in the core of the protein (Figure 4.23). This suggests a possible binding site or structural contribution to maintain the overall shape of the protein.

4.8 Structural similarity of BPSS0212+M to lipocalins

Visual analysis of the structure of BPSS0212+M-FormA revealed its similarity to the lipocalin family. Lipocalins are found in eukaryotes, but have also been found in bacteria (reviewed in Flower, 1996; Flower et al., 1995). They are small extracellular proteins that bind hydrophobic molecules (Godovac-Zimmermann, 1988; North, 1989; Pevsner et al., 1988). Lipocalins have a range of functions, including synthesis, transport, immune response regulation and coloration (reviewed in Flower, 1996). These functions are carried out by specific properties such as ligand binding, cell-surface receptor binding and complex formation (reviewed in Flower, 1996; Flower et al., 1993b). As shown in Figure 4.24, human complement protein C8γ is able to bind the pseudoligand lauric acid via a binding pocket (Chiswell et al., 2007), and retinol binding protein, as well as binding retinol is able to bind transtherytin (Zanotti et al., 2008). The structure of the family members consists of a 3_{10}-helix, followed by an eight-stranded antiparallel β-barrel and a C-terminal helix (Flower et al., 1993b), which is similar to the overall structure expected for BPSS0212, with the absence of the 3_{10}-helix. The lipocalin β-barrel contains an internal binding pocket that is not seen in BPSS0212+M, and has a large Ω (omega) loop connecting β-strands 1 and 2, which folds over the ligand-binding site (Flower et al., 1993b).

The amino acid sequence of lipocalins can contain up to three structurally conserved regions (SCRs), associated with regions of sequence conservation (Flower et al., 1991, Flower et al., 1993b; Godovac-



MMSEDLRVGLFPVRYLVGTGLPGAPQLVLDLMVDTVDHSVVGRAAVSQAVSPPL
NFHADVWGSYVFRLGPPPRRDGSGAIVQISLQGNQGGPQSNSMITFYGELLLKG
DGKTGVASYRYYSNGSWHEVENVPVKADPELVPIEPGPVIGQSSMSAIGSAAMY
GVAIQSAAASGDLAHMRTLSAYARQQLESRDEIAAALSELKAEIAKLESRQ

Figure 4.23. DUF1842 domain consensus sequence mapped onto the BPSS0212+M-FormA structure. The consensus sequence is defined at minimum 75% conservation, from the alignment shown in Figure 4.22. BPSS0212+M-FormA is shown as a 40% transparent cartoon, coloured blue, with conserved sidechains shown as orange sticks. The conserved residues are highlighted in yellow in the BPSS0212+M amino acid sequence. Alignment information from Pfam PF08896 and Jalview (Finn et al., 2016; Waterhouse et al., 2009). Images generated in PyMOL version 1.1r1.

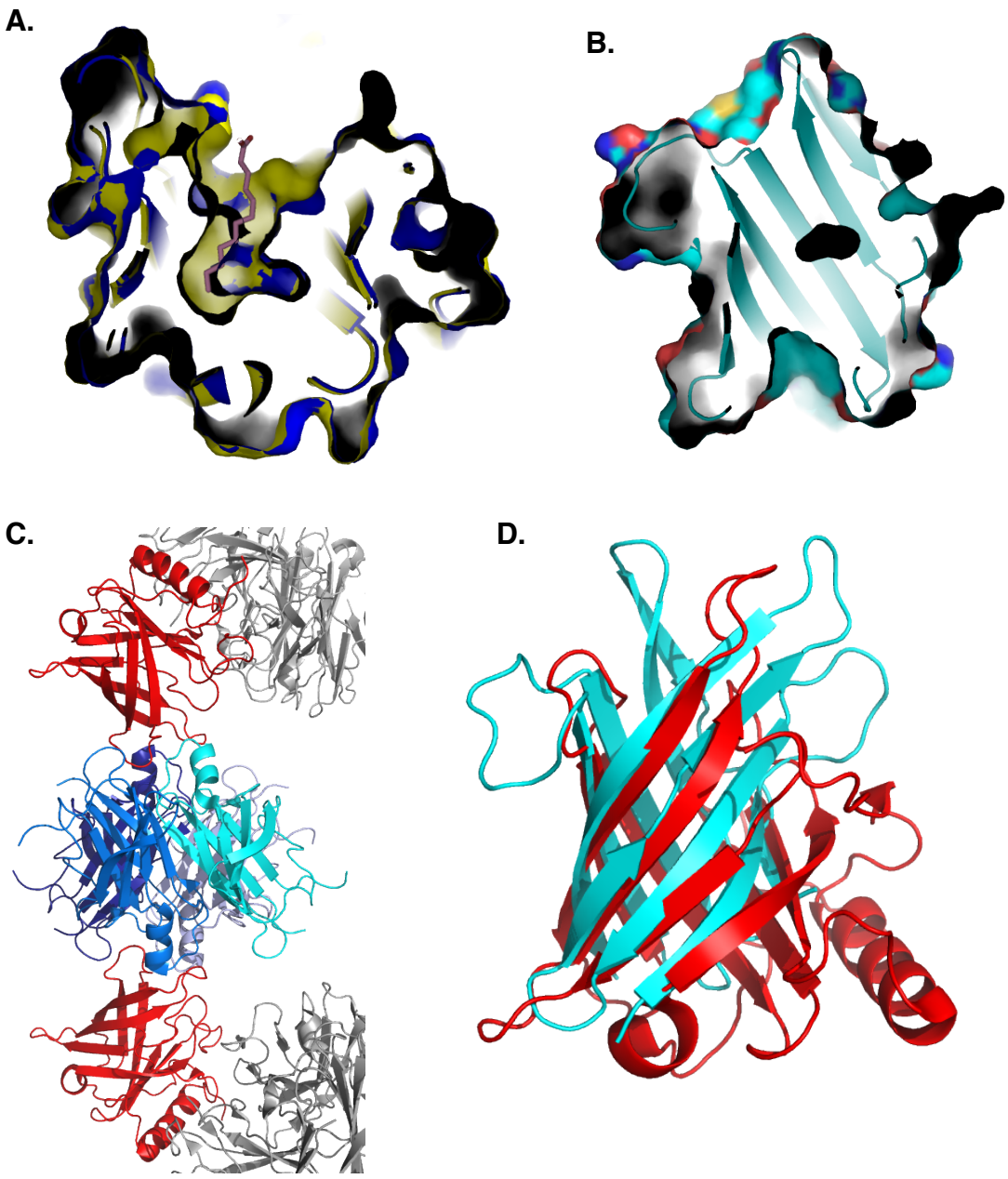


Figure 4.24. Lipocalin – structure vs function

A. A cross section of the human complement protein C8γ, shown as cartoon and surface, in the absence of ligand (PDB:2OVE) (Chiswell et al., 2007) (dark blue) contains a binding site, which changes conformation in order to accommodate ligand (pink) (PDB:2OVD) (yellow) (Chiswell et al., 2007). **B.** The cross section of BPSS0212+M-FormA, shown as cartoon and surface. **C.** Transthyretin in complex with plasma retinol-binding protein (RBP), and Fab antibody chains (PDB:3BSZ) (Zanotti et al., 2008), with four transthyretin coloured in shades of blue, two RBP in red and Fab proteins in grey. **D.** BPSS0212+M-FormA (blue) is aligned with RBP (from PDB:3BSZ) (Zanotti et al., 2008) (red). Structures aligned in SSM – *Coot* (Emsley et al., 2010; Krissinel and Hendrick, 2004) and images generated in PyMOL version 1.1r1.

Zimmermann, 1988; Pervaiz and Brew, 1985; Pevsner et al., 1988). SCR-1 has GXW(F/Y) in the region of the C-terminal 3_{10}-helix and first strand, SCR-2 is in the β-strands 6-7 and SCR-3 is in the last strand and loop of the barrel (Flower et al., 1993b; Godavac-Zimmerman, 1988; reviewed in Flower 1996). SCR-1 is present in all lipocalins and is used as an identifying feature (reviewed in Flower, 1996). One sub-group of lipocalins, the kernel lipocalins, contain all three SCR, whereas the other subgroup, outlier lipocalins, only contain one or two SCRs (Flower et al., 1993b; reviewed in Flower, 1996). BPSS0212 contains a β-barrel reminiscent of a lipocalin, however it does not contain any of the SCR or an identifiable binding site comparable to those found in other lipocalins.

4.9 Structural analysis of BPSS0212+M-FormA using the Dali server

The model of BPSS0212+M-FormA was run through the Dali server (Holm and Rosenström, 2010), to see if the protein fold is similar to any of the other proteins present in the PDB. The results included outer-membrane proteins, globins, quinohemoprotein amine dehydrogenases and members of the avidin family, with the top 10 results of PDB90 shown in Table 4.3. The PDB90 is derived from the PDB, as a group of structures that only includes one chain of a protein, when more than one chain is present that is 90+% identical in sequence (Holm and Rosenström, 2010). The superpositions carried out by Dali and output in PDB90 results (Holm and Rosenström, 2010) were analysed to assess the similarity between the structures (Table 4.3).

4.9.1 Outer membrane proteins

The top PDB90 result from Dali is outer membrane protein NspA (Neisserial surface protein A), chain A (PDB:1P4T) (Vandeputte-Rutten et al., 2003) with a Z-score of 10.4 and RMSD of 2.6 Å (Holm and Rosenström, 2010). The comparison of BPSS0212+M-FormA with the outer-membrane protein, NspA, revealed the main similarity is the presence of a β-barrel, which shares a very similar curvature and trace of

No.	Chain	Z-score	RMSD (Å)	lali	nres	% id	Description
1	1P4T-A	10.4	2.6	105	155	3	Outer membrane protein NSPA
2	2GTL-M	10.4	3.3	100	217	7	Extracellular globin 4
3	2ERV-A	10.4	2.6	103	150	6	Hypothetical protein PAER030023600
4	2GTL-N	10.2	3.0	99	220	7	Extracellular globin 4
5	4U8U-b	10.2	3.2	99	222	9	Globin A chain
6	4U8U-c	10.1	3.1	99	219	9	Globin A chain
7	1QJP-A	9.9	2.7	99	137	2	Outer membrane protein A
8	3EW1-B	9.4	2.8	95	135	9	Rhizavidin
9	1JJU-A	9.4	2.6	92	489	9	Quinohemoprotein amine dehydrogenase
10	4JNJ-A	9.2	2.6	91	115	5	Streptavidin/Rhizavidin Hybrid

Table 4.3. Top 10 Dali server PDB90 results for BPSS0212+M-FormA structure. BPSS0212+M-FormA structure was run through the Dali server (Holm and Rosenström, 2010). The top ten PDB90 results are shown in the table. The columns refer to: no. - match number, chain - PDB code and corresponding chain, Z-score – Z-score of match, RMSD – RMSD of match, lali – number of aligned residues, nres – number of residues, %id - % sequence identity and the description of the PDB structure (Holm and Rosenström, 2010).

β-strands to that of BPSS0212+M-FormA (Figure 4.25A and B). NspA has a number of charged residues in it's core, whereas, the core of BPSS0212+M-FormA is strongly hydrophobic (Vandeputte-Rutten et al., 2003). NspA also has a number of aromatic residues arranged in two bands at the top and bottom of the barrel (Vandeputte-Rutten et al., 2003). BPSS0212+M-FormA has a few aromatic residues present on one side of its outer face but they form a cluster rather than a band. Another two outer membrane proteins were identified in the top 10 – PagL and OmpA (PDB:2ERV) (Rutten et al., 2006) (PDB:1QJP) (Pautsch and Schulz, 2000). The Dali alignment of BPSS0212+M-FormA with both NspA and PagL highlights the presence of larger flexible loops at the same end of all three barrels (Rutten et al., 2006; Vandeputte-Rutten et al., 2003). It is likely that BPSS0212+M-FormA is not an outer membrane protein, due to the unfavourable arrangement of amino acids into a hydrophobic core and charged outer surface, lack of external aromatic residues and no signal sequence possibly required for targeting to the outer membrane.

4.9.2 Globins

The second result in the PDB90 Dali results is Extracellular globin 4, chain M (PDB:2GTL) (Royer et al., 2006) with a Z-score of 10.4 and RMSD of 3.3 Å (Holm and Rosenström, 2010). Chain M is the hemoglobin linker chain L1 from *Lumbricus terrestris*, which is one of three types of linker chain involved in the formation of erythrocruorin – a hemoglobin complex – formed from 180 protein chains, with 36 linker chains and 12 hemoglobin dodecamers (144 chains) (Royer et al., 2006). All the linker proteins have an eight-stranded β-barrel with a N-terminal helix (Royer et al., 2006). The linker chains have two helices located N-terminal to the β-barrel and a cysteine-rich region, unlike in BPSS0212 where the helix is at the C-terminus and there are no cysteine residues in the protein (Royer et al., 2006). Another globin present in the top 10 is the giant hemoglobin from *Glossoscolex paulistus* (PDB:4U8U) (Ruggiero

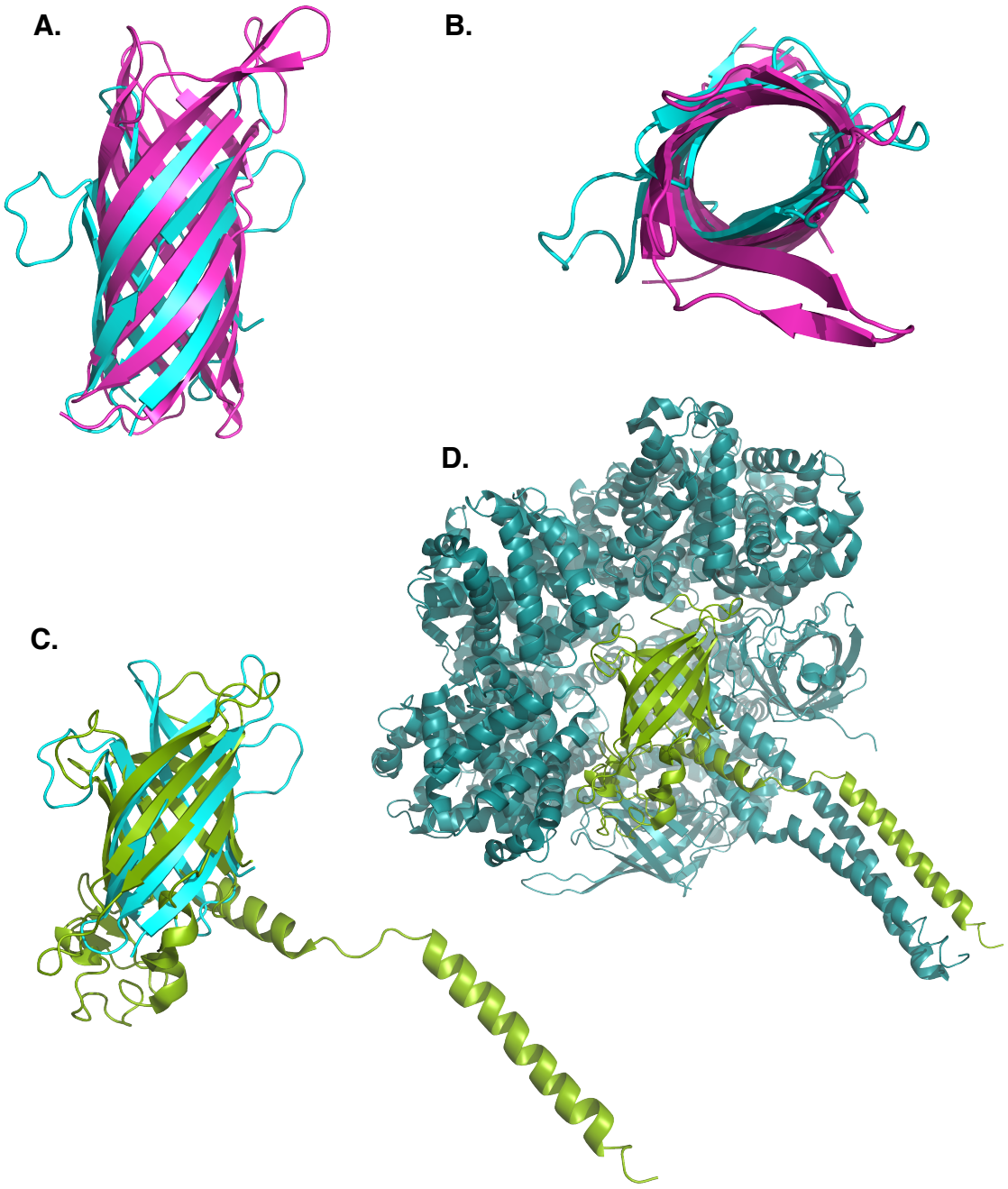


Figure 4.25. Structural comparisons between BPSS0212+M-FormA and the top two PDB90 Dali results. All structures are shown as cartoon. **A.** and **B.** BPSS0212+M-FormA (blue), with NpsA, chain A (magenta) (PDB:1P4T) (Vandeputte-Rutten et al., 2003). **C.** BPSS0212+M-FormA (blue), with *Lumbricus* erythrocruorin hemoglobin linker chain L1 (chain M) (PDB:2GTL) (Royer et al., 2006) in green. **D.** *Lumbricus* erythrocruorin hemoglobin dodecamer (teal) and the linker chain L1, (chain M) (green) (PDB:2GTL) (Royer et al., 2006). Alignments carried out by the Dali server (Holm and Rosenström, 2010). Images generated in PyMOL version 1.1r1..

Bachega et al., 2015), with a Z-score of 10.2-10.1 and RMSD of 3.2-3.1 Å. Similarly to 2GTL, BPSS0212+M-FormA is aligned with linker chain L1, but also linker chain L2 (Holm and Rosenström, 2010). The two linker chains share ~30% homology (Ruggiero Bachega et al., 2015). The core of the L1 protein β-barrel is hydrophobic, which is similar to that of BPSS0212+M, and in both cases the trace of the β-barrel is similar (Royer et al., 2006; Ruggiero Bachega et al., 2015).

4.9.3 Avidins

BPSS0212+M-FormA was aligned with rhizavidin, chain B (PDB:3EW1, Meir et al., 2009) with a Z-score of 9.4 and RMSD of 2.8 Å (Holm and Rosenström, 2010) (Figure 4.26A and B). Members of the avidin family bind the vitamin biotin, via a binding pocket located at the barrel end with the larger loops (Meir et al., 2009). The avidin family is part of the larger calycin superfamily, along with the lipocalin family (Flower, 1993a). The full Dali PDB results include avidin, rhizavidin, zebavidin, streptavidin, xenavidin, tamavidin and bradavidin, with Z-scores ranging from 9.5-3.0 (Holm and Rosenström, 2010). Although BPSS0212+M-FormA structure is found to be similar to that of rhizavidin, through the location of the extended loops and the tracing of the β-barrel, it does not contain a binding pocket in the region expected when aligned with 3EW1 or any of the conserved residues required for biotin binding (Meir et al., 2009).

4.9.4 Quinohemoprotein amine dehydrogenase

Another type of protein present in the Dali PDB90 top 10 results is a quinohemoprotein amine dehydrogenase. BPSS0212+M-FormA was aligned with chain A of quinohemoprotein amine dehydrogenase PDB:1JJU (Datta et al., 2001), with a Z-score of 9.4 and RMSD of 2.6 Å (Holm and Rosenström, 2010). Specifically BPSS0212+M-FormA is aligned with the first β-barrel of chain A (Figure 4.26C and D). The first β-barrel has a predominantly hydrophobic core and a similar trace of β-strands to BPSS0212+M-FormA. However BPSS0212 is most likely not a

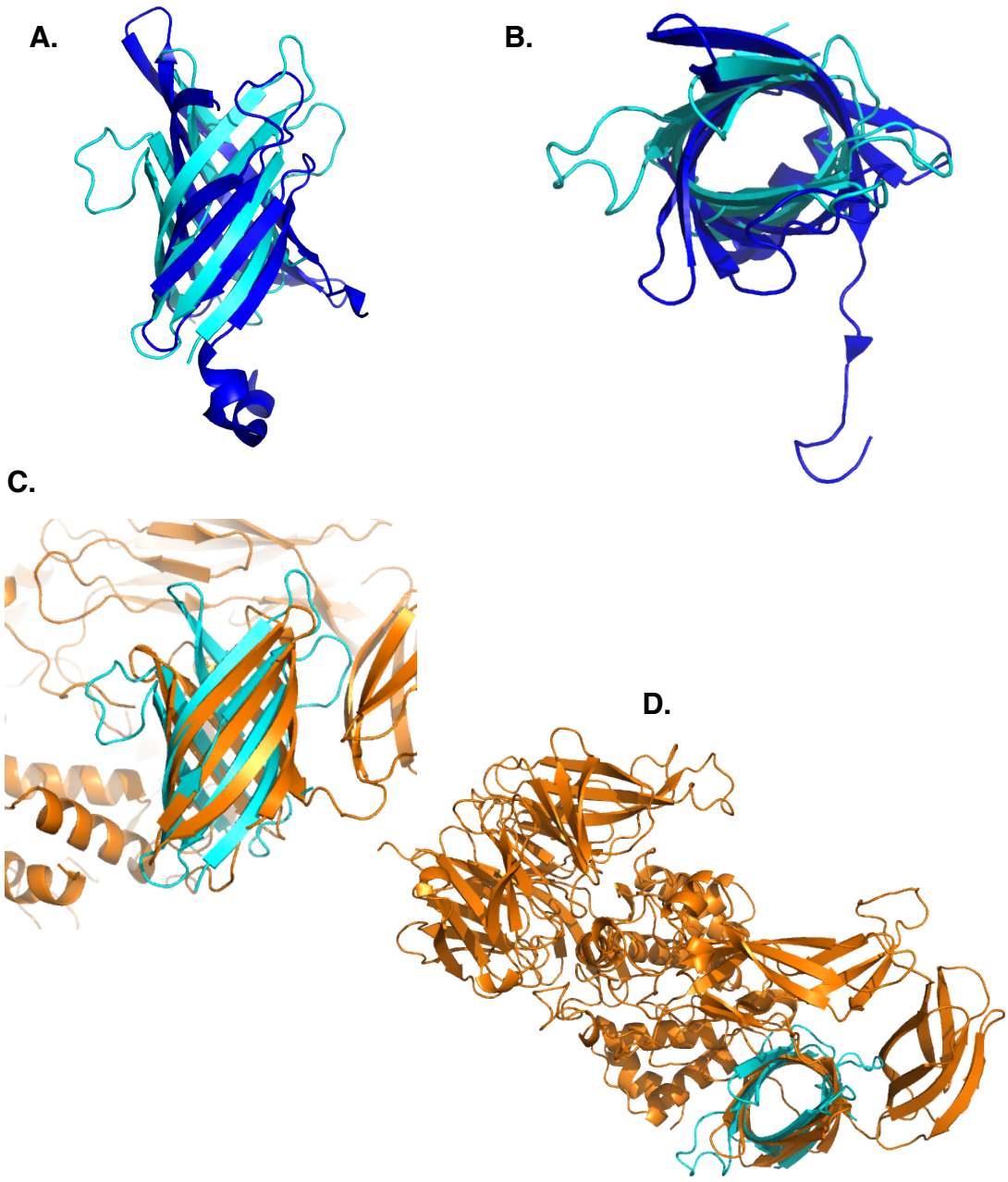


Figure 4.26. Structural comparisons between BPSS0212+M-FormA and the third and fourth PDB90 Dali results. All structures are shown as cartoons. **A.** and **B.** BPSS0212+M-FormA (blue), with rhizavidin, chain B (dark blue) (PDB:3EW1) (Meir et al., 2009). **C.** and **D.** BPSS0212+M-FormA (blue), with quinohemoprotein amine dehydrogenase, chain A (orange), (PDB:1JJU) (Datta et al., 2001). Alignments carried out by the Dali server (Holm and Rosenström, 2010). Images generated in PyMOL version 1.1r1.

quinohemoprotein amine dehydrogenase, due to the lack of other subunits present in the protein.

4.10 Functional studies

The analysis of BPSS0212 protein has thus far revealed a structure, which is unable to confirm a function for the protein. Therefore crystallisation studies were carried out with a range of ligands to identify possible interacting molecules.

4.10.1 Soaking biotin into crystals of BPSS0212

As BPSS0212 is found to be structurally similar to avidins and streptavidins, crystals of BPSS0212 were incubated with biotin to see if the ligand binds anywhere on the protein. Crystals were soaked with biotin, and sent to the Diamond Light Source Synchrotron, Oxford for data collection. Unfortunately no biotin was found in the structures solved from the soaked crystals of BPSS0212.

4.10.2 Use of a metabolic cocktail to identify possible ligands

Research carried out by Shumilin et al., 2012 used a number of metabolic cocktails soaked into crystals of proteins of unknown function in order to identify any possible binding molecules. The metabolites chosen (total of 87) were split into 11 groups of related compounds resulting in 11 metabolic cocktails (Shumilin et al., 2012). This approach produced metabolite bound structures of the two protein families tested and provided information to help predict the function of one and assigned a function to another (Shumilin et al., 2012). Therefore a similar approach was taken to test a number of metabolic compounds to see if any bound to BPSS0212.

BPSS0212 protein was purified and crystallised in sitting-drop trays with the PACT and PEGs crystallisation screens, known to produce crystals. The compounds chosen for this experiment were based on availability

and number of BPSS0212 crystal containing drops, along with literature and structural information. The molecules were added individually as a solid directly to the drop, not as a cocktail. The drops were left for 2 hours before crystals were cryoprotected, cooled and sent to the Diamond Light Source Synchrotron, Oxford for data collection.

A number of sugars were chosen, based on the assignment of the BPSS0211-BPSS0214 operon to a cluster of proteins including a protein involved in carbohydrate metabolism (Ooi et al., 2013; Winsor et al., 2008). The sugars used were: D-Glucose, D-Fructose, Sucrose, L-Fucose, L-Xylose, D-Galactose and L-Arabinose. As BPSS0212+M-FormA and avidins share structural similarity a repeat soak with biotin was carried out. Other metabolites used were 6-phosphogluconic acid, FAD, NAD and betaine. At least one of each of the 12 ligand soak experiments was screened and data sets were collected for crystals soaked with nine out of the 12 ligands. The data were solved using molecular replacement with the BPSS0212+M-FormA model in the PhaserMR program (McCoy et al., 2007). Unfortunately the structures produced did not contain any of the soaked in ligands. The lack of bound metabolite may have been due to BPSS0212 not binding any of those tested, the length of the experiment may have been too short or the use of cryoprotectant may have back soaked out any weakly bound molecules.

4.11 Conclusion and further work on BPSS0212(+M)

The structure of the DUF1842 domain, the DUF1842 domain N-terminal extension and part of the DUF1842-DUF1843 interdomain linker has been solved for BPSS0212. This is, thus far, the first structure of the DUF1842 domain produced, after analysis of Pfam and the PDB (Finn et al., 2016; Berman et al., 2000; Berman et al., 2003). The DUF1842 domain forms a eight-stranded anti-parallel β-barrel. However the function of the domain and full-length BPSS0212 remains to be

determined. A more comprehensive ligand binding study could be performed with a larger range of ligands to establish any potential binding partners.

4.12 BPSS0213

The third gene in the BPSS0211-BPSS0214 operon encodes the protein BPSS0213. This protein contains an N-terminal DUF1842 domain and a C-terminal DUF1843 domain, connected by an interdomain linker. A full-length construct of BPSS0213 was cloned from *B. pseudomallei* D286 genomic DNA during previous investigations carried out by Dr. Matthew Day (Day, 2012). Initial purification of BPSS0213 was carried out, as outlined by Dr. Matthew Day (Day, 2012), however for later purifications the pH of the gel filtration purification was changed from pH 8.0 to pH 9.0. Compared to the purification of BPSS0212 less proteolysis could be observed, suggesting the protein is more stable, also seen by Dr. Matthew Day (Day, 2012). Purification was also performed in the presence of protease inhibitors and EDTA. However despite multiple purifications of BPSS0213 and extensive crystallisation trials using sitting-drop crystallisation, protein could not be crystallised under the conditions tested.

Chapter 5. BPSS0212/BPSS0213 complex

5.1 BPSS0211, BPSS0212+M and BPSS0213 complex formation investigation

5.1.1 Investigations into the DUF1843 domain as an interaction domain

The domain analysis of the four proteins in the operon revealed that BPSS0211, BPSS0212 and BPSS0213 all contain a C-terminal DUF1843 domain (Figure 5.1). There are two distinct structures of the DUF1843 domain solved previously from BPSS0211, one is formed of two helices, the shorter of which, bends round and interacts with the larger helix (the closed conformation) (Day, 2012), the other has the shorter helix in an extended or open conformation. The closed conformation forms a dimer of dimers creating a four-helix bundle, with two zinc ion sites at the interface of the dimers and another four zinc ions forming interactions between the tetramer and symmetry related molecules (Day, 2012) (Figure 5.2). The formation of the dimer of dimers between the molecules of BPSS0211 shows the domain is able to interact with itself. Therefore this raised the possibility that the DUF1843 domains from different proteins could interact with each other, with the DUF1843 domain acting as a protein interaction motif (Day, 2012). This was investigated by mixing purified samples of the three proteins, BPSS0211, BPSS0212+M and BPSS0213 and running them down a gel filtration column together to identify any possible complex formation. BPSS0212+M is the construct of BPSS0212 with an extra N-terminal methionine present, due to a cloning error.

5.1.2 Complex formation between BPSS0212+M and BPSS0213

BPSS0211, BPSS0212+M and BPSS0213 were concentrated separately and then combined in a ~1:1:1 mg ratio and run down a gel filtration column pre-equilibrated with 50 mM Tris pH 8.0, 500 mM NaCl. SDS-PAGE analysis of the fractions from the resulting elution chromatogram

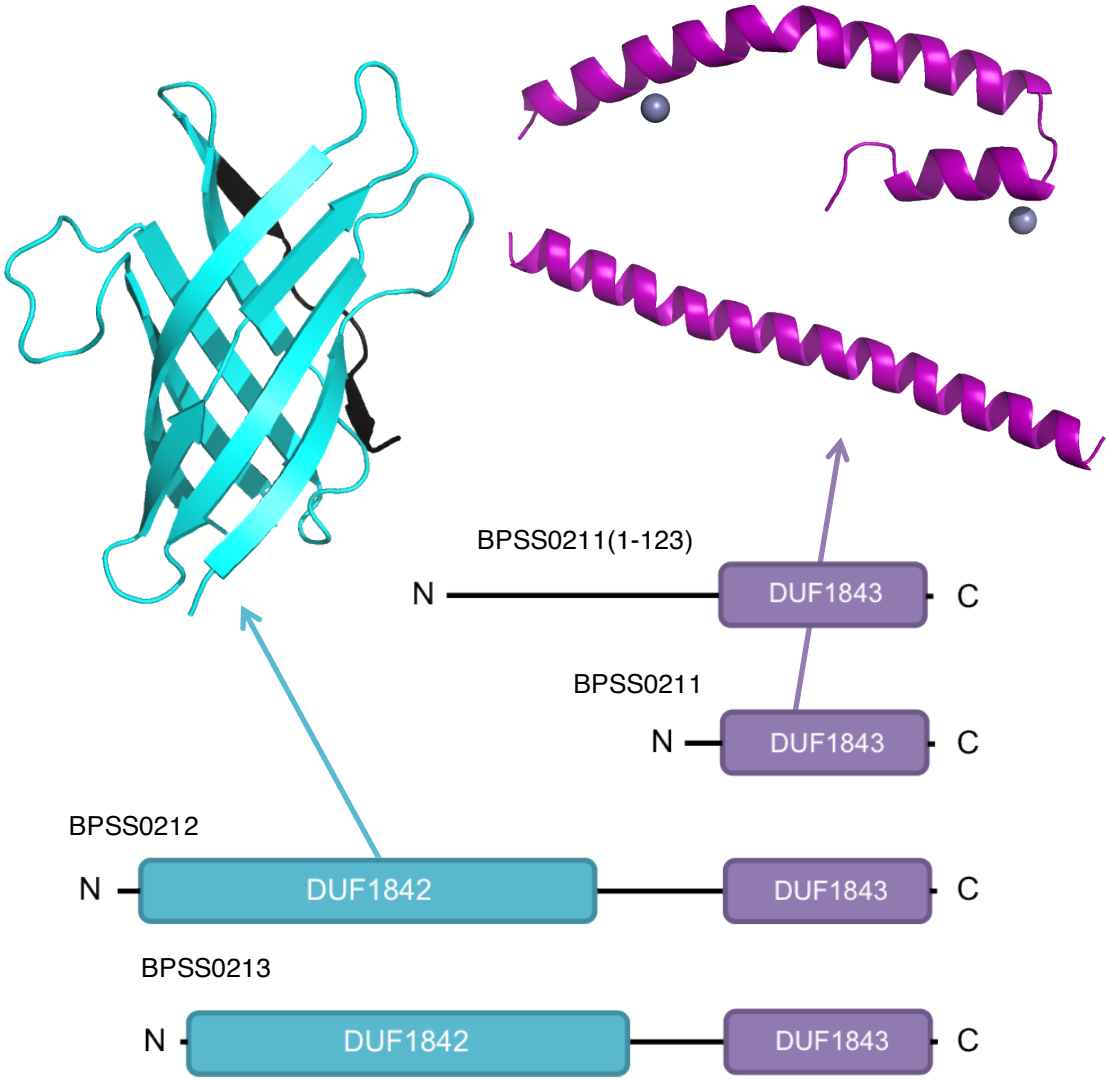


Figure 5.1. Select structures and schematic of operon proteins BPSS0211(1-123), BPSS0211, BPSS0212 and BPSS0213. BPSS0212+M-FormA structure is shown in cartoon with the DUF1842 domain in blue and part of the interdomain linker in black. The structure of both the closed and open helical conformations of BPSS0211 are shown in purple. The two zinc ions in the closed helical structure are shown as grey spheres. Protein structure images generated in PyMOL version 1.1r1, using BPSS0212+M-FormA coordinates and both sets of BPSS0211 coordinates. BPSS0211 coordinates in the presence of zinc were produced by Dr. Matthew Day (Day, 2012).

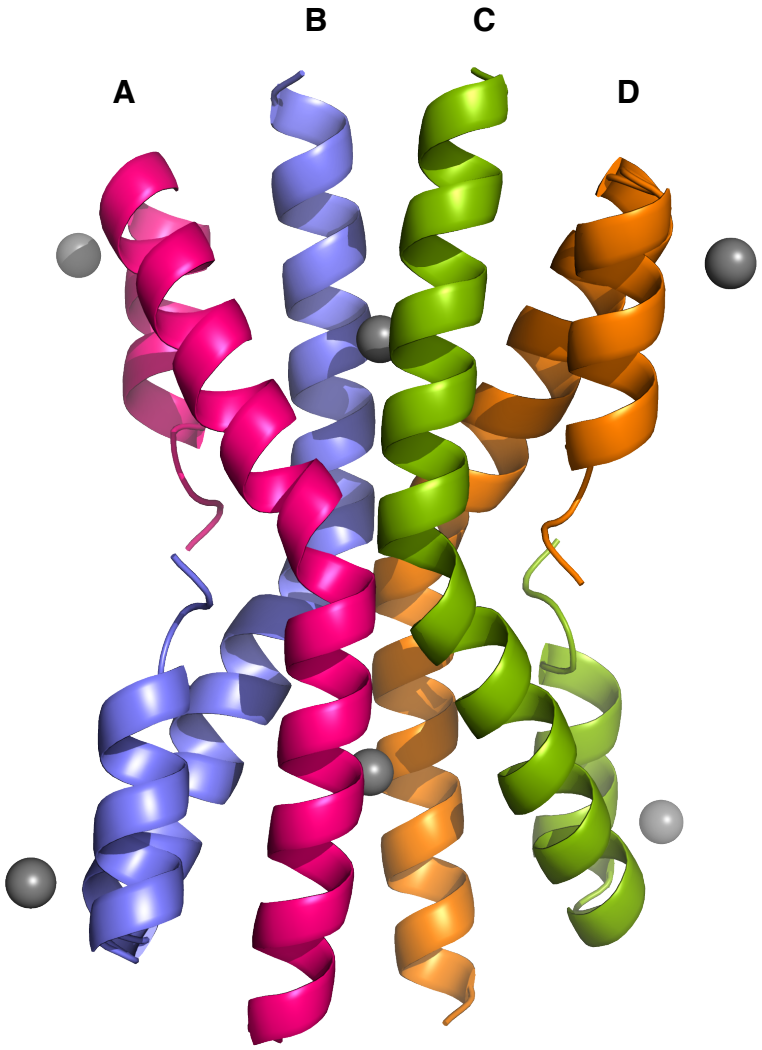


Figure 5.2. BPSS0211 dimer of dimers.
A dimer of dimers of BPSS0211 formed from four monomers. The two dimers are – pink/blue and orange/green. There are two zinc ion-binding sites per monomer, with one site shared across the 2-fold axis. The other site is present on the external surface of the dimer of dimers. The zinc atoms present in the structure are shown in grey. The monomers are labelled A, B, C and D. Image generated in PyMOL version 1.1r1 using BPSS0211 coordinates produced by Dr. Matthew Day (Day, 2012).

showed four protein containing peaks (Figure 5.3 and Figure 5.4). The first peak contains BPSS0212+M and BPSS0213 together, the second peak is BPSS0212+M, and the third is BPSS0211. The fourth peak is a low level contaminant.

The chromatogram and SDS-PAG indicate that BPSS0212+M and BPSS0213 elute together in the same elution fractions, whilst BPSS0211 elutes alone. The absence of any BPSS0211 from any peak with BPSS0212+M or BPSS0213, suggests no interaction with any of the other proteins. It is apparent from the SDS-PAGE analysis of the fractions containing BPSS0212+M and BPSS0213 that the proteins are present in equal amounts.

In gel filtration chromatography, the volume accessible to the protein depends on the protein's size and the size of the holes within the bead matrix (McGuinness, 1973). The gel filtration elution volume can be used to estimate the molecular weight of the protein, by using a K_{av} vs LogMW calibration graph (Figure 5.5). The estimated molecular weight for the potential BPSS0212+M/BPSS0213 complex is ~84 kDa, from an elution volume of ~75 mL. The average molecular weight of BPSS0212+M and BPSS0213 is ~22 kDa, this, along with the predicted ratio of BPSS0212+M and BPSS0213 from SDS-PAGE analysis, suggests the complex might be a $\alpha_2\beta_2$ tetramer, with BPSS0212+M as α and BPSS0213 as β. However gel filtration chromatography cannot be interpreted so simply as running position depends on shape (McGuiness, 1973).

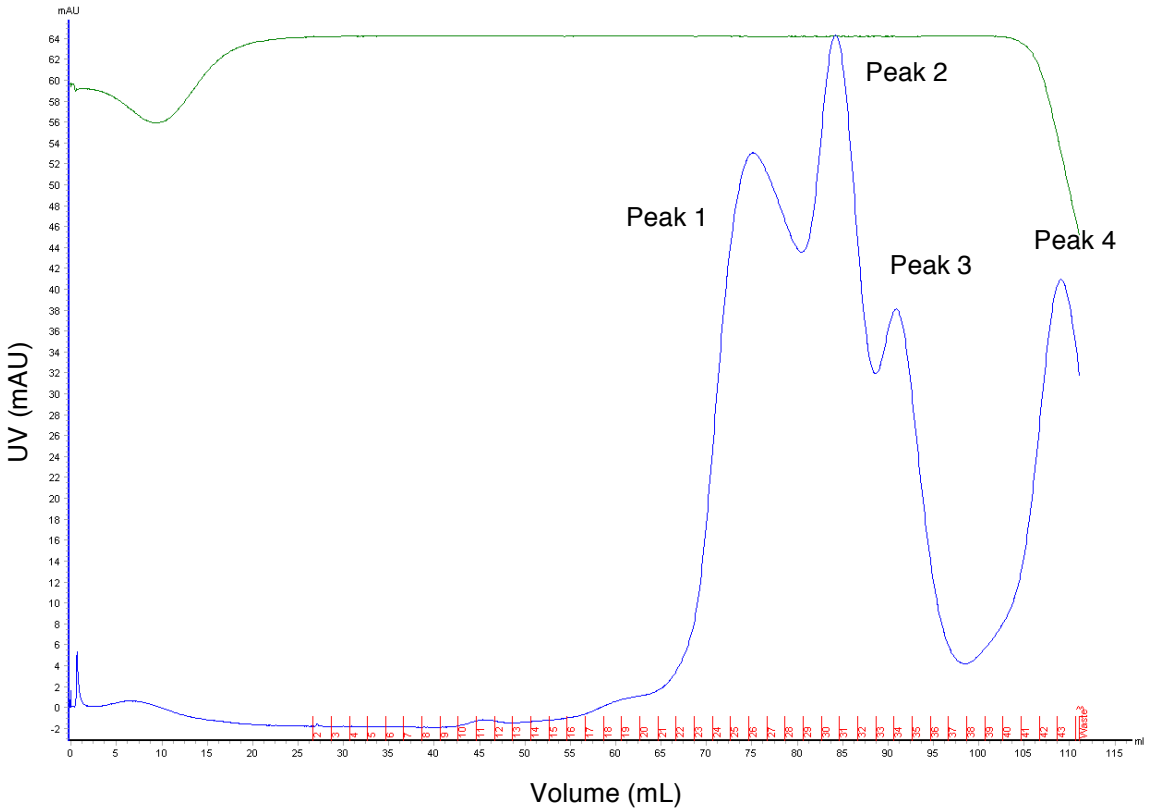


Figure 5.3. Chromatogram of BPSS0211, BPSS0212+M and BPSS0213 gel filtration experiment, run with 50 mM Tris pH 8.0, 500 mM NaCl. mAU at A_{280} – blue, conductivity - green and elution fractions, in 2 mL – red numbers. The four elution peaks are labelled 1-4 and the corresponding elution volumes are noted here. Peak 1 – 74.65 mL, Peak 2 – 84.21 mL, Peak 3 – 91.05 mL, and Peak 4 – 109.10 mL.

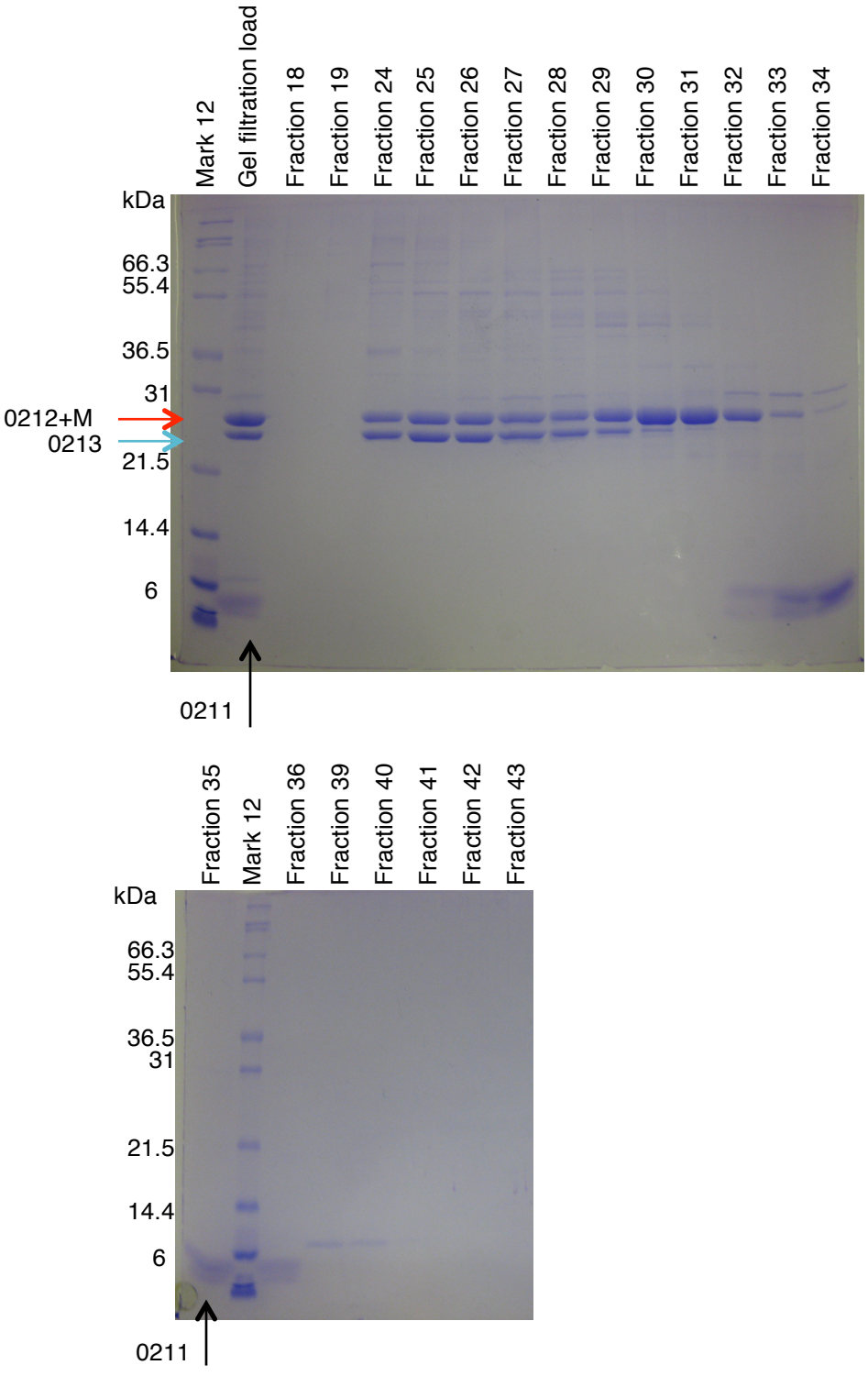


Figure 5.4. SDS-PAGE analysis of BPSS0211, BPSS0212+M and BPSS0213 gel filtration run. BPSS0211 is 6794 Da (black arrow), BPSS0212+M is 22597 Da (red arrow) and BPSS0213 is 21668 Da (blue arrow). Fractions 24-29 - peak 1, fractions 29-32 - peak 2, fractions 33-36 - peak 3 and fractions 39-43 - peak 4. Fractions 24-28 were taken and crystallised. However there is some overlapping of the elution peaks.

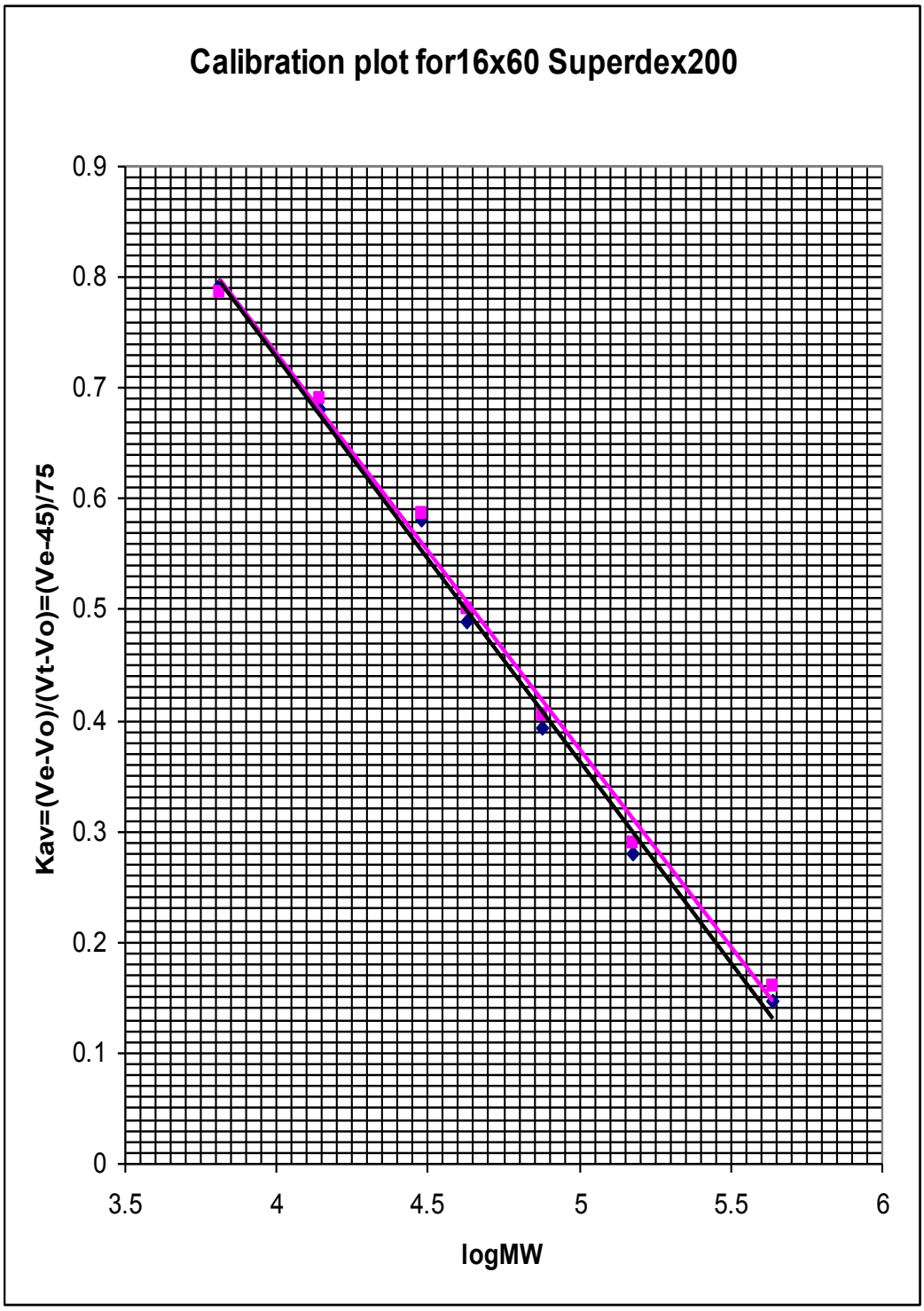


Figure 5.5. Calibration graph for Superdex200 gel filtration column. Calibration graph generated by Dr. Svetlana Sedelnikova (University of Sheffield), using seven proteins – RibonucleaseA, Ovalbumin, Aldolase, Ferritin, Aprotinin, Carbonic anhydrase and Conalbumin. The pink line of best fit was used.

5.1.3 Degradation of BPSS0212+M is reduced in complex with BPSS0213

When BPSS0212+M is purified using gel filtration chromatography, the protein is seen to degrade which produces two distinct fragments and a number of smaller degradation products (Chapter 4). The load samples and elution fractions from the BPSS0211/BPSS0212+M/BPSS0213 gel filtration run were analysed by SDS-PAGE, 15 and 41 days post experiment (Figure 5.6). The samples were stored at ~4°C during this time. The SDS-PAGE analysis shows that over the two time points, BPSS0212+M degrades more rapidly when alone, and this degradation is reduced when BPSS0212+M is in the presence of BPSS0213. This suggests that the part of BPSS0212+M protein that is susceptible to degradation is involved in complex formation with BPSS0213. In the SDS-PAGE analysis, the gel filtration chromatography fractions containing BPSS0212+M alone, compared to when it is purified in the absence of other operon proteins, also shows a reduction in degradation. This may be due to a dynamic equilibrium that takes place between BPSS0212+M present in the complex and excess free BPSS0212+M. Therefore the temporary binding of free BPSS0212+M to BPSS0213 offers protection from degradation. The structure of BPSS0212+M-FormA, (Figure 5.1), only contains the DUF1842 domain residues with an N-terminal extension and part of the interdomain linker, but not the DUF1843 domain, potentially due to degradation of BPSS0212+M. Therefore the DUF1843 domain may be involved in forming the interaction between BPSS0212+M and BPSS0213 and is thus protected from degradation.

5.1.4 Initial crystallisation of BPSS0212+M/BPSS0213 complex yields crystals of BPSS0212+M

The potential BPSS0212+M/BPSS0213 complex was concentrated to ~8 mg/mL and was set down in sitting-drop crystallisation trials. The crystallisation screens used were PACT, JCSG+, Proplex and PEGs, each consisting of 96 crystallisation conditions. Crystal hits were found in

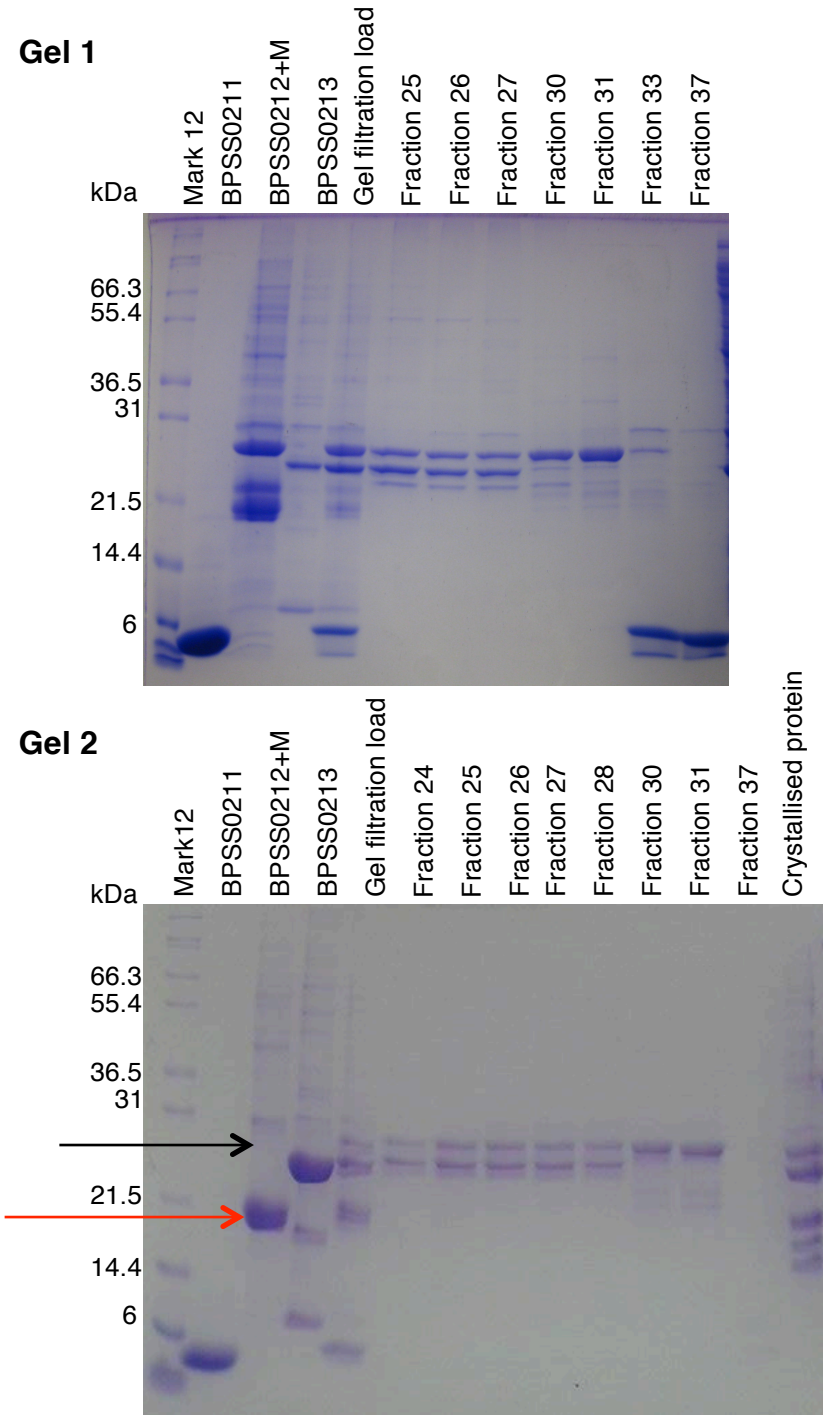


Figure 5.6. SDS-PAGE analysis of time-dependent storage of BPSS0212+M alone and in complex with BPSS0213. BPSS0211 is 6794 Da, BPSS0212+M is 22597 Da and BPSS0213 is 21668 Da. **Gel 1** analyses gel filtration run samples 15 days post-complex formation experiment. Fractions 25-27 - peak 1, fractions 30-31 - peak 2 and fractions 33-37 - peak 3. **Gel 2** analyses the gel filtration run samples 41 days post-complex formation experiment. Fractions 24-28 - peak 1, fractions 30-31 - peak 2 and fraction 37 - peak 3. The red arrow highlights degraded BPSS0212+M, and the black arrow highlights where the full-length BPSS0212+M is predicted to be.

all four trays. Crystals were sent to the Diamond Light Source Synchrotron, Oxford to be screened. The data collected were solved and found to contain only BPSS0212+M. These structures are discussed in section 4.6.

5.2 Investigation into possible complex formation between BPSS0211/BPSS0212 and BPSS0211/BPSS0213

During the initial complex formation experiment all three proteins - BPSS0211, BPSS0212+M and BPSS0213 - were run down the gel filtration column together. The formation of the complex between BPSS0212+M and BPSS0213 may have blocked any potential interaction between BPSS0211 with either BPSS0212+M or BPSS0213 separately. Therefore BPSS0211 and BPSS0212, and BPSS0211 and BPSS0213 were mixed separately and run down the gel filtration column pre-equilibrated with 50 mM Tris pH 8.0, 500 mM NaCl, in two separate experiments. In the initial complex experiment (described in section 5.1) BPSS0212+M was used, but following this BPSS0212 was recloned to remove the extra N-terminal methionine (section 2.3.6).

5.2.1 BPSS0211 does not form a complex with either BPSS0212 or BPSS0213

SDS-PAGE analysis of the protein containing peaks from both gel filtration experiments shows that BPSS0211 does not elute in the same fractions as BPSS0212 or BPSS0213 (Figures 5.7 and 5.8). BPSS0211 elution corresponds to peak 2 in both gel filtration experiments, at a volume of ~87 mL, at around the same volume as when it is purified alone. Therefore this suggests that BPSS0211 does not interact with either BPSS0212 or BPSS0213. However this may have been due to the interactions between the monomers of BPSS0211 (Figure 5.2), which then blocks any possible interactions that may form between BPSS0211 and other proteins.

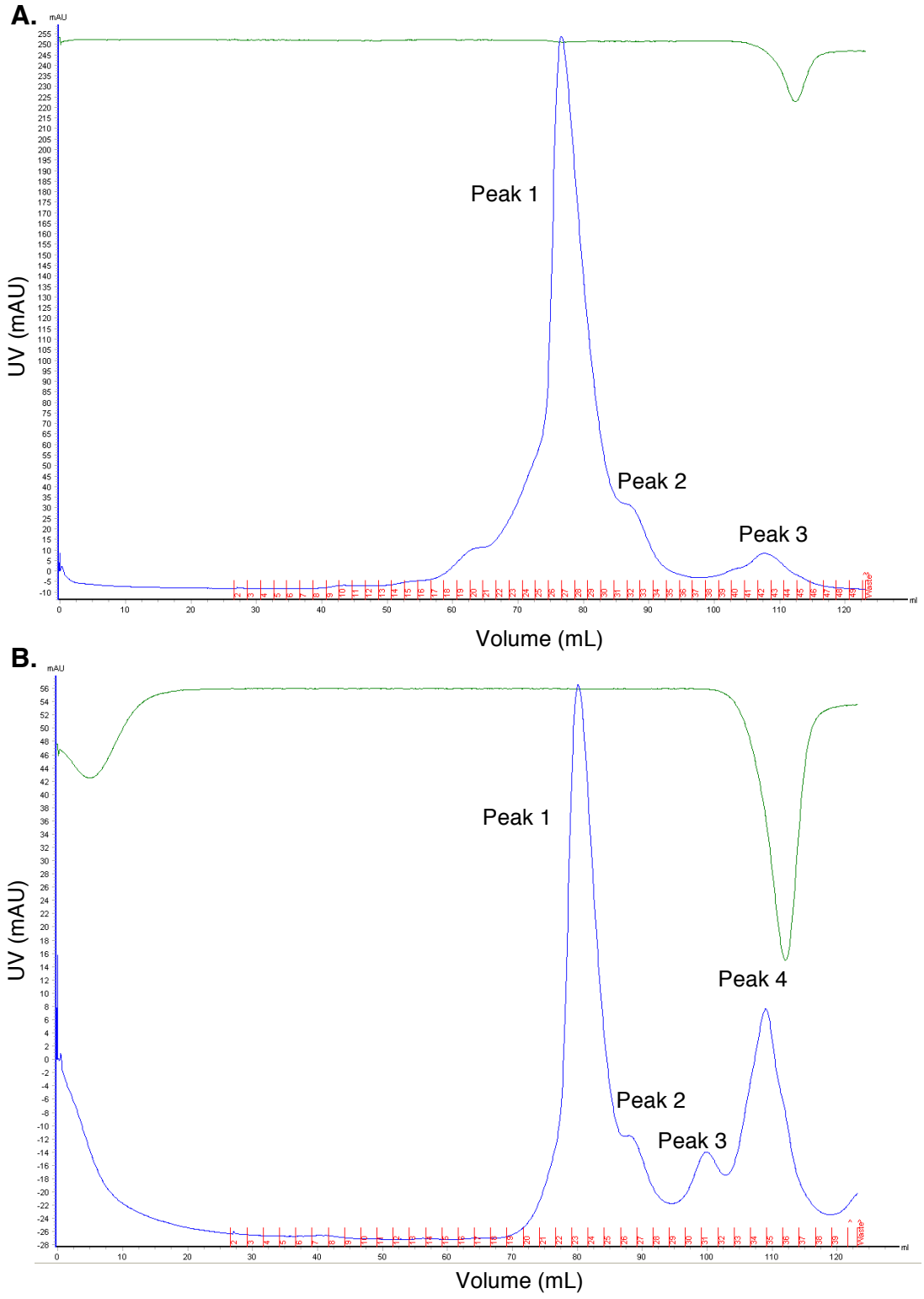


Figure 5.7. Chromatograms of BPSS0211 and BPSS0212 gel filtration experiment and BPSS0211 and BPSS0213 gel filtration experiment. mAU at A_{280} – blue, conductivity – green, elution fractions – red numbers. The elution peaks are labelled and the corresponding elution volumes noted. **A.** BPSS0211 and BPSS0212 - Peak 1 – 76.64 mL, Peak 2 – ~87 mL and Peak 3 – 107.77 mL. **B.** BPSS0211 and BPSS0213 - Peak 1 – 80.13 mL, Peak 2 – ~88 mL, Peak 3 – 99.85 mL, and Peak 4 – 109.02 mL.

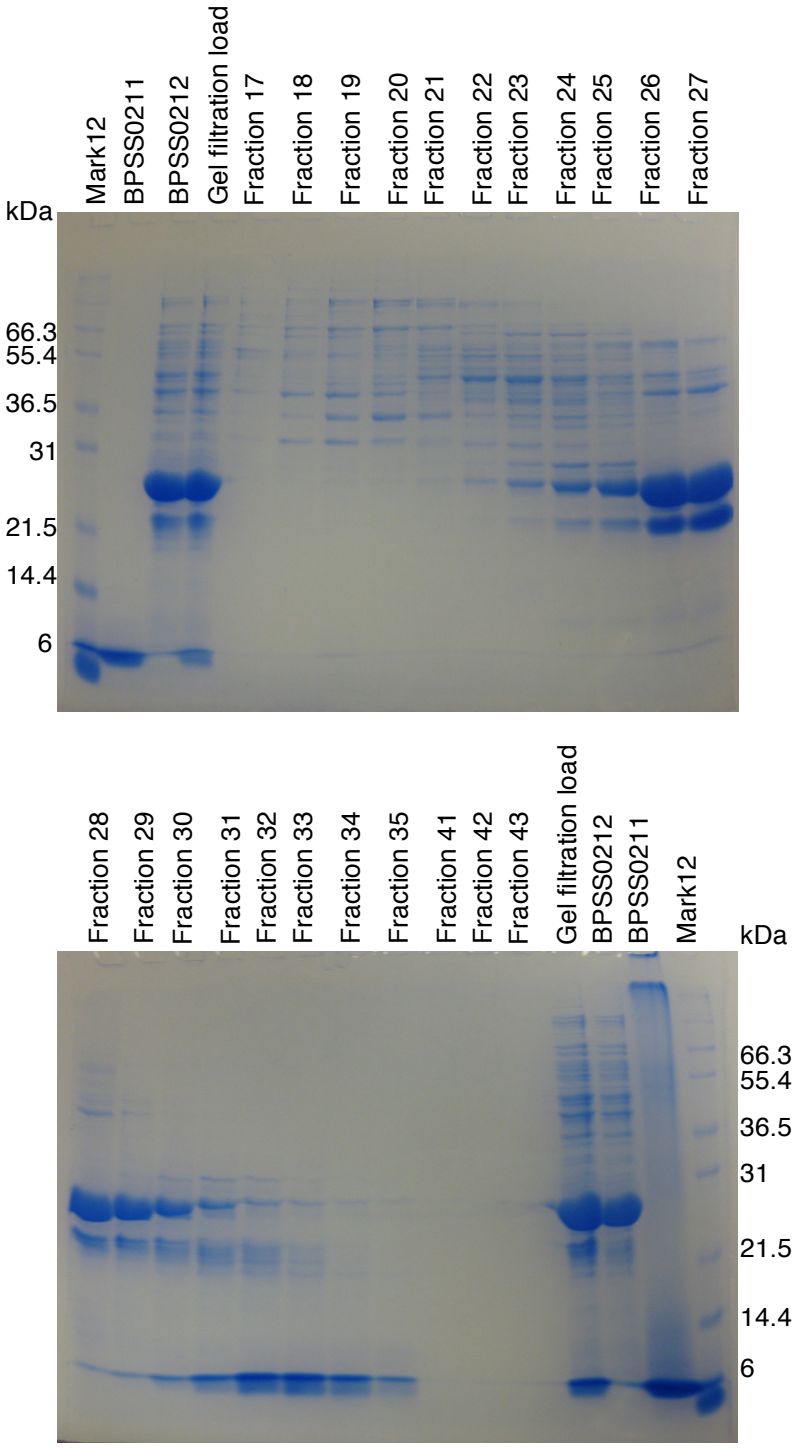


Figure 5.8a. SDS-PAGE analysis of the BPSS0211 and BPSS0212 gel filtration run. BPSS0211 is 6794 Da and BPSS0212 is 22465 Da. Fractions 21-30 - peak 1 and fractions 30-35 - peak 2. However there is some overlapping of the elution peaks. Degradation of BPSS0212 can be seen in fractions 23-33.

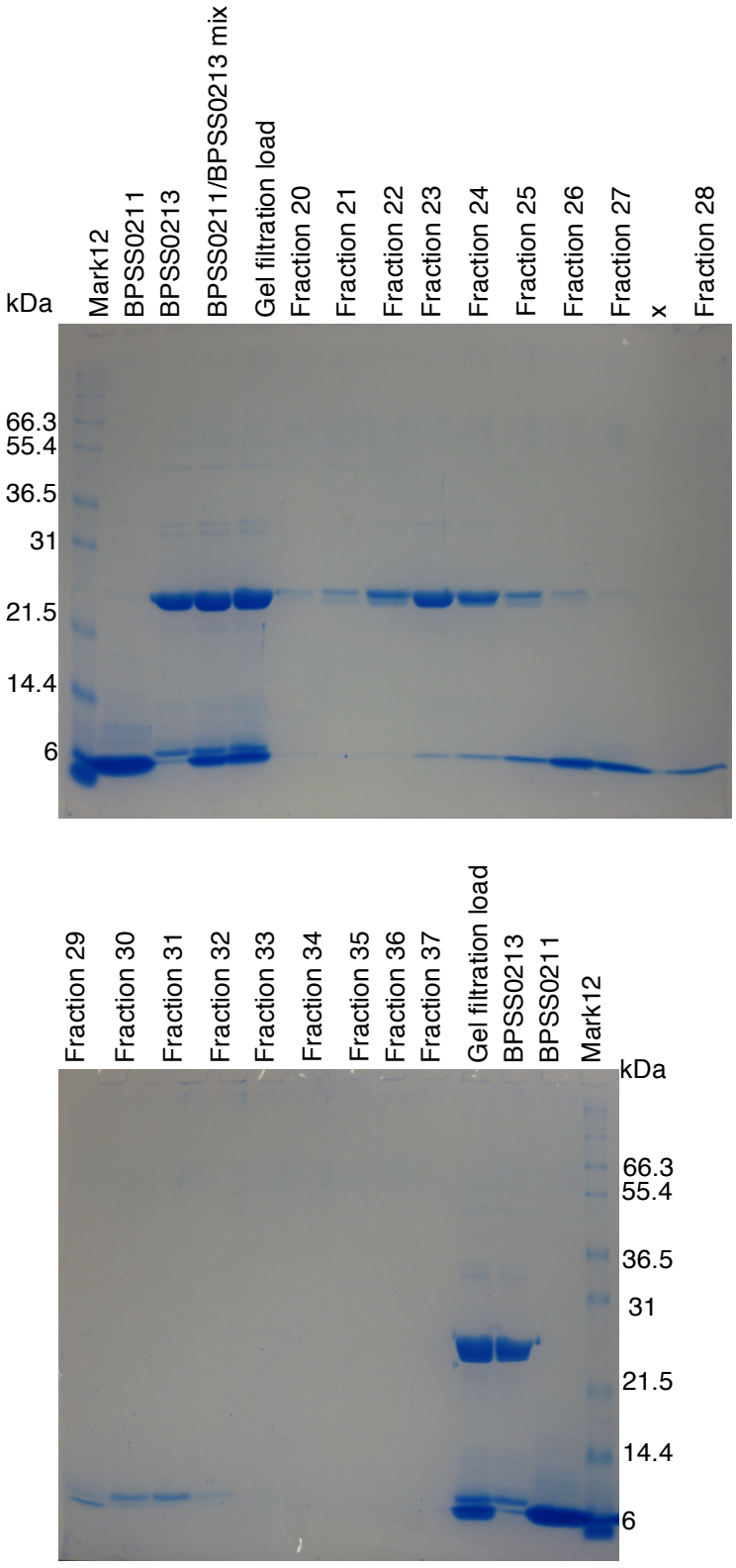


Figure 5.8b. SDS-PAGE analysis of the BPSS0211 and BPSS0213 gel filtration run. BPSS0211 is 6794 Da and BPSS0213 is 21668 Da. x indicates an empty lane. Fractions 20-25 - peak 1, fractions 25-29 - peak 2 and fractions 29-32 - peak 3. However there is some overlapping of the elution peaks.

5.3 Complex formation experiment with BPSS0212 and BPSS0213

The previous complex formation experiments show the predicted complex contains only BPSS0212 and BPSS0213 protein. The experiment was repeated but without BPSS0211. The complex was reproduced by a gel filtration run of combined purified BPSS0212 and BPSS0213. In order to compare more accurately the gel filtration elution profile of the complex versus the gel filtration elution profiles of the individual proteins involved, BPSS0212 and BPSS0213 were run separately on the same gel filtration column, under the same conditions as the complex formation experiment.

5.3.1 BPSS0212/BPSS0213 complex formation can be reproduced

The control gel filtration experiments for each protein individually show that BPSS0212 elutes at ~78 mL and BPSS0213 elutes at ~78 mL (Figure 5.9). The predicted molecular weight of both BPSS0212 and BPSS0213 from the control run was ~35 kDa, calculated using the K_{av} vs LogMW calibration graph shown in Figure 5.10. The molecular weight of BPSS0212 and BPSS0213 is 22465 Da and 21668 Da, respectively (ProtParam – Gasteiger et al., 2005). It is not clear whether the individual proteins are running as a monomer or a homodimer.

BPSS0212 and BPSS0213 were combined and loaded onto a gel filtration column pre-equilibrated with 50 mM Tris pH 8.0, 500 mM NaCl, as detailed in section 2.7.4. The gel filtration elution profile for both BPSS0212 and BPSS0213 together has two protein containing peaks (Figure 5.11). Analysis of the peak fractions by SDS-PAGE confirmed that the first peak contains BPSS0212 and BPSS0213, whilst the second peak is excess BPSS0212 (Figure 5.12). The peak elution volumes for the complex formation compared to each of the individual protein gel filtration elution profiles confirms a shift in elution volume, from ~78 mL to ~67 mL. The BPSS0212/BPSS0213 complex, at ~67 mL, has a predicted molecular weight of ~89 kDa (Figure 5.10), with the average BPSS0212 and BPSS0213 molecular weight of ~22 kDa. The size of the complex,

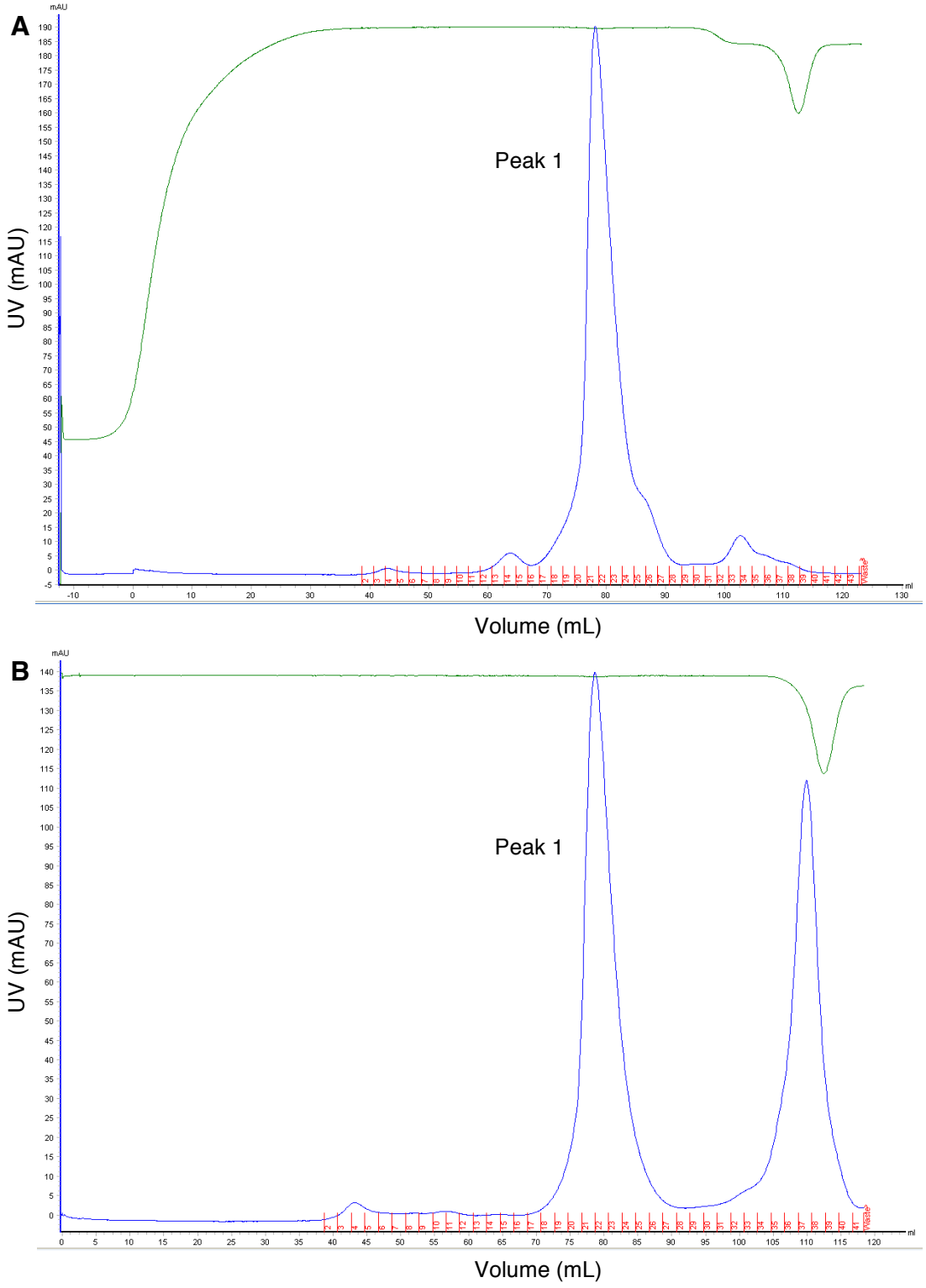


Figure 5.9. Chromatograms of gel filtration control runs of BPSS0212 and BPSS0213 separately, run with 50 mM Tris pH 8.0, 500 mM NaCl. mAU at A_{280} – blue, conductivity – green, elution fractions – red numbers. The elution peaks are labelled and the corresponding elution volumes noted. **A.** BPSS0212, Peak 1 – 78.14 mL. **B.** BPSS0213, Peak 1 – 78.66 mL.

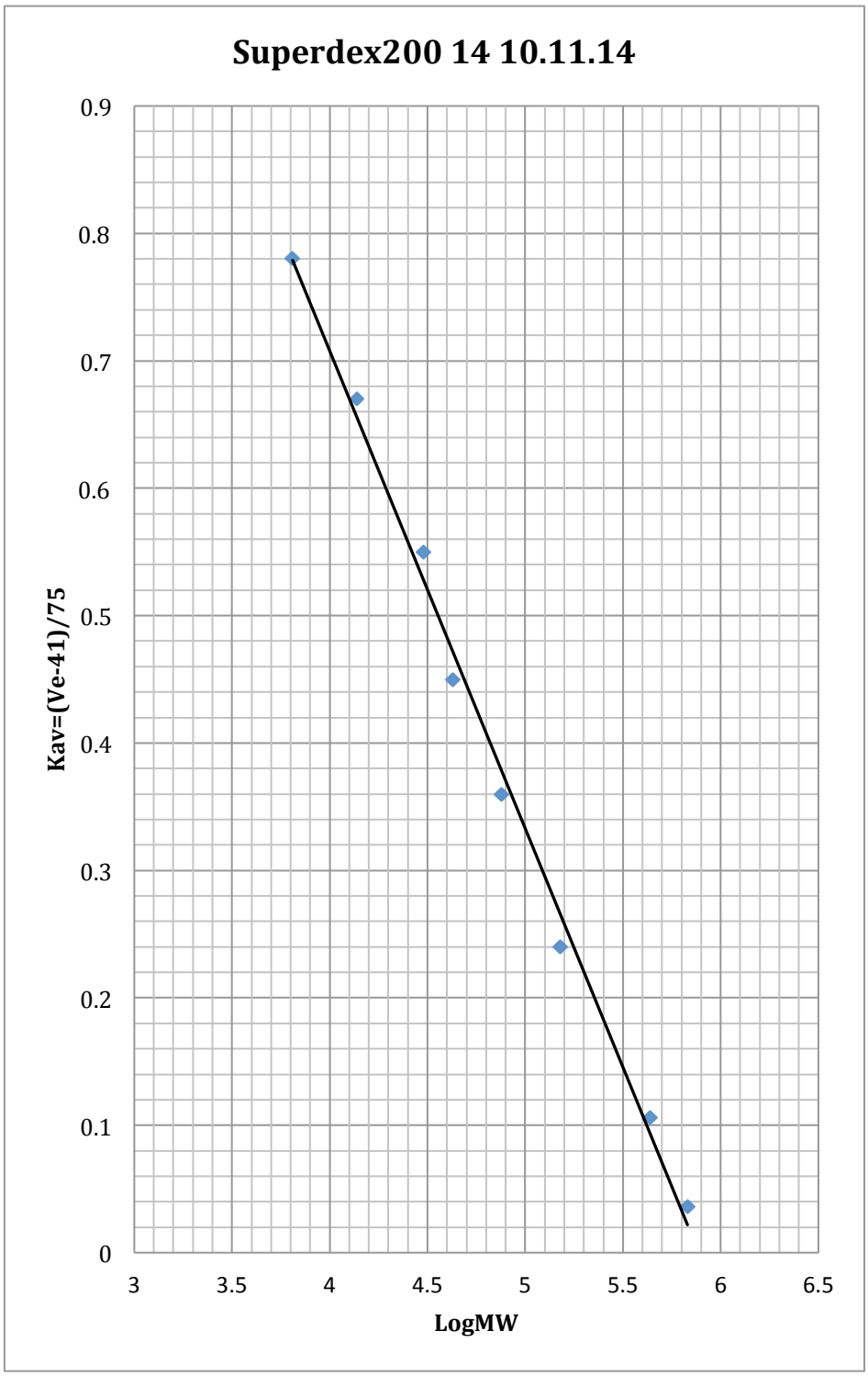


Figure 5.10. Calibration graph for Superdex200 gel filtration column. Calibration graph generated by Dr. Svetlana Sedelnikova (University of Sheffield) using eight proteins – RibonucleaseA, Ovalbumin, Aldolase, Ferritin, Aprotinin, Carbonic anhydrase, Conalbumin and Thyroglobulin.

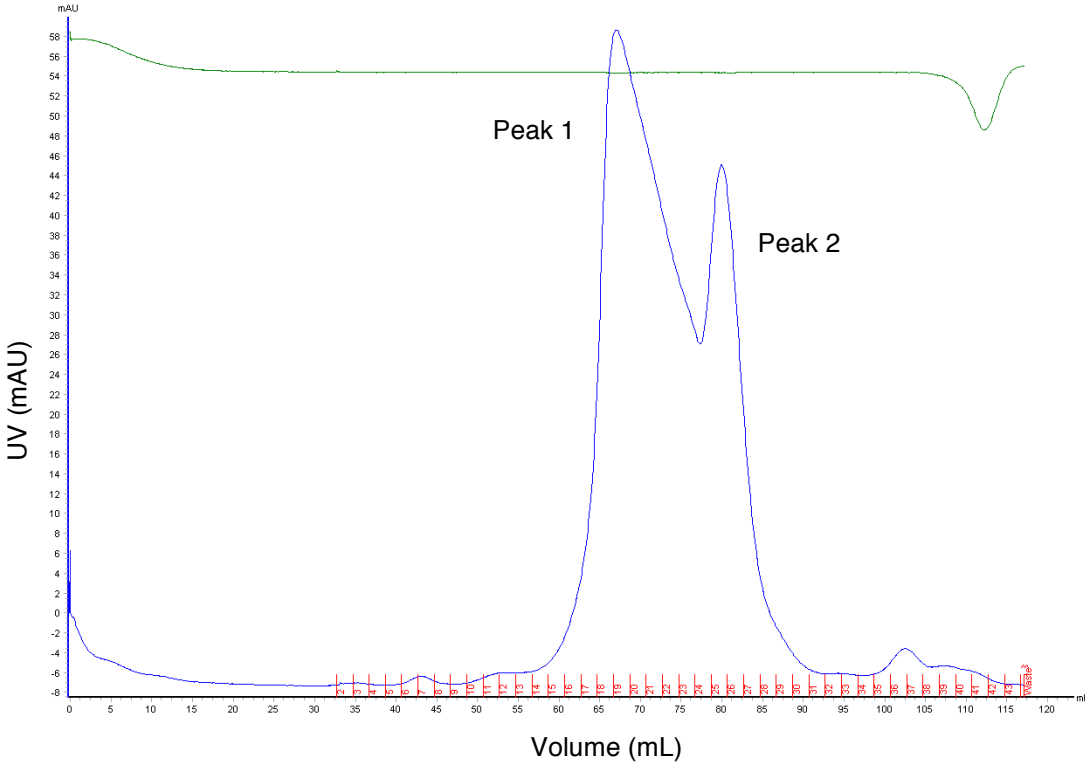


Figure 5.11. Chromatogram of BPSS0212 and BPSS0213 gel filtration experiment, run with 50 mM Tris pH 8.0, 500 mM NaCl. mAU at A_{280} – blue, conductivity – green, elution fractions – red numbers. The elution peaks are labelled and the corresponding elution volumes noted. Peak 1 – 67.0 mL and Peak 2 – 79.99 mL.

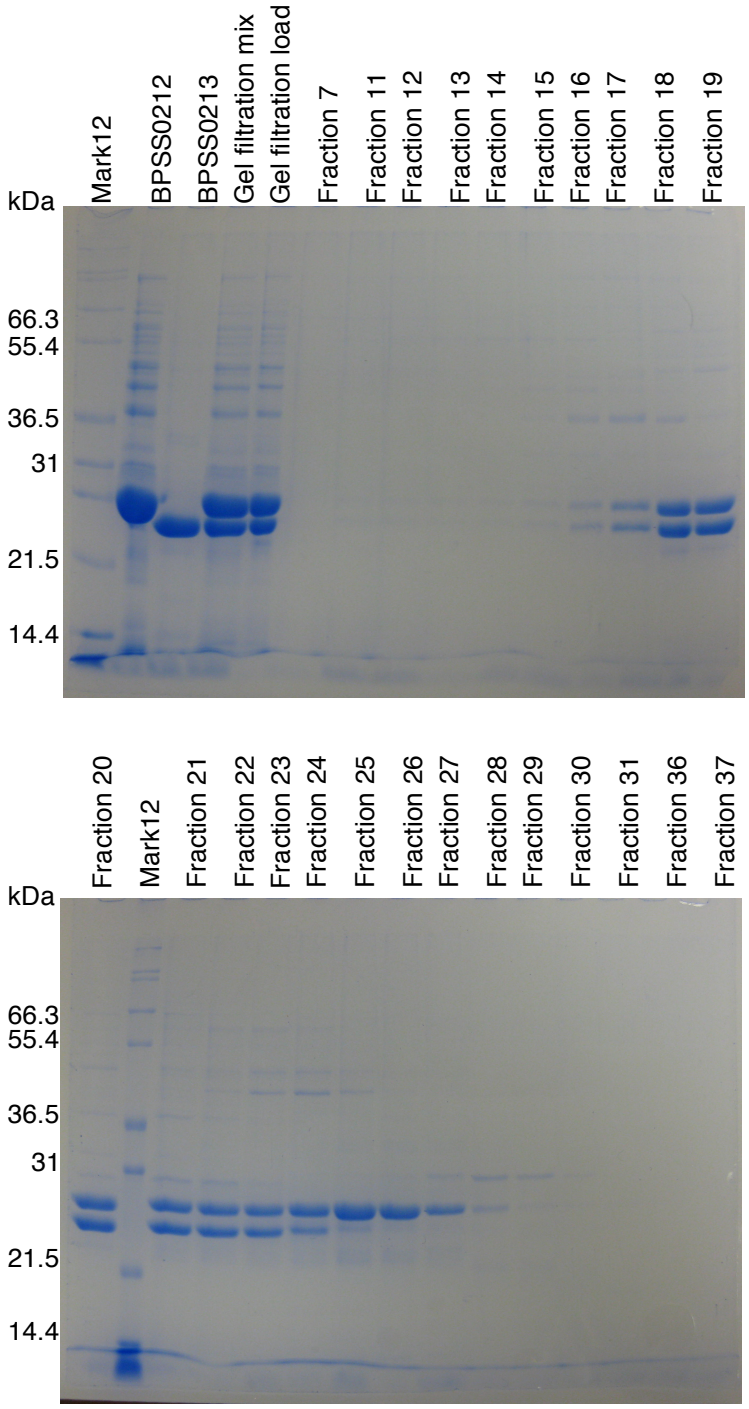


Figure 5.12. SDS-PAGE analysis of the BPSS0212 and BPSS0213 gel filtration run. BPSS0212 is 22465 Da and BPSS0213 is 21668 Da. Fractions 14-23 - peak 1 and fractions 24-30 - peak 2. Fractions 19-22 were taken and crystallised. However there is some overlapping of the elution peaks.

formed here, ~89 kDa, is similar to that in the previous experiment (~84 kDa), predicting ~four molecules in the complex. The difference in complex size between the initial experiment and this repeat experiment may be due to the difference in gel filtration column void volume. These results indicate that the complex of BPSS0212/BPSS0213 identified previously can be reproduced.

5.3.2 Crystallisation of BPSS0212/BPSS0213 complex

The BPSS0212/BPSS0213 complex was set down in sitting-drop crystallisation trials, at ~10 mg/mL in 10 mM Tris pH 8.0, 100 mM NaCl, using screens PACT, Proplex, and JCSG+. PACT and Proplex trays were set up with 200 nL:200 nL drops. Column 1 in the JCSG+ screen was set up with 1 µL:1 µL drops and columns 2-7 had 700 nL:700 nL drops. Plates were stored at 17°C. Crystal hits were found in a number of conditions, Figure 5.13, the largest crystals were in the Proplex screen, condition B4 - 0.2 M NaCl, 0.1 M MES pH 6.5, 10% (w/v) PEG4000 (Figure 5.13A).

5.3.3 Data collection of crystals formed from BPSS0212/BPSS0213 protein

Crystals grown in Proplex B4 sitting-drop crystallisation condition, were cryoprotected in ethylene glycol, cooled in liquid nitrogen and sent to the Diamond Light Source Synchrotron, Oxford, beamline I02 for screening. Native data were collected over 180°, with a 0.1° phi oscillation, 0.05 second exposure and transmission of 40% (Figure 5.14). The data were processed in xia2 3d (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Sauter et al., 2004; Winter, 2010; Zhang et al., 2006), to a resolution of 3.15 Å and spacegroup I4 with cell dimensions of a=b=117.6 Å, c=120.5 Å and α=β=γ=90° (Table 5.1). The Matthews calculation predicted four molecules, of 22 kDa, in the asymmetric unit, with a solvent content of 48% (Matthews, 1968; Kantardjieff and Rupp, 2003).

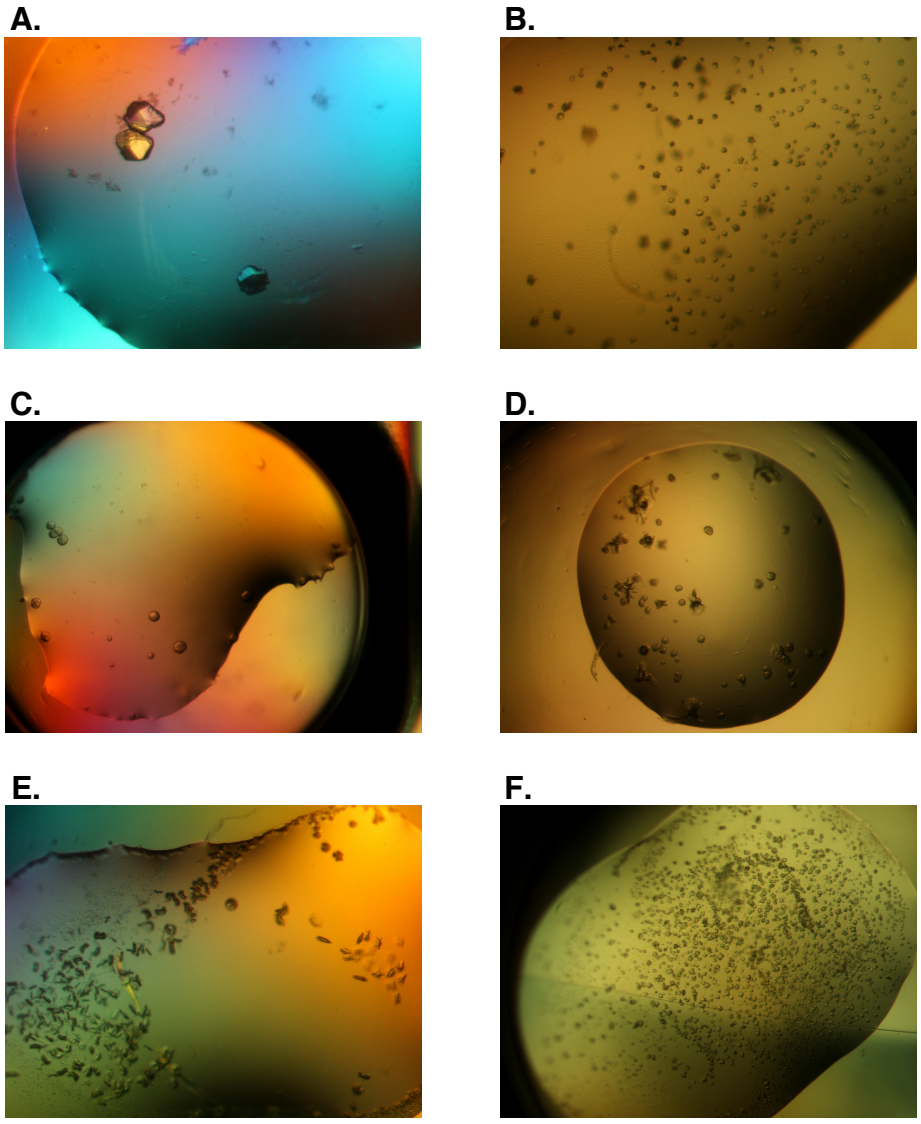
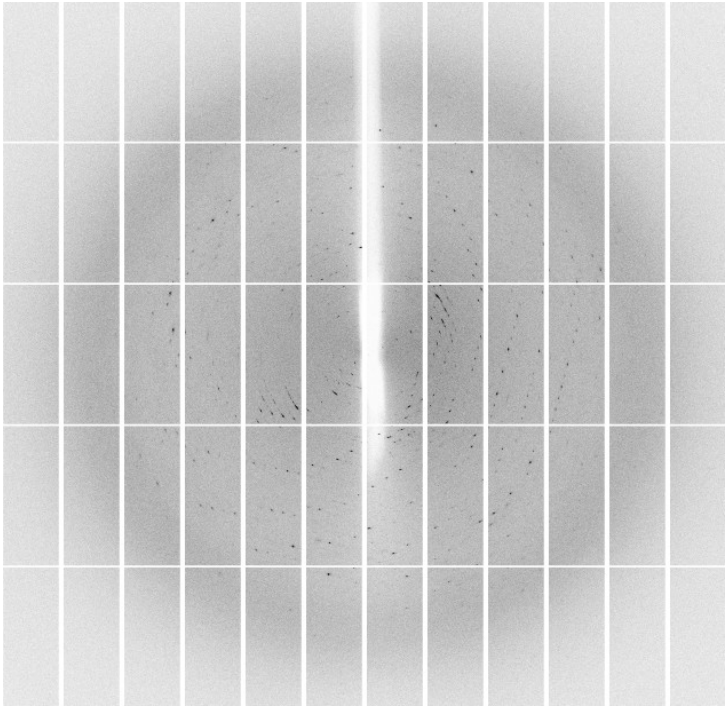


Figure 5.13. Crystal hits from sitting-drop crystallisation trials of BPSS0212/BPSS0213 complex. Protein set down at ~10 mg/mL in 10 mM Tris pH 8.0, 100 mM NaCl. **A.** Proplex B4 – 0.2 M NaCl, 0.1 M MES pH 6.5, 10% (w/v) PEG4000. **B.** Proplex F3 – 0.1 M magnesium acetate, 0.1 M MES pH 6.5, 10% (w/v) PEG1000. **C.** PACT D4 – 0.1 M MMT buffer pH 7.0, 25% (w/v) PEG1500. **D.** Proplex E7 – 0.1 M sodium phosphate pH 6.5, 12% (w/v) PEG8000. **E.** Proplex C1 - 0.15 M ammonium sulphate, 0.1 M Tris pH 8.0, 15% (w/v) PEG4000. **F.** JCSG+ A2 – 0.1 M tri-sodium citrate pH 5.5, 20% (w/v) PEG3000.

A.



B.

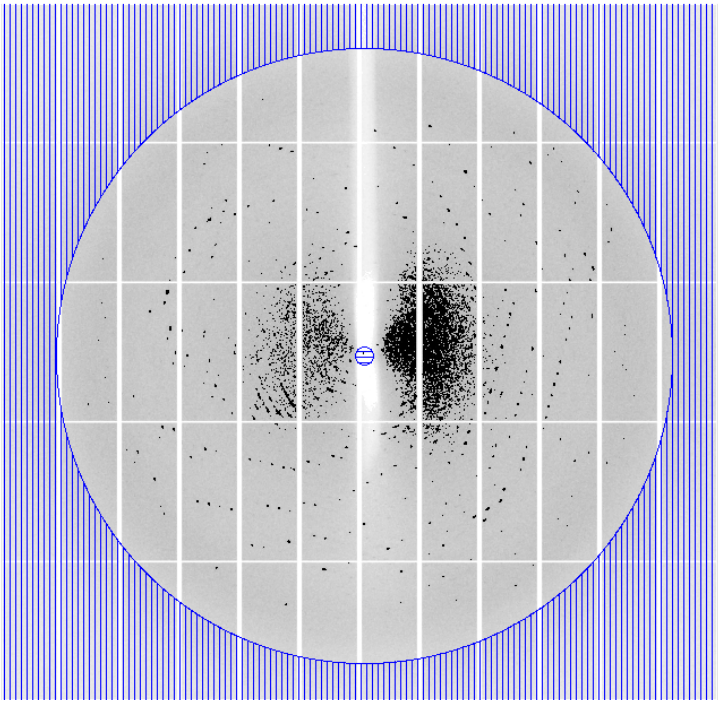


Figure 5.14. Diffraction test image of a crystal of BPSS0212/BPSS0213 from condition Proplex B4. A. Diffraction test image taken prior to data collection, with the edge of the detector plate set to 2.8 Å. Image generated from Diamond Light Source JPEGS. **B.** The lower image shows a resolution ring set to 3.15 Å. Image generated in iMosflm (Battye et al., 2011), and collected at the Diamond Light Source Synchrotron, Oxford.

Data collection	**BPSS0212/BPSS0213 (Complex 5.3)**
Wavelength (Å)	0.97946
Resolution range (Å)	60.26-3.15 (3.23-3.15)
Space group	I 4
Unit cell parameters	
a (Å)	117.6
b (Å)	117.6
c (Å)	120.5
α (°)	90
β (°)	90
γ (°)	90
Total reflections	97667 (7303)
Unique reflections	14251 (1047)
Multiplicity	6.9 (7.0)
Completeness (%)	99.9 (100.0)
Mean I/σ	19.7 (2.5)
Wilson B factor ($Å^2$)	74
R_{merge}	0.081 (0.800)
R_{pim} (I)	0.037 (0.353)

Table 5.1. Data collection processing statistics for BPSS0212/BPSS0213 complex 5.3. xia2 3d data processing statistics for BPSS0212/BPSS0213 native collected data (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Sauter et al., 2004; Winter, 2010; Zhang et al., 2006). The high-resolution shell information is shown in brackets.

5.3.4 Structure determination of the BPSS0212/BPSS0213 complex by molecular replacement

The native complex data was phased by molecular replacement using the program PhaserMR (McCoy et al., 2007) with the BPSS0212+M-FormA protein structure as the search model. PhaserMR (McCoy et al., 2007) produced a single solution, with four monomers of BPSS0212+M-FormA modelled in the asymmetric unit. The four monomers are arranged into two groups of two monomers.

If the crystals produced from the complex of BPSS0212/BPSS0213 protein contained both BPSS0212 and BPSS0213, with four molecules in the asymmetric unit, it would be expected that there would be two molecules of BPSS0212 and two molecules of BPSS0213. In order to confirm that this is the case and to determine which of the protein chains were BPSS0212 and which were BPSS0213, the electron density surrounding the modelled BPSS0212+M-FormA residues was analysed in order to identify any differences between the model and the 2Fo-Fc map density and any peaks in the Fo-Fc map, which might indicate the misassignment of residues.

BPSS0212 and BPSS0213 amino acid sequences were aligned using Jalview (Waterhouse et al., 2009), and the sidechains with significant structural differences were identified (Figure 5.15). Figures 5.16 and 5.17 show two examples of the specific residues identified and the difference seen in surrounding electron density between the four asymmetric unit chains. The results show that two of the molecules, chain A and B, have electron density that fits the residues in BPSS0212. Whereas, chains C and D suggest that the BPSS0213 residues would be a better fit for the density.

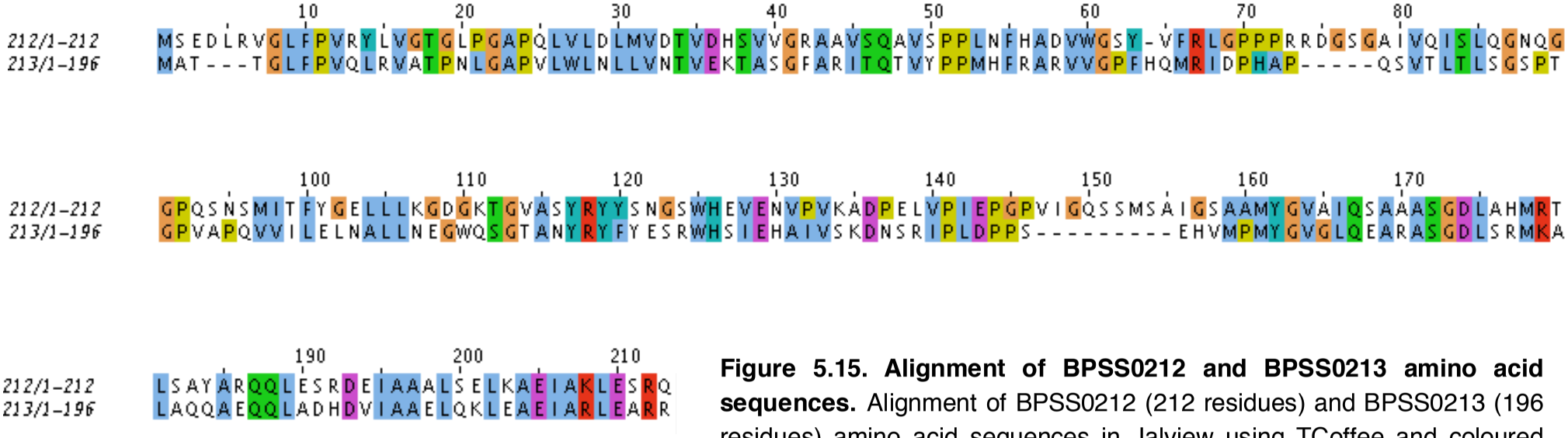


Figure 5.15. Alignment of BPSS0212 and BPSS0213 amino acid sequences. Alignment of BPSS0212 (212 residues) and BPSS0213 (196 residues) amino acid sequences in Jalview using TCoffee and coloured using Custalx (Larkin et al., 2007; Notredame et al., 2000; Waterhouse et al., 2009). Amino acid changes from small sidechains in BPSS0212 to larger sidechains in BPSS0213 include V27-W24, A44-R41, S50-Y47, G109-W102 and S120-Y113. Amino acid changes from larger sidechains in BPSS0212 to smaller sidechains in BPSS0213 include Y14-L11, W60-V57, F65-M63, F99-L92, and Y100-E93.

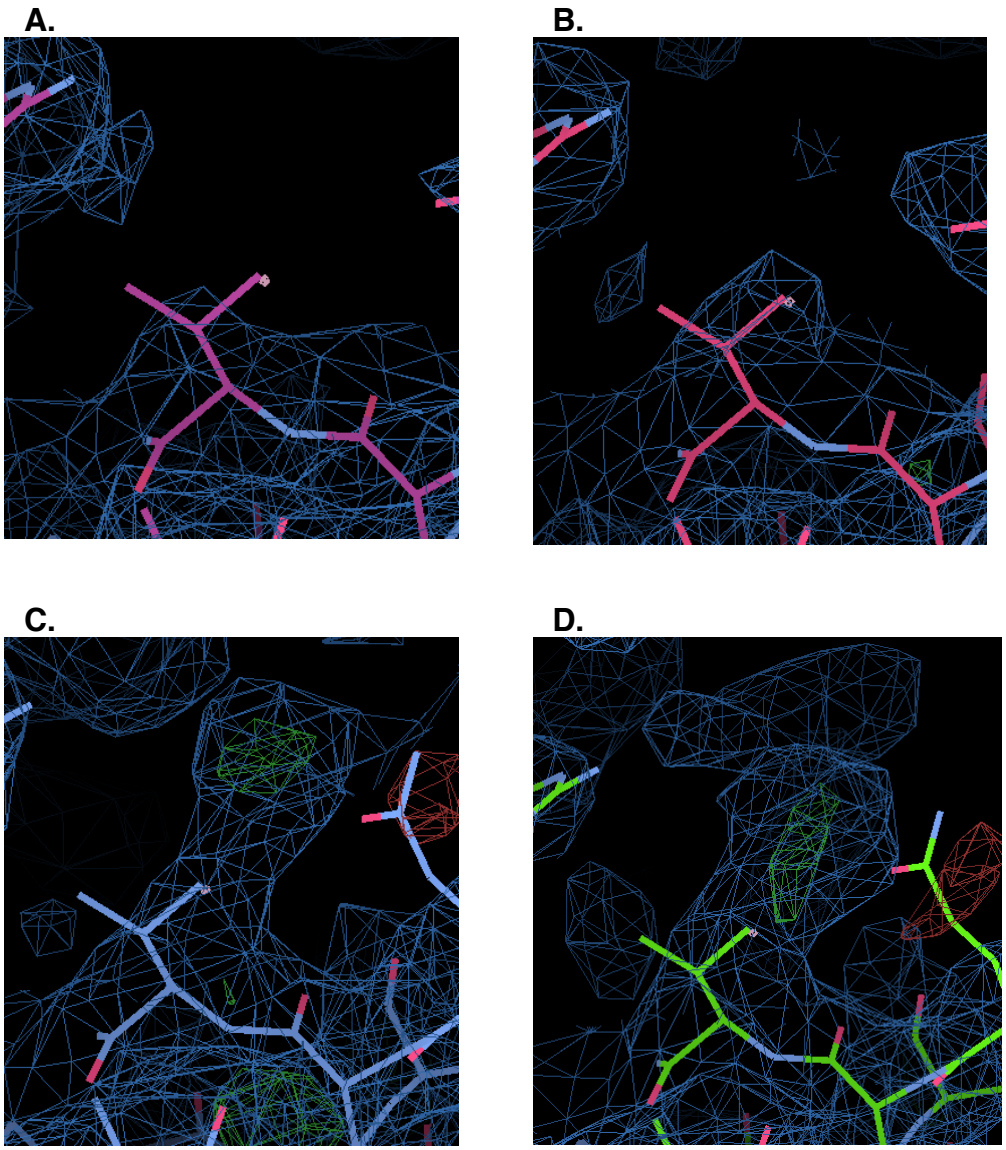


Figure 5.16. Analysis of the electron density around sidechain Valine27 in all four chains of BPSS0212+M-FormA in the asymmetric unit. 2Fo-Fc map – blue, Fo-Fc map – green/red, chains A, B, C and D are coloured – purple, pink, blue and yellow, respectively. V27 is BPSS0212 numbering. In the sequence alignment carried out between BPSS0212 and BPSS0213, the residue at this position is V27 in BPSS0212 and W24 in BPSS0213. The positive Fo-Fc density surrounding V27 in chains C and D suggests the residue occupying this electron density should be larger. In contrast the valine residues in chains A and B fill the electron density sufficiently. Images generated in *Coot*, with map level = 0.1539e/A^3 (1.01rmsd) (Emsley et al., 2010).

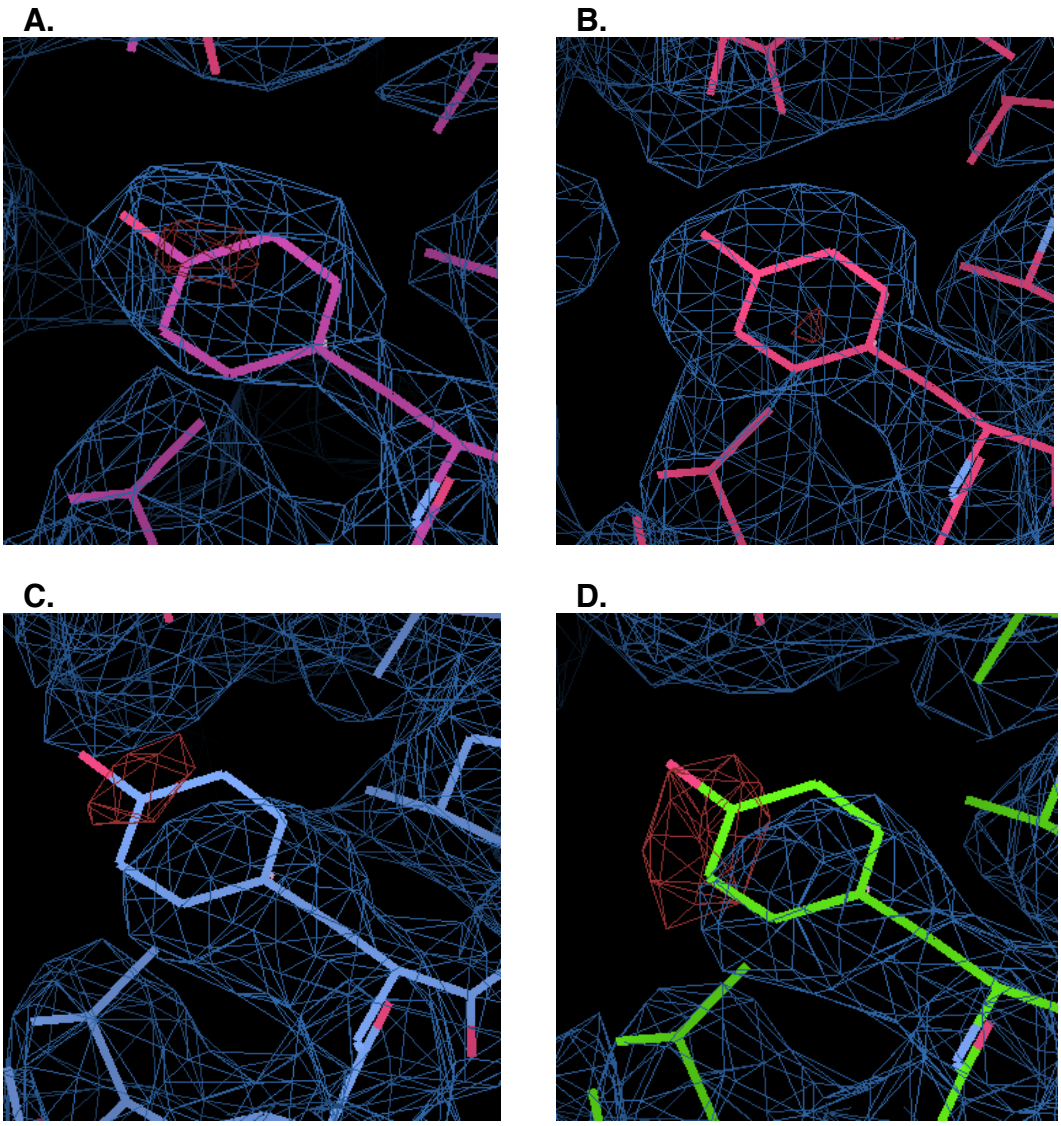


Figure 5.17. Analysis of the electron density around sidechain Tyrosine14 in all four chains of BPSS0212+M-FormA in the asymmetric unit. 2Fo-Fc map – blue, Fo-Fc map – green/red, chains A, B, C and D are coloured – purple, pink, blue and yellow, respectively. Y14 is BPSS0212 numbering. In the sequence alignment carried out between BPSS0212 and BPSS0213, the residue at this position is Y14 in BPSS0212 and L11 in BPSS0213. The negative Fo-Fc density surrounding Y14 in chains C and D suggests the residue occupying this electron density should be smaller. In contrast the tyrosine residues in chains A and B fill the electron density sufficiently. Images generated in *Coot*, with map level = 0.1539e/A^3 (1.01rmsd) (Emsley et al., 2010).

5.3.5 Structure building and refinement of initial BPSS0212/BPSS0213 complex structure

The structure was built in *Coot* (Emsley et al., 2010) and underwent rounds of refinement against native data, to 3.15 Å, using Refmac5 (Murshudov et al., 1997; Murshudov et al., 2011). The initial model of the BPSS0212/BPSS0213 complex contains four chains – two chains of BPSS0212 and two chains of BPSS0213 (Figure 5.18). The residues modelled (BPSS0212 – residues 7-145 and BPSS0213 – residues 5-132/133) in both proteins correspond to the DUF1842 domains and part of the interdomain linker between the DUF1842 domain and DUF1843 domain. In all four subunits there are a number of unmodelled residues due to lack of electron density. For both BPSS0212-A and -B residues 1-6 and 67-77 could not be built, with loop residue 78 not built in BPSS0212-A and residues 19, 20, 93 and 94 not present in BPSS0212-B. In both BPSS0213 molecules the first four residues and residues 67-69 and 84-87 are not present, with residues 64-66, 83, 88 and 133 not modelled in BPSS0213-A. The model R_{factor} is 0.2465 and R_{free} is 0.3162, with refinement to native data at 3.15 Å.

5.3.6 Complex structure vs predicted molecular weight

The molecular weight calculated for the complex from the gel filtration experiment suggests the formation of a $\alpha_2\beta_2$ tetramer, along with the 1:1 stoichiometry indicated by the SDS-PAGE analysis. Only the dimer formed between one molecule of BPSS0212 and one molecule of BPSS0213 was supported by PDBePISA analysis (Krissinel and Hendrick, 2007). The difference in stoichiometry between the predicted $\alpha_2\beta_2$ tetrameric complex from the gel filtration run and $\alpha\beta$ dimeric complex structure (α is BPSS0212 and β is BPSS0213) may be due to the BPSS0212/BPSS0213 complex being of a non-spherical nature, which then affects the elution volume (McGuiness, 1973). The nature of the interaction between BPSS0212 and BPSS0213 could be further explored using analytical ultracentrifugation.

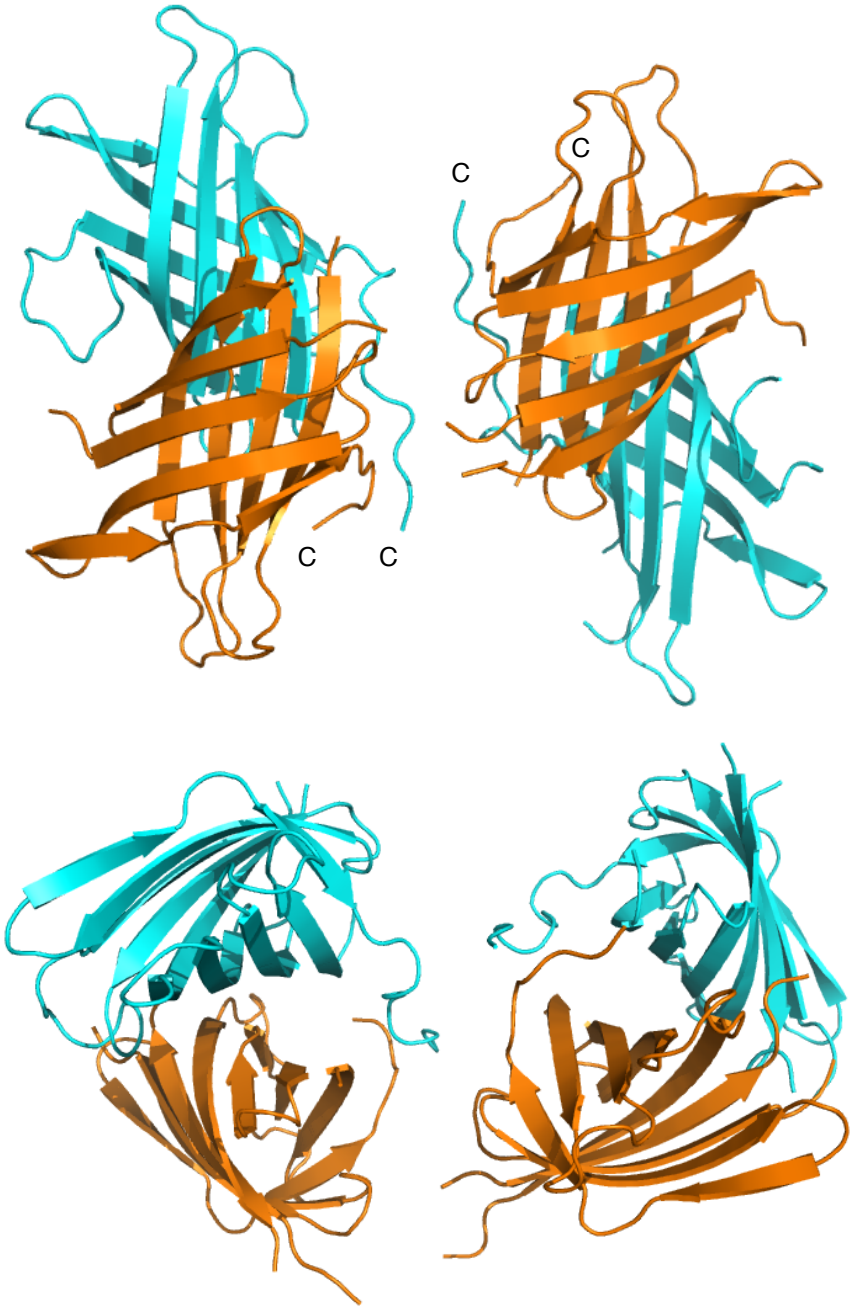


Figure 5.18. Initial structure of BPSS0212/BPSS0213 complex
Asymmetric unit of BPSS0212/BPSS0213 initial complex structure contains two dimers of BPSS0212 (blue) and BPSS0213 (orange), shown as cartoon. For each protein in the dimer, only the DUF1842 domain and part of the DUF1842-DUF1843 domain linker is present. The C-terminal end of each protein in the dimer clusters together. Figure generated in PyMOL version 1.1r1.

5.3.7 Weak electron density in the structure identifies the possible presence of the DUF1843 domain

In the electron density map, there is only interpretable electron density present for the DUF1842 domain and part of the interdomain linker region. However on building and refinement of the DUF1842 domains, significant electron density was found to be present that could not be accounted for by the DUF1842 domain residues (Figure 5.19). The size and shape of the electron density and distance between the regions of density suggested it could accommodate a number of α-helices. This would correspond to the missing DUF1843 domain, which had previously been suggested to be helical in the structure of BPSS0211 (Day, 2012). The electron density was not of a high enough quality to model in any of the DUF1843 domain, although the electron density map supports the presence of the domain.

5.4 BPSS0212/BPSS0213 complex crystal optimisation

The formation of the BPSS0212/BPSS0213 protein complex was repeated to produce more protein for crystallisation optimisation trials. The aim of these trials was to produce better quality crystals and thus improve the electron density to aid the building of the DUF1843 domain into the BPSS0212/BPSS0213 complex model. BPSS0212 and BPSS0213 were combined and loaded on a gel filtration column pre-equilibrated with 50 mM Tris pH 8.0, 500 mM NaCl, as detailed in section 2.7.5.

5.4.1 Repeat sitting-drop crystallisation trials of BPSS0212/BPSS0213 complex

The protein complex was put down into robot sitting-drop crystallisation trials to see if crystals could be produced in any other conditions and reproduced in the same conditions that had previously produced crystals. Protein was set down at ~10 mg/mL in 10 mM Tris pH 8.0, 100 mM NaCl. The 96-condition crystallisation screens used were PACT, JCSG+,

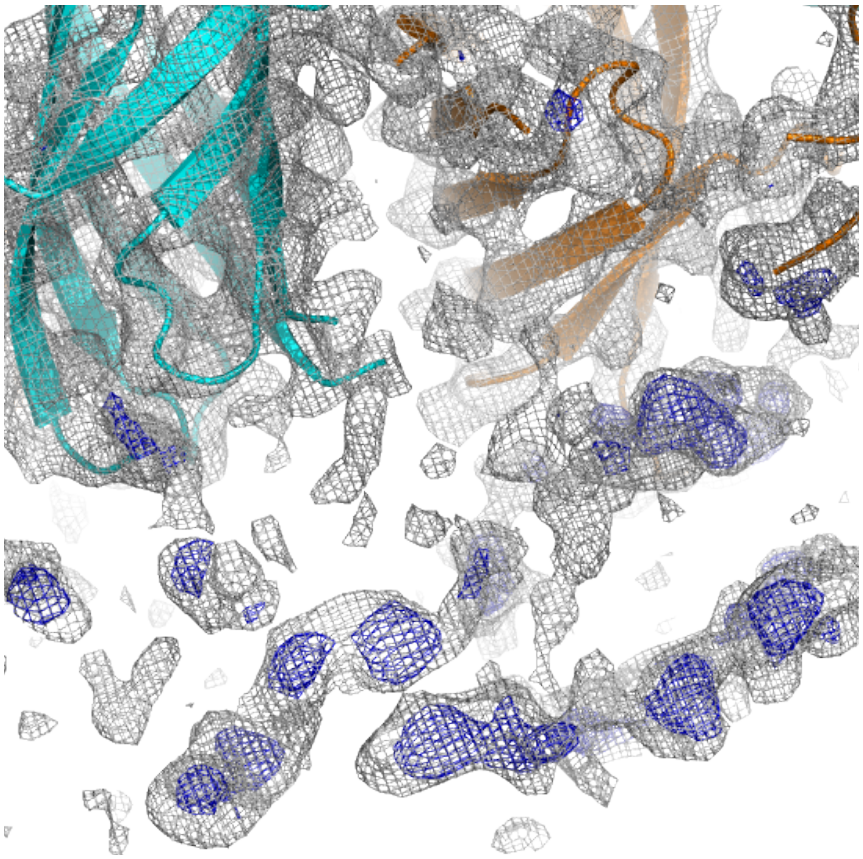


Figure 5.19. Weak electron density in BPSS0212/BPSS0213 complex map. BPSS0212/BPSS0213 structure shown as cartoon, BPSS0212 (blue) and BPSS0213 (orange). The 2Fo-Fc (FWT, PHWT) map, at 1.0 sigma (grey) and Fo-Fc (DELFWT, PHDELFWT) map, at 3.0 sigma (dark blue). Map generated in FFT (Winn et al., 2011). Figure generated in PyMOL version 1.1r1.

Proplex and PEGs.

5.4.2 Hanging-drop crystallisation optimisation trials of BPSS0212/BPSS0213 complex

The crystallisation condition Proplex B4 (0.2 M NaCl, 0.1 M MES pH 6.5, 10% (w/v) PEG4000), which previously produced crystals used to solve the initial structure, was optimised in order to see if the diffraction resolution of the crystals could be improved. Hanging-drop vapour diffusion optimisation trials were set up for Proplex B4, by varying the percentage PEG4000, from 6-16% and 20% (w/v) PEG4000 in 1% increments, with 0.2 M NaCl and 0.1 M MES pH 6.5. The concentration of NaCl was varied from 0.05 M-0.35 M in 0.05 M increments, with 0.1 M MES pH 6.5 and 10% (w/v) PEG4000. Crystallisation trials were stored at 17°C and 7°C. Further trials were set down at 7°C with 7-11% (w/v) PEG4000, with 0.2 M NaCl and 0.1 M MES pH 6.5 and one well of 0.35 M NaCl, 10% (w/v) PEG4000 and 0.1 M MES pH 6.5. Trays were set down with 1000 µL reservoir and both 2 µL:2 µL protein:reservoir drops.

Hanging drop optimisation trials were set down, at 17°C, for crystallisation condition Proplex F3, (0.1 M magnesium acetate, 0.1 M MES pH 6.5, 10% (w/v) PEG10000) which previously produced crystals. The pH of MES used was pH 6.0, pH 6.5 and pH 7.0, with 6-14% (w/v) PEG10000 in increments of 1% and 0.1 M magnesium acetate. Trays were set down with 1000 µL reservoir and 2 µL:2 µL protein:reservoir drops, for all three pH values, and also 2 µL:1 µL drops for pH 6.0 and pH 7.0.

5.4.3 Data collection of BPSS0212/BPSS0213 crystals

Crystals were grown in 0.2 M NaCl, 0.1 M MES pH 6.5 and 8% (w/v) PEG4000. The crystals were cryoprotected in 0.2 M NaCl, 0.1 M MES pH 6.5, 10% (w/v) PEG4000 and 25% ethylene glycol, cooled in liquid nitrogen and sent to the Diamond Light Source Synchrotron, Oxford, beamline I24. Native data were collected over 120°, with a phi oscilliation

of 0.1°, an exposure of 0.1 seconds, and a transmission of ~52%. 1-800 images of data were processed in xia2 3d (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Winter, 2010), in spacegroup I422 with cell dimensions of a=b=117.4 Å, c=146.1 Å and α=β=γ=90° and the resolution of the data is 2.83 Å (Table 5.2). The cell dimensions are similar to those from the previous data collection (a=b=117.6 Å, c=120.5 Å and α=β=γ=90°), however the length of c is ~25 Å longer.

5.4.4 Structure determination of the BPSS0212/BPSS0213 complex by molecular replacement

The Matthews coefficient calculation predicted two molecules in the asymmetric unit with 55% solvent when looking for 408 residues (one molecule of BPSS0212 and one molecule of BPSS0213) (Matthews, 1968, Kantardjieff and Rupp, 2003). The new higher resolution complex data was phased by molecular replacement using PhaserMR (McCoy et al., 2007) using coordinates for the BPSS0212/BPSS0213 dimer from the structure described in section 5.3 (subsequently referred to as complex 5.3) as the search model. The asymmetric unit was found to contain one αβ heterodimer of BPSS0212/BPSS0213; this complex structure will be subsequently referred to as complex 5.4.

5.4.5 Structure building and refinement of BPSS0212/BPSS0213 complex 5.4

The structure was built in *Coot* (Emsley et al., 2010) and underwent rounds of refinement against native data, to 2.83 Å, using Refmac5 (Murshudov et al., 1997; Murshudov et al., 2011). The model of complex 5.4 contains one chain of BPSS0212 and one chain of BPSS0213 (Figure 5.18). As in the previous structure, complex 5.3, the electron density was sufficient to build both DUF1842 domains and the part of the interdomain linker of BPSS0212 and BPSS0213. Residues 4-133 were modelled for BPSS0213, except for residues 85-86, due to the lack of electron density

Data collection	BPSS0212/BPSS0213 (Complex 5.4)
Wavelength (Å)	0.96861
Resolution range (Å)	82.98-2.83 (2.90-2.83)
Space group	I 4 2 2
a (Å)	117.4
b (Å)	117.4
c (Å)	146.1
α (°)	90.00
β (°)	90.00
γ (°)	90.00
Total reflections	70981 (5388)
Unique reflections	12499 (910)
Multiplicity	5.7 (5.9)
Completeness (%)	99.9 (100.0)
Mean I/σ	28.2 (2.3)
Wilson B factor ($Å^2$)	89
R_{merge}	0.032 (0.650)
R_{pim} (I)	0.016 (0.321)

Table 5.2. Data collection processing statistics for BPSS0212/BPSS0213 complex 5.4. xia2 3d data processing statistics for 1-800 images of BPSS0212/BPSS0213 native collected data (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Winter, 2010). The high-resolution shell information is shown in brackets.

in the loop region. For BPSS0212 residues 8-146 were modelled, except for the loop section of residues 69-78, due to the lack of electron density. The model R_{factor} is 0.2496 and R_{free} is 0.2873.

5.4.6 Analysis of BPSS0212/BPSS0213 crystal packing

PDBePISA predicted the presence of a $\alpha_4\beta_4$ complex in the complex 5.4 structure crystal lattice (Krissinel and Hendrick, 2007), unlike in complex 5.3, where PDBePISA predicted a αβ complex. The CSS (Complexation Significance Score), which describes the significance of the interface for complex or assembly formation, for the interfaces involved in the formation of the $\alpha_4\beta_4$ in complex 5.4 is 1.0, which is the highest score (Krissinel and Hendrick, 2007). The interactions between the heterodimer pair from complex 5.3, shown in Figure 5.18, is 1.0, whereas the other interactions required to form the $\alpha_4\beta_4$ in complex 5.3 have a low score of 0.03, indicating this interface is less significant (Krissinel and Hendrick, 2007). The analysis of the crystal lattice revealed a $\alpha_4\beta_4$ complex produced by the 4-fold crystal symmetry (Figure 5.20). Further comparisons showed that the arrangement of subunits around the 4-fold axis in the crystal of complex 5.4 (spacegroup I422) was similar to that seen around the 4-fold axis in complex 5.3 (spacegroup I4).

5.5 Investigation into complex formation between BPSS0211(1-123), BPSS0212 and BPSS0213

5.5.1 BPSS0212 and BPSS0213 produce a high molecular weight complex under low salt conditions

The newly annotated protein-coding region of BPSS0211 – BPSS0211(1-123) indicates the protein is twice as long as the original annotation of BPSS0211 (section 3.4). Therefore this opened up the possibility that the longer construct might interact with BPSS0212 and BPSS0213, in a way not seen with BPSS0211. BPSS0211(1-123) was combined with BPSS0212 and BPSS0213, in a 1:1:1 mg ratio and concentrated. Unfortunately some of the BPSS0211(1-123) protein precipitated prior to

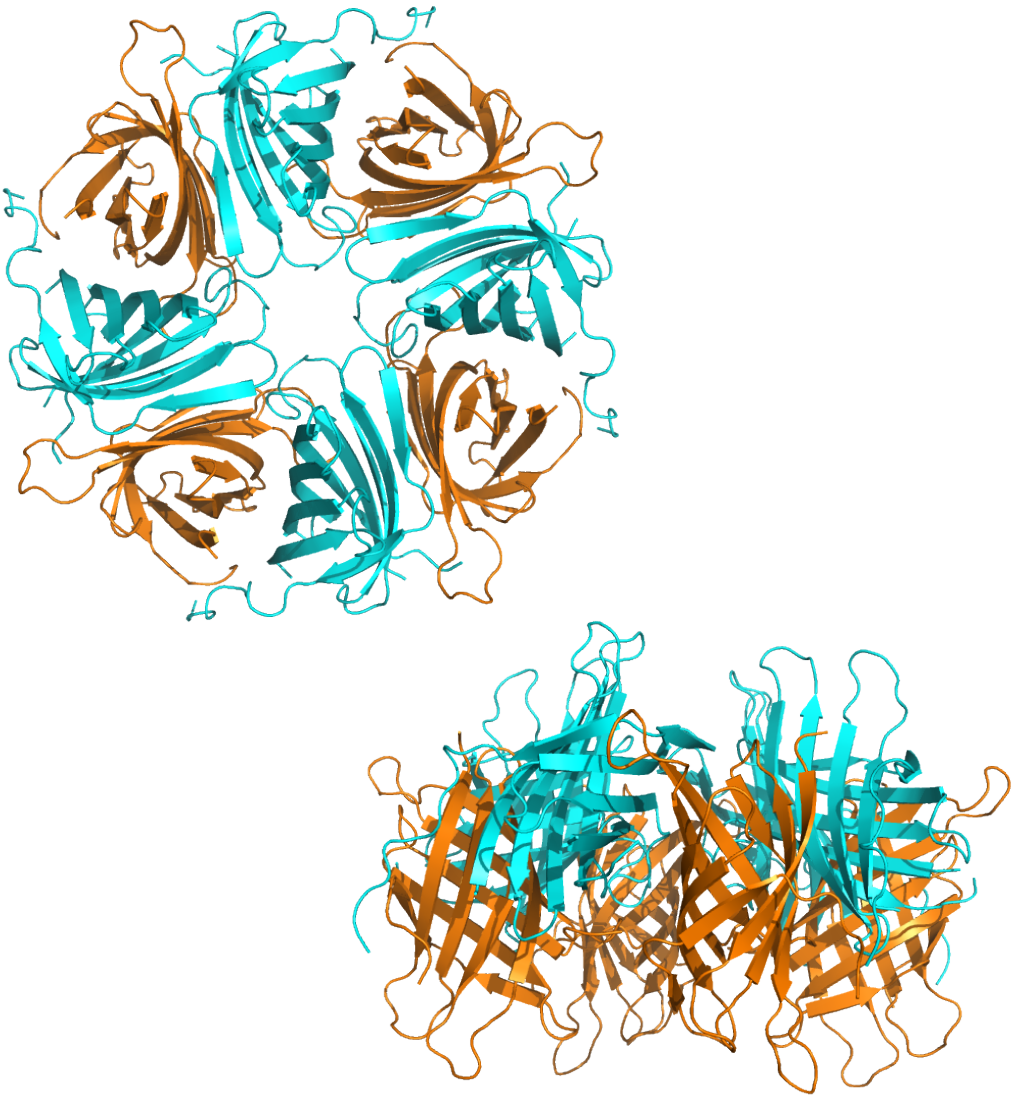


Figure 5.20. Predicted $\alpha_4\beta_4$ stoichiometry of BPSS0212/BPSS0213 complex 5.4. The $\alpha_4\beta_4$ complex structure generated by the 4-fold crystallographic cell symmetry. BPSS0212 (blue) and BPSS0213 (orange) alternate to form a ring structure, with BPSS0212 forming an upper layer and BPSS0213 forming the lower layer. Figure generated in PyMOL version 1.1r1.

loading the concentrated combined proteins onto the gel filtration column. The proteins were applied to a gel filtration column, pre-equilibrated with 50 mM HEPES pH 7.5, 100 mM NaCl. The gel filtration column was run at a lower pH to accommodate the pI of the new construct of BPSS0211(1-123) at 9.73 (ProtParam – Gasteiger et al., 2005).

The chromatogram showed five peaks (Figure 5.21). The first peak was at a lower elution volume, ~51 mL, not seen in any elution profiles from previous experiments run at the higher salt concentration of 50 mM Tris pH 8.0, 500 mM NaCl. This indicates that at a lower salt concentration, a larger molecular weight complex of BPSS0211(1-123), BPSS0212 and BPSS0213 could be present. However, this first peak was only found to contain BPSS0212 and BPSS0213 protein indicating that BPSS0212 and BPSS0213 elute together at a lower volume when run under these low salt conditions. Another BPSS0212/BPSS0213 complex is seen in peak 2 (~63 mL), this may correspond to the $\alpha_4\beta_4$ complex. The third peak contains BPSS0213 and BPSS0211(1-123), eluting at ~79 mL. The fourth is a low-level contaminant and no significant protein band was found in the fifth peak (Figures 5.22 and 5.23).

5.5.2 Molecular weight analysis of elution products

In order to estimate the molecular weight of this larger complex under low salt conditions, a calibration graph of K_{av} vs LogMW was produced by running a calibration mix containing proteins: Ferritin, Aldolase, Ovalbumin and Ribonuclease A, on the same gel filtration column pre-equilibrated with 50 mM HEPES pH 7.5, 100 mM NaCl (Figure 5.24). The elution volume of each protein was noted to calculate the K_{av} (section 2.1.14).

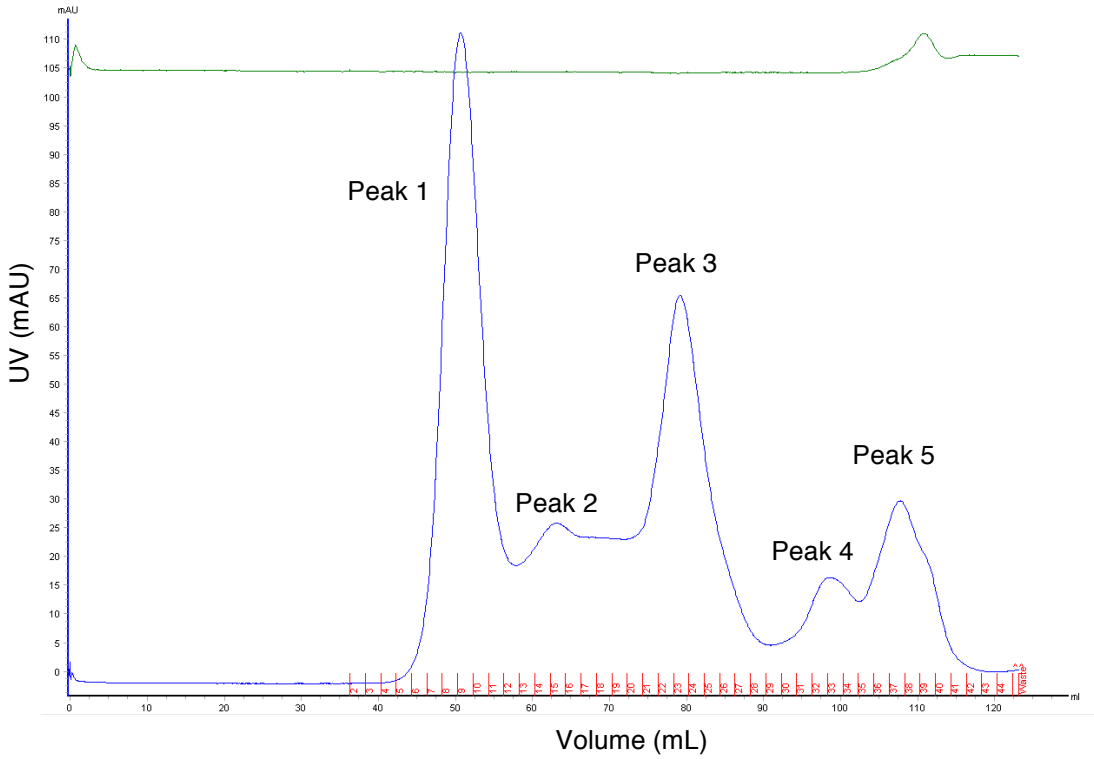


Figure 5.21. Chromatogram of BPSS0211(1-123), BPSS0212 and BPSS0213 gel filtration experiment, run with 50 mM HEPES pH 7.5, 100 mM NaCl. mAU at A_{280} – blue, conductivity – green, elution fractions – red numbers. The elution peaks are labelled and the corresponding elution volumes noted. Peak 1 – 50.68 mL, Peak 2 – 63.26 mL, Peak 3 – 79.19 mL, Peak 4 – 98.89 mL and Peak 5 – 107.83 mL.

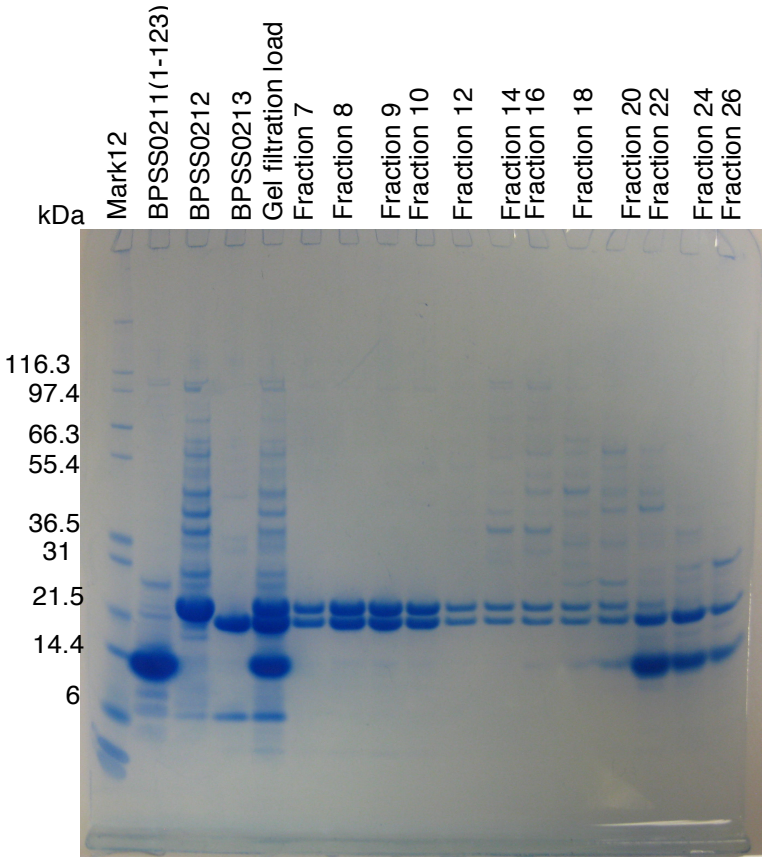


Figure 5.22. SDS-PAGE analysis of the BPSS0211(1-123), BPSS0212 and BPSS0213 gel filtration run, peaks 1-3. BPSS0211(1-123) is 13236 Da, BPSS0212 is 22465 Da and BPSS0213 is 21668 Da. Fractions 7-12 - peak 1, fractions 14-20 - peak 2 and fractions 22-26 - peak 3. However there is some overlapping of the elution peaks. Fractions 6-10 were taken and crystallised.

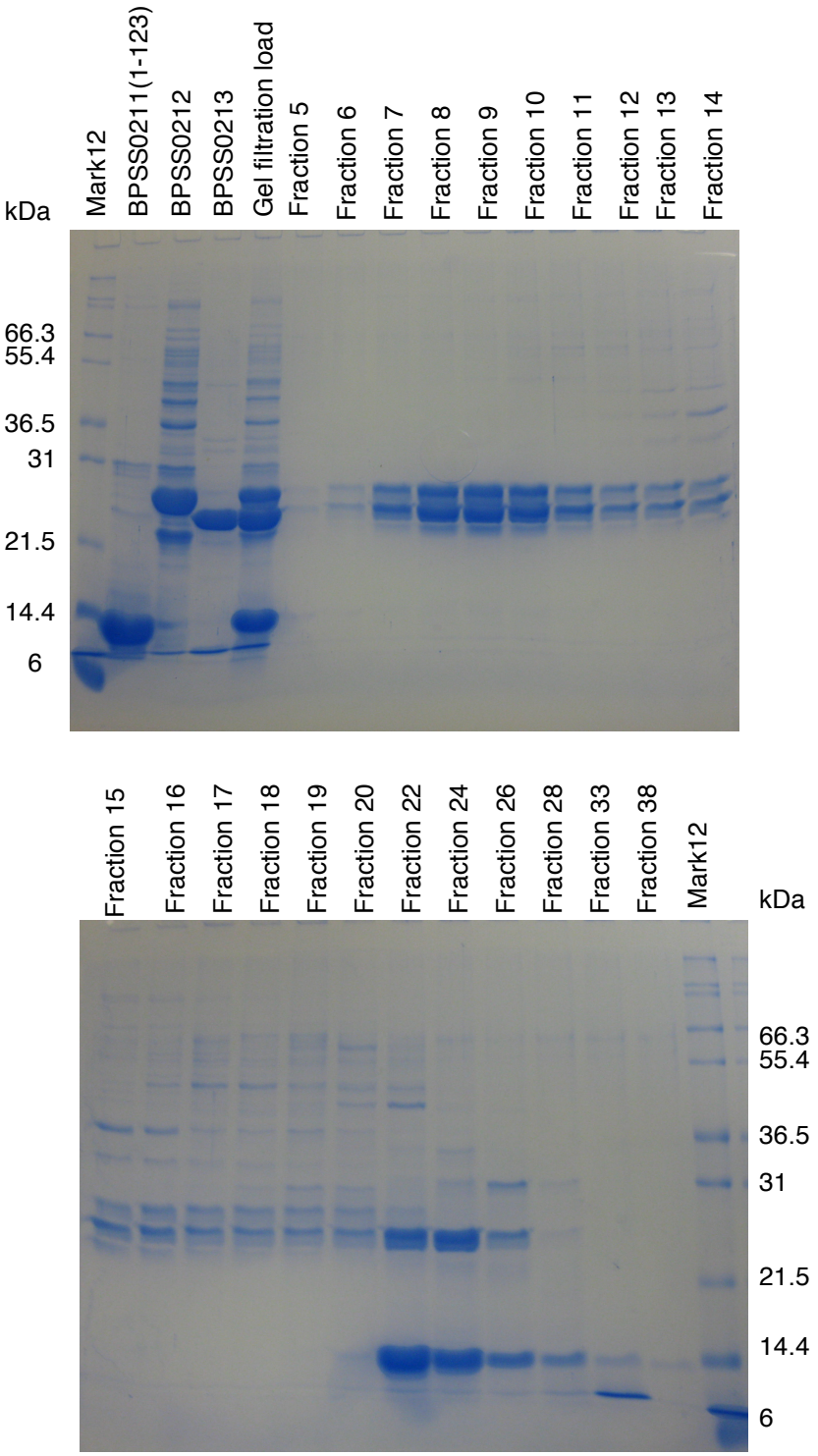


Figure 5.23. SDS-PAGE analysis of the BPSS0211(1-123), BPSS0212 and BPSS0213 gel filtration run, including peaks 4 and 5. BPSS0211(1-123) is 13236 Da, BPSS0212 is 22465 Da and BPSS0213 is 21668 Da. Fractions 7-12 - peak 1, fractions 13-20 - peak 2, fractions 22-28 - peak 3, fraction 33 - peak 4 and fraction 38 - peak 5. However there is some overlapping of the elution peaks.

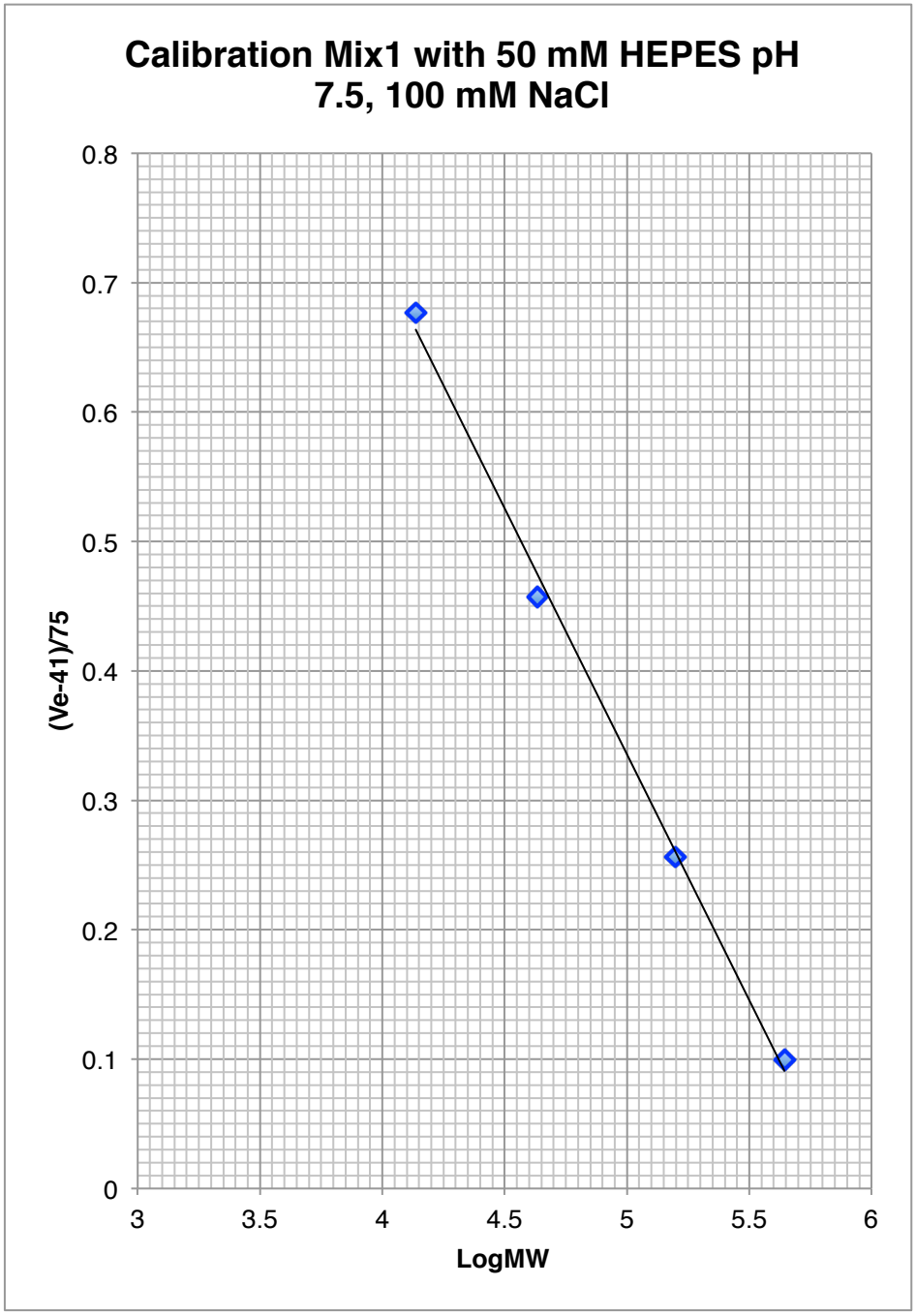


Figure 5.24. Calibration graph for Superdex200 gel filtration column, run with 50 mM HEPES pH 7.5, 100 mM NaCl. Calibration graph produced by running Ferritin, Aldolase, Ovalbumin and Ribonuclease A on a Superdex200 pg gel filtration column (GE) with 50 mM HEPES pH 7.5, 100 mM NaCl.

The calculated molecular weight of the larger complex in the first peak estimated from the elution volume of ~51 mL, corresponds to ~355 kDa. The average molecular weight of BPSS0212 and BPSS0213 is ~22 kDa, therefore the larger complex is predicted to contain ~16 protein molecules.

The previously solved structure of BPSS0212/BPSS0213, complex 5.4, suggested a stable stoichiometry of $\alpha_4\beta_4$, with the presence of a heterodimer in the structure and SDS-PAGE analysis, this suggests a 1:1 ratio of BPSS0212/BPSS0213. Therefore the complex of 16 proteins is predicted to have a $\alpha_8\beta_8$ stoichiometry. The second peak at ~63 mL has an estimated molecular weight of ~130 kDa, this complex peak has an unknown stoichiometry, but may correspond to the octameric structure seen in complex 5.3 and 5.4. The difference in salt concentration, between the low salt and high salt gel filtration experiments may have affected how this lower molecular weight complex runs, as there is no peak identified corresponding to ~90 kDa seen previously in the complex 5.1 and 5.3 formation experiments.

The SDS-PAGE analysis indicates fractions corresponding to peak 3 contained BPSS0213 and BPSS0211(1-123). The intensities of the bands on the gel indicate the two proteins are not eluting together. Peak 3 elution volume is ~79 mL, which corresponds to ~37 kDa. The predicted molecular weight for this peak is similar to that seen for BPSS0213 when run alone under conditions of 50 mM Tris pH 8.0, 500 mM NaCl, ~35 kDa. BPSS0211(1-123) elutes at ~76 mL when run at 50 mM Tris pH 8.0, 500 mM NaCl, corresponding to a molecular weight of ~45 kDa. Therefore under high salt conditions the two proteins run similarly in gel filtration. Further follow up is required to confirm whether BPSS0213 and BPSS0211(1-123) interact under the lower salt conditions.

5.5.3 Formation of BPSS0212/BPSS0213 large complex via a smaller molecular weight complex

The presence of a possible 16-mer formed under low salt conditions, suggests the ~90 kDa complex seen under high salt conditions does not reflect the final quaternary structure of the complex. However the presence of two types of complex under low salt conditions could represent two different stages of complex formation. A smaller complex forms first, which then assembles into a 16-mer complex of $\alpha_8\beta_8$, with the low salt conditions promoting the formation of the larger complex.

5.5.4 Identification of a 16mer in the crystal lattice

Further analysis of the crystal packing in structures of both 5.3 and 5.4 complexes, identified the same repeating unit of molecules producing the lattice (Figure 5.25). The repeating unit is formed of two $\alpha_4\beta_4$ complexes that are stacked top to bottom, or $\alpha_8\beta_8$. Therefore the presence of this repeating unit of $\alpha_8\beta_8$ suggests this maybe the stoichiometry of the complex.

5.6 Selenomethionine-incorporated BPSS0212+M/BPSS0213 complex

The structure of the BPSS0212/BPSS0213 complex has thus far been solved by molecular replacement, with either BPSS0212+M-FormA representing the DUF1842 domain, and part of the interdomain linker or the BPSS0212/BPSS0213 dimer complex 5.3. On both occasions there was insufficient electron density present for the DUF1843 domain of both BPSS0212 and BPSS0213 to be confidently modelled. Therefore, in order to improve the electron density around the DUF1843 domains, the structure of the complex was solved experimentally using selenomethionine-incorporated BPSS0212+M (BPSS0212+M-SeMet) and native BPSS0213 complex crystals. The BPSS0212+M-SeMet/BPSS0213 complex was produced under high salt conditions (50 mM Tris pH 8.0, 500 mM NaCl) and low salt conditions (50 mM HEPES

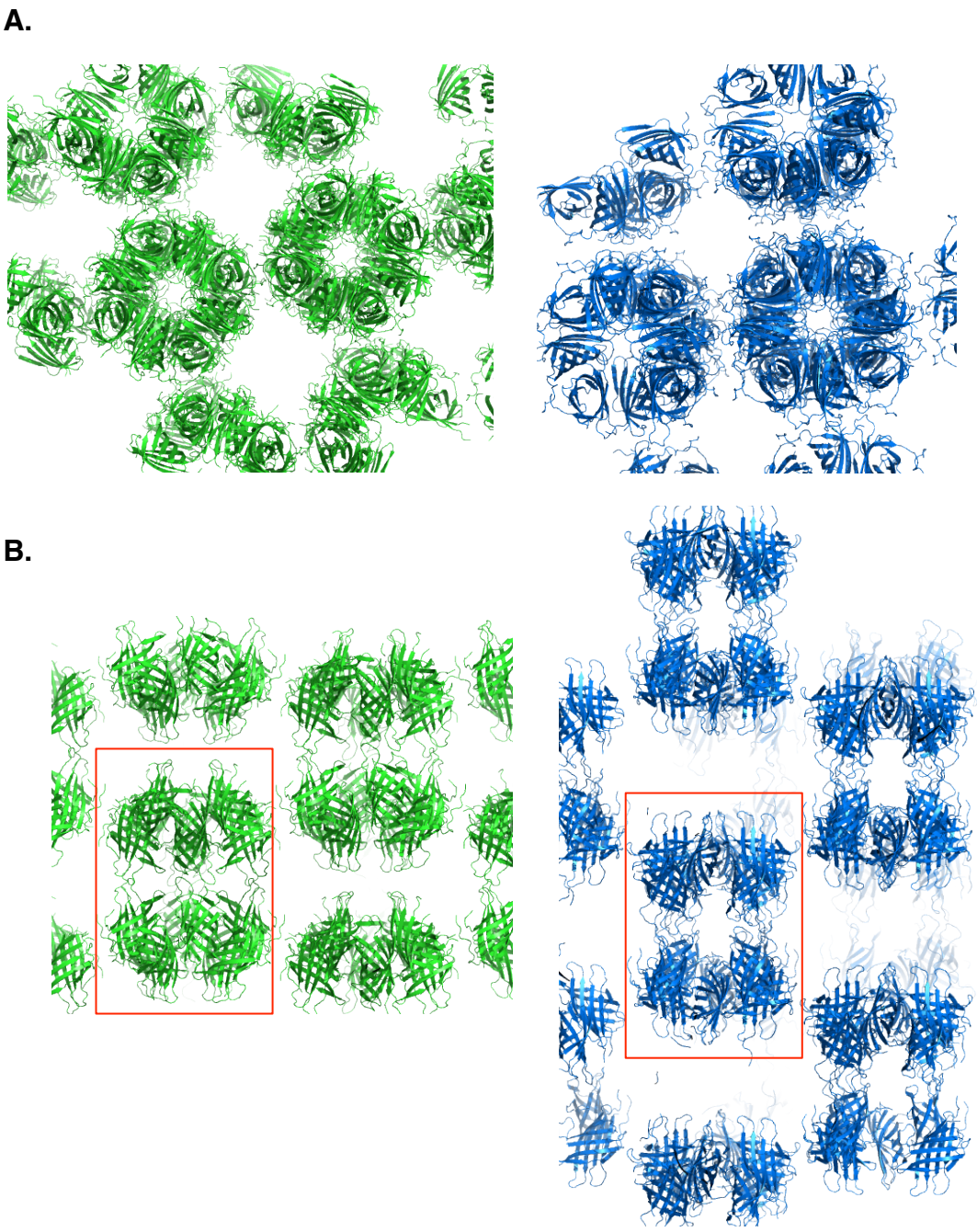


Figure 5.25. Repeating unit found in crystal lattices of complex 5.3 and 5.4. A repeating unit of $\alpha_8\beta_8$ is present in both complex 5.3 (green) and complex 5.4 (blue), highlighted by the red box. The unit is formed from two $\alpha_4\beta_4$ complexes stacked top to bottom, and is generated by the 4-fold and, 2-fold crystallographic symmetry axes in complex 5.4 or the complex 2-fold in complex 5.3. **A.** The top view (down the 4-fold) and **B.** the side view (down the 2-fold or complex 2-fold) of the $\alpha_8\beta_8$ unit is shown for both complexes. Images generated in PyMOL version 1.1r1.

pH 7.5, 100 mM NaCl), as outlined in section 2.7.7.

5.6.1 Large and small molecular weight BPSS0212+M-SeMet/BPSS0213 complexes are formed

The gel filtration experiment performed under high salt conditions produced two main protein-containing peaks (Figure 5.26). A smaller complex (~140 kDa) was represented followed by the excess BPSS0212+M-SeMet. The low salt gel filtration run produced the larger complex, with predicted molecular weight of ~316 kDa, assumed to be the $\alpha_8\beta_8$ 16-mer (Figure 5.27). These results reinforce those seen previously when the complex is formed in low salt conditions, with the identification of the possible 16mer. A smaller complex is also present under the low salt conditions, ~130 kDa (Figure 5.27). The predicted molecular weights are calculated from the previous calibration graph in Figure 5.24. The difference seen between elution volumes between the previous low salt gel filtration experiment (section 5.5) and this experiment may come as a result of general use of the column by others in between experiments. However, overall a similar elution profile was produced for both low salt gel filtration experiments, and a similar subunit composition is predicted.

5.6.2 Crystallisation of both BPSS0212+M-SeMet/BPSS0213 complexes produced in low and high salt conditions

The complex produced in high salt was set down in crystallisation trials at ~9 mg/mL in 10 mM Tris pH 8.0, 100 mM NaCl. The complex produced in low salt was set down at ~9 mg/mL in 50 mM HEPES pH 7.5, 100 mM NaCl. Both complex samples were set down into sitting-drop crystallisation trials using the Proplex and JCSG+ crystallisation suites, with the low salt complex also set down in the PACT crystallisation suite. Hanging-drop optimisation trials were set down for condition Proplex B4, with the complex produced in high salt, with 6-15% (w/v) PEG4000 in increments of 1%, 0.2 M NaCl and 0.1 M MES buffer pH 6.5. Also 0.05

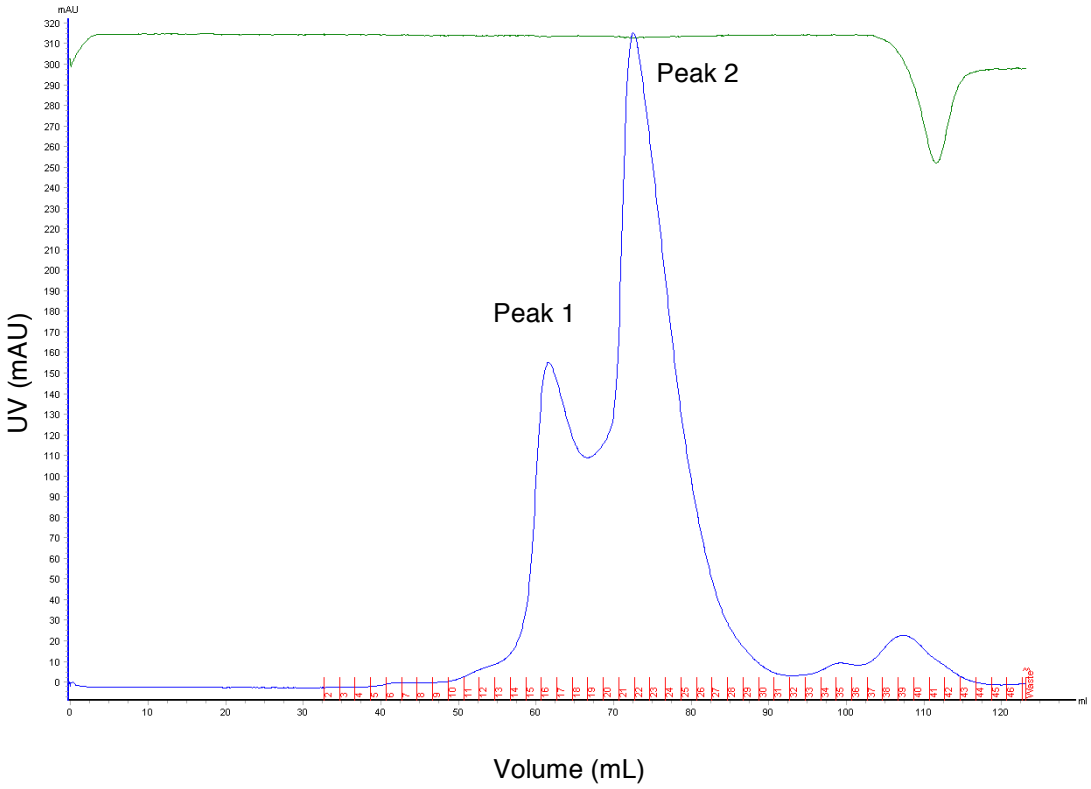


Figure 5.26. Chromatogram of BPSS0212+M-SeMet and BPSS0213 gel filtration experiment, run with 50 mM Tris pH 8.0, 500 mM NaCl. mAU at A_{280} – blue, conductivity – green, elution fractions – red numbers. The elution peaks are labelled and the corresponding elution volumes noted. Peak 1 – 61.62 mL and Peak 2 – 72.55 mL.

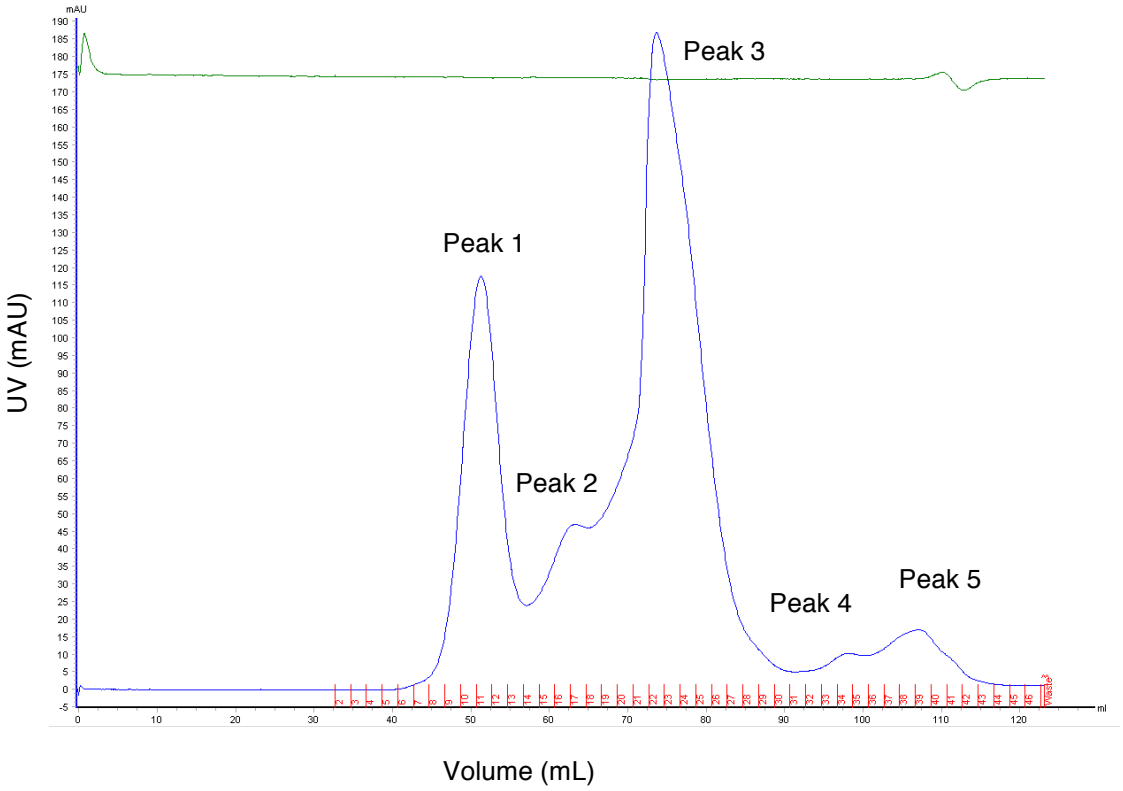


Figure 5.27. Chromatogram of BPSS0212+M-SeMet and BPSS0213 gel filtration experiment, run with 50 mM HEPES pH 7.5, 100 mM NaCl. mAU at A_{280} – blue, conductivity – green, elution fractions – red numbers. The elution peaks are labelled and the corresponding elution volumes noted. Peak 1 – 51.33 mL, Peak 2 – 63.21 mL, Peak 3 – 73.69 mL, Peak 4 – 98.55 mL and Peak 5 – 107.28 mL. Fractions 9-12 were pooled for crystallisation.

M-0.35 M NaCl, 10% (w/v) PEG4000 and 0.1 M MES buffer pH 6.5. These optimisation trials were stored at 7°C.

5.6.3 Sitting-drop crystal hits and hanging-drop crystallisation optimisation trials

For high salt complex formation, crystals were found in Proplex D11, D12 and F3 conditions (Figure 5.28). Hanging-drop optimisation trays were set up for Proplex D12. The low salt complex had crystals hits in Proplex B9, D7, D11, D10, E10, F3 and JCSG+ E10 conditions (Figure 5.29). Hanging-drop optimisation trays were set up for Proplex E10, Proplex D7 and Proplex B10 – from a sitting-drop hit from the previous low salt complex experiment.

5.6.4 Data collection of BPSS0212+M-SeMet/BPSS0213 protein crystals

Crystals produced, from the complex produced in high salt, in the hanging-drop optimisation of condition Proplex B4 – 0.1 M MES pH 6.5, 0.3 M NaCl, 10% (w/v) PEG4000 grown at 7°C, were cryoprotected with 0.1 M MES pH 6.5, 0.3 M NaCl, 12% (w/v) PEG4000 and 30% ethylene glycol, cooled and sent to Diamond Light Source Synchrotron, Oxford, beamline I02 for screening. A fluorescence scan was carried out at the selenium K edge, to test for incorporation of selenium in the protein crystals and to identify the energies at which to collect the peak and inflection data sets. *CHOOCH* (Evans and Pettifer, 2001) produced f'' and f' values of 5.68 e/-7.63 e, respectively, for the peak. Peak data were collected at 0.97625 Å wavelength. Data were collected over 900°, with 0.15° phi oscillation for 0.1 seconds, with 20% transmission. Data were processed in xia2-3d (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Sauter et al., 2004; Winter, 2010; Zhang et al., 2006) to a resolution of 2.95 Å, in I422 with cell dimensions of a=b=116.5 Å, c=146.8 Å and α=β=γ=90°. The processing statistics are shown in Table 5.3.

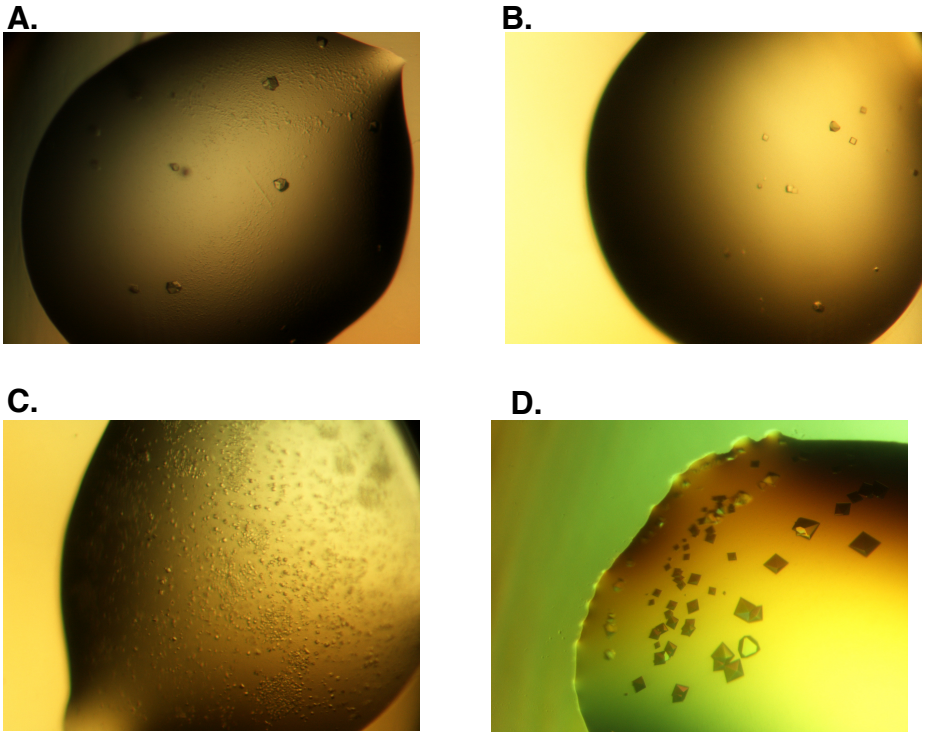


Figure 5.28. Crystal hits from crystallisation trials for BPSS0212+M-SeMet/BPSS0213 complex formed in high salt.
Sitting-drop trial - **A.** Proplex F3 – 0.1 M magnesium acetate, 0.1 M MES pH 6.5, 10% (w/v) PEG10000. **B.** Proplex D11 – 0.1 M Tris pH 8.5, 15% (w/v) PEG6000. **C.** Proplex D12 – 0.1 M Tris pH 8.5, 20% (w/v) PEG6000. Hanging-drop optimisation – **D.** Proplex B4 optimisation conditions.

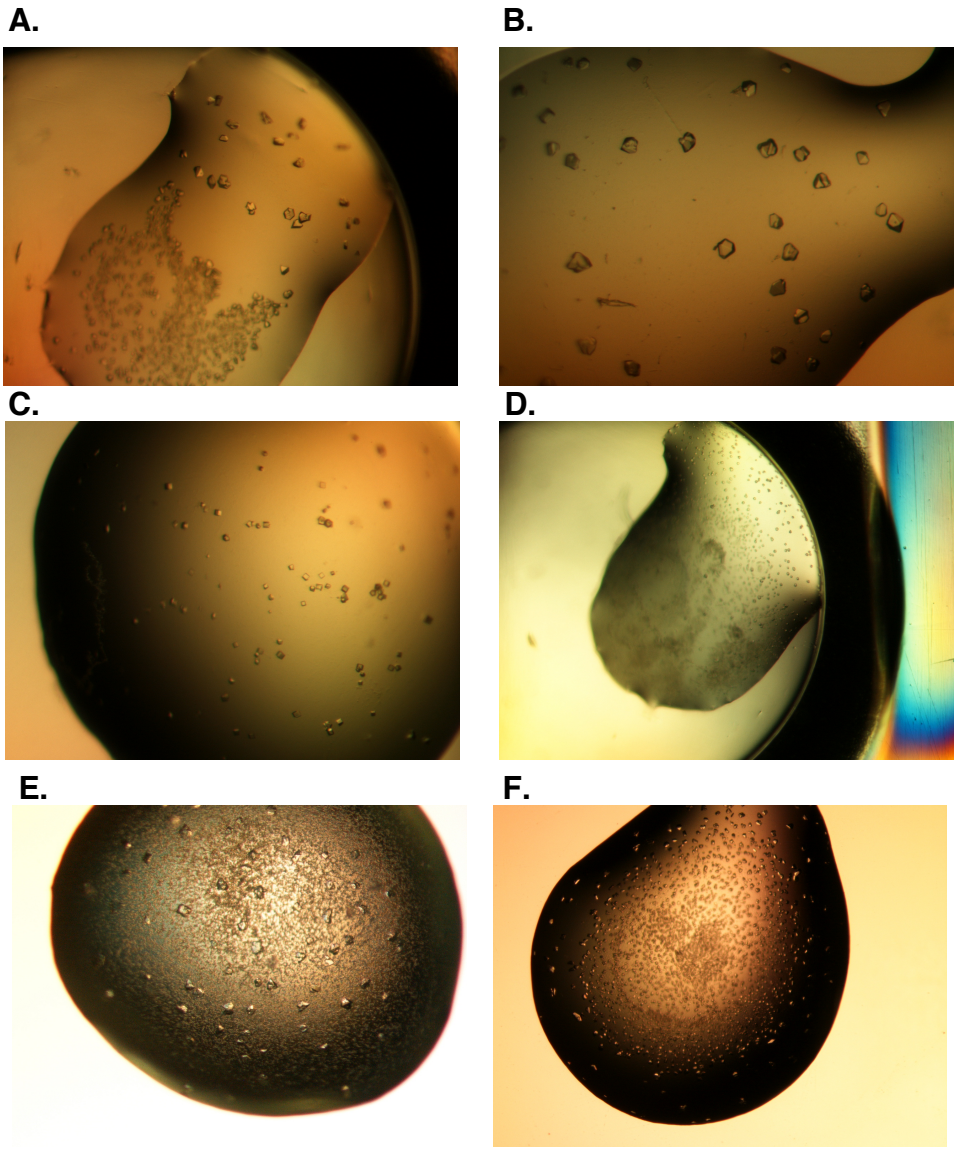


Figure 5.29. Crystal hits from crystallisation trials for BPSS0212+M-SeMet/BPSS0213 complex formed in low salt

Sitting-drop trial - **A.** Proplex D7 – 0.1 M sodium citrate pH 5.5, 15% (w/v) PEG6000. **B.** Proplex F3 – 0.1 M magnesium acetate, 0.1 M MES pH 6.5 and 10% (w/v) PEG10000. **C.** JCSG+ E10 – 0.1 M BICINE pH 9.0, 10% (w/v) PEG6000. **D.** Proplex B9 – 0.1 M sodium cacodylate pH 6.0, 15% (w/v) PEG4000. Hanging-drop optimisation - **E**. Proplex B10 optimisation – 0.15 M ammonium sulphate, 0.1 M MES pH 6.0, 14% (w/v) PEG4000. **F.** Proplex D7 optimisation – 0.1 M sodium citrate pH 5.5, 15% (w/v) PEG6000.

Data collection	Native	Peak
Wavelength (A)	0.96861	0.97625
Resolution range (Å)	82.98-2.83 (2.90-2.83)	91.24-2.95 (3.03-2.95)
Space group	I 4 2 2	I 4 2 2
Unit cell parameters		
a (Å)	117.4	116.5
b (Å)	117.4	116.5
c (Å)	146.1	146.8
α (°)	90.00	90
β (°)	90.00	90
γ (°)	90.00	90
Total reflections	70981 (5388)	717107 (55282)
Unique reflections	12499 (910)	10988 (797)
Multiplicity	5.7 (5.9)	65.3 (69.4)
Completeness (%)	99.9 (100.0)	100.0 (99.9)
Mean I/σ	28.2 (2.3)	40.2 (2.1)
Wilson B factor ($Å^2$)	89	103
R_{merge}	0.032 (0.650)	0.1 (3.301)
R_{pim} (I)	0.016 (0.321)	0.013 (0.400)
Anomalous completeness	-	100.0 (99.9)
Anomalous multiplicity	-	34.8 (36.0)
Anomalous correlation	-	0.832 (-0.016)
Anomalous slope	-	1.344

Table 5.3. Data collection processing statistics for BPSS0212+M-SeMet/BPSS0213 complex 5.6, and native data collected previously for complex 5.4. xia2 3d data processing statistics for BPSS0212+M selenomethionine SAD collected data (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Sauter et al., 2004; Winter, 2010; Zhang et al., 2006), and the native data (Table 5.2). The high-resolution shell information is shown in brackets.

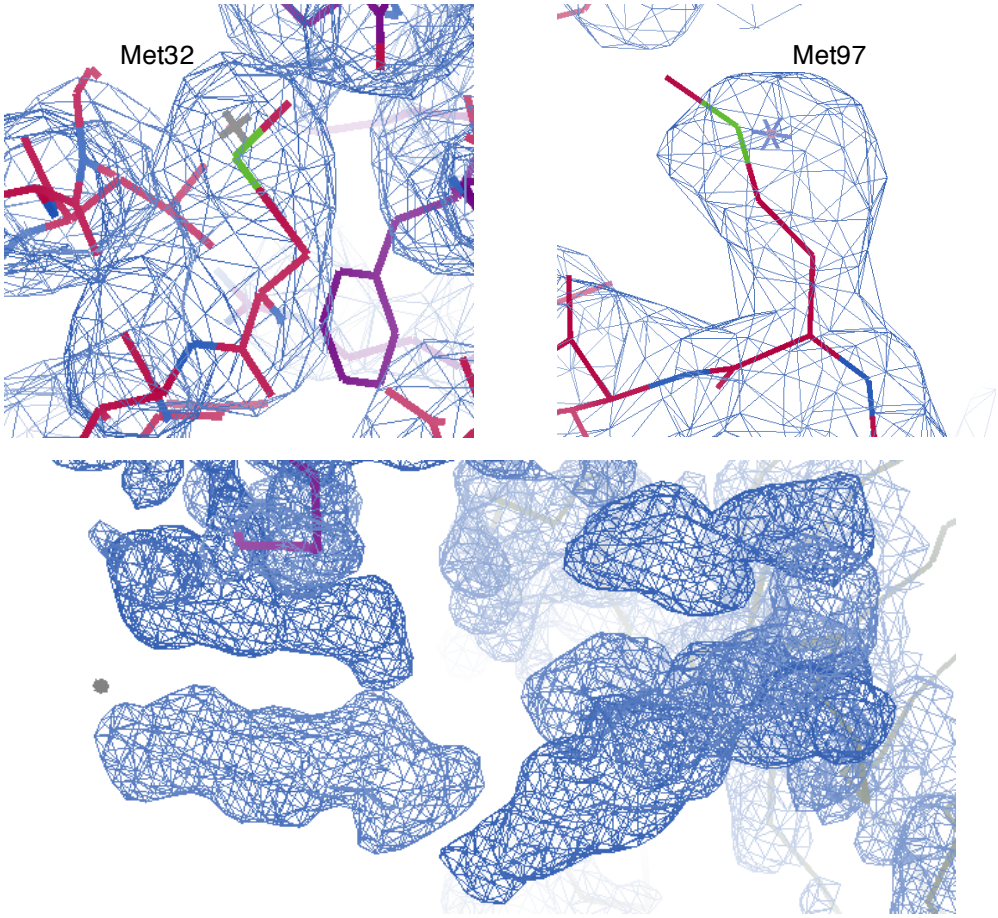
5.6.5 Structure determination of BPSS0212+M-SeMet/BPSS0213 by experimental phasing

SAD experimental phasing was carried out using the *SHELX* programs with the *HKL2MAP* interface (Sheldrick, 2008; Pape and Schneider, 2004). *SHELXC* (Sheldrick, 2008) was run for a SAD experiment using the peak data set alone, anomalous signal was identified to ~3.4 Å. *SHELXD* (Sheldrick, 2008) was used to search for six selenium sites, as there are six methionine residues in BPSS0212, up to a resolution of 3.5 Å over 1000 trials. *SHELXE* (Sheldrick, 2008) was run with solvent content of 55% and five cycles of autobuilding, in both the original and inverted hands. Seven selenium sites were identified for both hands. The original hand provided a higher estimated mean FOM (0.635) and pseudo-free CC value (68.41%) than the inverted hand (Sheldrick, 2008).

5.6.6 Building of the complex DUF1842 domains

The electron density map produced from *SHELXE* (Sheldrick, 2008) for the original hand was analysed in *Coot* (Emsley et al., 2010), along with the seven selenium sites identified, (Figure 5.30). As BPSS0212+M is the only protein in the complex containing selenomethionine, the selenium sites were used to confirm the electron density that represented BPSS0212+M and that which represented BPSS0213.

The top two selenium sites had occupancies of above 0.5 and could be mapped to two sites within the electron density representing one of the two DUF1842 domain β-barrels, most likely BPSS0212+M. Complex 5.4 was docked into the relevant section of map. The two selenium atoms mapped to methionine residues Met32 and Met97 in the DUF1842 domain of BPSS0212+M-SeMet (Figure 5.30). The orientation of the complex 5.4-BPSS0212 into the map placed BPSS0213 into the other unoccupied β-barrel density, in a similar location to that seen in previous complex structures. Thus the structure of the core of the $\alpha_8\beta_8$ was assigned. Analysis of the remaining density in the map identified



Selenium site	Occupancy	Amino acid (BPSS0212+M)
1	1.000	Met 32
2	0.7608	Met 97
3	0.3061	Met 178
4	0.2919	-
5	0.2049	-
6	0.2039	-
7	0.1528	-

Figure 5.30. Location of the top three selenium sites in the electron density map, and occupancies of the seven selenium sites identified. The top three selenium sites correspond to three methionine residues in BPSS0212+M-SeMet in the complex. Site 1 is associated with Met32 and site 2 with Met97. Site 3 is located near the predicted helical density, and was located at the conserved DUF1843 domain methionine site, when BPSS0211 was modelled into the density. The occupancies of the seven sites identified are shown in the table. Electron density map and selenium sites were produced by *SHELX* (Sheldrick, 2008). Images generated in *Coot*, with map level = 0.1314e/A^3 (0.96rmsd) (Emsley et al., 2010).

detached sections of electron density on the periphery of the $\alpha_8\beta_8$ complex, which could not be explained by the modelled DUF1842 domain (Figures 5.30 and 5.31). The density is similar to that seen in the electron density maps of previously solved structures, complex 5.3 and 5.4, and has been predicted to represent the missing DUF1843 domain helices. The third selenium site was found near to the unexplained density, which may represent one of the methionine sites in the BPSS0212+M-SeMet DUF1843 domain (Figure 5.30). The other four sites were not located in density, but in solvent regions between the protein density.

5.6.7 Building of the complex DUF1843 domains

A number of poly-alanine α-helical fragments were docked into the unassigned electron density, to account for the residues of the DUF1843 domain (Figure 5.32). The arrangement of the α-helices docked into the density, showed a similar structure to that of the BPSS0211 dimer, formed between monomers A and B, in the closed helical conformation produced when crystallised in the presence of zinc (Day, 2012) (Figure 5.2).

To further test the possibility that the conformation of the helical fragments of the DUF1843 domain is similar to that seen in BPSS0211, the A-B dimer structure of BPSS0211 (Day, 2012) was docked into the electron density (Figure 5.33). The docking of the BPSS0211 A-B dimer placed Met28 in BPSS0211 near to the third BPSS0212+M-SeMet selenium site identified in the complex, as predicted. In BPSS0211, Met28 is located close to the turn between the α1 and α2 helices (Day, 2012); and is conserved in BPSS0212 and BPSS0213 (Met177 and Met161, respectively). This suggests the third selenium site is Met178 from BPSS0212+M-SeMet.

The presence of the selenium site in only one of the two DUF1843 domains modelled in the density suggests that one DUF1843 domain in

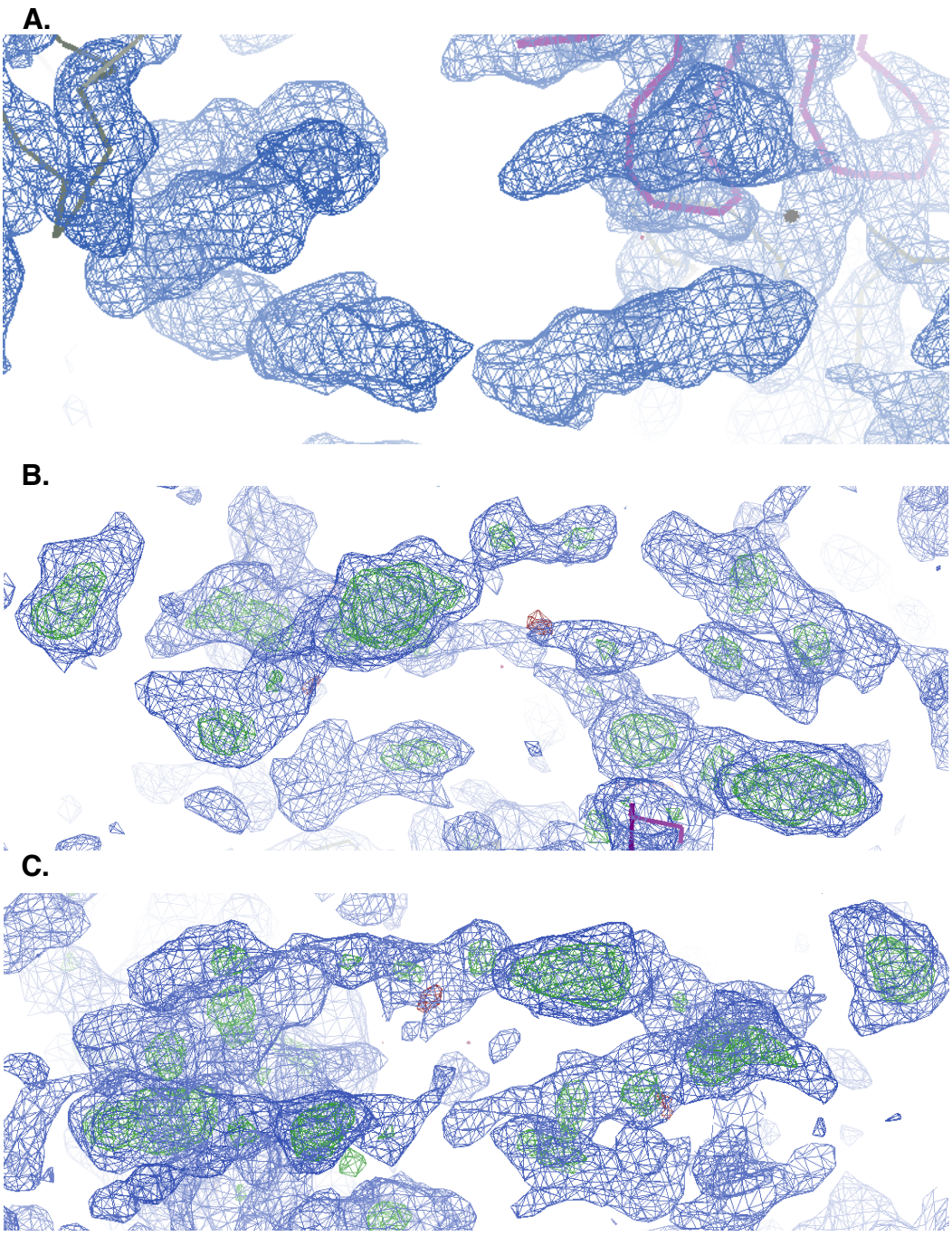
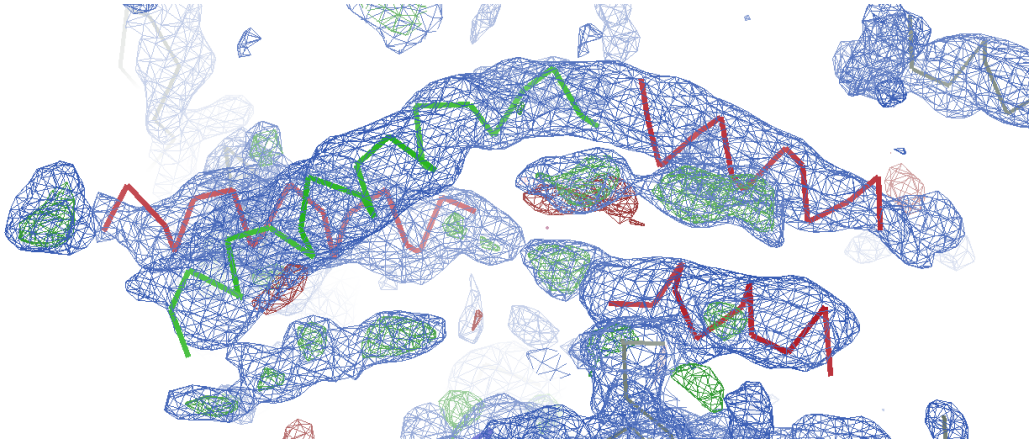


Figure 5.31. Unmodelled electron density in the complex.
A. Unmodelled electron density in map produced by *SHELX* (Sheldrick, 2008) for BPSS0212+M-SeMet/BPSS0213 complex. **B.** and **C.** Unmodelled electron density, showing positive Fo-Fc map peaks, with 2Fo-Fc map – blue, Fo-Fc map – green/red. Images generated in *Coot* (Emsley et al., 2010).

A.



B.

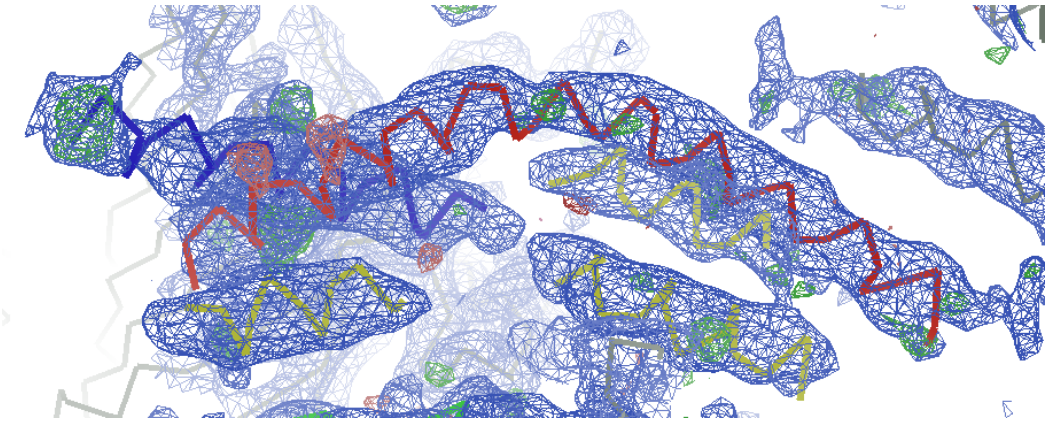


Figure 5.32. Multiple poly-alanine helices progressively modelled into unmodelled complex density. Helices are shown in multiple colours, with 2Fo-Fc map – blue, Fo-Fc map – green/red. **A.** Four separate poly-alanine helices modelled into the density, with two further regions of unmodelled density with positive electron density. Map level = 0.1148e/A^3 (1.04rmsd) (Emsley et al., 2010). **B.** Subsequent modelling of two extra helices into the positive density from Figure A, and the connection of two helices from Figure A (green and red), to form one longer helix (red). Map level = 0.1129e/A^3 (0.99rmsd) and images generated in *Coot* (Emsley et al., 2010).

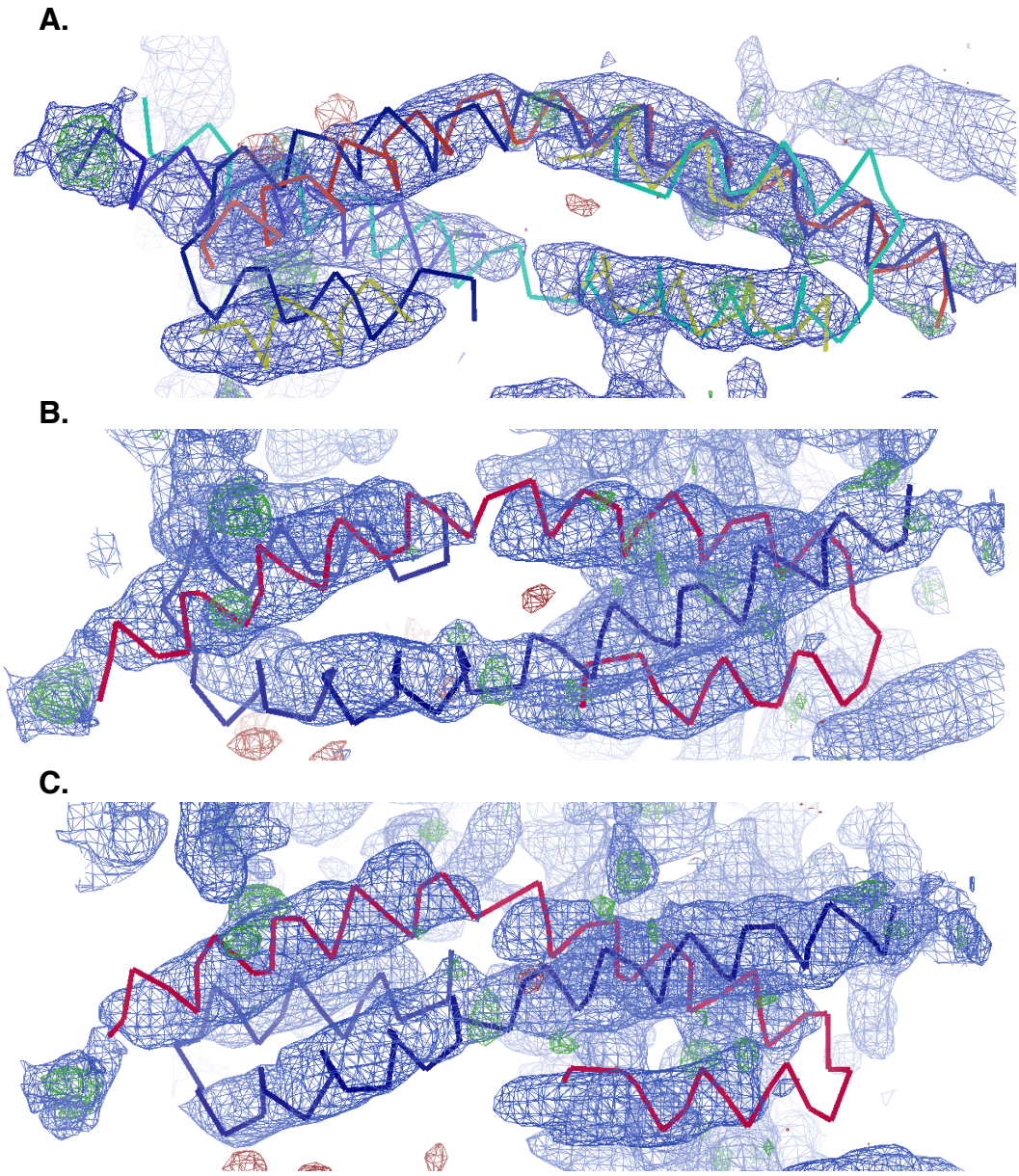


Figure 5.33. BPSS0211 closed conformation dimer structure docked into complex helical density. Helices are shown in multiple colours, with 2Fo-Fc map – blue, Fo-Fc map – green/red. The similarity between the modelled helices and the structure of BPSS0211 closed conformation helix A-B dimer, was predicted. The BPSS0211 closed conformation dimer was oriented into the extra density within the BPSS0212/BPSS0213 complex. **A.** BPSS0211 closed conformation helix A-B dimer (coloured dark blue and light blue) is in a similar arrangement to helical fragments (yellow, pink and dark blue)**. B.** and **C.** BPSS0211 closed conformation helix A-B dimer (coloured blue and red) docked into the density. Images generated in *Coot*, with map level = 0.1129e/A^3 (0.99rmsd) (Emsley et al., 2010). BPSS0211 closed conformation dimer coordinates were produced by Dr. Matthew Day (Day, 2012).

the dimer belongs to BPSS0212+M-SeMet, with the other DUF1843 domain as BPSS0213. The location of Met178 was used as a starting point and the register of the helices of the BPSS0212+M-SeMet DUF1843 domain was modelled to fit the density. The modelling location and orientation of the sidechains was aided using the BPSS0211 closed conformation structure (Day, 2012). The main chain helix was modelled into the density for BPSS0213 based on the BPSS0211 closed conformation A-B dimer (Day, 2012) arrangement, however the density alone could not be used to assign the sidechain locations. The sequences of BPSS0211, BPSS0212 and BPSS0213 DUF1843 domain regions can be confidently aligned (Figure 5.34), and as the structure of the two DUF1843 domains in the BPSS0212/BPSS0213 complex is similar to that of the BPSS0211 closed conformation AB dimer (Day, 2012), the positions of the sidechains for BPSS0213 can be tentatively assigned. For both BPSS0212+M-SeMet and BPSS0213, the density for the loop between α1 and α2 of the DUF1843 domains was not sufficient to model the loop residues confidently therefore these residues were omitted. The model was built in *Coot* (Emsley et al., 2010) and subjected to rounds of refinement using Refmac5 (Murshudov et al., 1997; Murshudov et al., 2011), followed by density modification using Parrot (Cowtan, 2010). Iterative rounds of refinement were then carried out in Refmac5 (Murshudov et al., 1997; Murshudov et al., 2011) against the higher resolution native data from complex 5.4, with resolution of 2.83 Å, and model building in *Coot* (Emsley et al., 2010). Therefore during the rounds of building to the native data the selenomethionine sites were exchanged for sulphur methionine. The crystal used for the native data collection contained BPSS0212 without the N-terminal methionine; therefore the residues were numbered to accommodate for this. This complex will be referred to as complex 5.6 containing BPSS0212/BPSS0213, not BPSS0212+M-SeMet/BPSS0213, from here on.

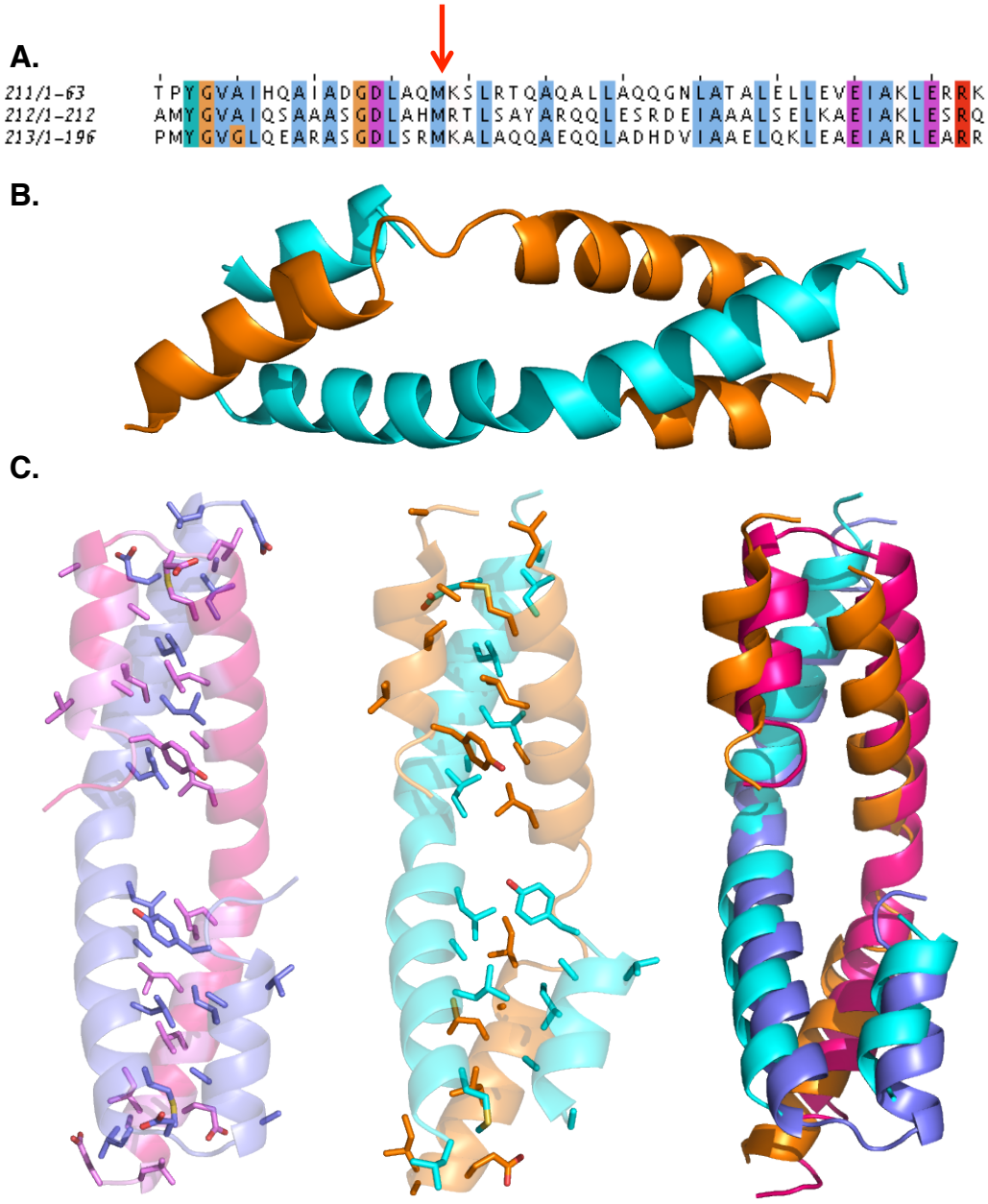


Figure 5.34. Modelling of the BPSS0212/BSS0213 DUF1843 domain helices. A. The alignment of the DUF1843 domains from BPSS0211, BPSS0212 and BPSS0213, with the conserved methionine residue highlighted (red arrow). Alignment carried out in Jalview with ClustalOmega coloured using ClustalX (Larkin et al., 2007; Sievers et al., 2011; Waterhouse et al.,2009). **B.** The BPSS0212/BPSS0213 complex DUF1843 domains – BPSS0212 is blue and BPSS0213 is orange. **C.** The DUF1843 domain 100% conserved residues are mapped onto the BPSS0211 closed conformation A-B dimer (pink/purple) and the BPSS0212/BPSS0213 DUF1843 domains (blue/orange). The BPSS0211 closed conformation A-B dimer (pink/purple) and BPSS0212/BPSS0213 DUF1843 domains (blue/orange) are overlaid. Figure generated in PyMOL version 1.1r1, the BPSS0211 closed conformation coordinates were produced by Dr. Matthew Day (Day, 2012).

5.6.8 Structure of BPSS0212/BPSS0213 complex 5.6

The current complex model contains one molecule of BPSS0212 and one molecule of BPSS0213, and has an R_{factor} = 0.2556 and R_{free} = 0.3089. Of the 212 expected residues in BPSS0212, 181 are modelled, with the first six and last four residues missing, along with residues 69-72 missing in loop β4-β5, residues 147-160 missing in the interdomain linker between the DUF1842 and DUF1843 domain and residues 171-173 in the linker between α1 and α2 of the DUF1843 domain. Of the 196 expected residues in BPSS0213, 176 are modelled, with the first three and last six residues missing, along with residues 86-87 missing in loop β5-β6, residues 136-142 missing in the interdomain linker between the DUF1842 and DUF1843 domains and residues 156-157 in the DUF1843 domain. The validation statistics are shown in Table 5.4.

5.6.9 Structural analysis of BPSS0212/BPSS0213 complex 5.6

The current model of the BPSS0212/BPSS0213 complex 5.6 has a core of 16 DUF1842 domains forming two layers of eight subunits – each layer has a $\alpha_4\beta_4$ formation of alternating BPSS0212 and BPSS0213 proteins. On the outside of the core DUF1842 domains are the DUF1843 domain helices. There are eight pairs of DUF1843 helices located in two layers around the $\alpha_8\beta_8$ DUF1842 domain core. The layers do not form solid rings of helices but result in alternating up/down helical pairs. Each pair is formed of one DUF1843 domain from each of BPSS0212 and BPSS0213 (Figure 5.35). The core of the complex contains a main central cross-shaped cavity, as well as a hole running through the length of the complex. Each of the four points of the cross is concealed on the outside by part of two DUF1843 domain dimers. The four BPSS0213 domains from the top $\alpha_4\beta_4$ layer and the four BPSS0213 domains from the bottom $\alpha_4\beta_4$ layer form four pillars and help shape the cavity (Figure 5.36). Analysis of the surface charge of the complex revealed strong patches of both positive and negative density on the surface of the sides of the complex, but noticeably less at either end. The internal cavity contains a

Refinement	BPSS0212/BPSS0213
R_{factor}	0.2556
R_{free}	0.3089
No. of non-H atoms	
Protein	2730
Protein residues	357
RMSD (bonds) (Å)	0.0108
RMSD (angles) (°)	1.7241
Ramachandran favoured (%)	92.08
Ramachandran outliers (%)	1.47
Favoured rotamers (%)	91.00
Poor rotamers (%)	2.77
MolProbity score	1.88 (99^{th} percentile, N=4376, 2.83Å ± 0.25Å)
Average B factors	
Main chain ($Å^2$)	106
Side chains ($Å^2$)	106

Table 5.4. Refinement and validation statistics for BPSS0212/BPSS0213 complex 5.6. Statistics generated in Refmac5 (Murshudov et al., 1997; Murshudov et al., 2011), MolProbity (Chen et al., 2010) and Baverage (Winn et al., 2011).

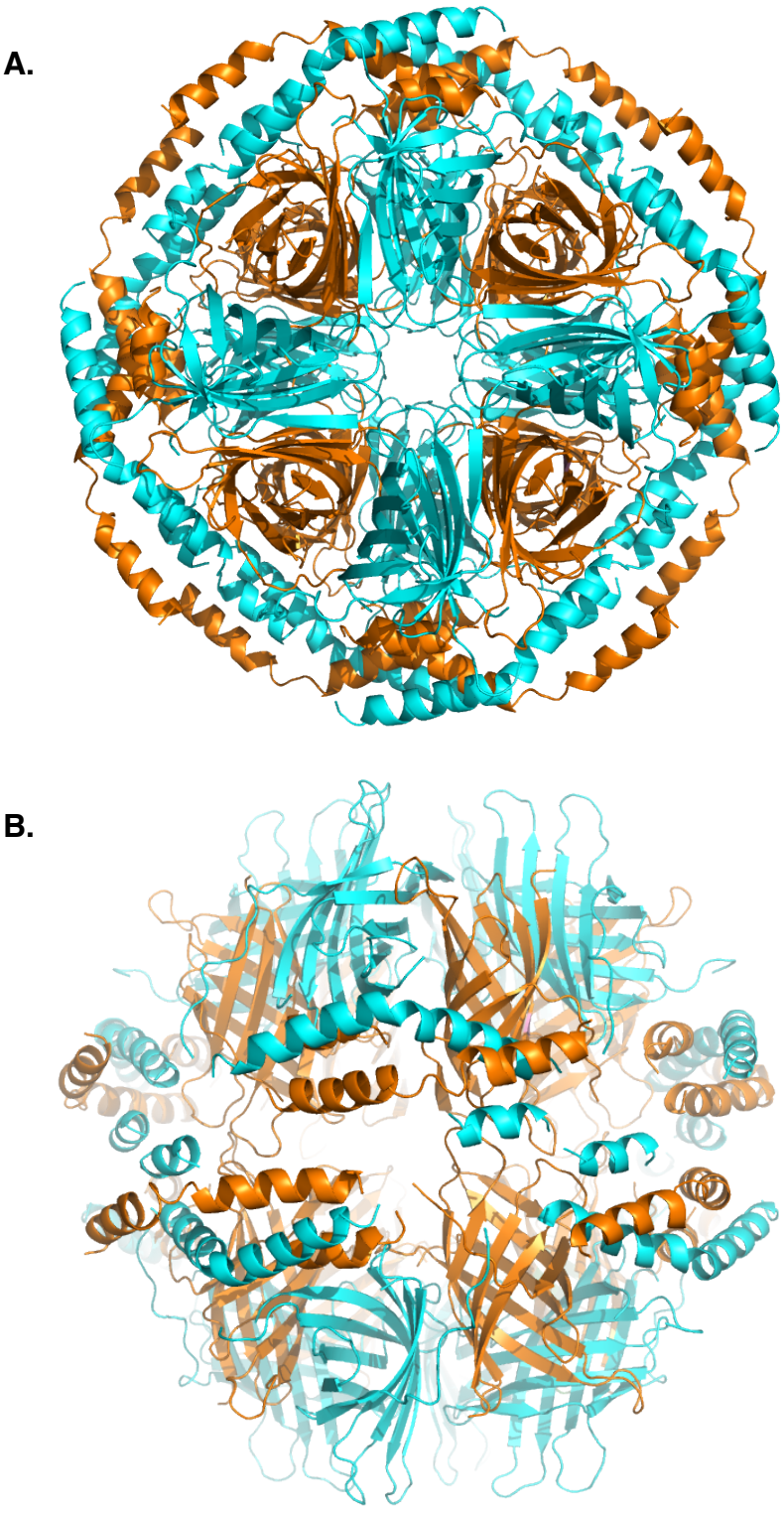


Figure 5.35. The current BPSS0212/BPSS0213 complex structure
The current BPSS0212/BPSS0213 $\alpha_8\beta_8$ complex 5.6 structure, with BPSS0212 (blue) and BPSS0213 (orange) shown as cartoon. **A.** Top view, parallel to 4-fold symmetry axis and **B.** Side view, perpendicular to 4-fold axis and parallel to 2-fold axis. Figure generated in PyMOL version 1.1r1.

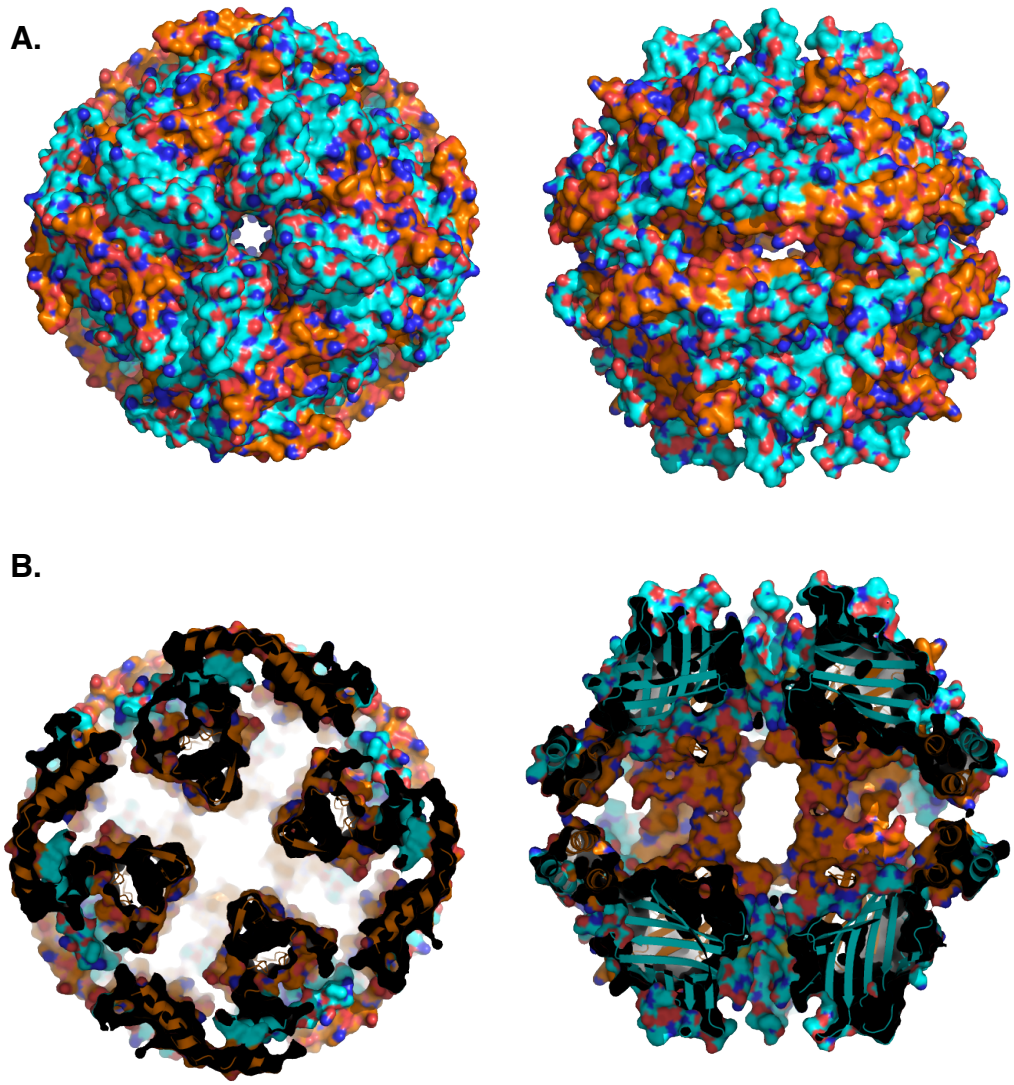


Figure 5.36. The current BPSS0212/BPSS0213 complex 5.6 structure surface and cavity. The current BPSS0212/BPSS0213 $\alpha_8\beta_8$ complex 5.6 structure, with BPSS0212 (blue) and BPSS0213 (orange) coloured by element, shown as cartoon, with the surface representation. **A.** Top view, parallel to 4-fold symmetry axis and side view, perpendicular to 4-fold axis and parallel to 2-fold axis. **B.** The cross-shaped internal cavity seen in the complex, top view - parallel to 4-fold axis. BPSS0213 (orange) is seen in the internal cavity, side view - perpendicular to 4-fold axis and parallel to 2-fold axis. Figure generated in PyMOL version 1.1r1.

strong patch of negative density, and a strong patch of positive density is seen along the side of the DUF1843 domain facing the complex (Figure 5.37).

The alternating arrangement of the BPSS0212 and BPSS0213 within the core results in two interacting faces for each subunit, on both sides of the β-barrel, in head-to-head and tail-to-tail interactions (Figure 5.38). This is consistent with the PDBePISA analysis, along with additional contacts involving the DUF1843 domain (Krissinel and Hendrick, 2007). The β5-β6 loops in both proteins, seem to form a loop-loop handshake in one of the two BPSS0212 and BPSS0213 interactions (Figure 5.39). Interestingly this same loop was shown to be flexible and adopt different conformations in structures of BPSS0212 described in section 4.6. One of these conformations in BPSS0212+M-FormA is the same as that observed in the complex. Ignoring the interaction of the DUF1843 domains and focusing on interactions involving the DUF1842 domains of BPSS0212 and BPSS0213, the residues involved in each BPSS0212/BPSS0213 interface are shown in Figures 5.40 and 5.41. Interface 1 has an area of ~1250 $Å^2$, compared to ~820 $Å^2$ for interface 2 (Krissinel and Hendrick, 2007). Of the residues involved in each interface, ~17-33% of residues are conserved between BPSS0212 and BPSS0213.

The interactions between the BPSS0212 and BPSS0213 DUF1842 domains in the $\alpha_4\beta_4$ core lie along axes of an approximate pseudo-2-fold NCS (non-crystallographic symmetry), which relate the BPSS0212 and BPSS0213 subunits. The head-to-head and tail-to-tail interactions lie every 45° around the 4-fold symmetry axis. However, this pseudo-2-fold symmetry is not reflected in the orientation of DUF1843 domain helical pairs. The DUF1843 domains are related by pseudo-2-fold symmetry, which is parallel to, but not coincident with the pseudo-symmetry relating the DUF1842 domains. The residues from BPSS0212 and BPSS0213 involved in interface 1 are contributed from equivalent regions in each

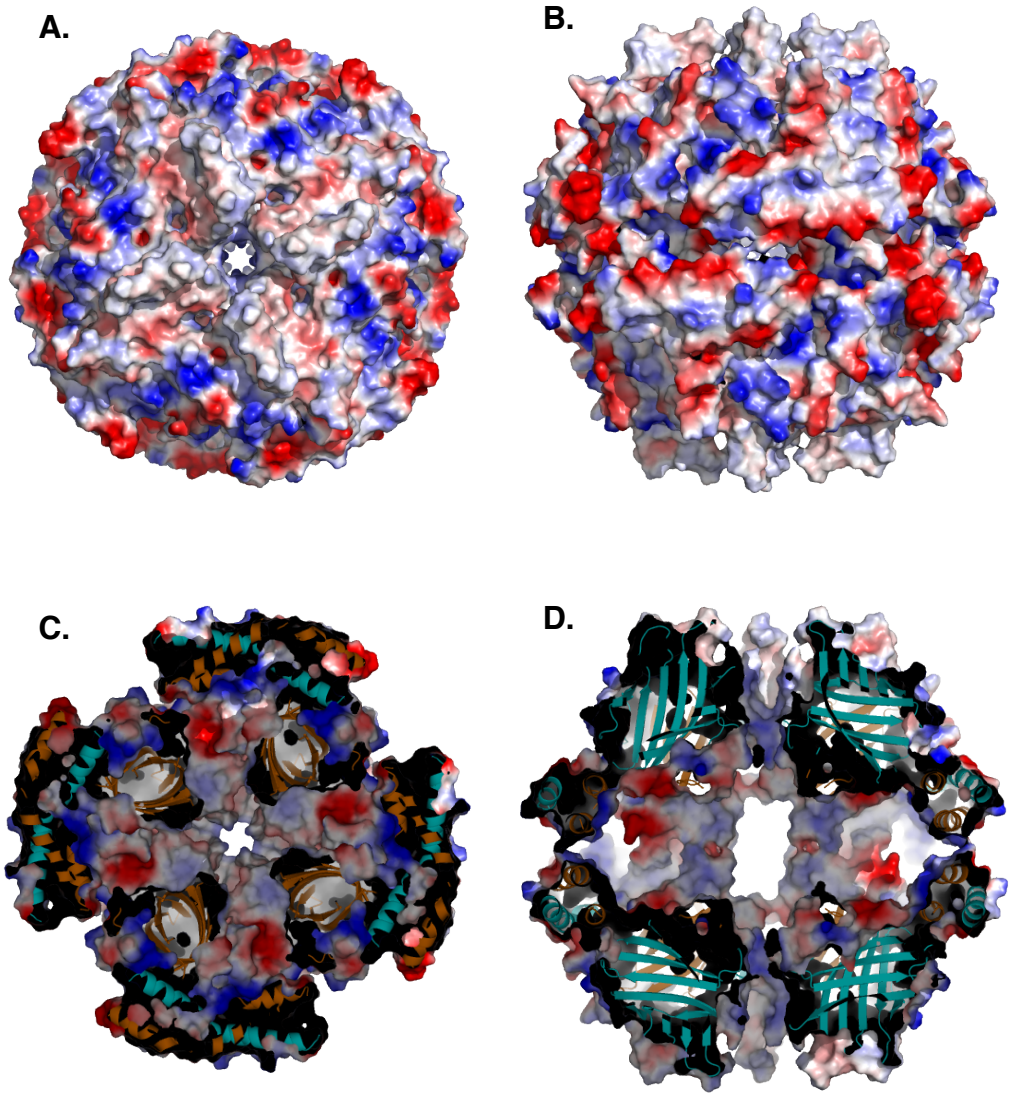


Figure 5.37. The surface charge of the current BPSS0212/BPSS0213 complex 5.6 structure and cavity. The current BPSS0212/BPSS0213 $\alpha_8\beta_8$ complex 5.6 structure, with BPSS0212 (blue) and BPSS0213 (orange) shown as cartoon, with surface representation. **A.** Top view, parallel to 4-fold symmetry axis, shows fewer strong charge patches than the side view **B**., perpendicular to 4-fold axis and parallel to 2-fold axis. **C.** The top view of the cross-shaped internal cavity seen in the complex, which shows strong patches of negative charge (red). **D.** The side view of the internal cavity, which shows no strong charge patches on the central cavity regions corresponding to BPSS0213. Figure generated in PyMOL version 1.1r1.

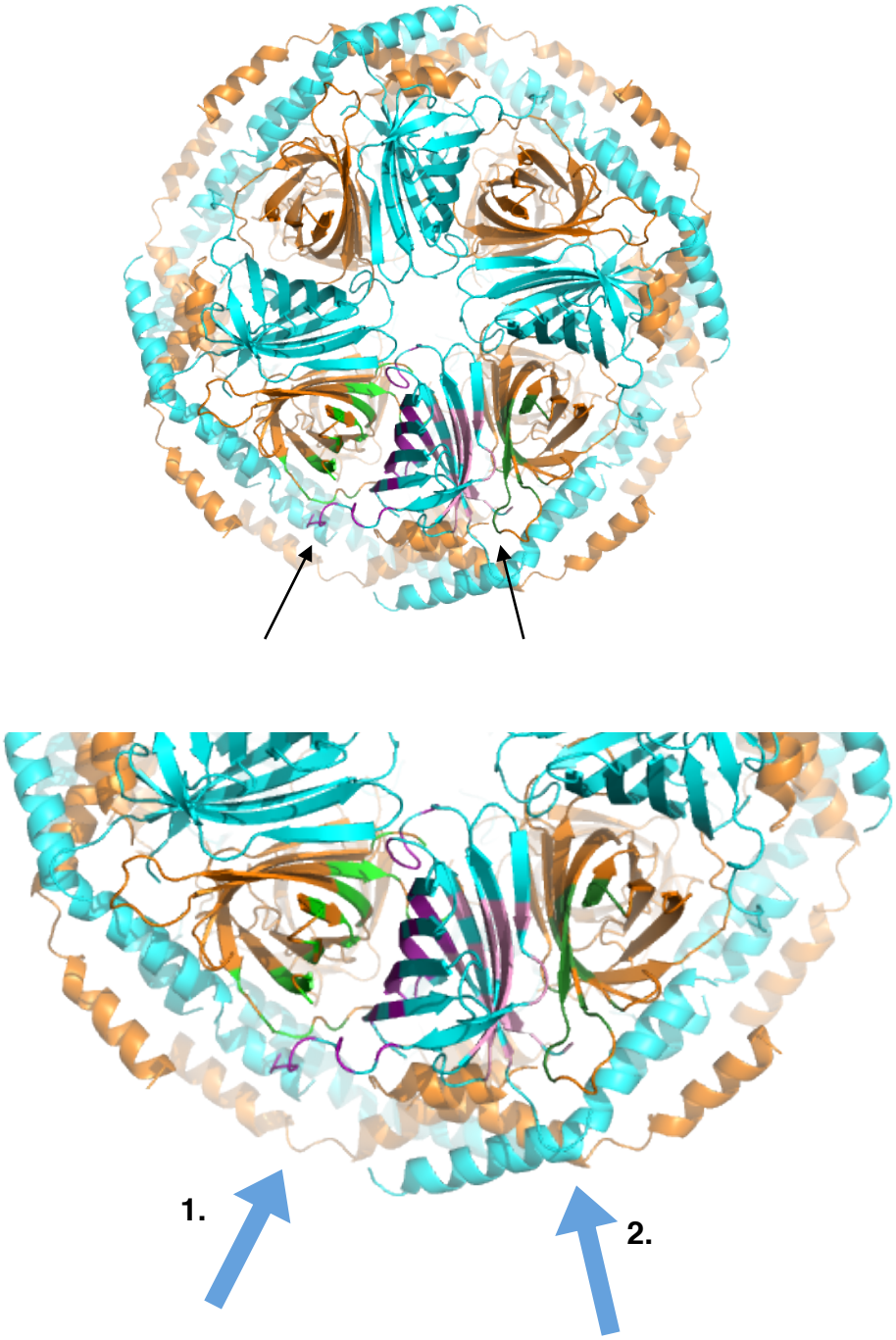


Figure 5.38. Two possible interfaces between BPSS0212 and BPSS0213 present on the pseudo-2-fold axes in the complex. The current BPSS0212/BPSS0213 $\alpha_8\beta_8$ complex 5.6 structure, with BPSS0212 (blue) and BPSS0213 (orange), shown as cartoon. The two interfaces are identified with arrows. Residues in interface 1 contributed by BPSS0212 are coloured purple and those from BPSS0213 are coloured light green. Residues in interface 2 contributed by BPSS0212 are coloured pink and those from BPSS0213 are coloured dark green. Interface information from PDBePISA (Krissinel and Hendrick, 2007) and images generated in PyMOL version 1.1r1.

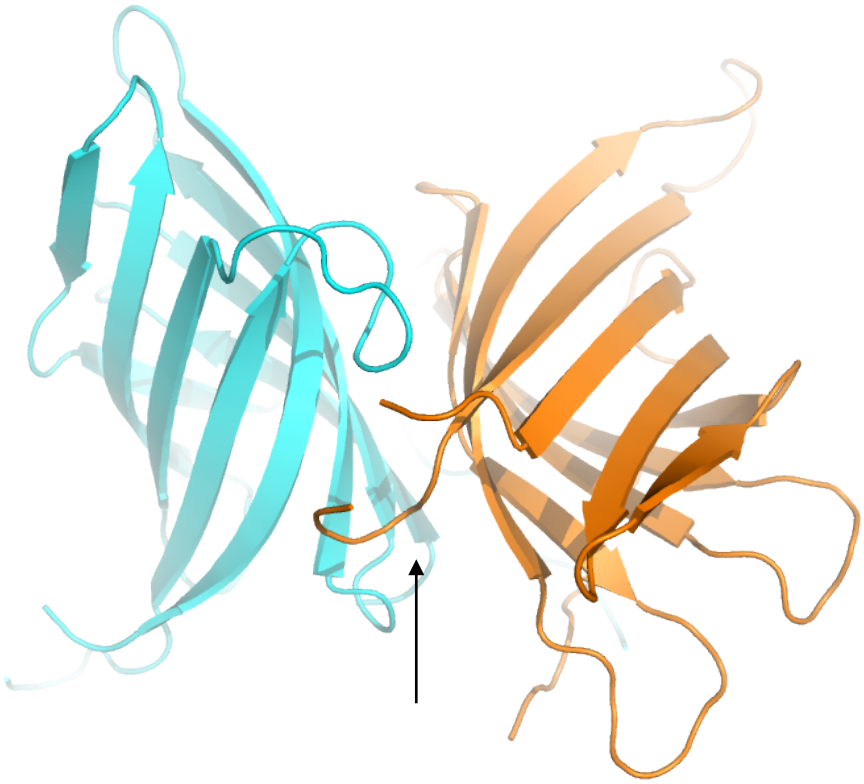


Figure 5.39. The β5-β6 loop handshake between BPSS0212 and BPSS0213. BPSS0212 (blue) and BPSS0213 (orange) shown as cartoon, with the β5-β6 loops in the foreground, which is not fully modelled in BPSS0213. Figure generated in PyMOL version 1.1r1.

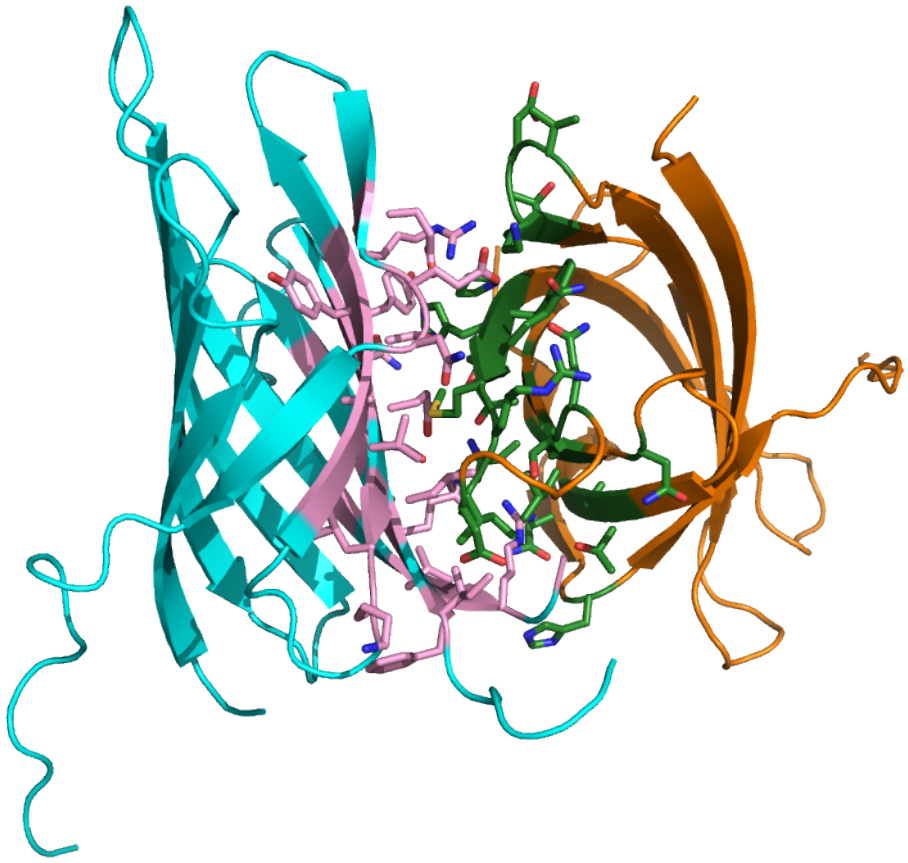
BPSS0212

~~MSEDLR~~VG**L**F**P**V**R**YLVGTGLPGAPQLVL**DLMVD**T**VD**H**SVVGR**A**A**V**SQAVSPPLN
F**H**A**DVWG**SYVFRLG~~PPPR~~RDGSGAIVQISLQGNQ**GGPQ**S**N**SMITFYGELLLKGD
GKTGVASYRYYSNGSWHEVENVPVKA**D**P**ELV**P**IEPGPV**~~IGQSSMSAIGSAAM~~YG
VAIQSAAAS~~GD~~LAHMRTLSAYARQQLESRDEIAAALSELKAEIAKL~~ESRQ~~

BPSS0213

~~MA~~**T**G**LFPVQ**LRVATPNLGAPVLWL**NLL**V**N**TV**E**K**T**A**SGF**ARITQTVYPPMHF**R**A
RVVGPFHQMRIDPHAPQSVTLTL**SGS**P**TG**PV~~AP~~**Q**V**V**ILELNALLNEGWQSGTAN
YRYFYESRWHSIEHAIVS**K**D**NSR**I**P**LD~~PPSEHVM~~PMYGVGLQEARAS~~GD~~LSRMK
ALAQQAEQQLADHDVIAAELQKLEAEIA~~RLEARR~~

Figure 5.40. Interface 1 between BPSS0212/BPSS0213 in the complex core. The current BPSS0212/BPSS0213 structure, from complex 5.6 with BPSS0212 (blue) and BPSS0213 (orange), shown as cartoon. Residues in interface 1 contributed by BPSS0212 are coloured purple and from BPSS0213 are coloured light green, and are shown as sticks. The interfacing residues are highlighted in the BPSS0212 and BPSS0213 sequences, with crossed through residues absent in the structure. Interface information from PDBePISA (Krissinel and Hendrick, 2007) and image generated in PyMOL version 1.1r1.



BPSS0212

~~MSEDLRV~~GLFPVRYLVGTGLPGAPQLVLDLMVDTVDHSVVGRAAVSQAVSPPLN
FHADVWG**S**Y**VFR**L**G**~~PPPR~~RDGSGA**I**V**Q**I**S**L**Q**GNQGGPQSNSMITF**YGE**L**LLK**GD
GKT**GVASYR**YYSNGSWH**E**V**EN**VPVKADPELVPIEPGPV~~IGQSSMSAIGSAAM~~YG
VAIQSAAASGDLAHMRTLSAYARQQLESRDEIAAALSELKAEIAKL~~ESRQ~~

BPSS0213

~~MAT~~TGLFPVQLRVATPNLGAPVLWLNLLVNT**VEKT**ASGFARITQTVYPPMHFRA
RVVG**PFHQMRID**PHAP**QS**V**T**L**T**LSGSPTGPV~~AP~~QVVILEL**NAL**L**N**EGWQSG**TAN**
YRYFYESRWHSIE**H**AIVSKDNSRIP~~LDPPSEHVM~~PMYGVGLQEARASGDLSRMK
ALAQQAEQQLADHDVIAAELQKLEAEIA~~RLEARR~~

Figure 5.41. Interface 2 between BPSS0212/BPSS0213 in the complex core. The current BPSS0212/BPSS0213 structure, from complex 5.6 with BPSS0212 (blue) and BPSS0213 (orange), shown as cartoon. Residues in interface 1 contributed by BPSS0212 are coloured pink and those from BPSS0213 are coloured dark green, and are shown as sticks. The interfacing residues are highlighted in the BPSS0212 and BPSS0213 sequences, with crossed through residues absent in the structure. Interface information from PDBePISA (Krissinel and Hendrick, 2007) and image generated in PyMOL version 1.1r1.

protein (Figure 5.40). This is exhibited less strongly for interface 2 (Figure 5.41).

5.6.10 Comparison of BPSS0212 and BPSS0213 structures

The structure of the BPSS0212/BPSS0213 complex has provided the structure of BPSS0213, which did not crystallise when purified and set down into crystallisation trials alone. Therefore two DUF1842 domain structures, with partial interdomain linkers, have now been solved. The overall fold of both structures is very similar (Figure 5.42), with slight difference in the length of some of the β-strands and the presence of a break in strand β3 of BPSS0213, which is also present in the structure of BPSS0212+M-FormA. The structure of BPSS0212+M-FormA, of the DUF1842 domain and part of the interdomain linker, was compared to the equivalent region BPSS0212 in the complex. The overall fold is the same, with an extra 11 residues at the C-terminal of the complex BPSS0212. In the complex BPSS0212 strands β3, β8 and β9 are slightly longer, and the break in strand β3 is no longer present.

5.6.11 Comparison of BPSS0211 dimer and BPSS0212/BPSS0213 DUF1843 domain structures

The folds of the DUF1843 domains of both BPSS0212 and BPSS0213 in the complex are closely related to that seen in the BPSS0211 closed helical structure produced previously (Day, 2012), and as such could be modelled from it. The BPSS0212/BPSS0213 dimer interface was compared to the BPSS0211 closed conformation dimer structure interface (Day, 2012), using PDBePISA (Krissinel and Hendrick, 2007), which showed that similar residues are involved in the interface (Figure 5.43). Therefore, this suggests that the DUF1843 domain acts a protein:protein interaction domain, as predicted previously (Day, 2012). The overall formation of the helices and helical packing is conserved between the BPSS0211 closed conformation dimer (Day, 2012) and the BPSS0212/BPSS0213 dimer.

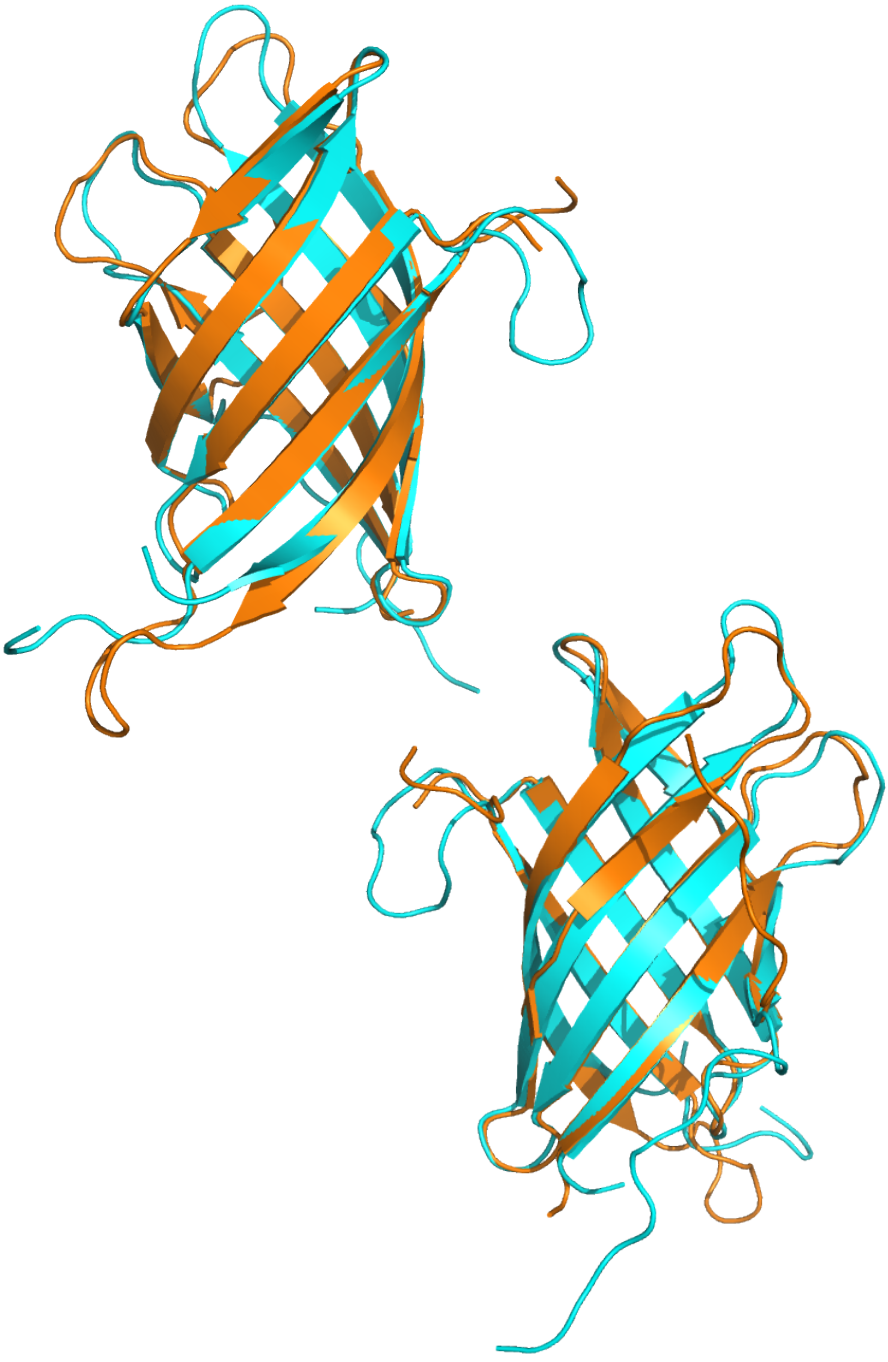


Figure 5.42. BPSS0212 vs BPSS0213
The structural comparison of BPSS0212 and BPSS0213, using structures from complex 5.6, with BPSS0212 (blue) and BPSS0213 (orange). Structures aligned in *Coot* using SSM (Emsley et al., 2010; Krissinel and Hendrick, 2004). Figure generated in PyMOL version 1.1r1.

A.

BPSS0211

P-YGVAIHQAIADGDLAQMKSLRTQAQALLAQQGNLATALELLEVEIAKLE~~RRK~~

BPSS0212

~~AM~~YGVAIQSAAA~~SGD~~LAHMRTLSAYARQQLESRDEIAAALSELKAEIAKL~~ESRQ~~

BPSS0213

PMYGVGLQEARAS~~GD~~LSRMKALAQQAEQQLADHDVIAAELQKLEAEIA~~RLEARR~~

B.

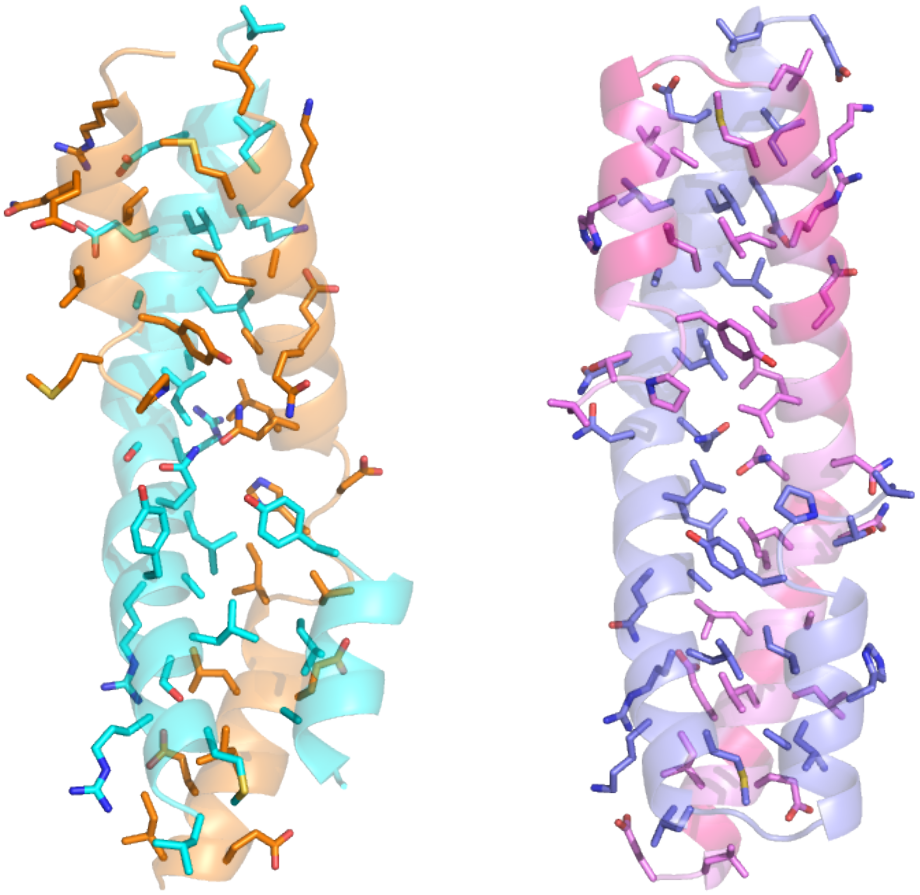


Figure 5.43. Comparison of BPSS0211, BPSS0212 and BPSS0213 DUF1843 domains. A. The DUF1843 domain sequences from BPSS0211, BPSS0212 and BPSS0213 are shown, with missing residues in the structure crossed through and the residues involved in the predicted interface (by PDBePISA – Krissinel and Hendrick, 2007) highlighted in yellow. The BPSS0211 sequence has an insertion to allow the alignment of all three sequences. **B.** The interfacing residues between the three aligned DUF1843 domains are shown as sidechains in the BPSS0212/BPSS0213 complex (blue/orange, respectively) and the BPSS0211 closed conformation A-B dimer (pink/purple). Figure generated in PyMOL version 1.1r1. BPSS0211 closed conformation coordinates were produced by Dr. Matthew Day (Day, 2012).

5.7 Complex formation in the presence of EDTA

The structure of the BPSS0212/BPSS0213 complex thus far contains the DUF1842 domain, part of the interdomain linker and the DUF1843 domain for both proteins. Unfortunately there is a lack of electron density to complete the linker connection between the DUF1842 domain and DUF1843 domain for each protein. Therefore there is still some ambiguity as to which of the DUF1843 domains correspond to which DUF1842 domain of the appropriate protein. Given this, a more complete electron density map would be beneficial and so attempts were made to make better protein.

5.7.1 Complex formation with protease inhibitors and EDTA

The lack of density may be attributed to partial degradation of the complex affecting the quality of the crystals produced, leading to poorly diffracting crystals and low-resolution data. Therefore, to prevent degradation of the separate proteins, the purification of BPSS0212 and BPSS0213, and the formation of the BPSS0212/BPSS0213 complex were carried out in the presence of EDTA and EDTA-free protease inhibitors. A tablet of EDTA-free protease inhibitors was added to the lysis buffer and 2 mM EDTA was added to all other buffers during purification. BPSS0212 and BPSS0213 were purified separately using DEAE anion-exchange chromatography, followed by gel filtration chromatography, as outlined in section 2.7.8.

In previous experiments the smaller BPSS0212/BPSS0213 complex produced in 50 mM Tris pH 8.0, 500 mM NaCl was prepared for crystallisation by putting the protein into 10 mM Tris pH 8.0, 100 mM NaCl. Whether this change in salt concentration leads to the formation of the 16mer was analysed by producing the complex in high salt conditions (50 mM Tris pH 8.0, 500 mM NaCl, 2 mM EDTA) and then taking this complex and loading it onto a gel filtration column equilibrated with lower salt conditions of 50 mM HEPES pH 7.5, 100 mM NaCl, 2 mM EDTA.

5.7.2 Degradation analysis of BPSS0212 purified in the presence of protease inhibitors and EDTA

The addition of EDTA-free protease inhibitors to the cell lysis buffer and 2 mM EDTA in all further purification buffers proved to significantly decrease the degradation of BPSS0212, seen previously during the gel filtration purification step (Figure 5.44).

5.7.3 BPSS0212/BPSS0213 complex formation in high salt conditions

The formation of the complex under high salt conditions (50 mM Tris pH 8.0, 500 mM NaCl, 2 mM EDTA) lead to a prominent peak at ~62 mL, representing the smaller molecular weight complex seen in previous high salt gel filtration co-purifications, ~126 kDa, estimated from the calibration graph in Figure 5.10. The shoulder seen at the tail end of the first peak still contains both BPSS0212 and BPSS0213, but with a degradation band also seen in the BPSS0212 pre-combination sample. This degradation of BPSS0212 may be affecting the way the complex runs on the gel filtration column. The third peak seen at ~80 mL contains BPSS0212 (Figures 5.45 and 5.46).

5.7.4 BPSS0212/BPSS0213 complex formed in high salt conditions run on gel filtration in low salt conditions

5 mL from the combined elution fractions (15-20) from the first peak of the high salt gel filtration run was concentrated and loaded onto the gel filtration column equilibrated with low salt buffer – 50 mM HEPES pH 7.5, 100 mM NaCl, 2 mM EDTA. The chromatogram produced a symmetric peak at ~51 mL (peak 1) and a broader peak at ~71 mL (peak 2) (Figure 5.47). Analysis of the peak 1 fractions by SDS-PAGE showed both BPSS0212 and BPSS0213 proteins were present (Figure 5.46). This first peak elution volume corresponds to a molecular weight of ~355 kDa, using the low salt calibration graph (Figure 5.24). This suggests this peak contains the $\alpha_8\beta_8$ 16-mer complex of BPSS0212 and BPSS0213.

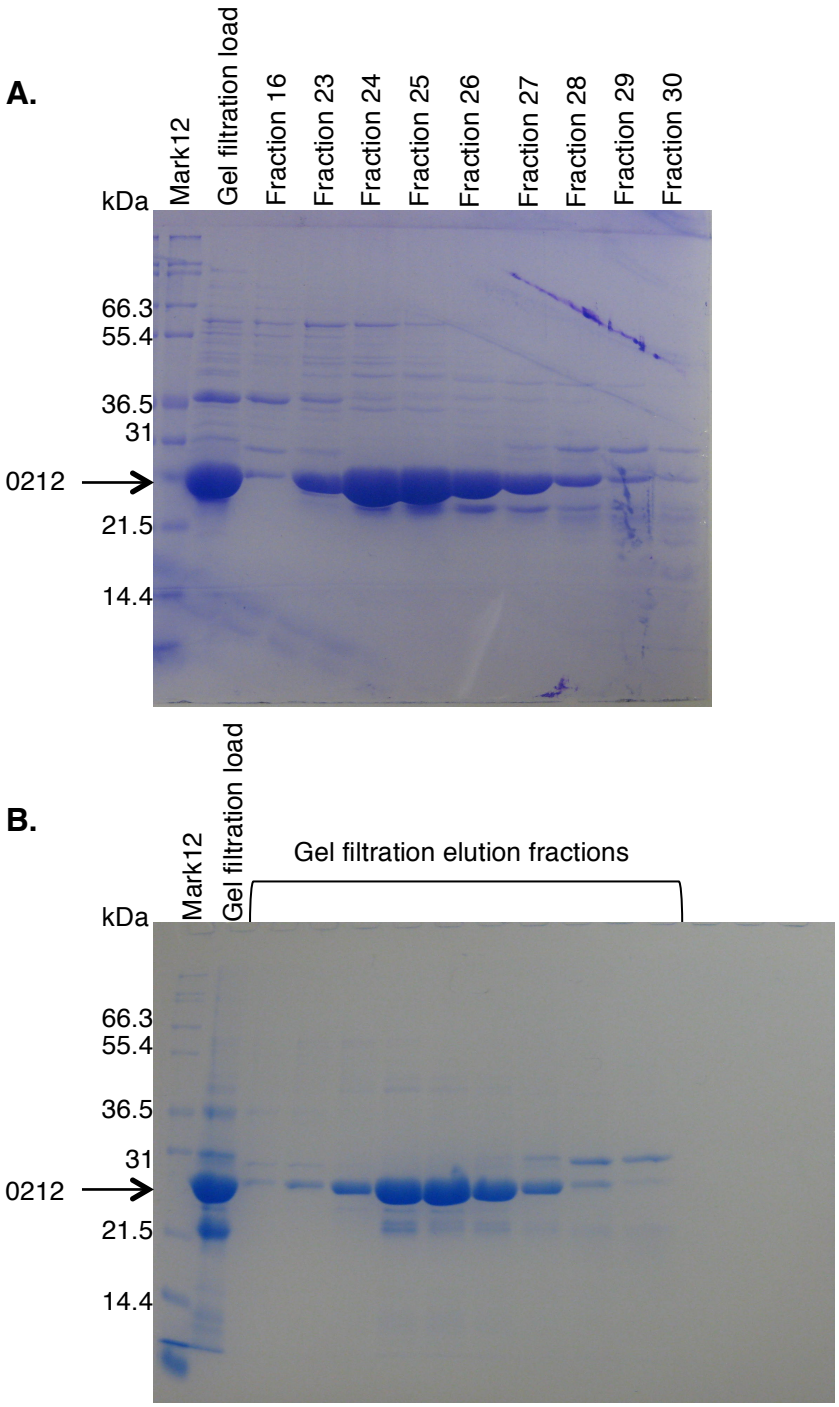


Figure 5.44. SDS-PAGE analysis of BPSS0212 gel filtration purification in the presence and absence of EDTA. A. Purification of BPSS0212 with protease inhibitors added to lysis buffer and 2 mM EDTA added to all buffers. Gel filtration run with 50 mM Tris pH 8.0, 500 mM NaCl, 2 mM EDTA. **B.** Purification in the absence of protease inhibitors and EDTA. Gel filtration run with 50 mM Tris pH 8.0, 500 mM NaCl.

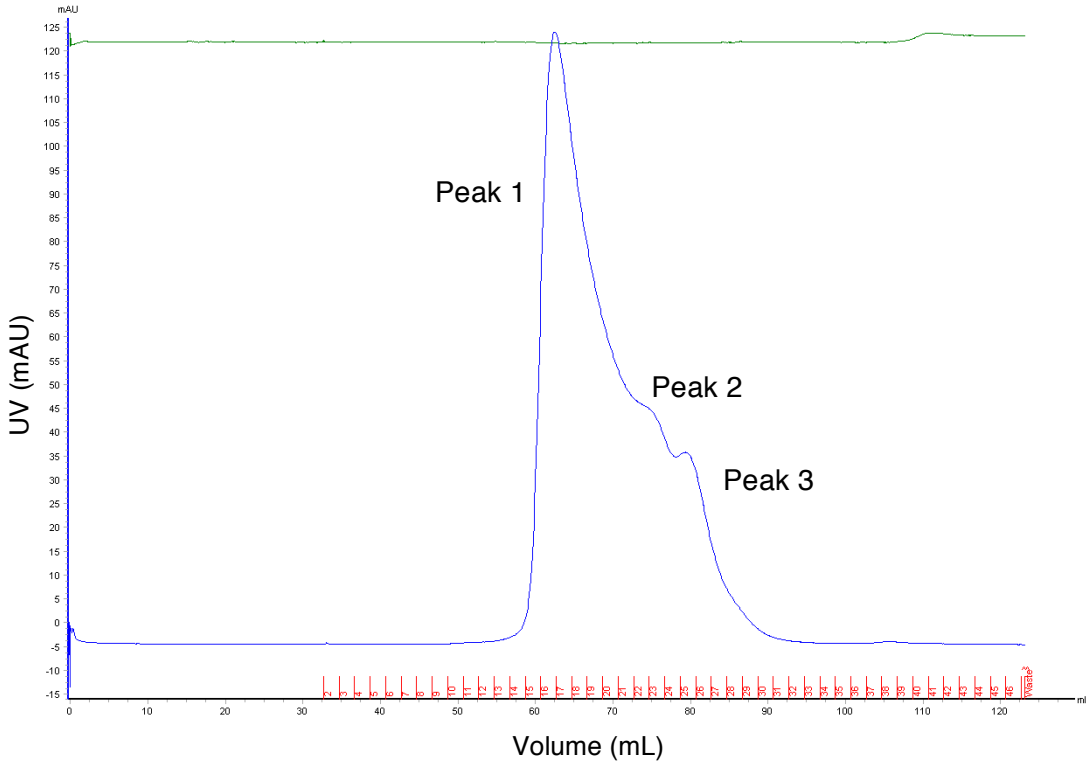


Figure 5.45. Chromatogram of BPSS0212 and BPSS0213 gel filtration experiment, run with 50 mM Tris pH 8.0, 500 mM NaCl, 2 mM EDTA. mAU at A_{280} – blue , conductivity – green, elution fractions – red numbers. The elution peaks are labelled and the corresponding elution volumes noted. Peak 1 - 62.49 mL, Peak 2 - ~75 mL and Peak 3 - ~79.4 mL. Fractions 15-20 were pooled, run on low salt gel filtration and used in crystallisation trials.

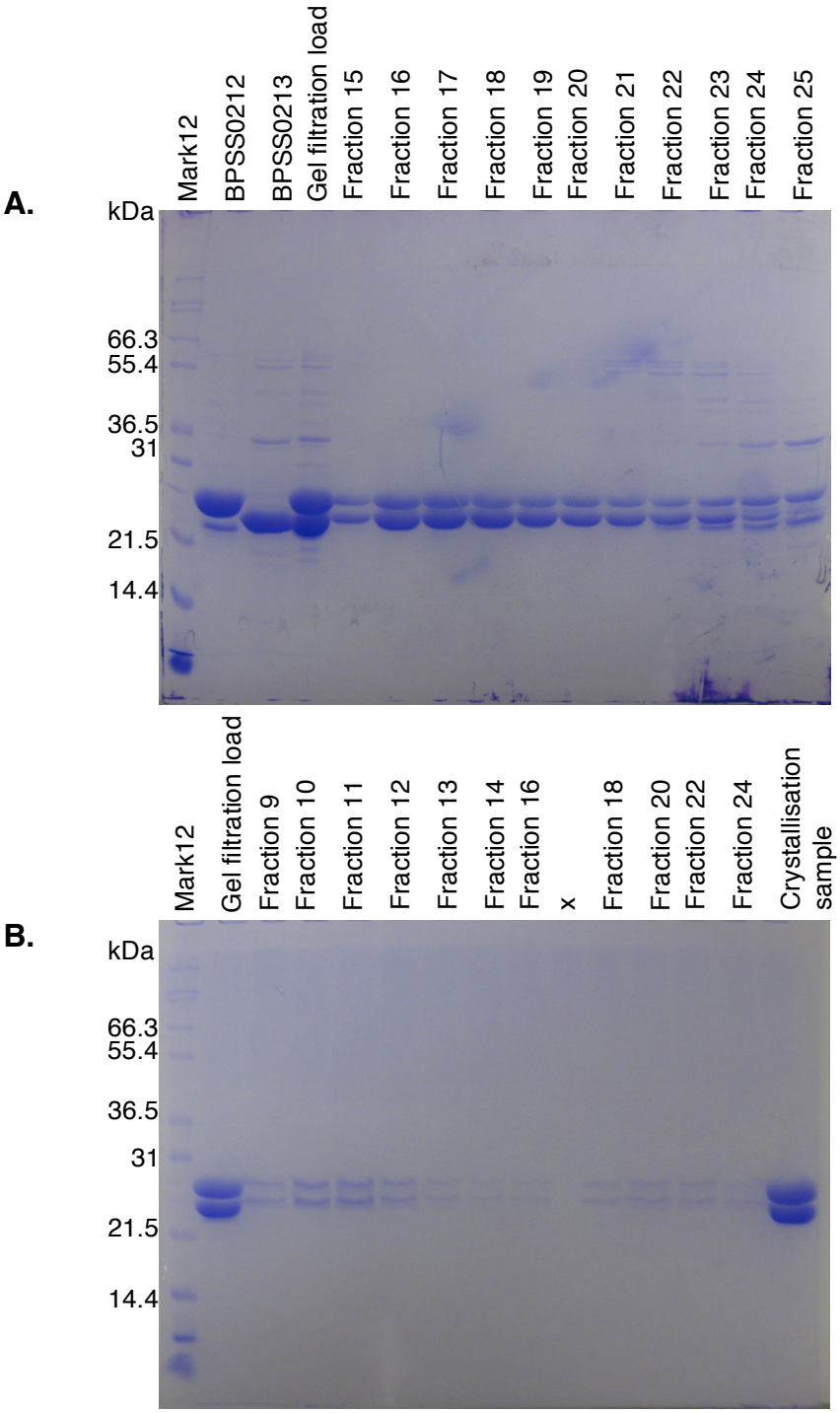


Figure 5.46. SDS-PAGE analysis of the BPSS0212 and BPSS0213 gel filtration co-purification, run at high and low salt concentration with 2 mM EDTA. BPSS0212 is 22465 Da and BPSS0213 is 21668 Da. x indicates an empty lane. **A.** High salt – elution fractions analysed cover peaks 1-3. Fractions 15-20 were pooled, and 5 mL run on the low salt gel filtration column. **B.** Low salt - Fractions 9-14 - peak 1, fractions 16-24 - peak 2, and the crystallisation sample – fractions 15-20 from the high salt gel filtration run. However there is some overlapping of the elution peaks.

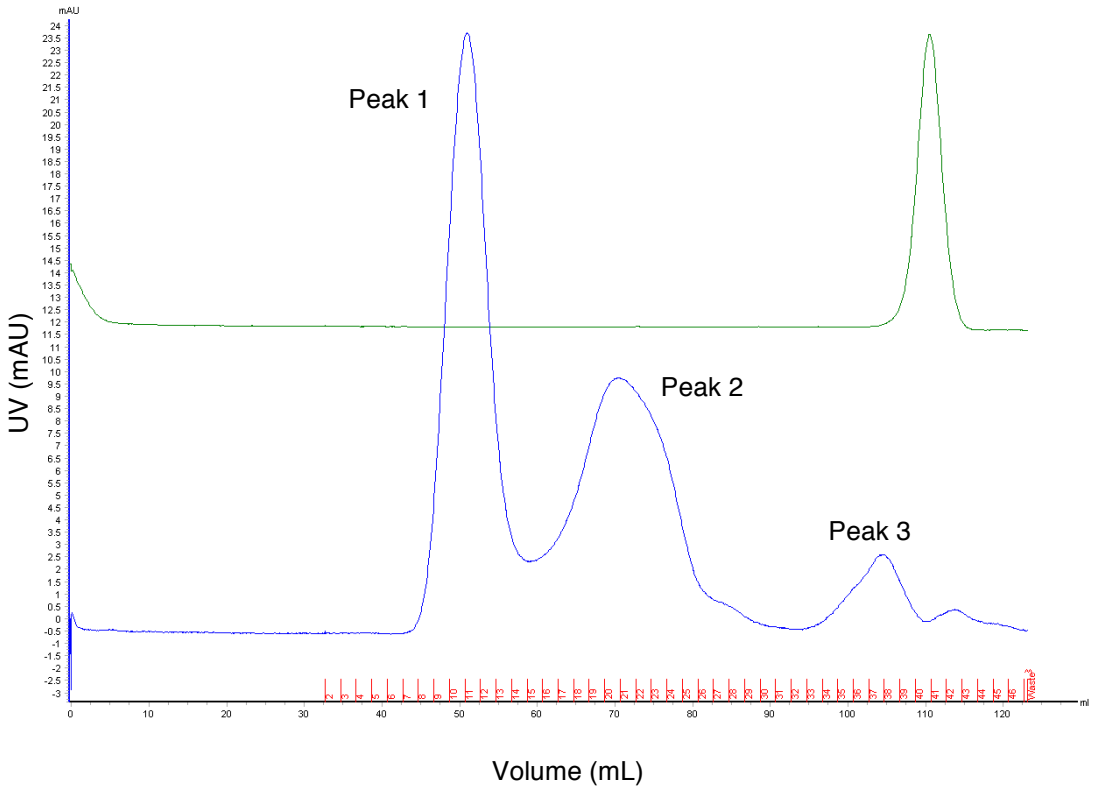


Figure 5.47. Chromatogram of BPSS0212 and BPSS0213 gel filtration experiment, run with 50 mM HEPES pH 7.5, 100 mM NaCl, 2 mM EDTA. mAU at A_{280} – blue , conductivity – green, elution fractions – red numbers. The elution peaks are labelled and the corresponding elution volumes noted. Peak 1 – 50.92 mL, Peak 2 – 70.60 mL and Peak 3 - 104.44 mL.

This is a similar molecular weight to that seen in the previous low salt gel filtration experiments of BPSS0212(+M-SeMet) with BPSS0213. The peak 2 fractions also contain both BPSS0212 and BPSS0213, the smaller elution volume suggests the complex is possibly starting to dissociate during the gel filtration run. These results suggest that when the complex formed in high salt, is transferred into low salt conditions, as in crystallisation preparations, the $\alpha_8\beta_8$ 16-mer complex is produced.

5.7.5 Crystallisation trials of the complex formed in high salt and in the presence of EDTA

The BPSS0212/BPSS0213 complex formed in 50 mM Tris pH 8.0, 500 mM NaCl, 2 mM EDTA was set down in sitting-drop crystallisation trials using the Proplex and Morpheus screens, at ~9 mg/mL. Crystal hits were produced in the Morpheus screen, which contains a selection of additive molecules (Gorrec, 2009). The conditions which produced crystals, E2, F2 and H2 (Figure 5.48), contain the same buffer and precipitant components, but with the addition of different small molecules – ethylene glycols, monosaccharides and amino acids, respectively (Gorrec, 2009). These small molecules are added to aid crystallisation (Gorrec, 2009) however they may bind to a certain part of the complex thus providing insight into the function of the BPSS0212/BPSS0213 complex.

These three crystal conditions from the Morpheus screen, E2, F2 and H2 were optimised to try and produce larger crystals. Sitting-drop optimisation experiments were set down for all three conditions using a 1:1:1 ratio of protein:reservoir:miliQ, 2:1 ratio of protein:reservoir and a 1:2 ratio of protein:reservoir, each with a 30 µL reservoir and 0.5 µL component volume.

Hanging-drop optimisation crystallisation trials of condition Proplex B4, which has produced the crystals used to solve all the complex structures previously, were set up. Optimisation trials were set up with a reservoir of

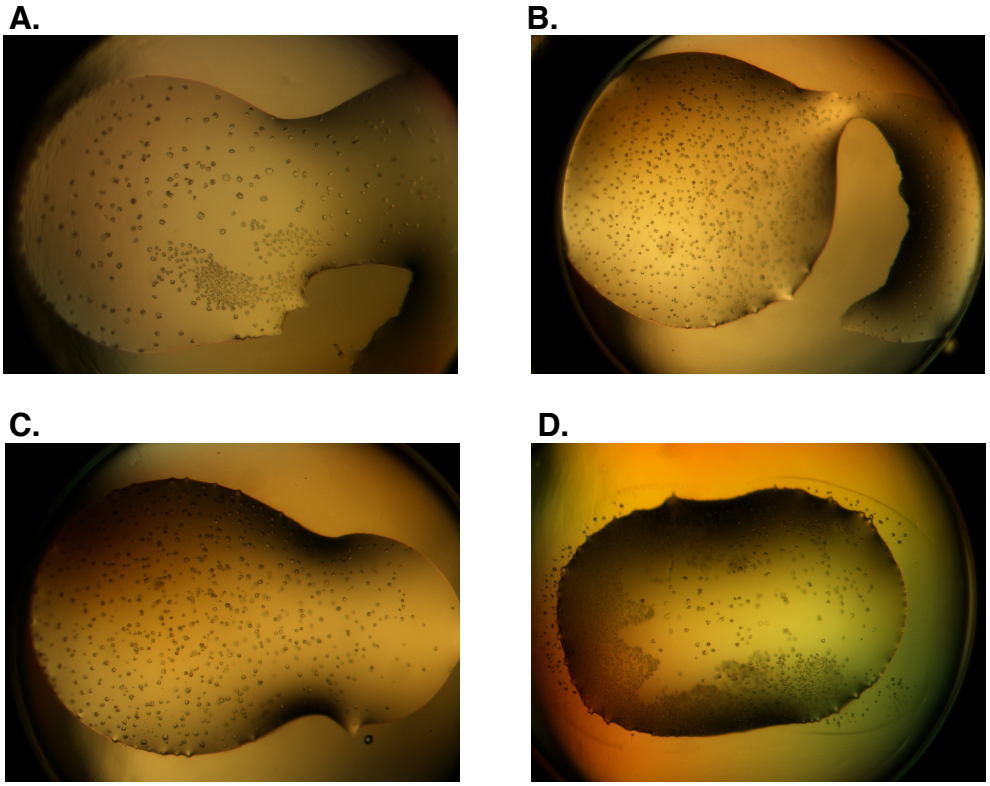


Figure 5.48. Crystal hits from sitting drop crystallisation trials for BPSS0212/BPSS0213 complex formed in the presence of EDTA.
A. Morpheus H2 – 0.1 M amino acids, 0.1 M Buffer System 1 pH 6.5 (0.1 M MES/imidazole pH 6.5) and 50% (v/v) Precipitant Mix 2 (10% (w/v) PEG8000, 20% (v/v) ethylene glycol). **B.** Morpheus F2 – 0.12 M monosaccharides, 0.1 M Buffer System 1 pH 6.5 (0.1 M MES/imidazole pH 6.5) and 50% (v/v) Precipitant Mix 2 (10% (w/v) PEG8000, 20% (v/v) ethylene glycol). **C.** Morpheus E2 – 0.12 M ethylene glycols, 0.1 M Buffer System 1 pH 6.5 (0.1 M MES/imidazole pH 6.5) and 50% (v/v) Precipitant Mix 2 (10% (w/v) PEG8000, 20% (v/v) ethylene glycol). **D.** Proplex F3 – 0.1 M magnesium acetate, 0.1 M MES pH 6.5, 10% (w/v) PEG10000.

0.2 M NaCl, 0.1 M MES buffer pH 6.5 and 6-15% (w/v) PEG4000 in increments of 1%, and 0.1 M MES buffer pH 6.5, 10% (w/v) PEG4000 and 0.05 M-0.35 M NaCl in increments of 0.05 M, with 1000 µL reservoir and 2 µL:2 µL protein:reservoir ratio drops. Optimisation trays were stored at 7°C. A number of crystals were produced in these conditions, notably two large crystals grew together in one drop (Figure 5.49).

Crystals from condition 0.1 M MES buffer pH 6.5, 0.35 M NaCl and 10%(w/v) PEG4000 (Figure 5.49) were cryoprotected, cooled and sent to the Diamond Light Source Synchrotron, Oxford, for screening. Unfortunately the data produced was of a similar quality to that collected previously and analysis of the electron density map revealed no further information as to the connectivity of the domains in the complex.

5.8 The effect of the removal of the DUF1843 domain on BPSS0212/BPSS0213 complex formation

Both proteins in the complex – BPSS0212 and BPSS0213 - contain a DUF1842 domain, part of an interdomain linker and a DUF1843 domain. The current model of the BPSS0212/BPSS0213 complex shows the DUF1843 domain helices, from both proteins, are located on the periphery of the DUF1842 domain complex core. The DUF1843 domains interact in 1:1 ratio of BPSS0212:BPSS0213. The DUF1842 domains and partial linker alone interact with each other via interfaces on both sides of each protein leading to the formation of a $\alpha_4\beta_4$ complex identified by PDBePISA (Krissinel and Hendrick, 2007), however no $\alpha_8\beta_8$ 16mer complex was identified as stable in solution for the DUF1842 domains alone in complex 5.3, 5.4 and DUF1842 domains alone of 5.6. This suggests little or no interaction between the DUF1842 domains in the top $\alpha_4\beta_4$ sub-complex and the DUF1842 domains in the bottom $\alpha_4\beta_4$ sub-complex, evident from the structure (Figure 5.50). Therefore the formation of the $\alpha_8\beta_8$ complex most likely involves the DUF1843 domains. The DUF1843 domain helices lie in two bands, around the middle of the

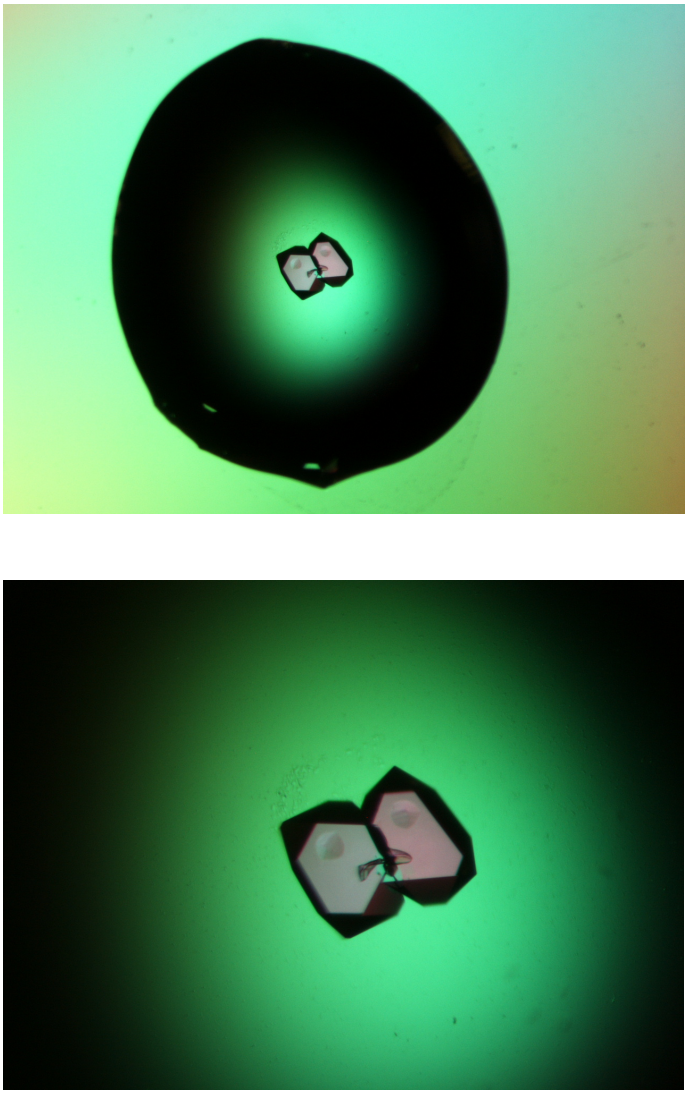


Figure 5.49. Crystal hits from hanging drop optimisation trials for BPSS0212/BPSS0213 complex formed in presence of EDTA. Proplex B4 optimisation conditions - 0.1 M MES pH 6.5, 0.35 M NaCl, 10% (w/v) PEG4000) and stored at 7°C.

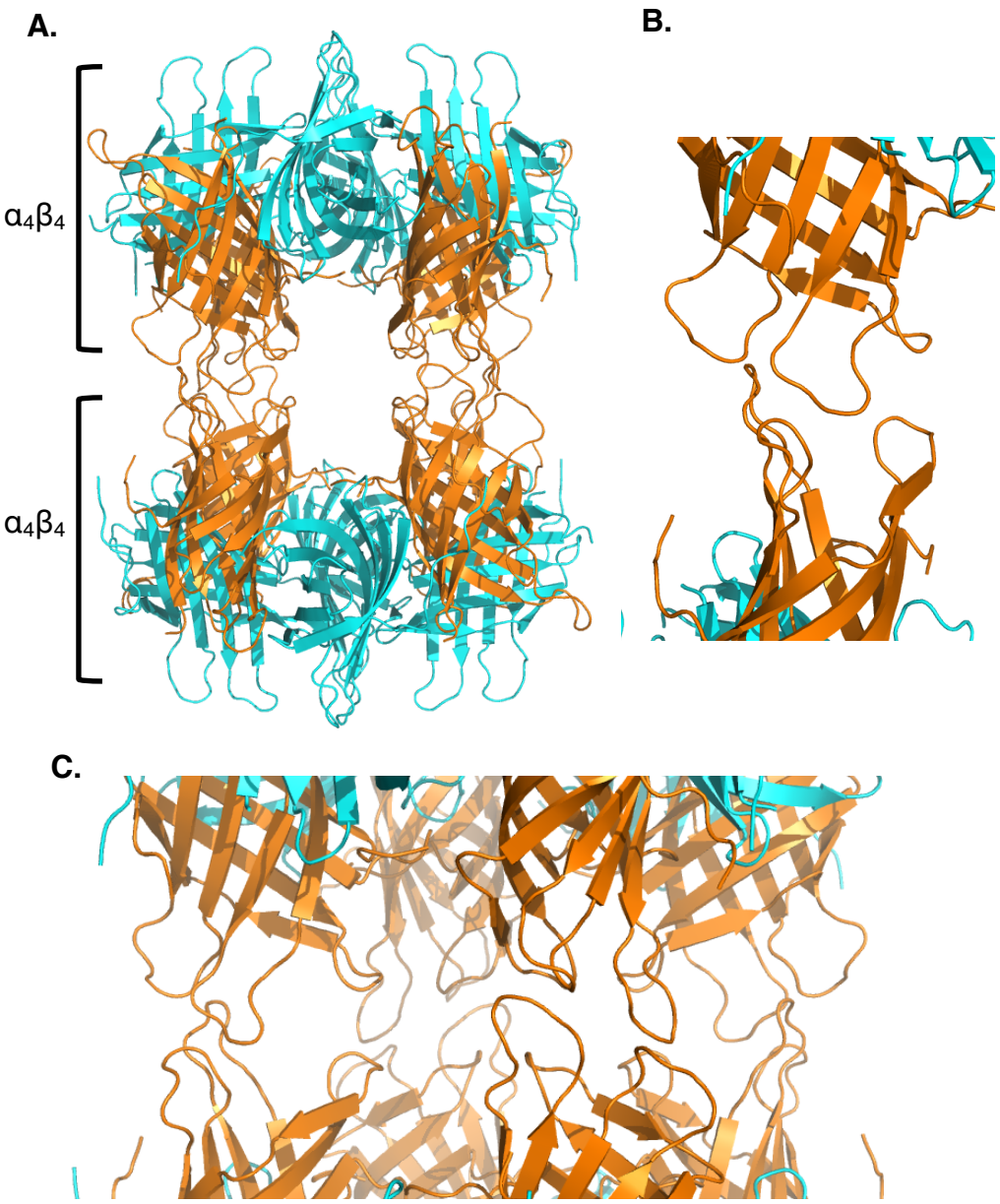


Figure 5.50. Interaction between the two $\alpha_4\beta_4$ subcomplexes
A. The two core $\alpha_4\beta_4$ sub-complexes, from the BPSS0212/BPSS0213 complex 5.6, with BPSS0212 (blue) and BPSS0213 (orange), shown as cartoon. The interaction between two BPSS0213 proteins of the two sub-complexes, and the cumulative interaction between all eight BPSS0213 of the two sub-complexes are shown in **B.** and **C**., respectively. Figure generated in PyMOL version 1.1r1.

complex. It may be that a number of DUF1843 domains connected to the top $\alpha_4\beta_4$ sub-complex layer interact with a number of DUF1843 domains from the bottom $\alpha_4\beta_4$ sub-complex layer. Therefore there may exist a cross over connection between the DUF1843 domains in the two layers of the complex.

5.8.1 Removal of the DUF1843 domain from BPSS0212 and BPSS0213

To determine whether formation of the $\alpha_8\beta_8$ complex is dependent on the interactions between the DUF1843 domain helices, the DUF1843 domain helix was removed from both BPSS0212 and BPSS0213, and the effect on complex formation investigated. Constructs were designed for BPSS0212 and BPSS0213 without the DUF1843 domain, based on the structure and the DUF1843 domain conservation sequence from Pfam entry PF08898 (Finn et al., 2016).

Both BPSS0212ΔDUF1843 and BPSS0213ΔDUF1843 constructs were cloned into pET21a, protein overexpressed in *E. coli* BL21 DE3 and protein purified using the methods outlined in the materials and methods sections 2.3-2.6. The complex formation experiment was carried out between BPSS0212ΔDUF1843 and BPSS0213ΔDUF1843 as in the materials and methods, section 2.7.9.

5.8.2 Complex formed between BPSS0212ΔDUF1843 and BPSS0213ΔDUF1843

During purification of the individual proteins, both were subject to a gel filtration chromatography purification step, in 50 mM HEPES pH 7.5, 100 mM NaCl and the elution volume for each protein was recorded as a control. BPSS0212ΔDUF1843 elutes at ~84 mL and BPSS0213ΔDUF1843 elutes at ~90 mL (Figure 5.51), these elution volumes correspond to ~25 kDa and ~15 kDa, respectively, using the

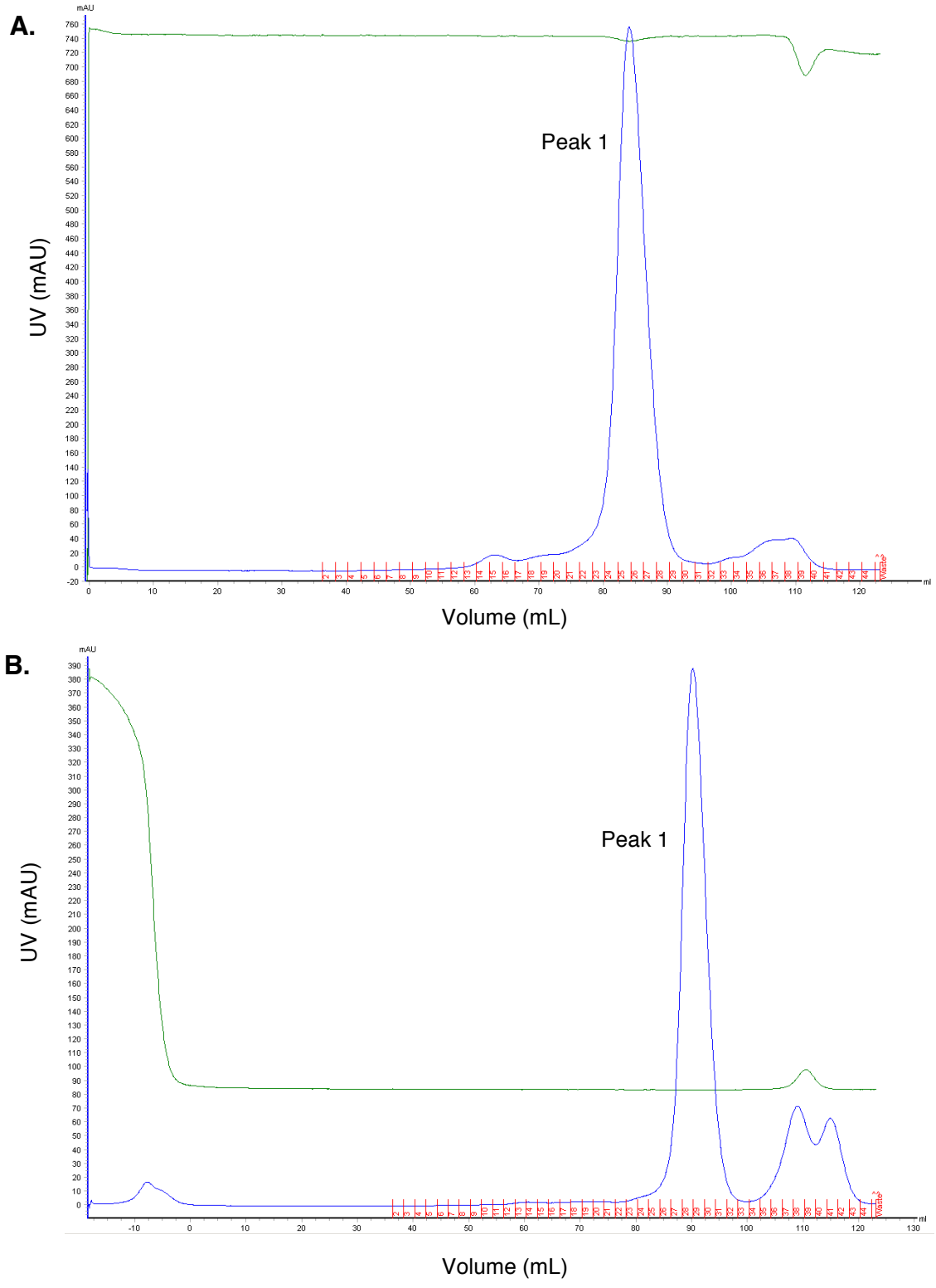


Figure 5.51. Chromatogram of BPSS0212ΔDUF1843 and BPSS0213ΔDUF1843 gel filtration experiment, each run at 50 mM HEPES pH 7.5, 100 mM NaCl. mAU at A_{280} – blue , conductivity – green, elution fractions – red numbers. The elution peaks are labelled and the corresponding elution volumes noted. **A.** BPSS0212ΔDUF1843, Peak 1 - ~84 mL, (fractions 22-29). **B.** BPSS0213ΔDUF1843, Peak 1 - ~90 mL, (fractions 26-32).

calibration graph in Figure 5.24. The gel filtration chromatogram for the BPSS0212ΔDUF1843 and BPSS0213ΔDUF1843 complex formation experiment showed an elution peak at ~69 mL, which is absent in the control experiments for the separate proteins (Figure 5.52). The contents of the elution fractions were analysed by SDS-PAGE. The resulting gel shows that this first peak contains both BPSS0212ΔDUF1843 and BPSS0213ΔDUF1843 (Figure 5.53).

The molecular weight calculated from this peak, using the calibration graph produced in 50 mM HEPES pH 7.5, 100 mM NaCl (Figure 5.24), is ~80 kDa, divided by the BPSS0212ΔDUF1843 and BPSS0213ΔDUF1843 protein average molecular weight of ~16300 Da, corresponds to ~five molecules. The second peak, at ~85 mL, corresponds to the excess BPSS0212ΔDUF1843 protein.

These results suggest that without the DUF1843 domain present in both proteins, a complex of 16 molecules cannot be produced. Therefore the DUF1843 domain plays a role in complex formation between BPSS0212 and BPSS0213.

5.8.3 Crystallisation of BPSS0212ΔDUF1843/BPSS0213ΔDUF1843 complex

The gel filtration elution fractions containing the BPSS0212ΔDUF1843/ BPSS0213ΔDUF1843 complex were pooled and concentrated for crystallisation trials, to ~8 mg/mL. The protein complex was set down in sitting-drop crystallisation trials using the Proplex, PACT and JCSG+ crystallisation screens. Crystal hits were produced in crystallisation conditions Proplex E4, JCSG+ E10 and PACT B10 (Figure 5.54). Crystals from conditions Proplex E4 and JCSG+ E10 were cryoprotected and sent to the Diamond Light Source Synchrotron, Oxford, beamline I03 for screening. Possible low-resolution protein diffraction was produced for crystals from condition Proplex E4, however no diffraction was produced

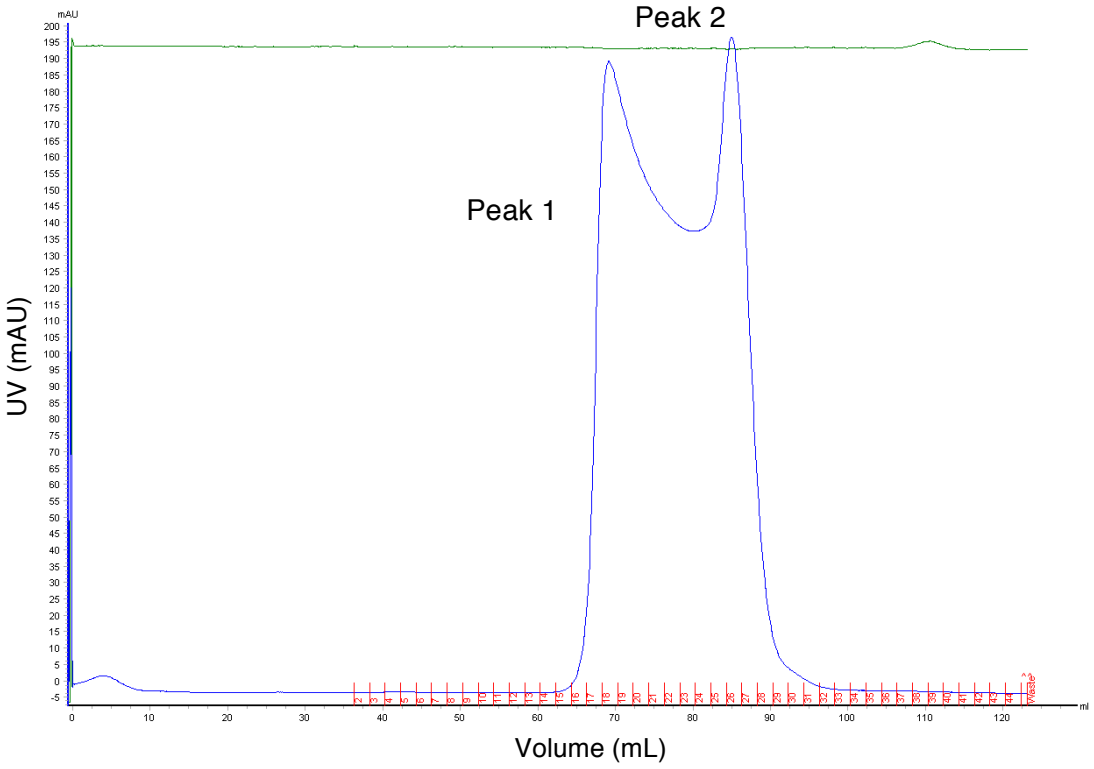


Figure 5.52. Chromatogram of BPSS0212ΔDUF1843 and BPSS0213ΔDUF1843 gel filtration experiment; run at 50 mM HEPES pH 7.5, 100 mM NaCl. mAU at A_{280} – blue , conductivity – green, elution fractions – red numbers. The elution peaks are labelled and the corresponding elution volumes noted. Peak 1 – 69.15 mL and Peak 2 – 84.99 mL.

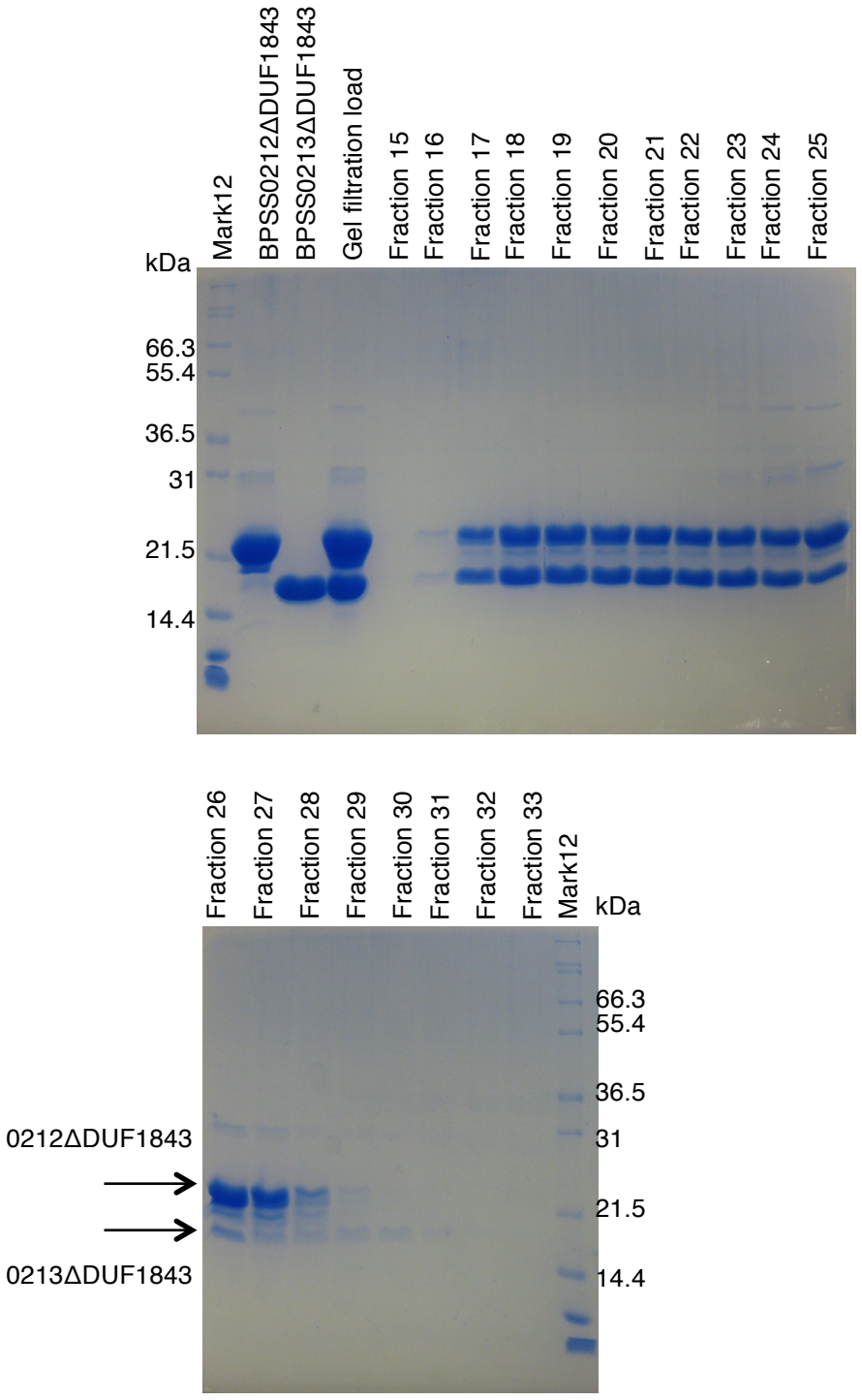


Figure 5.53. SDS-PAGE analysis of the BPSS0212ΔDUF1843 and BPSS0213ΔDUF1843 gel filtration experiment; run with 50 mM HEPES pH 7.5, 100 mM NaCl. BPSS0212ΔDUF1843 is 16762 Da and BPSS0213ΔDUF1843 is 15835 Da. Fractions analysed correspond to both peaks 1 and 2. Fractions 16-19 were taken and crystallised.

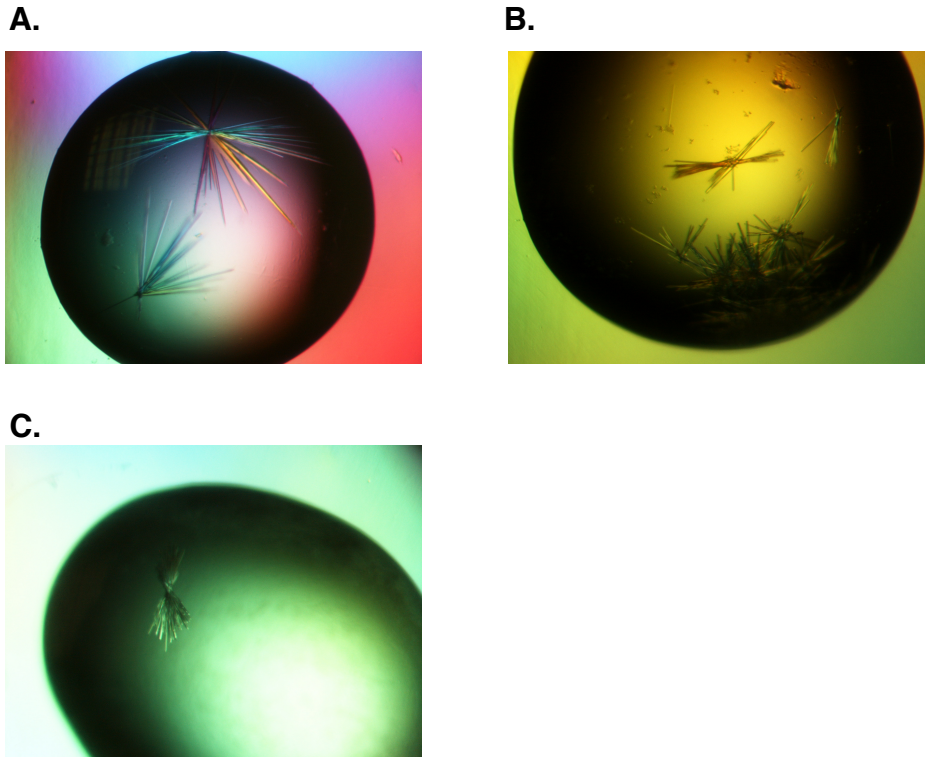


Figure 5.54. Crystal hits from sitting drop crystallisation trials for BPSS0212ΔDUF1843/BPSS0213ΔDUF1843

The protein complex was crystallised at ~8 mg/mL in 50 mM HEPES pH 7.5, 100 mM NaCl. **A.** Proplex E4 – 0.1 M sodium HEPES pH 7.0, 8% (w/v) PEG8000 **B.** JCSG+ E10 – 0.1 M BICINE pH 9.0, 10% (w/v) PEG6000 **C.** PACT B10 – 0.2 M magnesium chloride, 0.1 M MES buffer pH 6.0, 20% (w/v) PEG6000.

for crystals from condition JCSG+ E10, which prevents the structure elucidation and analysis of the core complex from which the DUF1843 domains have been removed.

5.8.4 Connectivity of DUF1843 domain helices to complex core

The BPSS0212/BPSS0213 complex has a $\alpha_8\beta_8$ stoichiometry, with two layers of $\alpha_4\beta_4$. The core of the complex contains alternating BPSS0212 and BPSS0213 DUF1842 domain β-barrels and is surrounded by two bands of DUF1843 domain dimers at the interface between the two $\alpha_4\beta_4$ layers. Unfortunately due to the lack of electron density between the C-terminus of the modelled DUF1842 domain and partial linker of BPSS0212 and BPSS0213 and the N-terminus of the DUF1843 domain helices, it cannot be established which BPSS0212 DUF1843 domain is associated with which BPSS0212 DUF1842 domain, and similarly for BPSS0213.

The linking of the DUF1842 and DUF1843 domains can occur within a $\alpha_4\beta_4$ layer or between the two $\alpha_4\beta_4$ layers. The different ways of connecting the domains were analysed to identify the most likely connection. For BPSS0212 five possible DUF1843 domains, which could link to the DUF1842 domain, were identified. Three of these DUF1843 domains are present in a different $\alpha_4\beta_4$ layer. Two of these appear to be inaccessible and can be ruled out. Another two DUF1843 domains are located in the same layer. The possible linkages between a single DUF1842 and the surrounding DUF1843 domains of BPSS0213 revealed three possible linkages, one in a different $\alpha_4\beta_4$ layer and two in the same, one of which appears to be inaccessible and can be ruled out. Therefore BPSS0212 has a possible three connections and BPSS0213 has two possible connections. This suggests that there potentially exists six different ways of connecting the DUF1842 and DUF1843 domains of each protein within the complex (Figures 5.55 and 5.56).

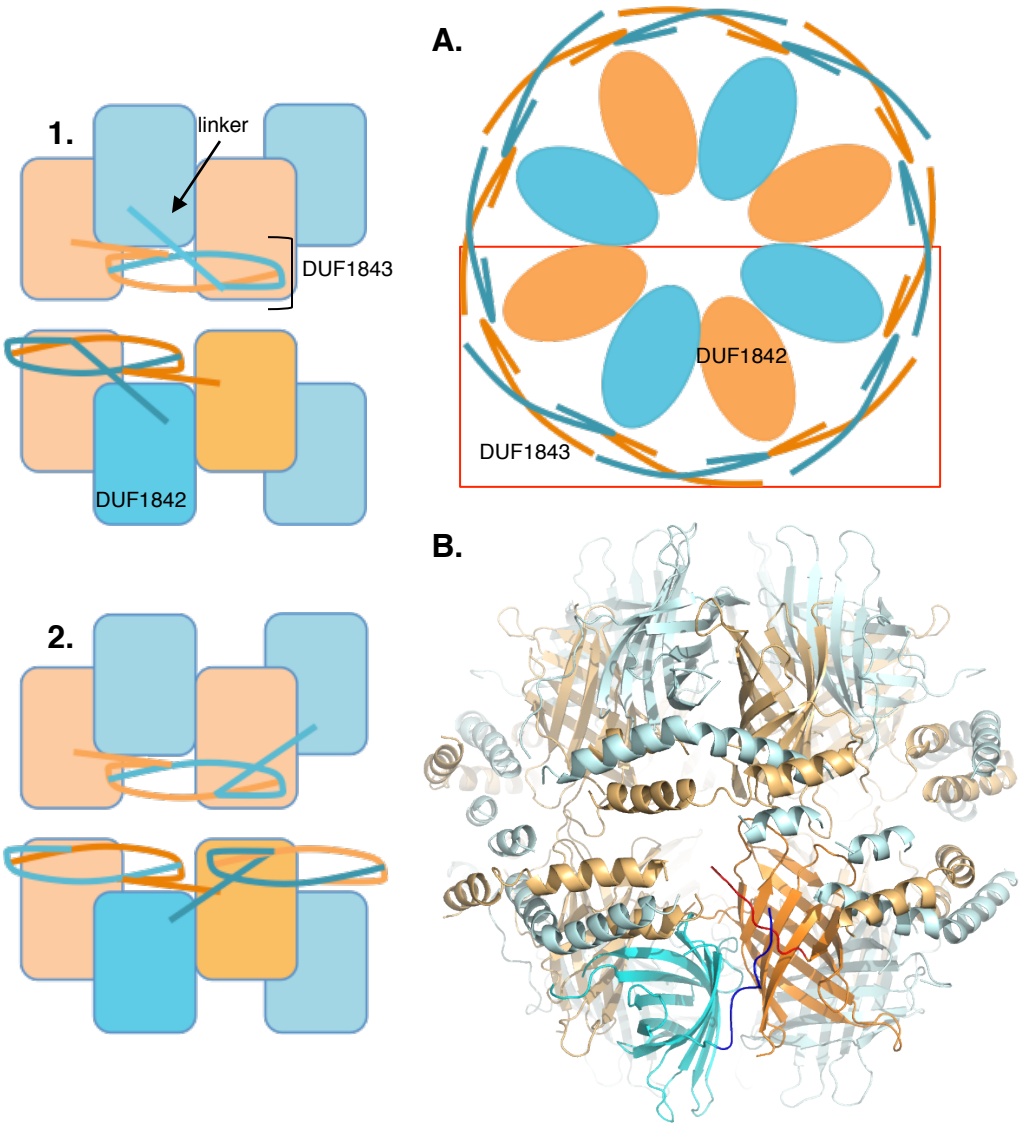


Figure 5.55. Complex 5.6 - DUF1842 – DUF1843 domain connectivity
BPSS0212 is blue and BPSS0213 is orange. A schematic top view of the $\alpha_8\beta_8$ is shown in **A.**, half of the $\alpha_8\beta_8$, in the red box, is shown in Models **1** and **2,** viewed perpendicular to the 4-fold axis. The connectivity for a pair of DUF1842 and DUF1843 domains of BPSS0212 and BPSS0213 are represented by darker shades of blue and orange, respectively (**B., 1.** and **2.**). **Model 1.** has both BPSS0212 and BPSS0213 DUF1842 domains connecting to the DUF1843 domains in the same $\alpha_4\beta_4$ layer. **Model 2.** has both BPSS0212 and BPSS0213 DUF1842 domains connecting to the DUF1843 domains in the same $\alpha_4\beta_4$ layer, with BPSS0212 connecting to a different DUF1843 domain than in **Model 1.** Figure B. generated in PyMOL version 1.1r1.

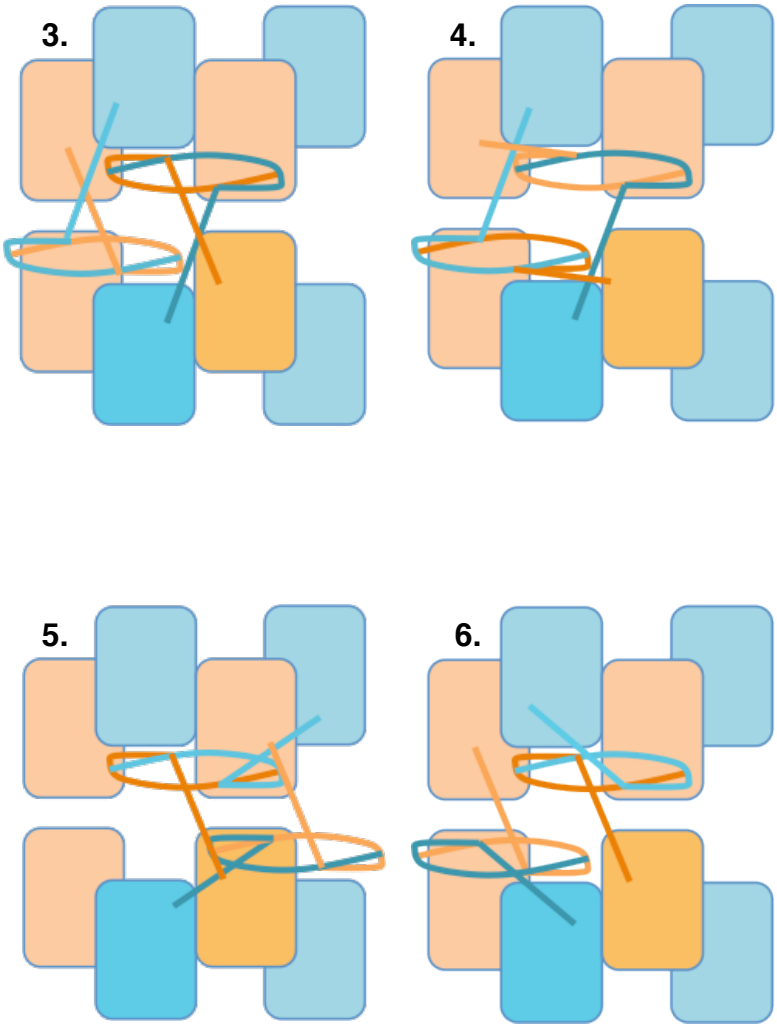


Figure 5.56. Complex 5.6 - DUF1842 – DUF1843 domain connectivity -2. BPSS0212 is blue and BPSS0213 is orange. The schematic half of the $\alpha_8\beta_8$ (Figure 5.55A), perpendicular to the 4-fold axis, is shown in Models **3., 4., 5. and 6.** The connectivity for a pair of DUF1842 and DUF1843 domains of BPSS0212 and BPSS0213 are represented by darker shades of blue and orange, respectively. **Model 3.** has both BPSS0212 and BPSS0213 DUF1842 domains connecting to the DUF1843 domains in a different $\alpha_4\beta_4$ layer. **Model 4.** has BPSS0212 connecting to a different layer and BPSS0213 connecting to the same $\alpha_4\beta_4$ layer. **Model 5.** has BPSS0213 connecting to a different layer and BPSS0212 connecting to the same $\alpha_4\beta_4$ layer and **Model 6.** is the same as Model 5 except BPSS0212 is connecting to a different DUF1843 domain in the same layer.

Models 1 and 2, which only have linkages present within a single $\alpha_4\beta_4$ layer, were discounted based on the observation that deletion of the DUF1843 domain helices results in the absence of the large molecular weight peak, indicating the $\alpha_8\beta_8$ formation, in the gel filtration purification experiment.

Therefore this suggests the linkages formed to the DUF1843 domains connect the two $\alpha_4\beta_4$ layers together. The remaining four models are 3, 4, 5 and 6. Model 3, has both DUF1842 domains in one $\alpha_4\beta_4$ layer connecting to DUF1843 domains in a different $\alpha_4\beta_4$ layer. Model 4, has BPSS0213 domains connecting in the same $\alpha_4\beta_4$ layer and BPSS0212 domains connecting from different $\alpha_4\beta_4$ layers. Models 5 and 6 are similar with two possible options of BPSS0212 domains connecting to the same $\alpha_4\beta_4$ layer, and BPSS0213 domains connecting from different $\alpha_4\beta_4$ layers.

The distances between the end of the DUF1842 domain and partial linker and the start of the possible DUF1843 domains were measured and compared to the number of residues available in the rest of the linker. The number of unmodelled residues in the linker between the domains for BPSS0212 is 14 residues and for BPSS0213 is seven residues. The remaining linker residues were also modelled into the complex to assess the space available for such linkers.

The two possible remaining linker residue distances for BPSS0213 are ~22 Å, with a linker distance of ~24.5 Å for seven residues in an extended conformation, and from analysis of the spatial arrangement, both linkages are possible. The possible linker distances of BPSS0212 is ~17 Å for models 3 and 4, ~28 Å for model 5 and ~42 Å for model 6. The extended conformation of the linker would be ~49 Å. Model 6 was discounted as less likely as the distance of the linker is close to the limit of the linker distance and there is a potential spatial packing issue. Model 5 was discounted due to a possible spatial packing issue.

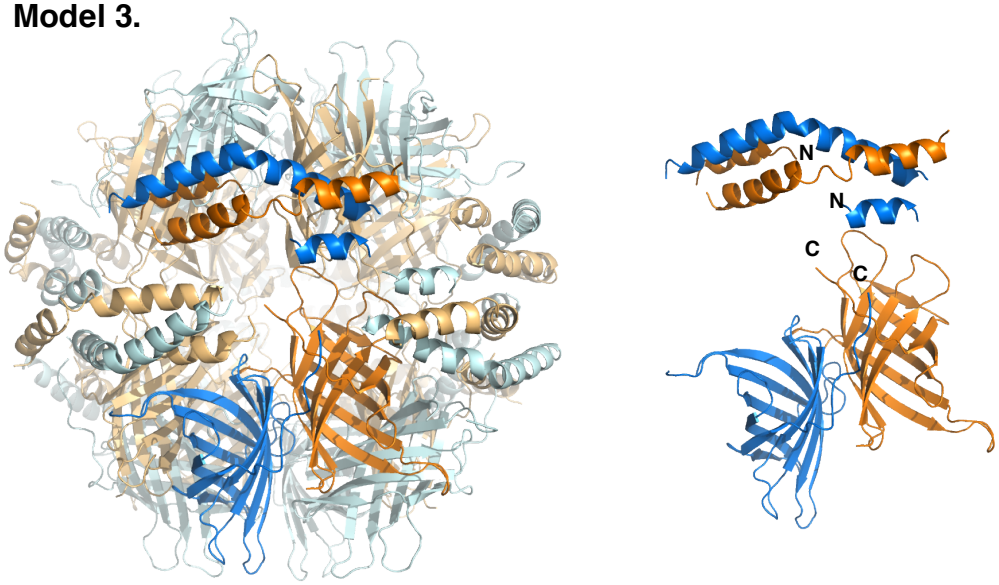
Overall the connection possibilities between the DUF1842 and DUF1843 domains in the complex described in model 3 and 4, stand out as most likely, both are consistent with the DUF1843 performing a role in the interaction between the two $\alpha_4\beta_4$ layers and the BPSS0211 conformational change which may be involved in dimer formation. The assignment of connectivity between the domains within the complex is not definite and an unambiguous assignment is not possible.

Model 3 has linkers from BPSS0212 and BPSS0213 in one layer, connecting to the DUF1843 domains involved in a dimer interaction in the other $\alpha_4\beta_4$ layer (Figure 5.56). This suggests the possible formation of a dimer between BPSS0212 and BPSS0213 (Figure 5.57), which would then assemble into the $\alpha_8\beta_8$ complex, along with seven other dimers. The second possibility is model 4, that again has the BPSS0212 DUF1842 domain in one $\alpha_4\beta_4$ layer involved in a linkage with a DUF1843 domain in a different $\alpha_4\beta_4$ layer. The BPSS0213 DUF1842 domain in one $\alpha_4\beta_4$ layer then links to a DUF1843 domain in the same layer, therefore a cross linkage is formed between the two $\alpha_4\beta_4$ layers.

5.8.5 Electron microscopy analysis of the BPSS0212/BPSS0213 structure

A sample of the isolated larger molecular weight complex, $\alpha_8\beta_8$, was taken and applied to a carbon grid and stained with 1% uranyl formate. The grid was analysed using a Phillips CM-100 transmission electron microscope and BPSS0212/BPSS0213 complex molecules were visualised (Figure 5.58). The diameter of the visualised complex was measured to be ~120 Å. The diameter of the core of the $\alpha_8\beta_8$ complex is ~90 Å, however the addition of the external DUF1843 domain helices produces a complex of diameter of ~120 Å. The overall dimensions of the current BPSS0212/BPSS0213 complex structure is approximately 120 Å x 120 Å. Further EM (electron microscopy) images were collected and analysis was performed, in order to try and produce a low-resolution mask in which to model the complex. Initial image analysis and further

Model 3.



Model 4.

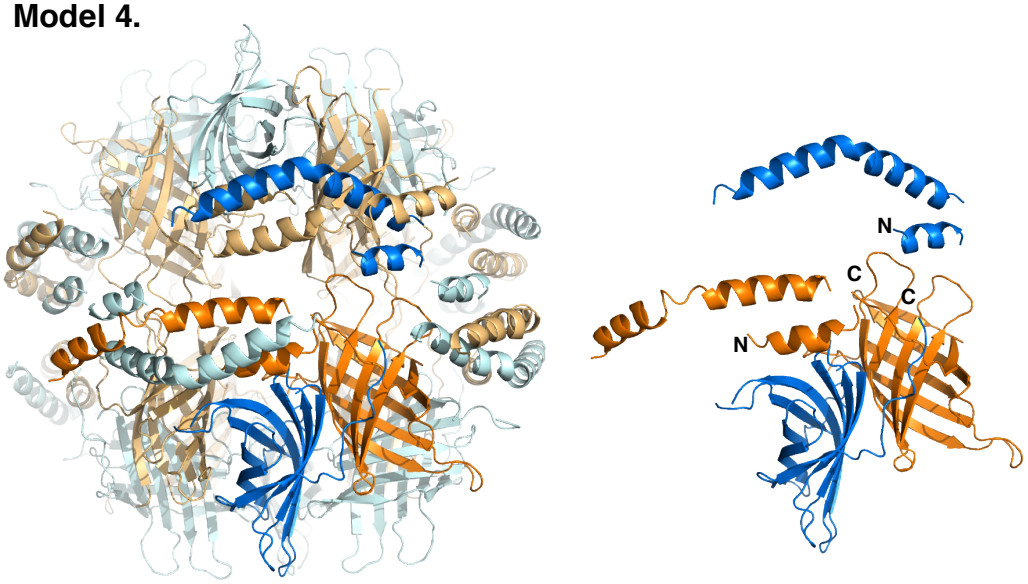


Figure 5.57. Complex 5.6 final connectivity possibilities
The current BPSS0212/BPSS0213 $\alpha_8\beta_8$ complex 5.6 structure, with BPSS0212 (blue) and BPSS0213 (orange), shown as cartoon. The connectivity for a pair of DUF1842 and DUF1843 domains of BPSS0212 and BPSS0213 are represented by darker shades of blue and orange, respectively. **Model 3.** has both BPSS0212 and BPSS0213 DUF1842 domains connecting to the DUF1843 domains in a different $\alpha_4\beta_4$ layer. **Model 4.** has BPSS0213 DUF1842 domain connecting to the DUF1843 domain in the same $\alpha_4\beta_4$ layer, with BPSS0212 connecting to a DUF1843 domain in a different $\alpha_4\beta_4$ layer. Figure generated in PyMOL version 1.1r1.

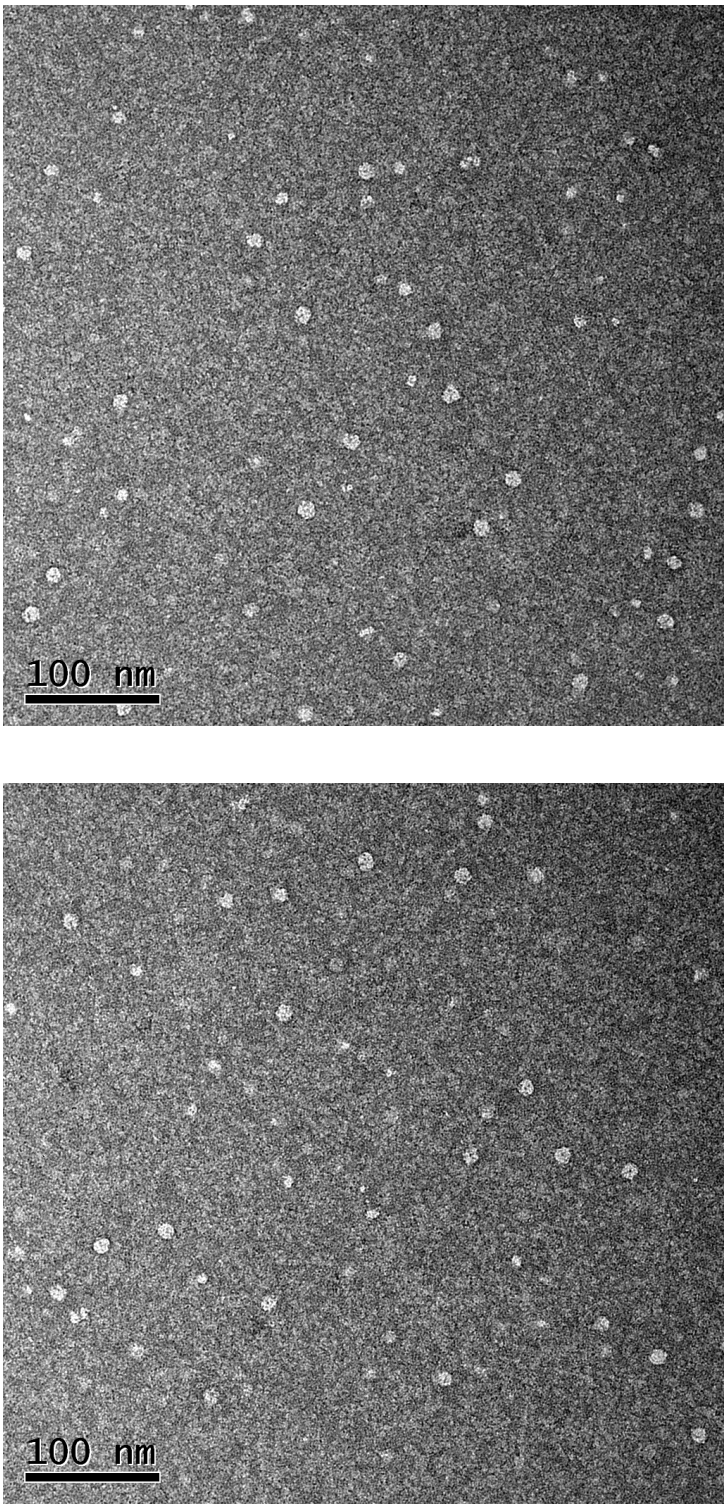


Figure 5.58. Electron micrograph of BPSS0212/BPSS0213 complex
BPSS0212/BPSS0213 complex produced under low salt conditions, at 0.1 mg/mL, 5 µL on a carbon grid stained with 1% uranyl formate. Images were taken on a Phillips CM100 microscope. Grids prepared and images taken by Jason Wilson.

work was carried out with Dr. Svetomir Tzokov (University of Sheffield). However, the orientation of the complex could not be established due to the quality of the preparation and the inability to discern between side and top views of the complex.

5.9 Functional experiments on BPSS0212/BPSS0213 complex

In order to analyse the potential function of the BPSS0212/BPSS0213 complex, in collaboration with the Universiti Kebangsaan Malaysia, Professor Sheila Nathan and students Yee-Chin Wong and Alice Cin Kong, constructed a double knock-out mutant of BPSS0212 and BPSS0213 in *B. pseudomallei* K96243 and carried out phenotype analysis experiments and expression profiling, which will be described below. The wildtype strain used was *B. pseudomallei* K96243. The region of DNA deleted to produce the ΔBPSS0212/BPSS0213 double knock-out mutant is shown in Figure 5.59.

5.9.1 Effect of ΔBPSS0212/BPSS0213 double mutation on growth

The growth of the ΔBPSS0212/BPSS0213 double mutant vs wildtype in LB broth and M9 medium was investigated by analysing the number of colonies produced on Ashdown agar at each time point. Each experiment was replicated twice. The results show that, in either media, the growth of the ΔBPSS0212/BPSS0213 double mutant was not significantly different to that of the wildtype. One of the replicate results is shown in Figure 5.60.

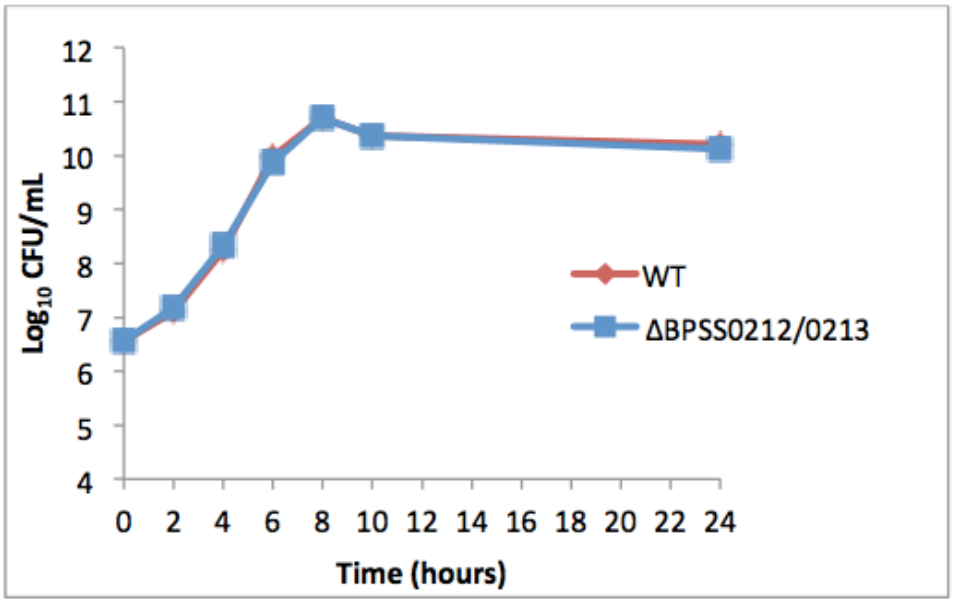
5.9.2 Effect of ΔBPSS0212/BPSS0213 double mutation on virulence

A *C. elegans* (*Caenorhabditis elegans*) killing assay was carried out with ΔBPSS0212/BPSS0213 double mutant vs wildtype, and *E. coli* OP50 as a negative control. The assay was replicated three times, the first experiment exhibited a possible increase in virulence of the double mutant, but this was not shown in later replicates. The results show that the double mutation of ΔBPSS0212/BPSS0213 does not reduce the

GACGTGCCGGTTCAGAGGGATAACACGAATGGCTAATCACGGGCGGCGAGCAGTGCGGCACGAGG
CACCGTGTCGCAGCATGGCGGCGAGCGCGACATTCATCGACGAAATTCCCCGATCCTGCACAAAC
ACGACAGCCGGATGCACTGCGCCGCCGCGGGGCAGATCGCAGGGCTGTCGGCCATCGTCCACTTT
CAGCGAGGCAAACATGAGTCAAGCACAGGGCCATCCGGTCACCCCTTATGGCGTCGCCATCCATC
AGGCGATCGCGGACGGCGATCTCGCGCAGATGAAATCGCTGCGCACGCAGGCGCAGGCGCTGCTC
GCCCAGCAGGGCAATCTCGCCACCGCGCTCGAGCTTCTCGAAGTCGAGATCGCCAAGCTCGAGCG
GAGGAAGTAGCCGCGAATCCGCCGCTTCGGCGAGCCGTGCCGCCGCGGCGGCGTTTCAGCGTGTC
AACACGTCGTTCAACCACTCACATCAAGGAGACTCACATGTCCGAAGATCTTCGCGTCGGTCTTT
TTCCGGTTCGCTATCTCGTCGGCACCGGCTTGCCCGGCGCGCCGCAGCTCGTGCTGGATCTGATG
GTCGACACGGTCGATCACAGCGTCGTCGGCCGCGCGGCGGTCTCGCAAGCCGTGAGCCCGCCGCT
CAATTTCCACGCCGACGTCTGGGGCTCCTACGTATTCCGGCTCGGCCCGCCGCCGCGCCGCGACG
GCAGCGGCGCGATCGTGCAGATCTCGCTGCAGGGCAACCAGGGCGGGCCGCAGTCGAATTCGATG
ATCACGTTCTACGGCGAGCTGCTGCTCAAGGGCGACGGCAAGACGGGCGTCGCATCGTACCGCTA
CTACTCGAACGGCTCCTGGCACGAAGTCGAGAACGTGCCGGTCAAGGCCGACCCGGAGCTCGTGC
CGATCGAGCCGGGCCCGGTCATCGGCCAATCGTCGATGAGCGCGATCGGATCGGCCGCGATGTAC
GGCGTGGCGATCCAGTCTGCGGCCGCTTCCGGCGATCTCGCGCACATGCGCACGCTCAGTGCCTA
TGCGCGGCAGCAGCTGGAAAGCCGTGACGAAATCGCGGCCGCGCTGAGCGAGCTGAAGGCGGAAA
TCGCCAAGCTCGAAAGCCGGCAGTGAGCGGCGCGGCCTTGCATCGCCATCGTTCGAACCTTTTCG
TGGAGCTATCGCCATGGCAACTACCGGTCTCTTTCCCGTCCAGCTTCGCGTCGCGACGCCCAATC
TCGGCGCGCCCGTTCTCTGGCTGAACCTGCTCGTCAACACGGTCGAGAAAACCGCGTCGGGCTTC
GCCCGCATCACGCAGACCGTCTATCCGCCGATGCATTTCCGCGCGCGCGTGGTCGGCCCGTTCCA
TCAAATGAGGATCGATCCTCACGCGCCGCAGTCGGTCACGCTCACCCTCTCGGGCAGCCCCACCG
GGCCCGTCGCGCCGCAGGTCGTGATCCTGGAGCTCAACGCGCTGCTGAACGAAGGCTGGCAATCG
GGTACGGCGAACTATCGCTACTTCTACGAGAGCCGCTGGCATTCGATCGAGCACGCGATCGTGTC
GAAGGACAACTCGCGGATTCCGCTCGATCCGCCGAGCGAGCACGTGATGCCGATGTACGGCGTCG
GGCTGCAGGAAGCGCGGGCGTCCGGCGATCTGAGCCGGATGAAGGCACTCGCGCAGCAGGCCGAG
CAGCAACTCGCCGATCATGACGTGATCGCCGCCGAGCTCCAGAAGCTCGAAGCGGAGATCGCGCG
GCTCGAAGCGCGCCGCTGAGCCCGCGCGCGCCGTTTGCCGCACGCGCGGGCCCGTTTCGCGCGTG
CGGCCGCCATCGCCCCACCAGGAGACATCGACATGAGCACAACGGACGCGCGCCCGGCCCGCTAC
CTGTTCGACAGCGACTATCAACGCTTCGTCCCGGTTCATGCCGTGTGGGAAATCACGCTCGCCTG
CGATCTCAAGTGCCTGCACTGCGGCTCGCGGGCCGGCCACCGGCGGACCAACGAACTGAGCACGG
CCGAATGCCTCGAAGTGATCGACGCGCTCGCGCGGCTCGGCACACGCGAAGTCTCGCTGATCGGC
GGCGAGGCGTATCTGCGCAAGGACTGGACGCAACTGATCCGCGCGATCCGCTCGCACGGCATGTA
TTGCGCGATCCAGACGGGCGGCCGCAATCTCACGCCGAAGCGCCTC

Figure 5.59. The sequence for the *B. pseudomallei* K96243 ΔBPSS0212/BPSS0213 double mutation clone, within the BPSS0211-BPSS0214 operon region. The operon gene DNA sequences are highlighted - BPSS0211 – yellow, BPSS0211(1-123) – orange and yellow, BPSS0212 – green, BPSS0213 – blue and BPSS0214 – pink. The underlined section was deleted in the ΔBPSS0212/BPSS0213 double mutation. A mutation of DNA base T to A was found in base position 287731, shown in bold, in the intergenic region between BPSS0213 and BPSS0214. Information produced by Yee-Chin Wong. Figure adapted from report produced by Yee-Chin Wong.

A.



B.

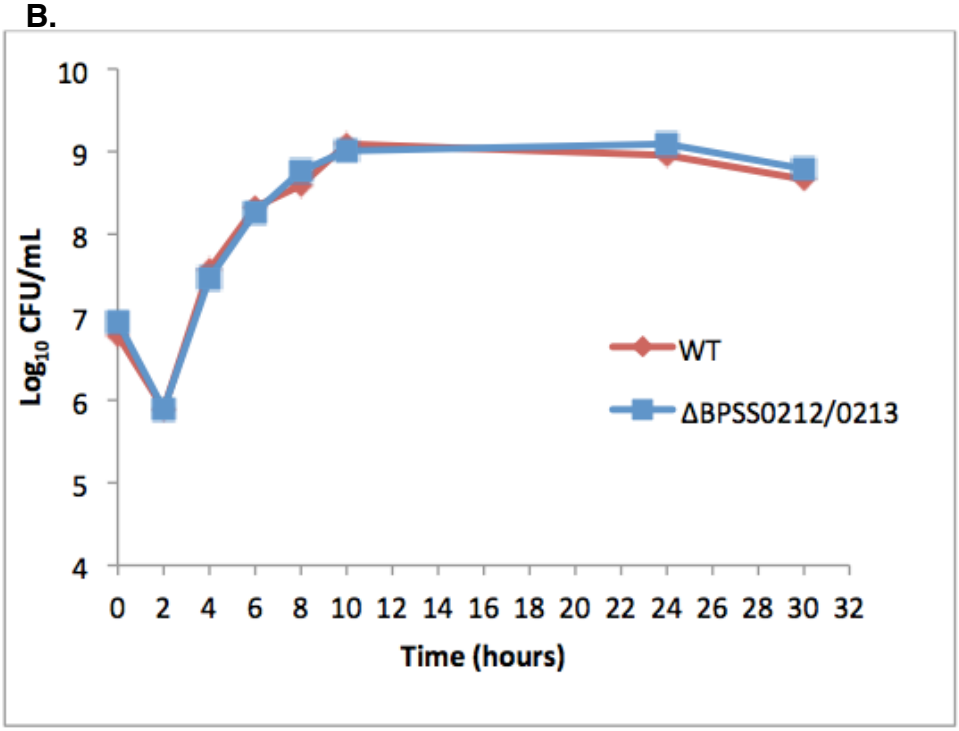


Figure 5.60. A representative result for growth studies of *B. pseudomallei* K96243 ΔBPSS0212/BPSS0213 double mutant. A. The growth of *B. pseudomallei* K96243 ΔBPSS0212/BPSS0213 double mutant vs wildtype *B. pseudomallei* K96243 compared in LB medium, and **B.** M9 medium. Data and figures produced by Mei-Perng Lim and Rui-Rui Wong.

killing rate of *C. elegans*, as the rate of killing is not significantly different between the ΔBPSS0212/BPSS0213 double mutant and the wildtype strain. One of the later replicate results is shown in Figure 5.61.

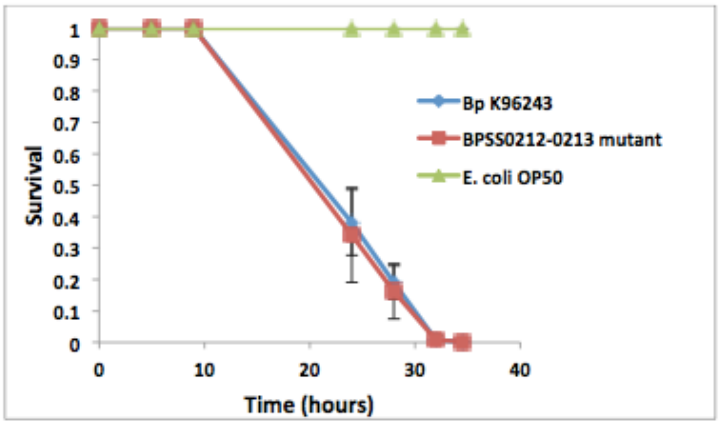
5.9.3 Expression of operon under different environmental conditions

An expression profile of the operon – BPSS0211, BPSS0212, BPSS0213 and BPSS0214 – was produced for *B. pseudomallei* when grown in LB broth, human plasma and soil-extract medium (Yoder-Himes et al., 2009). The results show that the operon is more highly expressed in the soil-extract medium, than the LB broth or human plasma. Subsequently the growth of the ΔBPSS0212/BPSS0213 mutant vs the wildtype in soil-extract medium was analysed with one replicate. There was no significant difference between the growth rates.

These results suggest that BPSS0212-BPSS0213, along with the other members of the operon may be involved in the survival of *B. pseudomallei* in the soil, and may not be involved in pathogenicity under the conditions analysed above.

5.10. Conclusions and further work

Investigations into the BPSS0211-BPSS0214 operon have revealed the formation of a large molecular weight $\alpha_8\beta_8$ complex between BPSS0212 and BPSS0213. The core of the complex is formed from alternating BPSS0212 and BPSS0213 DUF1842 domains with partial interdomain linkers, in two $\alpha_4\beta_4$ layers. The DUF1843 domains, one from each BPSS0212 and BPSS0213 subunit, are located on the exterior of the protein, packed against the core. The two core $\alpha_4\beta_4$ layers of the DUF1842 domains interact solely through contacts mediated by the DUF1843 domains, which are predicted to act as an interaction module. On removal of the DUF1843 domains from both BPSS0212 and BPSS0213, the larger molecular weight complex is not produced. The



Bacteria strains	TD mean (hours)	*P* value
E. coli OP50	-	
B. pseudomallei K96243 wild type	26.302 ± 0.293	
B. pseudomallei K96243 ΔBPSS0212/0213 mutant	26.021 ± 0.284	0.5136

Figure 5.61. A representative result for killing studies of *B. pseudomallei* K96243 ΔBPSS0212/BPSS0213 double mutant. Wildtype *B. pseudomallei* K96243, *B. pseudomallei* K96243 ΔBPSS0212/BPSS0213 double mutant and *E. coli* OP50 on killing of *C. elegans*. Data and figures produced by Cin Kong.

interaction between the DUF1843 domains of BPSS0212 and BPSS0213 is closely related to the interaction seen between two molecules of BPSS0211, when crystallised in the presence of zinc. Therefore BPSS0211 may be able to interact with both BPSS0212 and BPSS0213 via the DUF1843 domain. However BPSS0211 was not seen to interact with either BPSS0212 and BPSS0213 under the conditions investigated. The structure of BPSS0211 in the presence of zinc produces a closed structure of two helices, however in the absence of zinc the structure produces a single long helix. This conformational change seen in BPSS0211 may be mirrored by the BPSS0212 and BPSS0213 DUF1843 domains in the formation of the complex, to allow the intertwining of the domains. In vivo it is still possible that BPSS0211 may be involved in binding BPSS0212 and BPSS0213, via the DUF1843 domain, to prevent complex formation until it is required. The function of this complex is yet to be determined. The deletion of both genes for BPSS0212/BPSS0213 produced no effect on the growth rate of *B. pseudomallei* or its virulence towards *C. elegans*. Overall the structures of two DUF1842 domains and one additional structure of a DUF1843 domain have been produced. The role of BPSS0214 within the operon still remains to be established.

The connectivity between the domains within BPSS0212 and BPSS0213 in the complex requires further investigation and could be elucidated with higher resolution data. The other smaller molecular weight complexes containing BPSS0212 and BPSS0213 identified in the complex formation experiments in both high salt and low salt conditions could be analysed by analytical ultracentrifugation to assign a molecular weight and stoichiometry. Constructs of the DUF1843 domains of BPSS0212 and BPSS0213 could be made, in order to try to produce a higher resolution structure of the BPSS0212 and BPSS0213 DUF1843 domain heterodimer. The coexpression of BPSS0211 with BPSS0212 and BPSS0213 separately may able to indicate whether BPSS0211 is able to interact with BPSS0212 and BPSSS0213 under different conditions. Also

the coexpression of all four proteins in the operon may indicate a role for BPSS0214 and for BPSS0211 in relation to the BPSS0212/BPSS0213 complex.

Chapter 6. Type VI Secretion System

6.1 Project introduction

The structure of a number of the T6SS core components have yet to be determined including TssA, TssE, TssF, TssG and TssK. The aim of this project was to investigate a number of these components further. TssA from *A. hydrophila* and *B. cenocepacia* were investigated, along with TssK and TssE from *B. cenocepacia*.

TssA was cloned from *A. hydrophila* and *B. cenocepacia* H111. TssK and TssE were also cloned from *B. cenocepacia* H111. The construct design, cloning and overexpression of all protein constructs was carried out by collaborators in the University of Sheffield Medical School – Dr. Mark Thomas and PhD student, Ruyue Sun. The overexpression of the selenomethonine-incorporated *B. cenocepacia* $TssA_{CTD}$(303-373) protein construct was carried out as part of the work in this thesis. Dr. Svetlana Sedelnikova (University of Sheffield) carried out protein purification for crystallisation. *A. hydrophila* TssA and *B. cenocepacia* TssE and TssK protein were set down in crystallisation trials, however work on these proteins will not be discussed any further here. The alignment of a number of TssA proteins previously identified two distinct families of – TssA-1 and TssA-2, based on the CTD (C-terminal domain) sequences and part of the NTD (N-terminal domain) (Dr. Mark Thomas, personal communication). An alignment of TssA proteins is shown in Figure 6.1. The work outlined below will concern the *B. cenocepacia* TssA protein, with constructs designed for the NTD and the CTD (Figure 6.2).

6.2 $TssA_{NTD}$

6.2.1 Crystallisation of $TssA_{NTD}$(1-255) protein

$TssA_{NTD}$(1-255) protein construct was set down in sitting-drop vapour diffusion crystallisation trials, using crystallisation screens PACT, JCSG+, pHClear, Classics, MPD, $AmSO_4$, PEGs and Proplex. Crystals grew in a

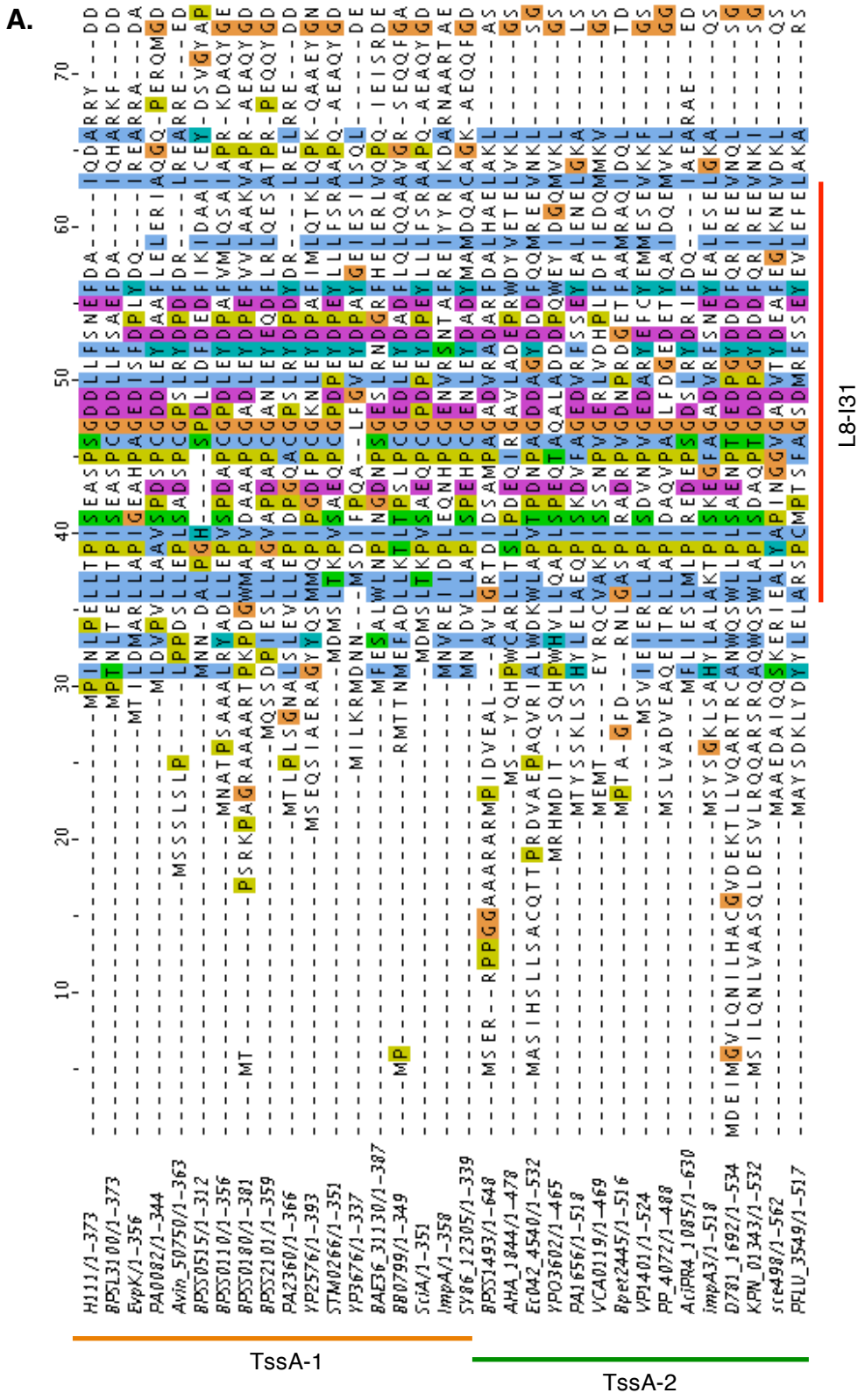
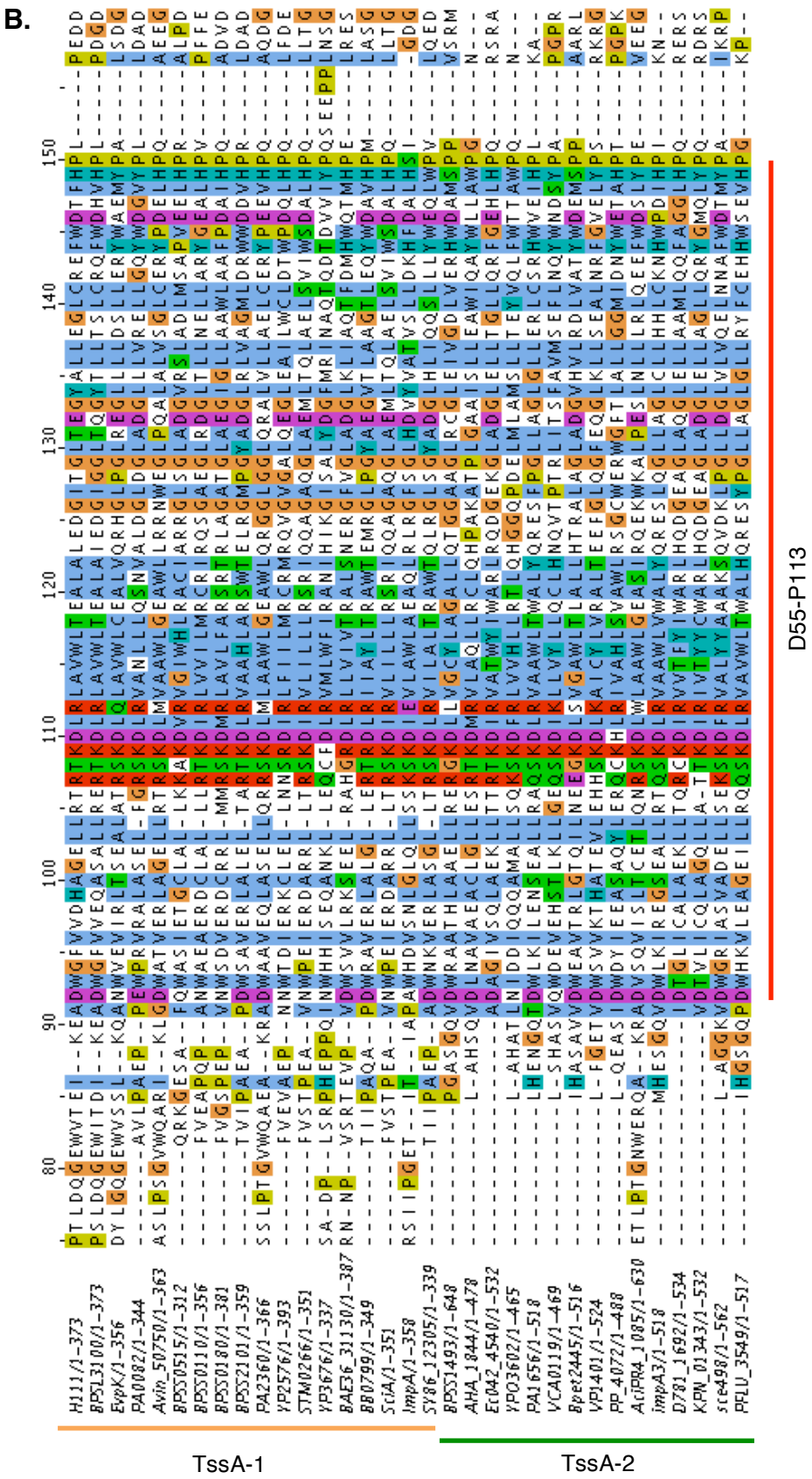
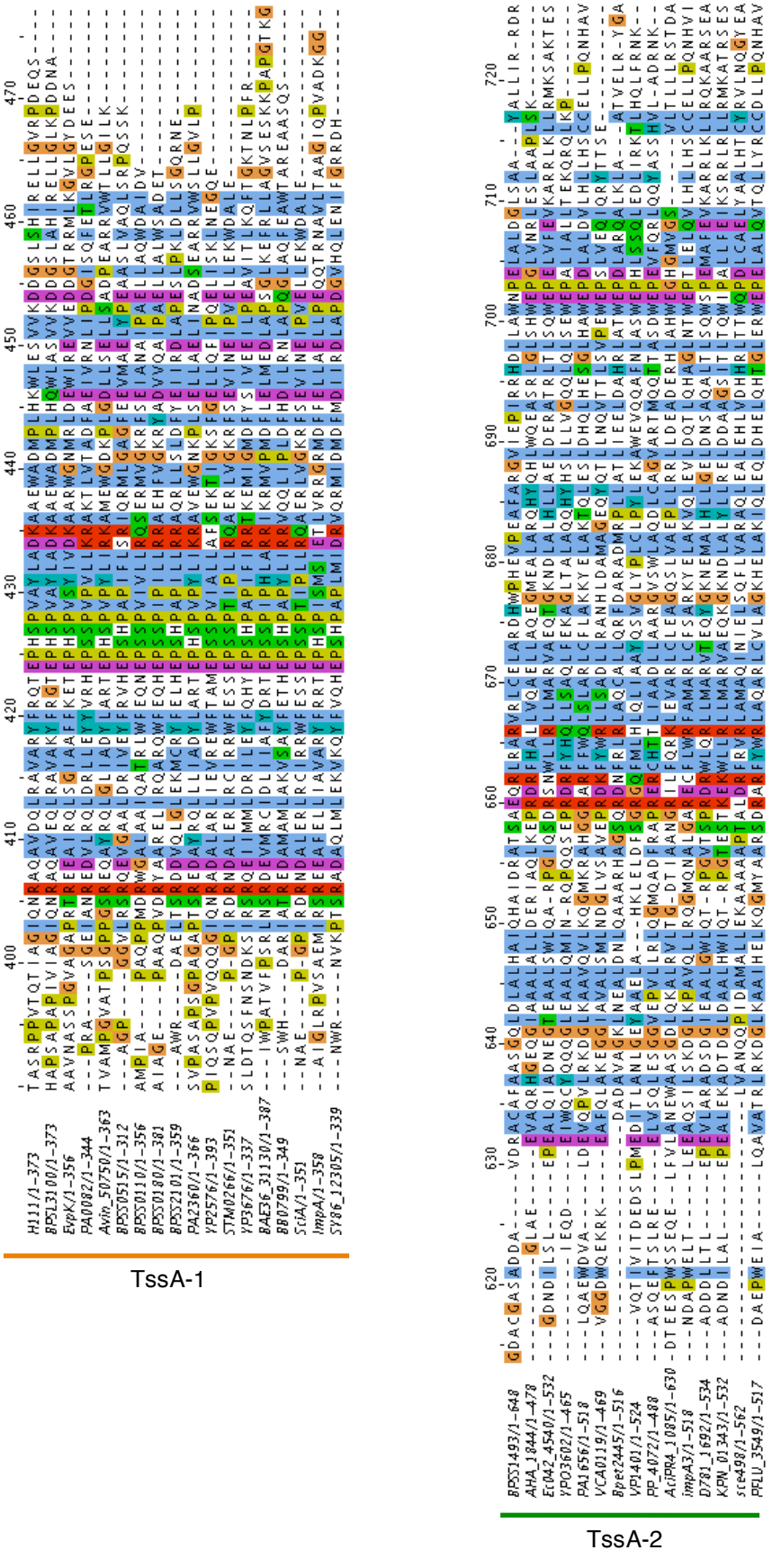


Figure 6.1. Alignment of the amino acid sequences of a selection of TssA proteins. A. Section of alignment showing the first conserved region – residues 8-31 (*B. cenocepacia* H111 TssA numbering). The two TssA groups are TssA-1 and TssA-2. Sequences aligned in Jalview using ClustalOmega with defaults, coloured using Clustalx (Larkin et al., 2007; Sievers et al., 2011; Waterhouse et al., 2009).



B. Section of alignment showing the second conserved region – residues 55-113 (*B. cenocepacia* H111 TssA numbering). Sequences aligned in Jalview using ClustalOmega with defaults, coloured using Clustalx (Larkin et al., 2007; Sievers et al., 2011; Waterhouse et al., 2009).

C.



C. Section of alignment showing the CTDs of TssA-1 and TssA-2. Sequences aligned in Jalview using ClustalOmega with defaults, coloured using Clustalx (Larkin et al., 2007; Sievers et al., 2011; Waterhouse et al., 2009).

MPINLPELLTPISEASPSGDDLLFSNEFDAIQDARRYDDPTLDQGEWVTEIKEA
DWGFVVDHAGELLRTRTKDLRLAVWLTEALALEDGITGLTEGYALLEGLCREFW
DTFHPLPEDDDIEHRLGNVAWLSGRTAELLRAVPLTDGASNAFSTLDWEVAQHV
AQSIKRDPEHADDIARGKPSIEQIDASRRVTSIAFYTALLANLKAFEFALDAFE
ERLVERAGDSAPSFRQARDAFETVYRLAERFAREQGYTGSAPHTQAAQQAQPER
IEPVFGQPIQTEETHVQQQTASRPPVTQTIAGIQNRAQAVDQLRAVARYFRQTE
PHSPVAYLADKAAEWADMPLHKWLESVVKDDGSLSHIRELLGVRPDEQS

Figure 6.2. *B. cenocepacia* TssA schematic and sequence
Schematic and associated sequence of *B. cenocepacia* H111 TssA, with the NTD construct region highlighted in blue and the CTD construct region highlighted in red. (Dr. Mark Thomas, personal communication; NCBI Resource Coordinators, 2016).

range of conditions (Figure 6.3), including JCSG+ E11, with a crystallisation condition of 0.16 M calcium acetate, 0.08 M sodium cacodylate buffer pH 6.5, 14.4% (w/v) PEG8000 and 20% (v/v) glycerol.

6.2.2 $TssA_{NTD}$(1-255) native data collection

Crystals grown in crystallisation condition JCSG+ E11 were cryoprotected with 0.16 M calcium acetate, 0.08 M sodium cacodylate buffer pH 6.5, 16.4% (w/v) PEG8000 and 30% glycerol, cooled in liquid nitrogen and sent to the Diamond Light Source Synchrotron, Oxford, beamline I02. Native data were collected at 1.0443 Å wavelength, with 0.2 second exposure over a phi oscillation of 0.2° for a total of 180°, at 60% transmission (Figure 6.4). The data were processed to 1.8 Å using 150° of data using the xia2 3da pipeline (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Winter, 2010). The spacegroup is $P2_12_12$ with cell dimensions a=49.7 Å, b=125.7 Å, c=45.6 Å, α=β=γ=90°. The full processing statistics are shown in Table 6.1.

6.2.3 $TssA_{NTD}$(1-255) experimental data collection

$TssA_{NTD}$(1-255) contains six tyrosine residues, therefore it might be possible to incorporate iodine into the crystals in order to phase the structure experimentally. Crystals taken from the same drop as the native crystals, in condition JCSG+ E11, were subjected to sublimation of elemental iodine crystals before being cryoprotected in 0.16 M calcium acetate, 0.08 M sodium cacodylate buffer pH 6.5, 16.4% (w/v) PEG8000 and 30% glycerol and cooled in liquid nitrogen, after two separate time points - 2 hours and overnight. Crystals were sent to beamline I03 at the Diamond Light Source Synchrotron, Oxford. Peak data were collected at 1.7 Å wavelength with 0.2 second exposures, over a phi oscillation of 0.2° for a total of 720°, at 2.5% transmission. The data processed in Fast DP (Winter and McAuley, 2011) to 2.04 Å in P222 with cell dimensions a=45.6 Å, b=49.8 Å, c=125.5 Å, α=β=γ=90°. The full processing statistics are shown in Table 6.1.

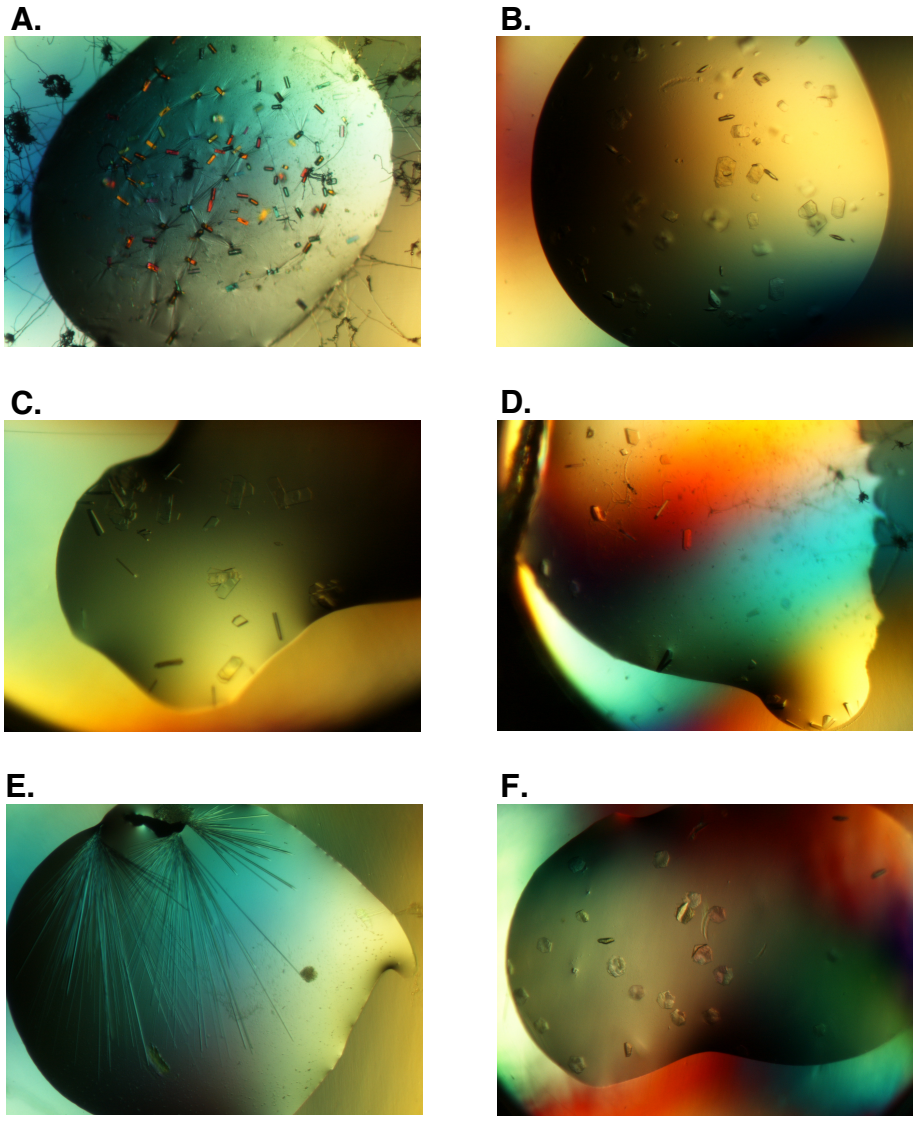


Figure 6.3. Crystal hits from sitting-drop crystallisation trials for $TssA_{NTD}$(1-255). Protein set down at 12 mg/mL in 10 mM Tris pH 8.0 **A.** Classic F7 – 0.2 M calcium acetate, 0.1 M sodium cacodylate buffer pH 6.5, 18% (w/v) PEG8000. **B.** Proplex E6 – 0.1 M calcium acetate, 0.1 M sodium cacodylate buffer pH 5.5, 12% (w/v) PEG8000. **C.** Proplex A2 – 0.1 M calcium acetate, 0.1 M MES buffer pH 6.0, 15% (v/v) PEG400 **D.** JCSG+ E11 – 0.16 M calcium acetate, 0.08 M sodium cacodylate buffer pH 6.5, 14.4% (w/v) PEG8000, 20% (v/v) glycerol. **E.** Classic F8 – 0.2 M magnesium acetate, 0.1 M sodium cacodylate buffer pH 6.5, 20% (w/v) PEG8000. **F.** Proplex B2 - 0.1 M calcium acetate, 0.1 M sodium acetate pH 4.5, 10% (w/v) PEG4000.

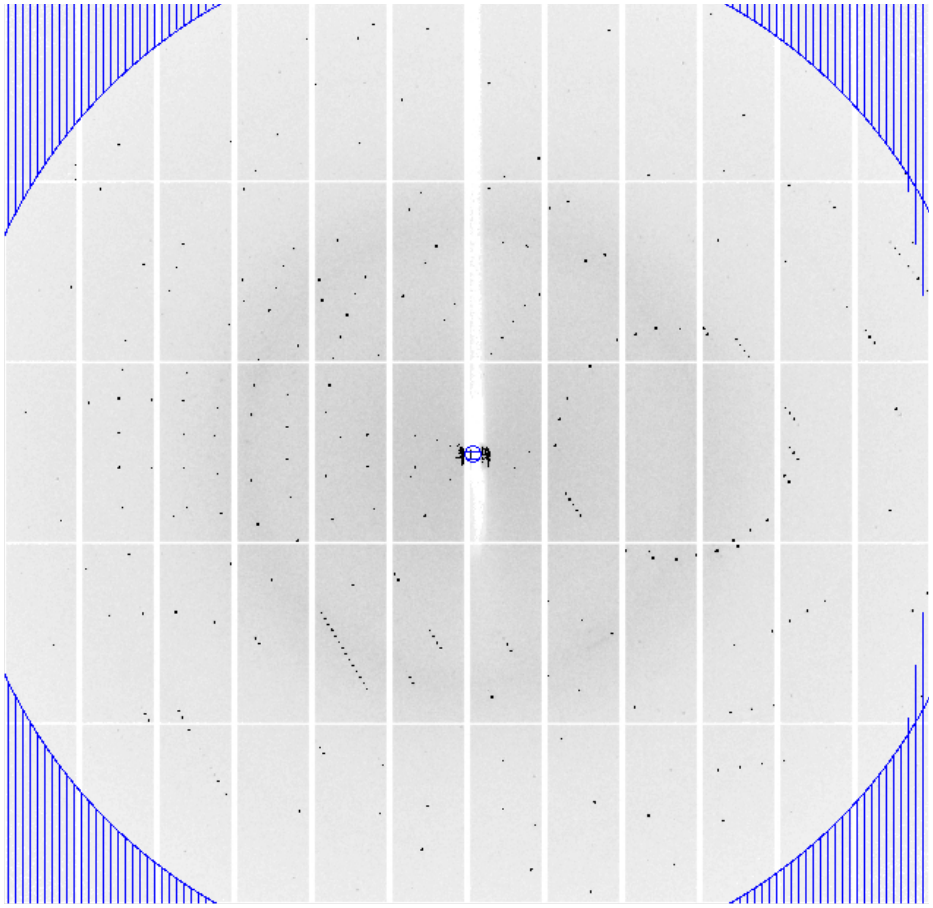


Figure 6.4. Diffraction test image of a crystal of $TssA_{NTD}$(1-255). Diffraction test image taken on a crystal of $TssA_{NTD}$(1-255) from condition JCSG+ E11 prior to data collection, with the resolution ring set to 1.8 Å. Data collected at the Diamond Light Source Synchrotron, Oxford. Image generated in iMosflm (Battye et al., 2011).

Data collection – $TssA_{NTD}$(1-255)	**Native**	**Iodine Peak**
Wavelength (Å)	1.04434	1.70000
Resolution range (Å)	49.67-1.80 (1.85-1.80)	29.64-2.04 (2.09-2.04)
Space group	$P2_12_12$	P 2 2 2
Unit cell parameters		
a (Å)	49.67	45.57
b (Å)	125.74	49.81
c (Å)	45.56	125.51
α (°)	90	90
β (°)	90	90
γ (°)	90	90
Total reflections	146307 (10493)	427376 (15999)
Unique reflections	27244 (1966)	18554 (1011)
Multiplicity	5.4 (5.3)	23.0 (15.8)
Completeness (%)	99.8 (99.0)	98.1 (74.2)
Mean I/σ	21.2 (4.4)	19.3 (2.8)
Wilson B factor ($Å^2$)	18	-
R_{merge}	0.043 (0.344)	0.169 (0.865)
R_{pim} (I)	0.022 (0.177)	0.038 (0.215)
Anomalous completeness	-	97.7 (69.2)
Anomalous multiplicity	-	12.2 (8.1)
Anomalous correlation	-	0.556 (0.033)
Anomalous slope	-	1.022

Table 6.1. Data collection processing statistics for $TssA_{NTD}$(1-255)
xia2 3d (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Winter, 2010) and Fast DP (Winter and McAuley, 2011) - information from Aimless (Collaborative Computational Project, Number 4, 1994), data processing statistics for $TssA_{NTD}$(1-255) native and iodine SAD collected data, respectively. The high-resolution shell information is shown in brackets.

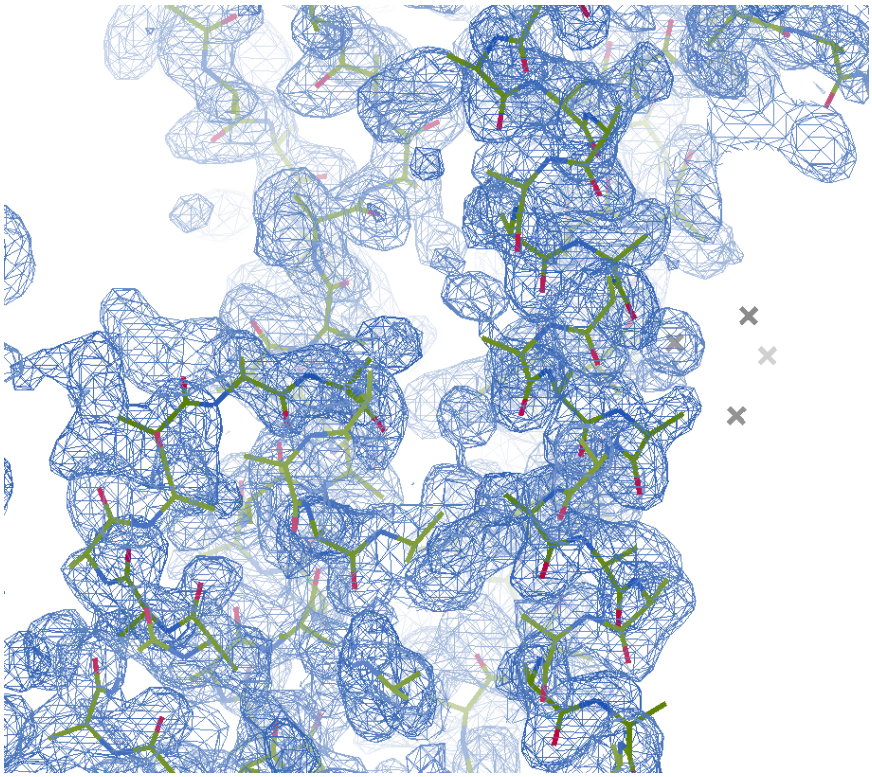
6.2.4 Experimental phasing of the $TssA_{NTD}$(1-255) structure

The structure of $TssA_{NTD}$(1-255) was solved using phases from iodine SAD. The Fast DP processed data were fed into the Fast EP pipeline (Diamond Light Source – automatic software pipeline), which uses the *SHELX* program to identify the spacegroup and iodine sites (Sheldrick, 2008). *SHELXE* (Sheldrick, 2008), was re-run, to produce a model, with a solvent content of 55%, looking for α-helices, (predicted by PSI-PRED – Buchan et al., 2013; Jones, 1999) and the seven sites from *SHELXD* (Sheldrick, 2008), and five autobuilding cycles. A map and initial model was produced containing 216 residues in six separate poly-alanine chains, in spacegroup $P22_12_1$ (Figure 6.5). The initial model was used to molecular replace the data previously collected on the native protein, at the same time resolving the cell axes definitions to the standard setting for the spacegroup.

The model of $TssA_{NTD}$(1-255) was built in *Coot* (Emsley et al., 2010) and underwent cycles of refinement against the 1.8 Å native data set in Refmac5 (Murshudov et al., 1997; Murshudov et al., 2011). The final R_{factor} is 0.177 and R_{free} is 0.226. The structure contains residues 1-254 except residues 39-51, which remain unmodelled due to the lack of electron density, also 119 water molecules, two calcium ions and an ethylene glycol molecule. Electron density is present for the first eight residues of the purification tag, residues -7 to 0. The location of the iodine sites were analysed in the final structure. The first two iodine sites are located next to Y241. The third and fourth iodine sites correspond to the two calcium sites modelled. The fifth is near H161, and the sixth and seventh were not in density in the initial *SHELXE* map. The validation statistics are shown in Table 6.2.

6.2.5 $TssA_{NTD}$(1-255) structure analysis

The $TssA_{NTD}$(1-255) structure is formed of 11 α-helices (Figure 6.6). The N-terminus of the protein has a loop region, containing the purification tag



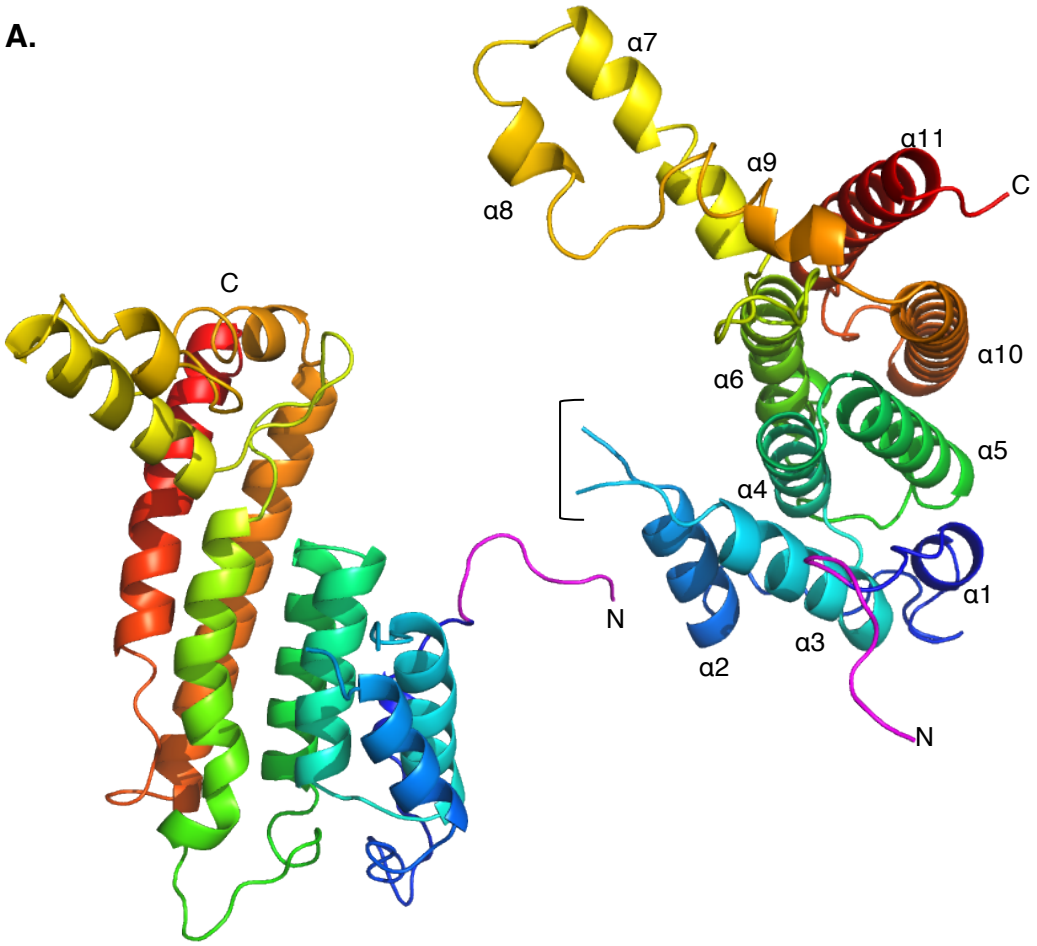
Iodine site	Occupancy
1	1.000
2	0.576
3*	0.569
4*	0.181
5	0.326
6	0.280
7	0.267

Figure 6.5. Map output from Fast EP in the original hand, with initial model and predicted iodine sites. Electron density map, model and sites output from Fast EP. The poly-alanine model contains 216 residues separated into six chains. A α-helix has been built into the helical density shown here. The predicted iodine sites are shown as grey pointer atoms. Map set to 1.6671e/A^3 = 1.00 rmsd. Image generated in *Coot* (Emsley et al., 2010). The number and occupancy of the seven iodine sites is shown in the table, from *SHELXE* (Sheldrick, 2008) in Fast EP with * indicating the sites corresponding to calcium sites.

Refinement	**$TssA_{NTD}$(1-255)**
R_{factor}	0.177
R_{free}	0.226
No. of non-H atoms	
Protein	1991
Metal ions	2 – 2 x Calcium ions
Ligands	4 – 1 x Ethylene glycol
Water	119
Protein residues	254
RMSD (bonds) (Å)	0.0157
RMSD (angles) (°)	1.5665
Ramachandran favoured (%)	98.79
Ramachandran outliers (%)	0.00
Favoured rotamers (%)	95.15
Poor rotamers (%)	0.00
MolProbity score	0.85 (100^{th} percentile, N=11444, 1.80Å ± 0.25Å)
Average B factors	
Main chain ($Å^2$)	24
Side chains ($Å^2$)	37
Metal ions ($Å^2$)	(25 + 16) average 20.5
Ligands ($Å^2$)	23
Water ($Å^2$)	29

Table 6.2. Refinement and validation statistics for $TssA_{NTD}$(1-255). Statistics generated in Refmac5 (Murshudov et al., 1997; Murshudov et al., 2011), MolProbity (Chen et al., 2010) and Baverage (Winn et al., 2011).

A.



B.

GLVPRGSHMPINLPELLTPISEASPSGDDLLFSNEFDAIQDARRYD~~DPTLD~~
~~QGEWVTEI~~KEADWGFVVDHAGELLRTRTKDLRLAVWLTEALALEDGITGL
TEGYALLEGLCREFWDTFHPLPEDDDIEHRLGNVAWLSGRTAELLRAVPL
TDGASNAFSTLDWEVAQHVAQSIKRDPEHADDIARGKPSIEQIDASRRVT
SIAFYTALLANLKAFEFALDAFEERLVERAGDSAPSFRQARDAFETVYRLA
ERFAREQGYT~~G~~

Figure 6.6. $TssA_{NTD}$(1-255) structure

A. The structure of the $TssA_{NTD}$(1-255) shown as cartoon, coloured by chainbows. The N-terminal His-tag linker is highlighted in magenta. The N- (blue) and C-termini (red) are labelled. **B.** The $TssA_{NTD}$(1-255) sequence. The section of the His-tag sequence still present in the structure is highlighted in magenta. The $TssA_{NTD}$ construct consists of residues 1-255. Residues 39-51 are absent in the structure (indicated by the square bracket in Figure 6.6A) due to a lack of electron density, and are crossed through in the sequence, along with the N-terminal G255. Images generated in PyMOL version 1.1r1..

linker, which leads to helix α1. Following α1 a long loop tucks under the protein and leads to four similar sized helices forming two helical pairs, α2-α3 and α4-α5. Helices α4 and α5 pack against α6 and α10. α11 packs against α6 and α10 to form a bundle of five helices (α4, α5, α6, α10 and α11), with a predominantly hydrophobic core. α7, α8 and α9, form a helical protrusion at right angles, to the other eight helices. Analysis of the crystal packing, by PDBePISA (Krissinel and Hendrick, 2007), did not identify any oligomerisation of the $TssA_{NTD}$ protein monomers in the crystal lattice. The two calcium sites form crystal contacts between the molecules and are presumed to come from the crystallisation solution, which contains calcium acetate.

6.2.6 TssA-1 and TssA-2 share a similar domain at the NTD

The sequence alignment carried out between TssA proteins revealed two areas of sequence conservation between the two major types of TssA – TssA-1 and TssA-2 (Dr. Mark Thomas, personal communication) (Figure 6.1). *B. cenocepacia* TssA is a member of the TssA-1 group (Dr. Mark Thomas, personal communication). The regions of sequence similarity are, using the residue numbering from *B. cenocepacia* TssA, residues L8-I31 and D55-P113. The sequence conservation suggests that TssA-1 and TssA-2 share a similar domain at the N-terminus. The two regions of conservation identified are mapped onto the $TssA_{NTD}$(1-255) structure (Figure 6.7). All the conserved regions are clustered around helices α1, α2, α3, α4 and α5. The region of variability between the L8-I31 region and the D55-P113 region corresponds to the loop between α2-α3 in $TssA_{NTD}$(1-255), which is not present in the structure. The second conserved region, shown in green, coincides with a region of negative charge on the surface of $TssA_{NTD}$ (Figure 6.8).

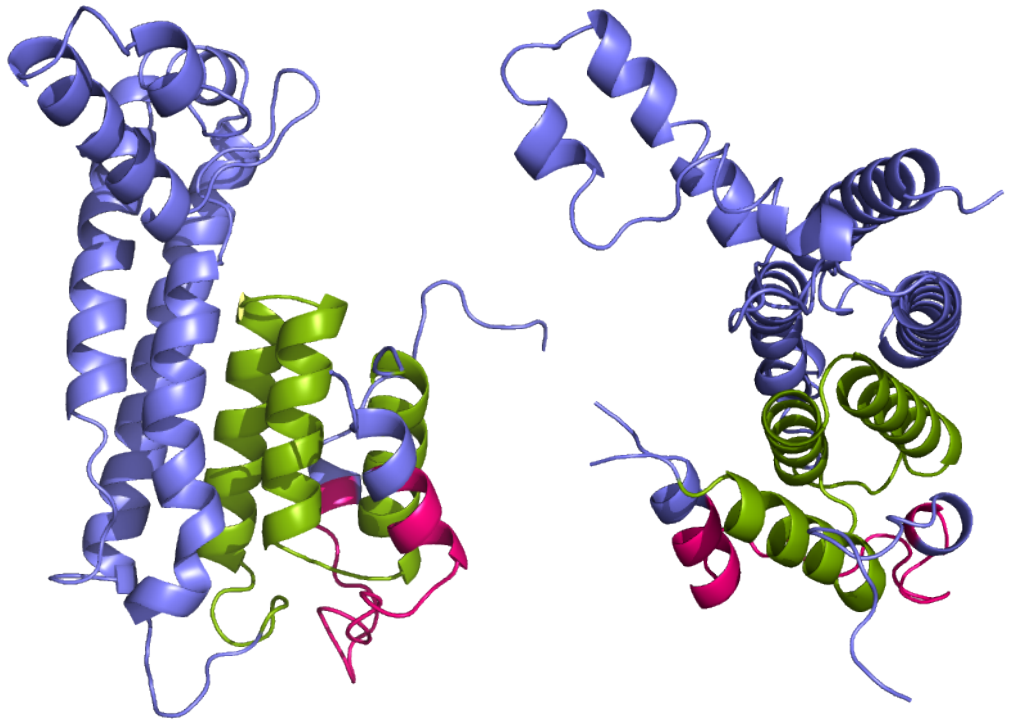


Figure 6.7. Conserved regions of sequence between TssA-1 and TssA-2 highlighted on the *B. cenocepacia* $TssA_{NTD}$ structure. The structure of $TssA_{NTD}$(1-255) is shown as cartoon and coloured purple, the N-terminal His-tag linker is shown. The first conserved sequence region covers residues 8-31 and corresponds to parts of helices α1 and α2 and the α1-α2 loop, shown in pink. The second conserved sequence region covers residues 55-113 and corresponds to helices α3-α5 and part of the α5-α6 loop, shown in green. Residue numbers are as they appear in $TssA_{NTD}$ sequence. Images generated in PyMOL version 1.1r1..

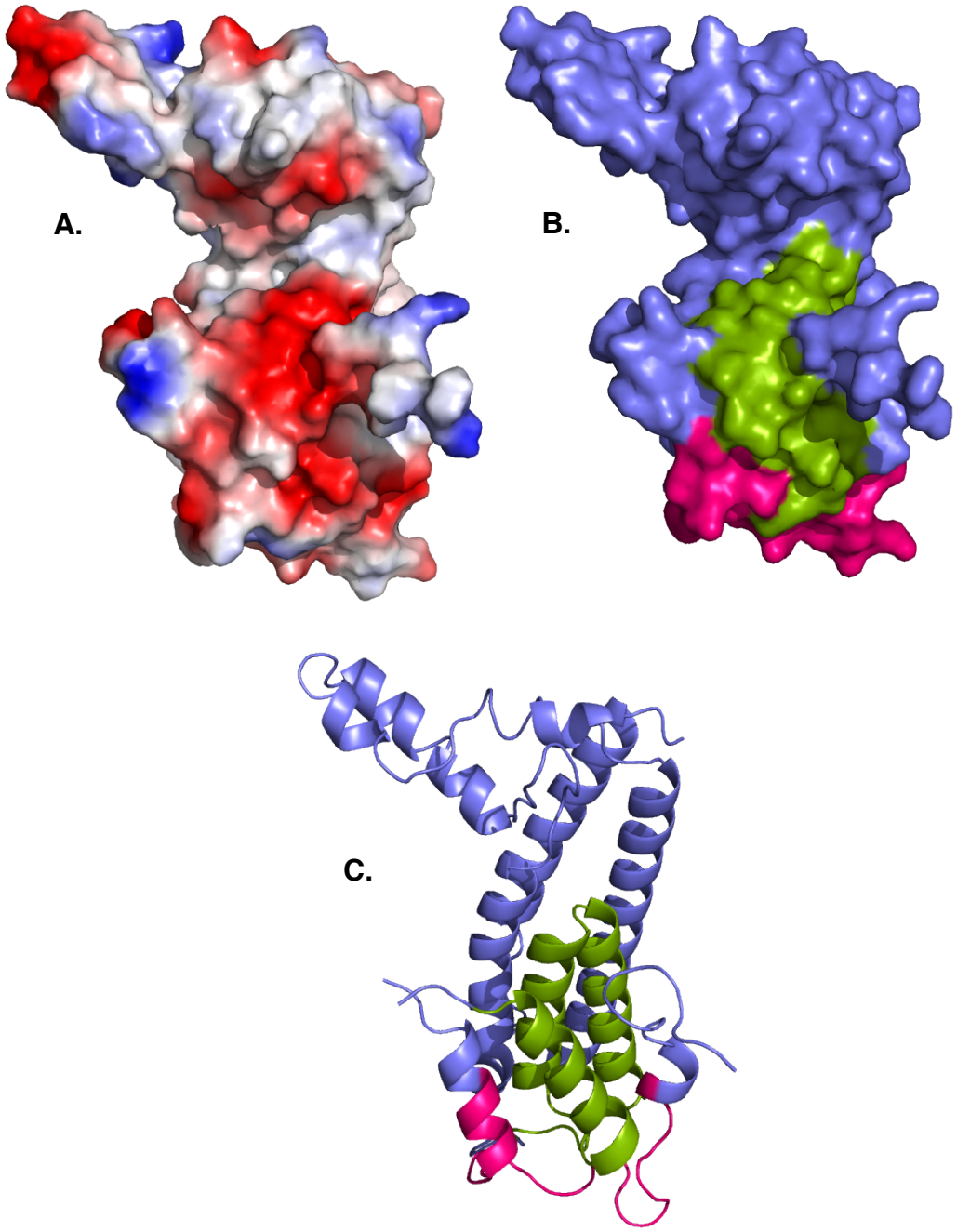


Figure 6.8. Conserved residues and surface charge of $TssA_{NTD}$(1-255). The structure of $TssA_{NTD}$(1-255) is shown as surface in **A**. and **B**. and cartoon in **C**., the N-terminal His-tag linker is shown. **A.** A large patch of negative surface charge identified on one side of $TssA_{NTD}$(1-255). **B.** The negatively charged region coincides with the second conserved region (residues 55-113, as in $TssA_{NTD}$(1-255)) shown in green. **C.** The second conserved region corresponds to helices α3-α5 and part of the α5-α6 loop. Vacuum electrostatics and images generated in PyMOL version 1.1r1.

6.2.7 $TssA_{NTD}$(1-255) structural comparison using the Dali server

The $TssA_{NTD}$(1-255) structure was run through the Dali Server (Holm and Rosenstrom, 2010), and the PDB90 results were analysed. The PDB90 is derived from the PDB, as a group of structures that only includes one chain of a protein, when more than one chain is present that is 90+% identical in sequence (Holm and Rosenström, 2010). The results indicate the structure is similar to protein unc13, heat shock proteins, TPR (tetratricopeptide repeat)-containing proteins, TssA N-terminal region from enteroaggregative *E. coli*, chaperones, anaphase-promoting complex subunit and regulatory proteins (Holm and Rosenstrom, 2010) (Table 6.3). The analysis of the Dali structural alignments revealed a number of these proteins aligned to the α2-α5 helices of the $TssA_{NTD}$ (Figure 6.9). Seven out of the top ten proteins from the PDB90 results contain a TPR domain or a TPR-like repeat domain (Holm and Rosenström, 2010). A TPR motif has 34 amino acids, which folds into a helix-turn-helix structure (Das et al., 1998; Hirano et al., 1990; Sikorski et al., 1990). The TPR motif has been identified to have a role in protein-protein interaction (Lunelli et al., 2009; Pallen et al., 2003). The $TssA_{NTD}$ α2, α3, α4 and α5 helices are two pairs of antiparallel helices, which align with the two TPR folds from TPR containing proteins. The $TssA_{NTD}$ sequence was analysed for the presence of the TPR motif, using the TPR prediction program TPRpred (Karpenahalli et al., 2007), however no TPR repeats were identified. Whether the similarity between the $TssA_{NTD}$ and TPR containing protein is indicative of a function is as yet unclear.

6.2.8 *B. cenocepacia* $TssA_{NTD}$ and the *E. coli* $TssA_{Nt2}$ share a similar fold

During the course of this work the structure of the enteroaggregative *E. coli* (referred to as *E. coli* in this chapter) TssA C-terminus and an N-terminal fragment, Nt2 (corresponding to residues 221-377, out of 401) were solved (Zoued et al., 2016). The N-terminal fragment, Nt2, of *E. coli* TssA (PDB:4YO3), helices α2 to α7 (residues 242-374), was aligned with

No.	Chain	Z-score	RMSD (Å)	lali	nres	% id	Description
1	3SWH-B	7.1	3.7	132	297	9	Protein Unc-13 Homolog A
2	3UQ3-A	6.9	4.7	110	258	10	Heat shock-protein STI1
3	3U4T-A	6.9	5.7	103	258	6	TPR Repeat-containing protein
4	3SZ7-A	6.8	4.6	121	151	10	Hsc70 cochaperone (SGT)
5	2PL2-A	6.7	8.8	107	194	10	Hypothetical conserved protein TTC0263
6	4YO3-A	6.7	3.1	101	156	10	TssA
7	4GCN-A	6.7	3.4	95	127	11	Protein STI-1
8	4J8F-A	6.7	3.6	84	551	7	Heat Shock 70kDa protein 1A/1B, Hsc70-interactin
9	4UI9-Y	6.6	5.1	71	496	14	Anaphase-promoting complex subunit 1
10	2XCB-A	6.6	4.2	76	133	9	Regulatory protein PcrH

Table 6.3. The top 10 Dali server PDB90 results for $TssA_{NTD}$(1-255) structure. The $TssA_{NTD}$(1-255) structure was run through the Dali server (Holm and Rosenström, 2010). The top 10 PDB90 results are shown in the table. The columns refer to: no. - match number, chain - PDB code and corresponding chain, Z-score – Z-score of match, RMSD – RMSD of match, lali – number of aligned residues, nres – number of residues, %id - % sequence identity and the description of the PDB structure (Holm and Rosenström, 2010).

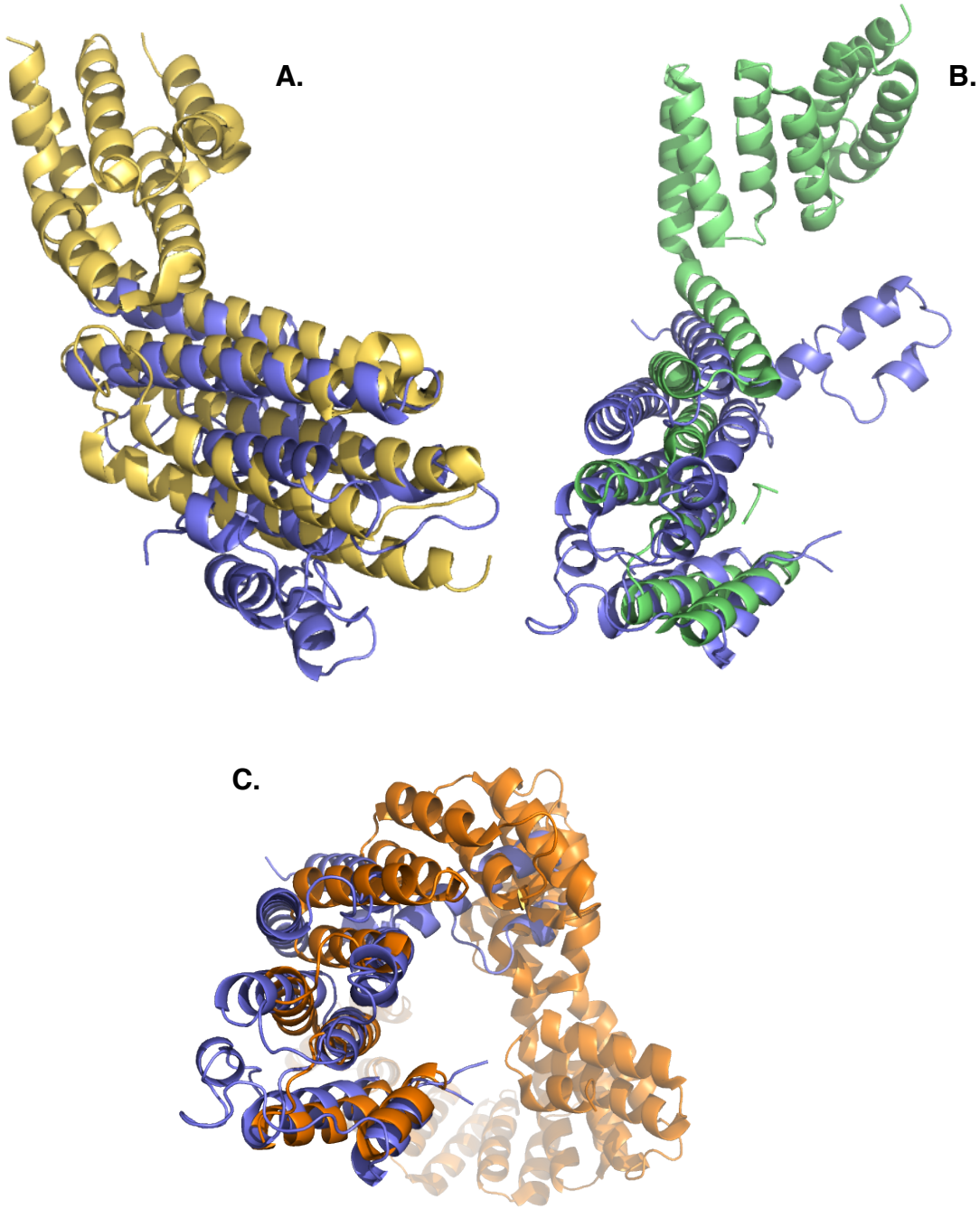


Figure 6.9. Dali server structural alignment of $TssA_{NTD}$(1-255) and the top three Dali PDB90 results. The structures are shown as cartoon. $TssA_{NTD}$(1-255) is coloured purple. Alignments carried out by the Dali Server (Holm and Rosenström, 2010). **A.** Protein unc-13 homolog A is coloured yellow (PDB:3SWH) (Li et al., 2011) – aligns with helices α4, α5, α6, α10 and α11 **B.** Heat shock protein STI1 is coloured green (PDB:3UQ3) (Schmid et al., 2012)**. C.** TPR containing protein is coloured orange (PDB:3U4T) (Vorobiev et al., unpublished). Images generated in PyMOL version 1.1r1.

the *B. cenocepacia* $TssA_{NTD}$ structure, helices α1 to half of α6 (residues 1-132), by the Dali server (Holm and Rosenstrom, 2010) (Figure 6.10A). The Dali alignment (Holm and Rosenstrom, 2010) shows, similarly to the other alignments mentioned above, that *E. coli* $TssA_{Nt2}$ α3, α4, α5 and α6 aligns with *B. cenocepacia* $TssA_{NTD}$(1-255) helices α2, α3, α4 and α5. This similarity between the structures suggests that the sequences in these two regions would show some conservation. However analysis of the sequence corresponding to the structurally aligned regions of *E. coli* $TssA_{Nt2}$ and of *B. cenocepacia* $TssA_{NTD}$ (Holm and Rosenstrom, 2010) found ~10% sequence based similarity (Figure 6.10B).

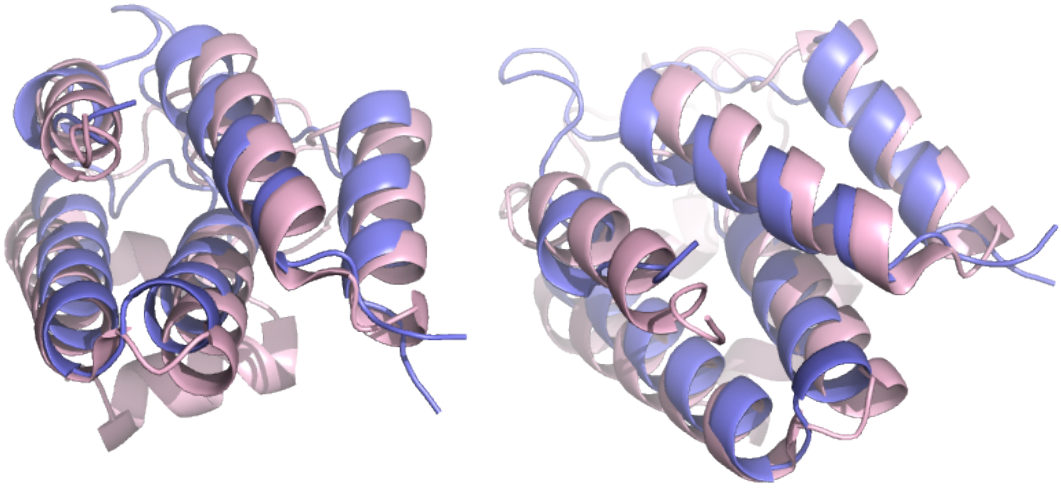
It has already been established by sequence conservation that the N-terminal region of TssA-1 and TssA-2 is similar between the two protein types (Figure 6.1A and B). Specifically the structure of residues 8-113 in *B. cenocepacia* TssA (half of α1 to half of the loop after α5 in $TssA_{NTD}$(1-255)), is predicted to be the same as the corresponding region of conservation in *E. coli* TssA N-terminus, residues 32-124, whose structure has yet to be solved. $TssA_{NTD}$(1-255) is shown to be structurally similar to the structure of *E. coli* $TssA_{Nt2}$, therefore it is predicted that the *E. coli* TssA N-terminus contains two sequential Nt2 domains – the first part of the NTD, followed by the Nt2. Therefore the fold of $TssA_{NTD}$ and $TssA_{Nt2}$ are related. The nature of the functional relationship between $TssA_{NTD}$ and $TssA_{Nt2}$ remains to be determined.

6.3 $TssA_{CTD}$

6.3.1 $TssA_{CTD}$ protein constructs

The constructs of $TssA_{CTD}$ provided by collaborators from the University of Sheffield Medical School discussed here are MBP(maltose-binding protein)-$TssA_{CTD}$(303-373) and MBP-6His-$TssA_{CTD}$(303-358). MBP-$TssA_{CTD}$(303-373) construct contains the CTD with three N-terminal residues of the interdomain linker. MBP-6His-$TssA_{CTD}$(303-358) covers the residues in helices α1, α2 and α3, of the CTD.

A.



B.

```
B.c GLVPRGSHMPINLPELLTPISEASPSGDDLLFSNEFDAIQDARRYDKEADWGFVVDHAGELLRTRTKDLRLAVWLTEALALEDG
E.c --DQPE---AHRL-XKSVR--------APPG-PDRRASLKRLYLQ---QNWLSLLEQCDDXFARGHLWLDLQWYIHQALLQTGK
    LLLLLLLLLLLLHHHHLLLLLLLLLLLLLLLHHHHHHHHHHLLLLLLLHHHHHHHHHHHHHHLLLLHHHHHHHHHHHHHHL
    --HLLL---HHHH-HHHHH--------LLLL-HHHHHHHHHHHHL---LLHHHHHHHHHHHHLLLLLLHHHHHHHHHHHLLL

    ITGLTEGYALLEGLCREFWDTFHPLPEDDDIEHRLGNVAWLSGRTAELLRAVPLTDGASNAFSTLDWEVAQHVAQSIKRDPEHAD
    ENYAAIIQYDLKGLLLRLPGL--ETLA------DDVTLSWIQQ----------------------------------------
    HHHHHHHHHHHHHHHHHLLLLLLLLLLLLLHHHHLHHHHHHHHHHHHHHLLLLLLLLLLLLLHHHHHHHHHHHHHHHHLHHHHH
    HHHHHHHHHHHHHHHHLHHH--HHLL------LHHHHHHHHH----------------------------------------

    DIARGKPSIEQIDASRRVTSIAFYTALLANLKAFEFALDAFEERLVERAGDSAPSFRQARDAFETVYRLAERFAREQGYT
    ----------------------------------------------------------------------------------
    HHHLLLLLHHHHHHHHHLLLHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHLLLLHHHHHHHHHHHHHHHHHHLLLL
    ----------------------------------------------------------------------------------
```

Figure 6.10. Dali server structural alignment of *B. cenocepacia* $TssA_{NTD}$(1-255) and *E. coli* TssA N-terminal region, Nt2 – chain A. A. $TssA_{NTD}$(1-255) (purple) and *E. coli* TssA Nt2 (light pink) (PDB:4YO3) (Zoued et al., 2016) are shown as cartoon. Helices α2-α5 from $TssA_{NTD}$(1-255) aligned with helices α3-α6 of *E. coli* TssA Nt2, with partial alignment between α1 of $TssA_{NTD}$(1-255) and α2 of *E. coli* TssA Nt2, as the helices are oriented in a similar position. **B.** Structural alignment with sequences from both *B. cenocepacia* $TssA_{NTD}$(1-255) (B.c) and *E. coli* TssA NTR (PDB:4YO3) (Zoued et al., 2016) (E.c), with the assigned secondary structure beneath - H (helices), L (loops). Alignments by Dali Server (Holm and Rosenström, 2010) and images generated in PyMOL version 1.1r1.

6.3.2 $TssA_{CTD}$(303-373) protein crystallisation

$TssA_{CTD}$(303-373) protein, the MBP was cleaved prior to crystallisation, was set down into sitting-drop vapour diffusion crystallisation trials, with crystallisation condition suites PACT, JCSG+, Proplex and pHClear. Crystallisation trays were set down with 50 µL reservoir and 200 nL:200 nL drops. Trays were stored at 17°C. Crystal hits were found in a large range of conditions (Figure 6.11). A selection of crystals was sent to the Diamond Light Source Synchrotron, Oxford for data collection.

6.3.3 Information provided by $TssA_{CTD}$(303-373) crystal data collection and processing

The data collected from a number of different crystals was indexed and processed into different spacegroups and cell dimensions, the majority containing a cell length of ~45-48 Å (Table 6.4). The Matthews coefficient (Kantardjieff and Rupp, 2003; Matthews, 1968) calculation for data sets 1, 2, 3, 6, 7, 8, 9, 10, 12 and 13, indicated the number of $TssA_{CTD}$(303-373) monomers in the asymmetric unit could not be confidently predicted. A self-rotation function was carried out on $TssA_{CTD}$(303-373) data set 2 (processed with 3daii pipeline). For data set 2 in spacegroup I222, the self-rotation analysis showed the three 2-fold axes expected, along with 16 additional 2-fold axes in the b/c or ω(omega) plane (Figure 6.12). The 2-fold axes are separated by 22.5° rotation, and a 16-fold axis can be seen down the a-axis, at ω=90° φ(phi)=0° at κ(kappa)=23° rotation (Figure 6.12). The 16-fold rotation axis was predicted to be the rotation around the ring, therefore the cell length of ~46-48 Å was predicted to be the height of the ring. These data were consistent with EM and size exclusion chromatography - multi-angle laser light scattering results (Dr. Mark Thomas, personal communication), albeit the 16-fold rotation and plane of 2-fold axes suggests 32 subunits of $TssA_{CTD}$(303-373). The gel filtration analysis also predicted a large molecular weight for $TssA_{CTD}$ (Dr Svetlana Sedelnikova, personal communication).

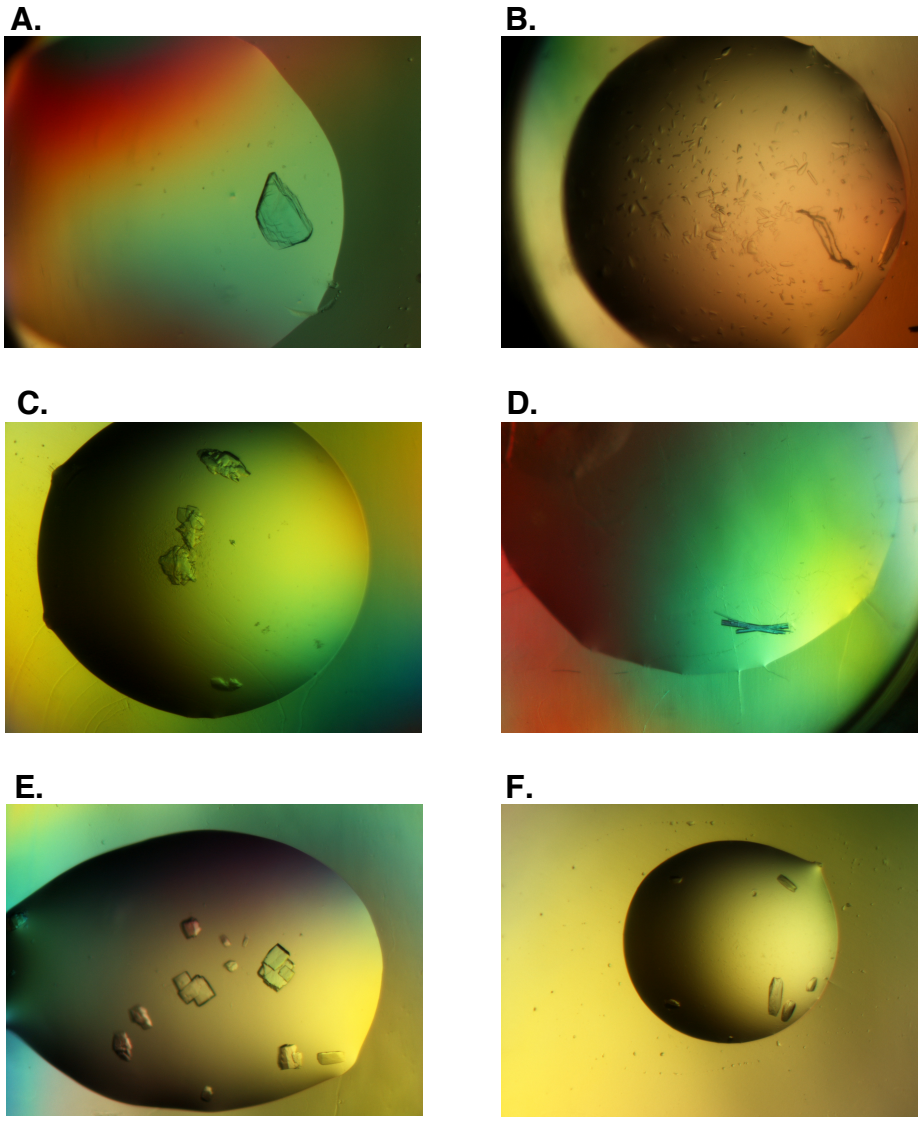


Figure 6.11. Crystal hits from sitting-drop crystallisation trials of $TssA_{CTD}$(303-373)

Protein was set down at 13 mg/mL. **A.** Proplex H12 – 0.1 M NaCl, 0.1 M Tris pH 8.0, 15% (v/v) ethanol, 5% (v/v) MPD. **B.** JCSG+ E3 – 0.2 M NaCl, 0.1 M HEPES pH 7.5, 10% (v/v) isopropanol. **C.** pHClear E4 – 0.1 M HEPES pH 7.0, 0.8 M ammonium sulphate. **D.** pHClear D2 – 0.1 M citric acid pH 5.0, 20% PEG6000. **E.** Proplex F9 – 0.1 M MES pH 6.5, 1 M ammonium sulphate. **F.** Proplex B9 – 0.1 M sodium cacodylate pH 6.0, 15% (w/v) PEG4000 - an umixed protein drop.

Data set	Space-group	Cell dimensions (Å and °)	Resolution (Å)	pH of crystallisation condition
1	I 2 2 2	**47.07** x 202.52 x 266.51, α=β=γ=90	3.83	JCSG+ E3 (Native) pH 7.5
2 (3da)	I 2 2 2	**46.33** x 201.7 x 263.66, α=β=γ=90	3.08	Proplex H12 (Iodine) pH 8.0
3	I 2 2 2	**46.02** x 200.25 x 262.97, α=β=γ=90	3.03	JCSG+ H2 Optimisation (SeMet) pH 6.0
4	I 2 2 2	**48.26** x 62.7 x 65.74, α=β=γ=90	1.78	pHClear D2 (Native) pH 5.0
5	I 2 2 2	**47.75** x 62.59 x 65.63, α=β=γ=90	1.71	JCSG+ B9 (SeMet) pH 4.0
6	C 1 2 1	614.22 x **46.23** x 308.71, α=γ=90 β=96.135	5.11	JCSG+ H2 (SeMet) pH 5.5
7	C 2 2 2	427.61 x 506.97 x **46.59**, α=β=γ=90	4.71	JCSG+ E3 (Iodine) pH 7.5
8	C 2 2 2_1	412.02 x 457.59 x **45.84**, α=β=γ=90	4.52	Proplex F9 (SeMet) pH 6.5
9	P 2_1 2_1 2_1	**46.05** x 305.45 x 312.35, α=β=γ=90	3.17	pHClear E4 (Native) pH 7.0
10	P 2 2 2	**45.32** x 203.32 x 223.67, α=β=γ=90	7.49	Proplex F10 (SeMet) pH 8.0
11	P 2_1 2_1 2	23.87 x 70.06 x 103.63, α=β=γ=90	2.93	Classic F10 (SeMet) pH 6.5
12	P2_1 2_1 2_1	**46.1** x 305.32 x 310.73 α=β=γ=90	3.44	pHClear E4 (Native) pH 7.0
13	P 4 2 2	309.28 x 309.28 x **46.52**, α=β=γ=90	5.83	JCSG+ H2 (SeMet) pH 5.5

Table 6.4. The range of spacegroups and cell dimensions for data collected on $TssA_{CTD}$(303-373) crystals, related to the pH of the crystallisation solution. The range of cell dimensions produced from data collected from crystals of $TssA_{CTD}$(303-373). The majority of cell dimensions have a length of ~45-48 Å, highlighted in bold. Data collections 4 and 5 correspond to the clipped structure of $TssA_{CTD}$(303-373). The smaller cell dimensions correspond to a crystallisation solution with a low pH, 4.0 or 5.0. The crystallisation condition (optimisation specified) and type of protein crystals are shown in brackets – native, selenomethionine-incorporated or iodine-soaked native. Data set 2 values from data processed with 3da pipeline. Spacegroup, cell dimensions and resolution provided by xia2 processing (Winter, 2010).

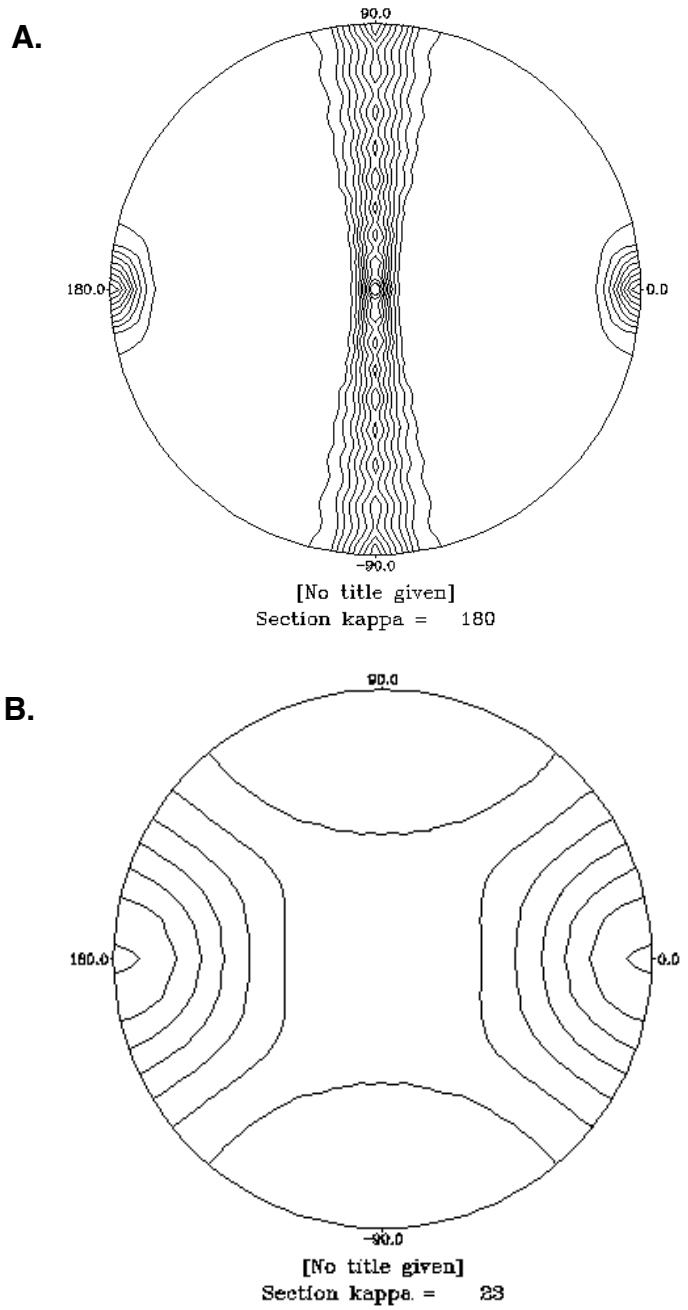


Figure 6.12. Self-rotation function of $TssA_{CTD}$(303-373) data
Self-rotation function performed using Polarrfn (Winn et al., 2011) on $TssA_{CTD}$(303-373) data from 80.61 Å-3.07 Å. 32 Å radius of integration was used. Results are viewed down the c-axis, with the edge of the circle as ϕ**. A.** At κ=180° rotation, the peaks corresponding to the crystallographic axes could be seen at ω=0° (ϕ=all), ω=90° ϕ=90° and ω=90° ϕ=0°, with heights equal to that of the origin. In addition non-crystallographic 2-fold axes can be found in the b/c plane at different values of ω and ϕ=90°. The 2-fold axes are separated by 22.5°. **B.** At κ=23° rotation, a 16-fold axis can be seen at ω=90° ϕ=0°. Figure generated in Polarrfn (Collaborative Computational Project, Number 4, 1994; Winn et al., 2011).

6.4 Initial $TssA_{CTD}$ structures

6.4.1 Data collection of $TssA_{CTD}$(303-373) iodine-soaked crystals

Attempts were made to phase the native data using iodine-sublimated crystals. Data were collected on a crystal from condition Proplex H12 – 0.1 M NaCl, 0.1 M Tris pH 8.0, 15% ethanol and 5% MPD. 360° of data were collected with 0.05 second exposure, 0.1° phi oscillation and 5.3% transmission (Figure 6.13). The wavelength was set to 1.7 Å, due to the potential incorporation of iodine into the crystals. The data processed to a resolution of 3.08 Å in spacegroup I222. However on processing of the data, insufficient anomalous signal was present thus suggesting no iodine was present, therefore this dataset was treated as native.

6.4.2 Selenomethionine-incorporated $TssA_{CTD}$(303-373) protein crystallisation

Selenomethionine-incorporated MBP-$TssA_{CTD}$(303-373) was produced based on the protein overexpression protocol designed by collaborators from University of Sheffield Medical School (section 2.4.4). The MBP was cleaved during purification. Initial sitting-drop crystallisation trials were set down for multiple purification preparations of selenomethionine-incorporated $TssA_{CTD}$(303-373) with 50 µL reservoir and 200 nL:200 nL drops, using a selection of crystallisation suites – PACT, JCSG+, pHClear, Classic, Proplex, $AmSO_4$. A number of crystal hits were produced from a range of conditions (Figure 6.14).

6.4.3 Analysis of $TssA_{CTD}$(303-373) selenomethionine structure – two-helix structure (residues 303-347)

Selenomethionine-incorporated $TssA_{CTD}$(303-373) protein crystals from condition JCSG+ B9 (0.1 M Citric acid pH 4.0, 20% (w/v) PEG6000) were sent to the Diamond Light Source Synchrotron, Oxford, for data collection. Data were collected using a selenomethionine MAD strategy – using native, peak, inflection and high-energy remote wavelengths. The data processed into I222 with average cell dimensions of a=47.7 Å,

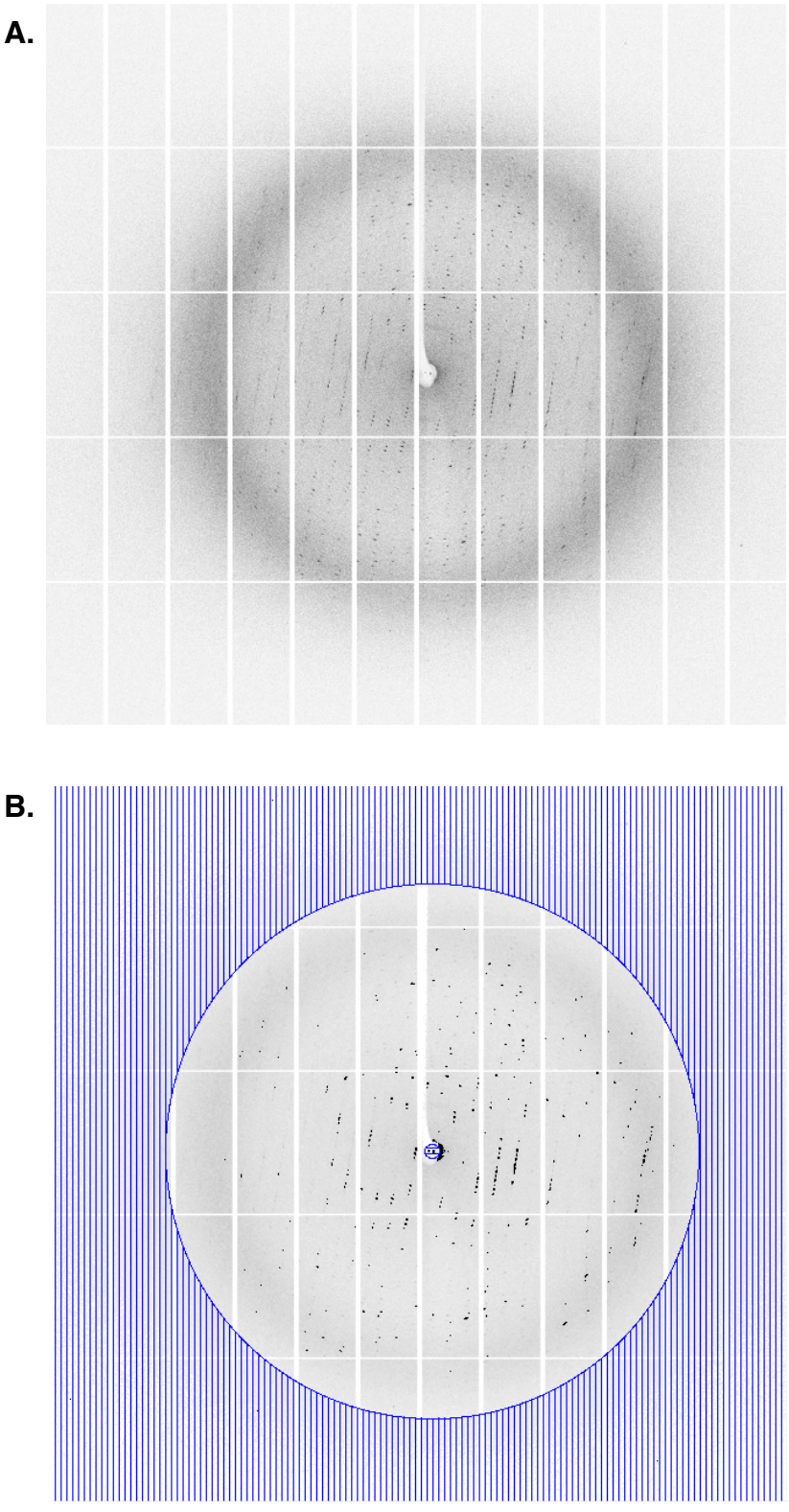


Figure 6.13. Diffraction test image of a crystal of $TssA_{CTD}$(303-373) from condition Proplex H12. Diffraction test image taken prior to data collection. **A.** The edge of the detector is set to 2.5 Å. Image from Diamond Light Source JPEGS. **B.** Resolution ring set to 3.08 Å. Image generated in iMosflm (Battye et al., 2011).

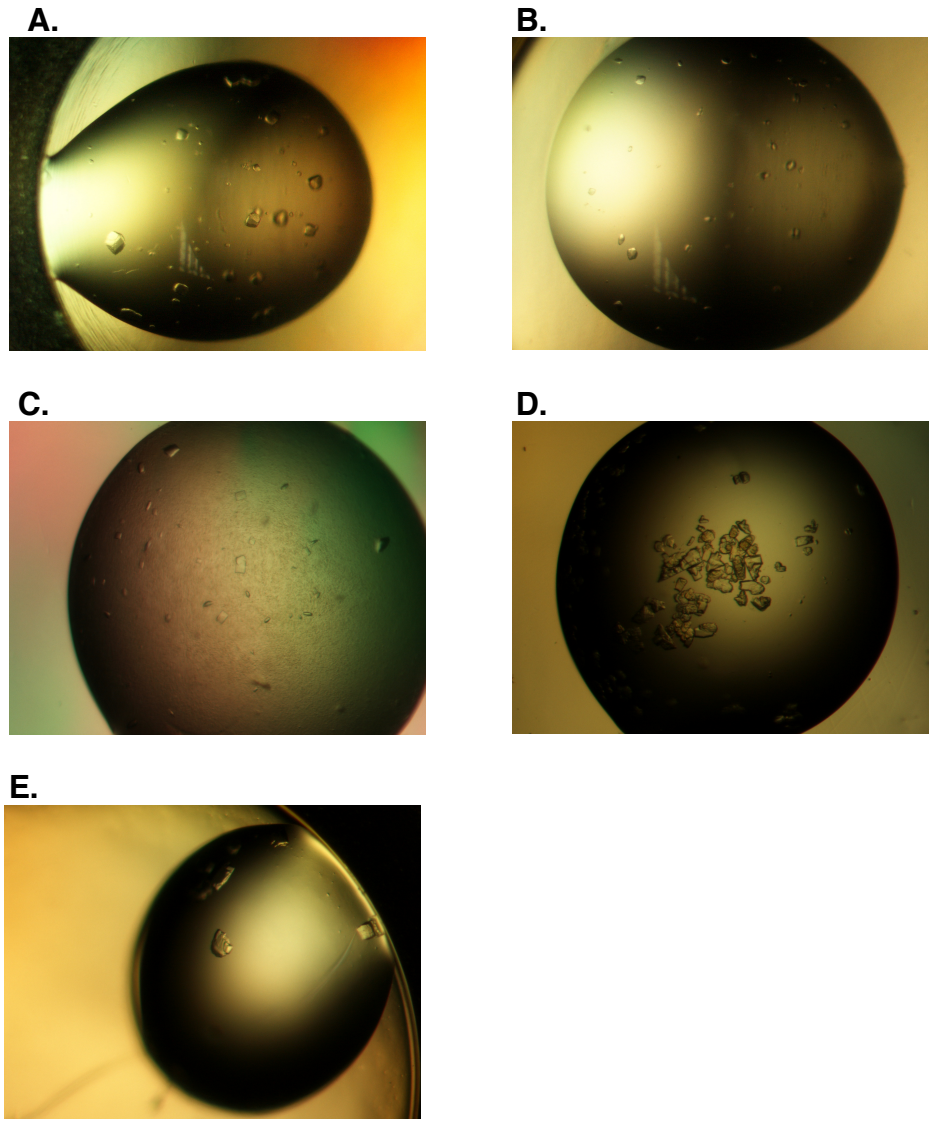


Figure 6.14. A selection of crystal hits from a number of sitting-drop crystallisation trials for selenomethionine-incorporated $TssA_{CTD}$(303-373). A. JCSG+ H2 – 1 M ammonium sulphate, 0.1 M Bis-Tris pH 5.5, 1% (w/v) PEG3350, **B.** pHClear E4 – 0.1 M HEPES, 0.8 M ammonium sulphate, pH 7.0 **C.** Proplex F10 – 0.1 M Tris pH 8.0, 1 M ammonium sulphate, **D.** Proplex F9 – 0.1 M MES pH 6.5, 1 M ammonium sulphate, **E.** Proplex F9 – 0.1 M MES pH 6.5, 1 M ammonium sulphate.

b=62.7 Å, c=65.7 Å, α=β=γ=90°. An experimentally phased map was produced using the *SHELX* programs, *SHELXC*, *D* and *E* (Sheldrick, 2008) and *HKL2MAP* interface (Pape and Shneider, 2004). The model was built in *Coot* (Emsley et al., 2010) and refined using Refmac5 (Murshudov et al., 1997; Murshudov et al., 2011), to a native data set from crystals grown in condition pHClear D2 (0.1 M Citric acid pH 5.0, 20% (w/v) PEG6000), collected previously, to 1.78 Å. The structure contains residues 303-347, representing the first two helices of the $TssA_{CTD}$(303-373) construct. It is assumed that the protein construct has undergone degradation during crystallisation, leading to the removal of 26 C-terminal residues of the protein construct. The asymmetric unit contains two molecules, which form a dimer through the interaction between the two helices in the structure. The arrangement of the protein in the crystal does not form a ring structure.

6.4.4 MBP-6His-$TssA_{CTD}$(303-358) structure - three-helix structure

The $TssA_{CTD}$ domain has been predicted to contain four helices, using PSI-PRED (Buchan et al., 2013; Jones 1999; Dr. Mark Thomas, personal communication) the first two of which had been identified in the structure of the degraded $TssA_{CTD}$(303-373) construct. A construct was produced of the first three helices, MBP-6His-$TssA_{CTD}$(303-358). This construct was purified with the MBP protein cleaved and set down into sitting-drop crystallisation trials of PACT, JCSG+ and Proplex at 10 mg/mL and pHClear and PEGs at 7.5 mg/mL. Crystals were produced in condition PEGs H12 (0.18 M tri-ammonium citrate, 20% (w/v) PEG3550), and data collected and processed into spacegroup $P6_2$, (a=65.3 Å, b=65.3 Å, c=66.1 Å, α=β=90° γ=120°) to a resolution of 2.35 Å. The $TssA_{CTD}$(303-358) structure was solved by molecular replacement in PhaserMR (McCoy et al., 2007) using the two helical $TssA_{CTD}$(303-373) structure. The structure contains residues 303-(354-358), with an extra N-terminal residue from the purification tag, in two molecules forming a dimer via the same interactions identified in the previous two-helix structure, and

similarly the arrangement of the protein in the crystal does not form a ring structure.

6.5 Structure for complete $TssA_{CTD}$

6.5.1 Selenomethionine-incorporated $TssA_{CTD}$(303-373) hanging-drop crystallisation trials

A number of sitting-drop crystallisation conditions that produced crystals, were optimised using hanging-drop optimisation trials (Figure 6.15). Selenomethionine-incorporated $TssA_{CTD}$(303-373) protein was purified and hanging drop optimisation trials were set down for condition JCSG+ H2, containing 1 M ammonium sulphate, 0.1 M Bis-Tris pH 5.5 and 1% (w/v) PEG3350. The PEG3350 concentration was varied from 0.25-2% (w/v) PEG3350 in increments of 0.25% and the 0.1 M Bis-Tris buffer pH set at pH 5.0, pH 5.5 and pH 6.0. Protein was set down in drops of 2 µL:2 µL and 2 µL:1 µL, protein:reservoir ratio.

6.5.2 Selenomethionine-incorporated $TssA_{CTD}$(303-373) crystal data collection

Crystal hits from optimised condition 1 M ammonium sulphate, 0.1 M Bis-Tris buffer pH 6.0 and 2% (w/v) PEG3350 were cryoprotected in 1-1.2 M ammonium sulphate, 0.1 M Bis-Tris buffer pH 6.0, 3-4% (w/v) PEG3350 and 30% glycerol, and cooled in liquid nitrogen. The crystals were sent to the Diamond Light Source Synchrotron, Oxford, beamline I03. A fluorescence scan was performed at the selenium K edge, to check for incorporation of selenium in the crystal. The energy at which to collect the data set was provided by *CHOOCH* (Evans and Pettifer, 2001), with peak energy of 12661.5 eV (0.9792 Å), with f'' = 5.28 e, f' = -7.53 e. Data were collected over 360° with 0.1° phi oscillation, 0.05 second exposure, and 80% transmission. The data were processed with the xia2 3d pipeline (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Winter, 2010) and the processing statistics are shown in Table 6.5.

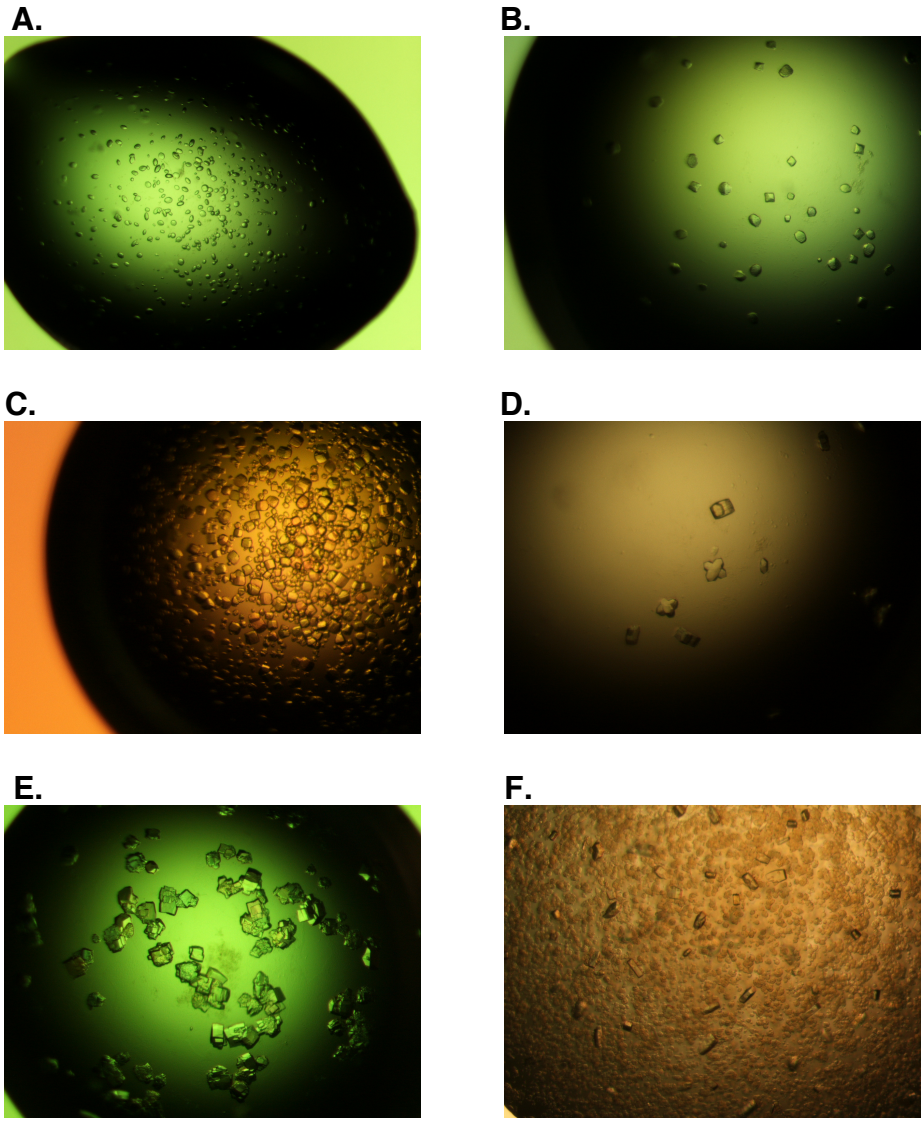


Figure 6.15. A selection of crystal hits from a number of hanging-drop crystallisation optimisation trials for selenomethionine-incorporated $TssA_{CTD}$(303-373). Ratio of protein:reservoir is specified for **B.** and **C. A.** pHClear E4 optimisation – 0.1 M HEPES pH 7.0, 0.6 M ammonium sulphate. **B.** Proplex F10 optimisation – 0.1 M Tris pH 8.0, 2.0 M ammonium sulphate, ratio 1:2. **C.** JCSG+ H2 optimisation – 1 M ammonium sulphate, 0.1 M Bis-Tris pH 5.0, 1.75% (w/v) PEG3350, ratio 2:2. **D.** JCSG+ H2 optimisation – 1 M ammonium sulphate, 0.1 M Bis-Tris pH 6.0, 2.0% (w/v) PEG3350. **E.** JCSG+ H2 optimisation – 1 M ammonium sulphate, 0.1 M Bis-Tris pH 6.0, 0.25% (w/v) PEG3350. **F.** Proplex F9 optimisation - 0.1 M MES pH 6.5, 1.25 M ammonium sulphate.

Data collection $TssA_{CTD}$(303-373)	**Native**	**Selenomethionine Peak**
Wavelength (Å)	1.70001	0.97922
Resolution range (Å)	41.49-3.08 (3.16-3.08)	100.12-3.03 (3.11-3.03)
Space group	I 2 2 2	I 2 2 2
Unit cell parameters		
a (Å)	46.3	46.0
b (Å)	201.7	200.3
c (Å)	263.7	263.0
α (°)	90	90
β (°)	90	90
γ (°)	90	90
Total reflections	297122 (20394)	315386 (23629)
Unique reflections	23602 (1709)	24385 (1759)
Multiplicity	12.6 (11.9)	12.9 (13.4)
Completeness (%)	99.9 (100.0)	99.9 (100.0)
Mean I/σ	17.5 (3.7)	9.8 (1.4)
Wilson B factor ($Å^2$)	53	87
R_{merge}	0.131 (0.786)	0.220 (2.144)
R_{pim} (I)	0.040 (0.251)	0.068 (0.622)
Anomalous completeness	-	100.0 (100.0)
Anomalous multiplicity	-	6.9 (7.0)
Anomalous correlation	-	0.333 (-0.045)
Anomalous slope	-	1.127

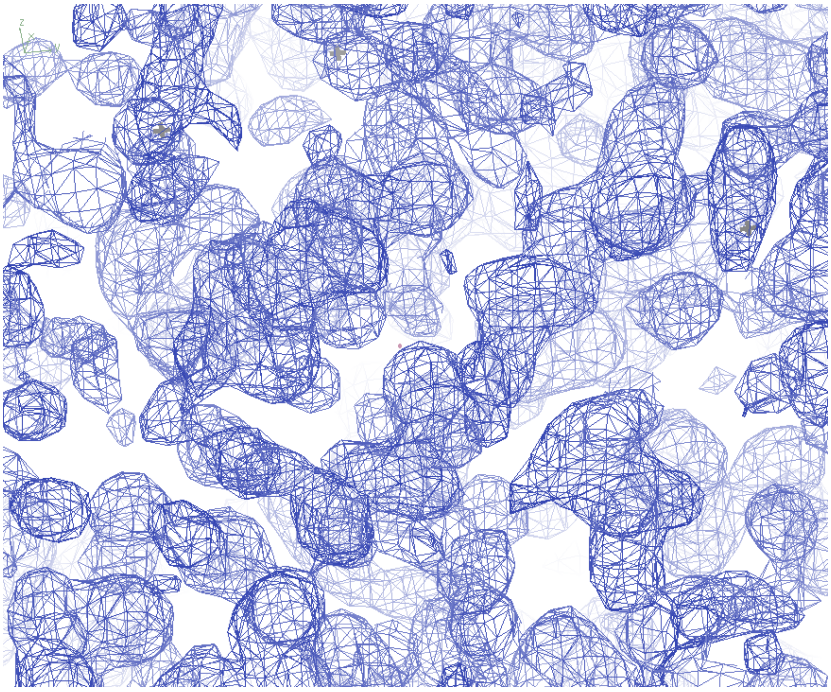
Table 6.5. Data collection processing statistics for $TssA_{CTD}$(303-373). xia2 3da data processing statistics for $TssA_{CTD}$(303-373) native collected data (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Sauter et al., 2004; Winter, 2010; Zhang et al., 2006), and xia2 3d data processing statistics for selenomethionine peak SAD collected data (Collaborative Computational Project, Number 4, 1994; Evans, 2006; Kabsch, 2010; Winter, 2010). The high-resolution shell information is shown in brackets.

6.5.3 Experimental phasing of $TssA_{CTD}$(303-373)

The Matthews coefficient calculation carried out for the peak data could not confidently identify the number of monomers in the asymmetric unit and percentage of solvent. The structure was solved using SAD experimental phasing, using the *SHELX* program (Sheldrick, 2008) with the *HKL2MAP* interface (Pape and Shneider, 2004). The peak dataset was analysed using *SHELXC* (Sheldrick, 2008) for the presence of anomalous signal. *SHELXD* (Sheldrick, 2008) was run to identify the location of the selenium atom sites, at 4.5 Å, looking for 16 selenium sites. 15 selenium sites were identified (Figure 6.16). *SHELXE* (Sheldrick, 2008) was run in the original and inverted hand with 50% solvent, searching for helices with five cycles of autobuilding. The heavy atom sites and both maps output by *SHELXE* (Sheldrick, 2008) were viewed in *Coot* (Emsley et al., 2010), with the inverted map chosen based on a higher estimated mean FOM (0.602) and pseudo-free CC value (65.59%) than the original hand, combined with analysis of the map.

6.5.4 Structure building and refinement of $TssA_{CTD}$(303-373)

The *SHELXE* (Sheldrick, 2008) inverted hand map and selenium sites were analysed in *Coot* (Emsley et al., 2010) (Figure 6.16). A number of helices were modelled into the electron density map, and then refined to the peak data using Refmac5 (Murshudov et al., 1997; Murshudov et al., 2011). Buccaneer (Cowtan, 2008; Winn et al., 2011) was then run and a model produced. The arrangement of the helices in the model was compared to the two- and three- helical structures solved previously. The three-helix structure could be modelled into the density, with slight remodelling of the third helix and the addition of a fourth helix. This four-helix model was repeated to fill the asymmetric unit, which contains eight molecules. The eight molecules are related by NCS. The phase and build program in Phenix (Echols et al., 2012) was used with the peak selenomethionine data, $TssA_{CTD}$(303-373) sequence, current $TssA_{CTD}$(303-373) model (four-helix model) and *SHELX* .phs map, with



Iodine site	Occupancy	Iodine site	Occupancy
1	1.0000	9	0.3643
2	0.8376	10	0.2946
3	0.8151	11	0.2671
4	0.7335	12	0.2666
5	0.7120	13	0.2148
6	0.6594	14	0.2127
7	0.6347	15	0.2102
8	0.5172		

Figure 6.16. Map produced by *SHELXE* for selenomethionine peak SAD experimental phasing of $TssA_{CTD}$(303-373). Electron density map is from the inverted hand map file output from *SHELXE* (Sheldrick, 2008). Helical shaped electron density can be seen. The selenium sites output from *SHELXD* (Sheldrick, 2008) are shown as grey pointer atoms. Map set to 1.6671e/A^3 = 1.00 rmsd. Image generated in *Coot* (Emsley et al., 2010).

eight NCS copies specified. The model produced was then used as the basis for building in the $TssA_{CTD}$(303-373) model, carried out in *Coot* (Emsley et al., 2010) and superpose (Krissinel and Hendrick, 2004; Winn et al., 2011). The model was initially refined to the selenomethionine peak data. Due to the presence of eight monomers of $TssA_{CTD}$ related by NCS in the asymmetric unit, further density modification, using averaging, could be carried out to improve the electron density map. Density modification using averaging was carried using the programs Sfall (Winn et al., 2011), SigmaA (Winn et al., 2011), NCSmask (Winn et al., 2011), and DM (Collaborative Computational Project, Number 4, 1994; Cowtan, 1994).

The model was refined against the native data to 3.08 Å using Refmac5 (Murshudov et al., 1997; Murshudov et al., 2011) and built in *Coot* (Emsley et al., 2010). The final R_{factor} is 0.1983 and R_{free} is 0.2424. The final model contains eight chains, with residues 299-302 to 367-369. The validation statistics are shown in Table 6.6.

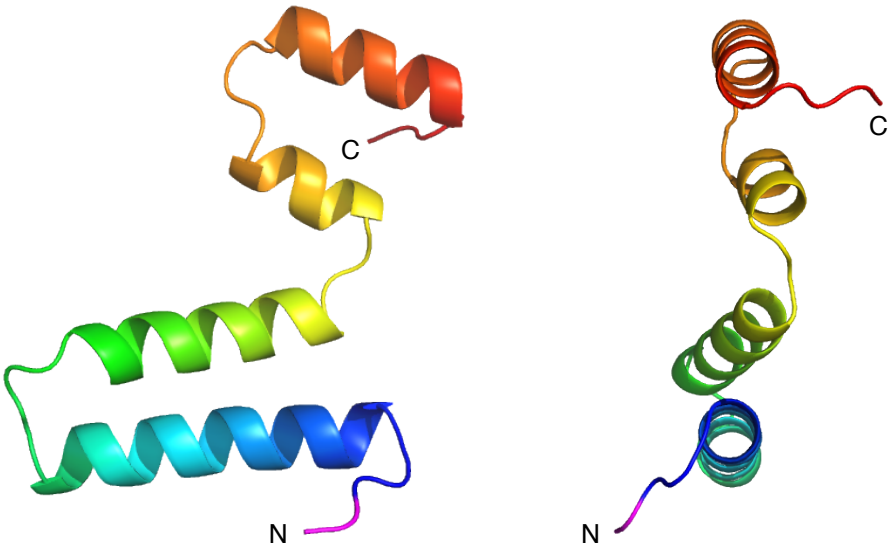
6.5.5 Structure analysis of $TssA_{CTD}$(303-373)

The monomer of $TssA_{CTD}$(303-373) is formed of four α-helices, the numbering of which will continue on from the $TssA_{NTD}$(1-255) structure, with helices α1-α11. The four helices form two pairs, α12-α13 and α14-α15 (Figure 6.17A). The asymmetric unit contains eight monomers of $TssA_{CTD}$(303-373). The last four residues of the N-terminal purification MBP fusion tag linker are residues 299-302. Six of the eight chains have H301 modelled as a full residue or with a truncated sidechain. One chain contains the residues 299-302 of the MBP fusion tag linker forming the ordered N-terminal end of helix α12, whilst in the other seven chains residues 301/302 form a loop. All eight chains contain residues M302-V367, with R368 and P369 present in seven of the eight chains (Figure 6.17B). The last four residues of the construct are unmodelled in all chains.

Refinement	**$TssA_{CTD}$(303-373)**
R_{factor}	0.1983
R_{free}	0.2424
No. of non-H atoms	
Protein	4401
Ligands/Metal ions	0
Water	0
Protein residues	548
RMSD (bonds) (Å)	0.114
RMSD (angles) (°)	1.6210
Ramachandran favoured (%)	96.05
Ramachandran outliers (%)	0.00
Favoured rotamers (%)	94.07
Poor rotamers (%)	0.66
MolProbity score	0.89 (100^{th} percentile, N=2115, 3.08Å±0.25Å)
Average B factors ($Å^2$)	
Main chain ($Å^2$)	67
Side chains ($Å^2$)	73

Table 6.6. Refinement and validation statistics for $TssA_{CTD}$(303-373). Statistics generated in Refmac5 (Murshudov et al., 2011; Murshudov et al., 1997), MolProbity (Chen et al., 2010) and Baverage (Winn et al., 2011).

A.



B.

Chains A, F, G
HMIAGIQNRAQAVDQLRAVARYFRQTEPHSPVAYLADKAAEWADMPLHKWLESVVK
DDGSLSHIRELLGVRP~~DEQS~~

Chain D, E
HMIAGIQNRAQAVDQLRAVARYFRQTEPHSPVAYLADKAAEWADMPLHKWLESVVK
DDGSLSHIRELLGVR~~PDEQS~~

Chains B, C
MIAGIQNRAQAVDQLRAVARYFRQTEPHSPVAYLADKAAEWADMPLHKWLESVVKD
DGSLSHIRELLGVRP~~DEQS~~

Chain H
ISHMIAGIQNRAQAVDQLRAVARYFRQTEPHSPVAYLADKAAEWADMPLHKWLESV
VKDDGSLSHIRELLGV~~RPDEQS~~

Figure 6.17. $TssA_{CTD}$(303-373) monomer structure
A. The structure of the $TssA_{CTD}$(303-373) chain A monomer shown as cartoon, coloured by chainbows. The remainder of the N-terminal MBP-fusion tag linker is highlighted in magenta. The N- (blue) and C-termini (red) are highlighted. Images generated in PyMOL version 1.1r1. **B.** The $TssA_{CTD}$(303-373) sequence in each monomer. The eight monomers are split into four groups, depending on the amount of N-terminal MBP-fusion tag linker seen and which C-terminal residues are visible in the structure. The linker is highlighted in magenta and the unmodelled residues are crossed through.

The asymmetric unit contains eight monomers, which results in a crystal lattice with a high solvent content of ~73%. The crystal lattice contains large solvent channels. The $TssA_{CTD}$(303-373) lattice is formed of interlocking layers of rings. Each ring interacts with four other rings in the lattice via four points around the initial ring (Figure 6.18).

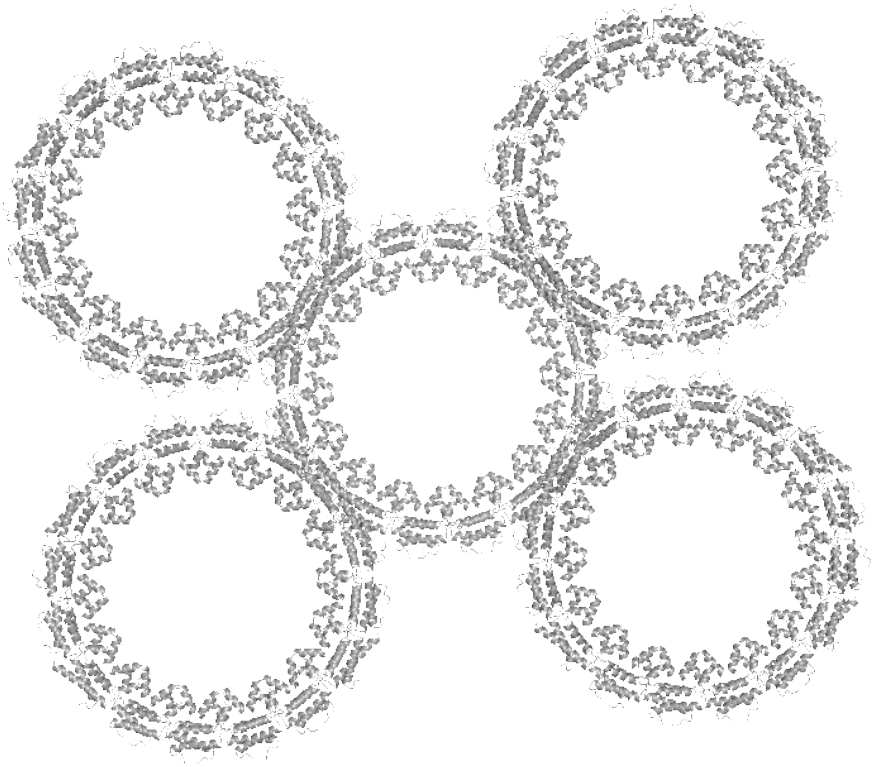
The $TssA_{CTD}$(303-373) monomers assemble into a ring of 32 subunits, forming two layers of 16 subunits, related by 2-fold symmetry. The arrangement of the $TssA_{CTD}$(303-373) monomers in the ring adopt pseudo-D16-fold symmetry. There is a 1-3 Å deviation between the monomers in the ring, leading to the pseudo symmetry (Figure 6.19). The pseudo symmetry may be a result of the flexibility of the ring within the crystal coupled with the very limited crystal contacts.

6.5.6 Interactions within the $TssA_{CTD}$ ring

Each monomer is involved in a number of interactions to enable the formation of the overall ring structure. Two monomers run antiparallel to each other, related by 2-fold symmetry, and interact to form a dimer with the majority of interfacing residues from helices α12-α13 forming hydrophobic interactions (Figure 6.20). The residues involved in hydrogen bonds and/or salt bridges, depending on the monomers involved in the dimer, are Q304, R306, Q312, Y319 and E324, identified in PDBePISA (Krissinel and Hendrick, 2007). A salt bridge is formed between E324 and R306 at both ends of the dimer interface, with each amino acid contributed from a separate monomer involved in the dimer (Figure 6.20). R306 is conserved among TssA proteins in the TssA-1 family.

The interface between dimers related by the pseudo-16-fold symmetry axis leaves helices α14-α15 unpaired. The two unpaired α14-α15 helices at each end of the dimer then form interactions with pseudo-16-fold related dimers on either side. The interaction is mainly hydrophobic with key residues involved in hydrogen bonds and/or salt bridges - V351,

A.



B.

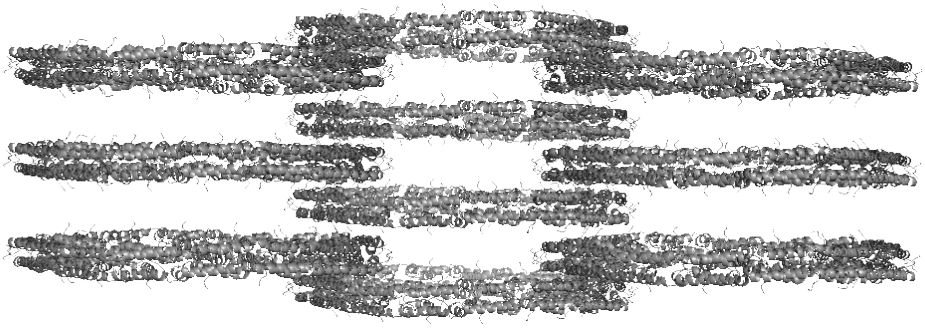


Figure 6.18. $TssA_{CTD}$(303-373) crystal lattice

The crystal lattice of $TssA_{CTD}$(303-373) solved structure. Large solvent channels can be seen running through the crystal. Contact between the rings forming the lattice takes place at four points around the ring. **A.** View parallel to the pseudo-16-fold symmetry axis. **B.** View perpendicular to the pseudo-16-fold symmetry axis. Images generated in PyMOL version 1.1r1.

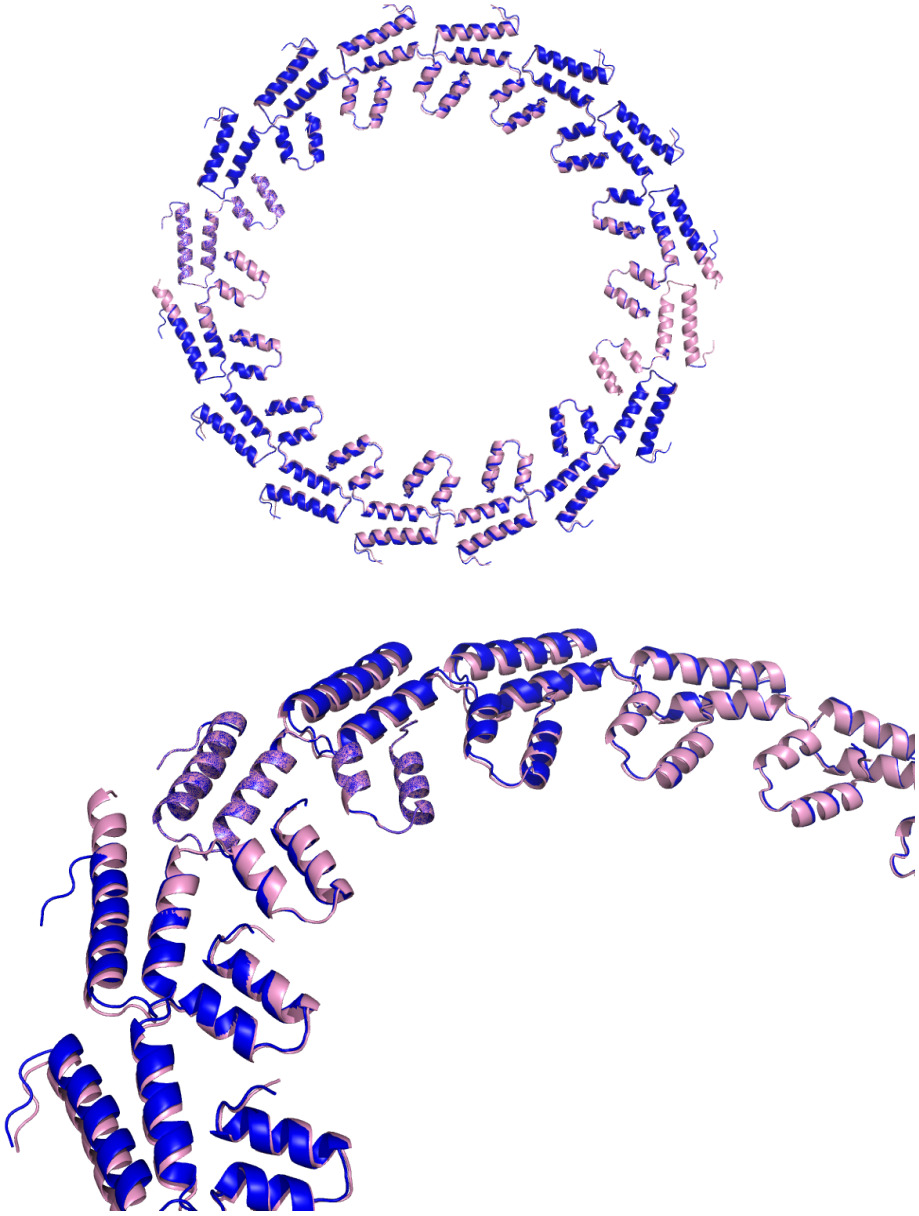


Figure 6.19. $TssA_{CTD}$(303-373) pseudo-D16-fold symmetry
Two copies of one layer of the $TssA_{CTD}$(303-373) ring are shown as cartoon. One monomer of $TssA_{CTD}$(303-373) was taken and rotated by 22.5° to create a $TssA_{CTD}$(303-373) ring with 16-fold symmetry (blue). The $TssA_{CTD}$(303-373) solved structure ring, has pseudo-D16-fold symmetry (pink). The two rings were then overlapped to show the difference between the ring generated by 16-fold rotational symmetry and the $TssA_{CTD}$(303-373) structure. Images generated in PyMOL version 1.1r1.

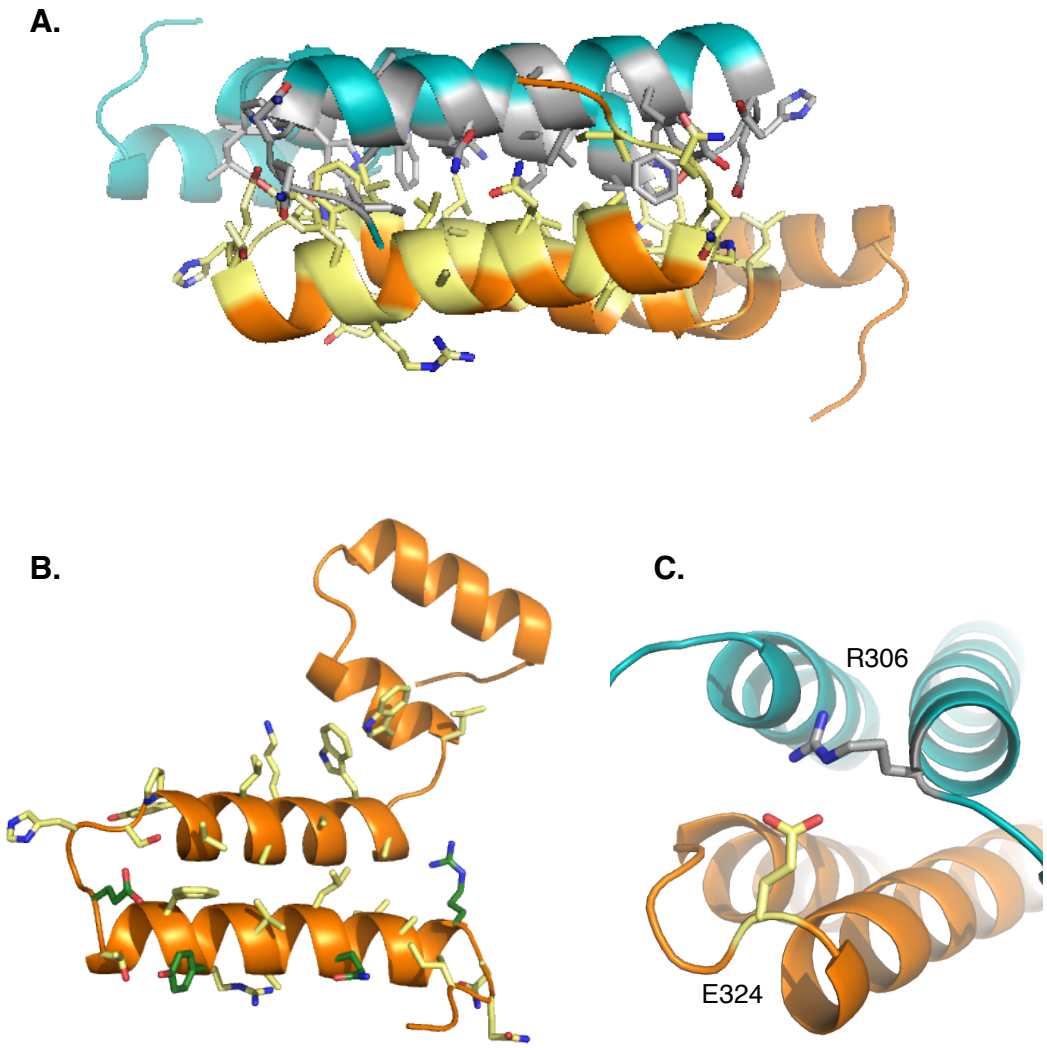


Figure 6.20. Interactions present between helices α12-α13 involved in dimer formation. A. Two subunits of $TssA_{CTD}$(303-373), orange and teal, forming a dimer are shown as cartoons. Interfacing residues are shown as sticks and are coloured silver in the teal monomer and yellow in the orange monomer. Residues are coloured by element. **B.** The interfacing residue sidechains of the orange monomer, shown as cartoon, are shown as yellow and green sticks. The green residues are involved in hydrogen bonds and/or salt bridges. Residues are coloured by element. **C.** A salt bridge formed between residues E324 of one monomer (orange) and R306 of another monomer (teal), and vice versa at the other end of the dimer interface. Interfacing residues identified by PDBePISA (Krissinel and Hendrick, 2007) and images generated in PyMOL version 1.1r1.

K353, D354, S357 and H360, identified by PDBePISA (Krissinel and Hendrick, 2007). This interaction completes the 16-fold ring formation through the interlocking of the dimers via the α14-α15 helices (Figure 6.21). As expected the dimer interface shows very little significant positive or negative surface charge (Figure 6.22).

The dimer interface, between pairs of helices α12-α13, has been seen in both previously solved structures of $TssA_{CTD}$ - the two helix and three helix structures (Figure 6.23). The absence of the full set of the second pairs of helices in both structures may have contributed to the absence of ring formation.

Further interactions, discussed below, were identified but were not deemed highly significant for assembly, by PDBePISA (Krissinel and Hendrick, 2007). The assembled ring highlights another two interactions between the dimers related by pseudo-16-fold symmetry. There are interactions between the α12-α13 loop of one dimer and α12-α13 loop of a pseudo-16-fold related dimer (Figure 6.24A). Another interaction is present between the α12-α13 loop end and α14 residues of one monomer and the open end of α12 and α13, along with residues from α14 and α15, of another monomer related by pseudo-16-fold symmetry in the same ring layer (Figure 6.24B). There is also an interaction between α15 of one monomer of dimer n-1, and α15 of one monomer in the other layer of the ring of dimer n+1 (Figure 6.24C).

6.5.7 Difficulties in crystallisation of $TssA_{CTD}$(303-373)

The initial $TssA_{CTD}$(303-373) structure, phased using selenomethionine-incorporated $TssA_{CTD}$(303-373) protein, underwent degradation and only the first two helices were present in the structure. Crystallisation of the same construct produced crystals containing almost full-length protein and thus the assembled ring, in both native and selenomethionine conditions. No clipping or degradation was seen in the $TssA_{CTD}$(303-373)

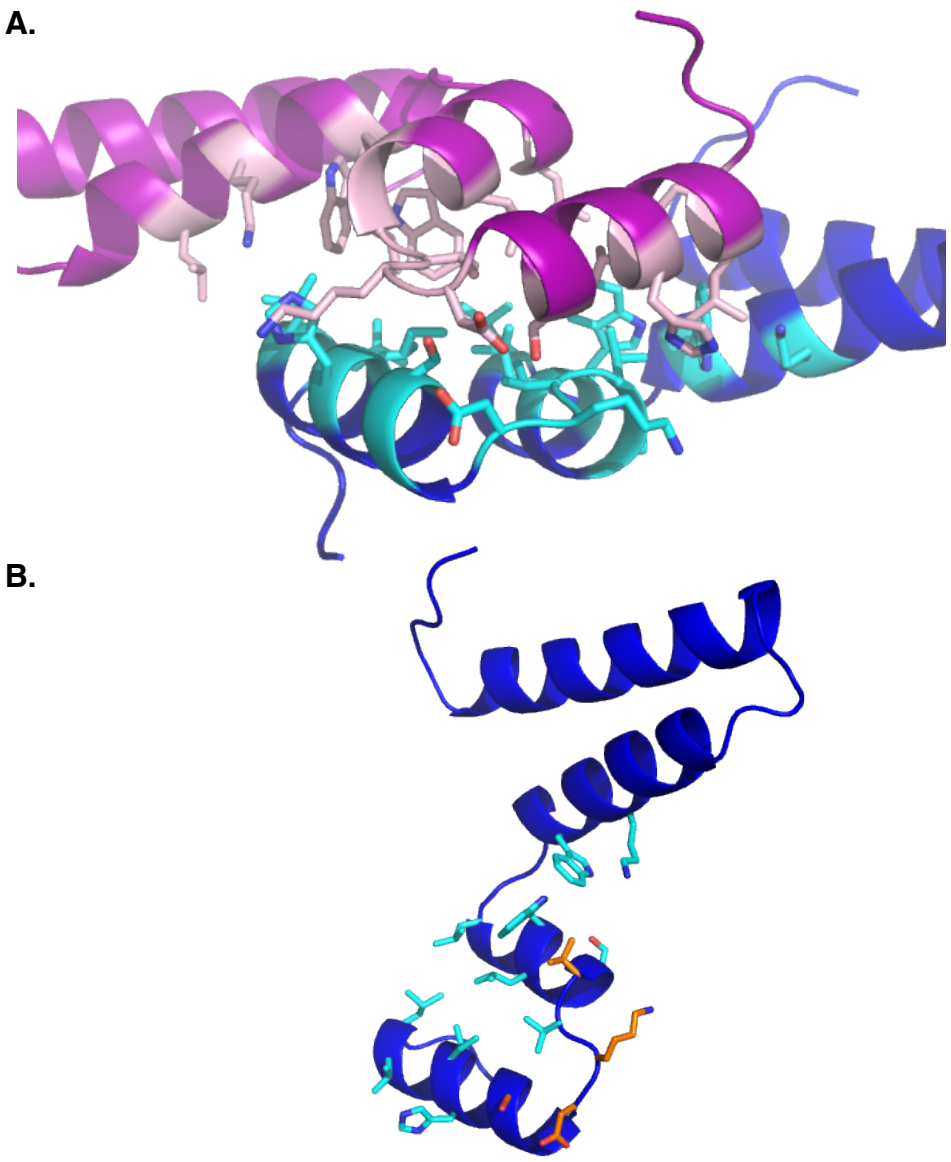


Figure 6.21. Interactions present between helices α14-α15 involved in ring formation. A. Two subunits of $TssA_{CTD}$(303-373), purple and dark blue, forming a dimer-dimer interaction are shown as cartoons. Interfacing residues are shown as sticks and are coloured pink in the purple monomer and cyan in the dark blue monomer. Residues are coloured by element. **B.** The interfacing residues sidechains of the dark blue monomer, shown as cartoon, are shown as cyan and orange sticks. The orange residues are involved in hydrogen bonds and/or salt bridges. Residues are coloured by element. Interfacing residues identified by PDBePISA (Krissinel and Henrick, 2007) and images generated in PyMOL version 1.1r1.

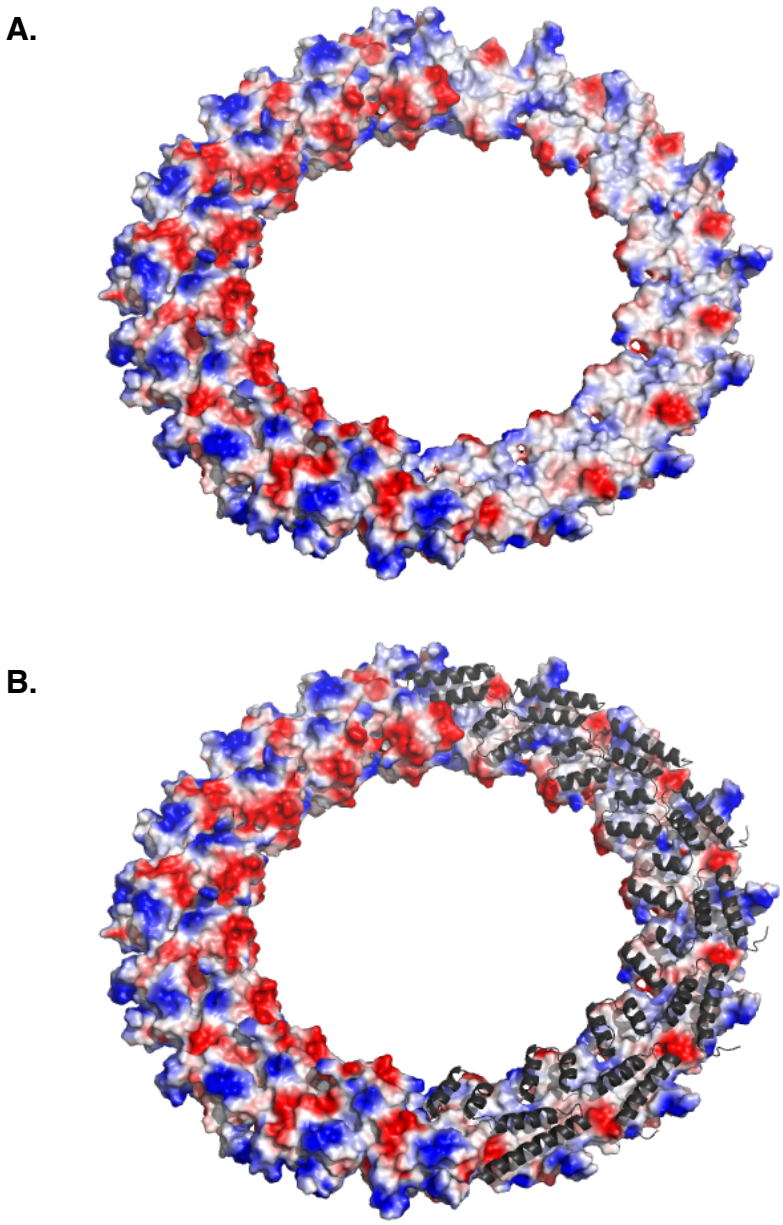


Figure 6.22. Surface charge of $TssA_{CTD}$(303-373) ring.
Electrostatic surface of $TssA_{CTD}$(303-373) ring was generated in PyMOL version 1.1r1. The $TssA_{CTD}$(303-373) ring is generated from symmetry of the asymmetric unit which consists of eight $TssA_{CTD}$(303-373) monomers. **A**. The electrostatic surface of the $TssA_{CTD}$(303-373) ring with one asymmetric unit of eight monomers removed. The charge on the internal interfacing surface is less charged than the exterior of the ring. **B.** The electrostatic surface of the $TssA_{CTD}$(303-373) ring with one asymmetric unit of eight monomers, shown as cartoon, included. The ring is viewed at an angle away from the plane parallel to the page. Images generated in PyMOL version 1.1r1.

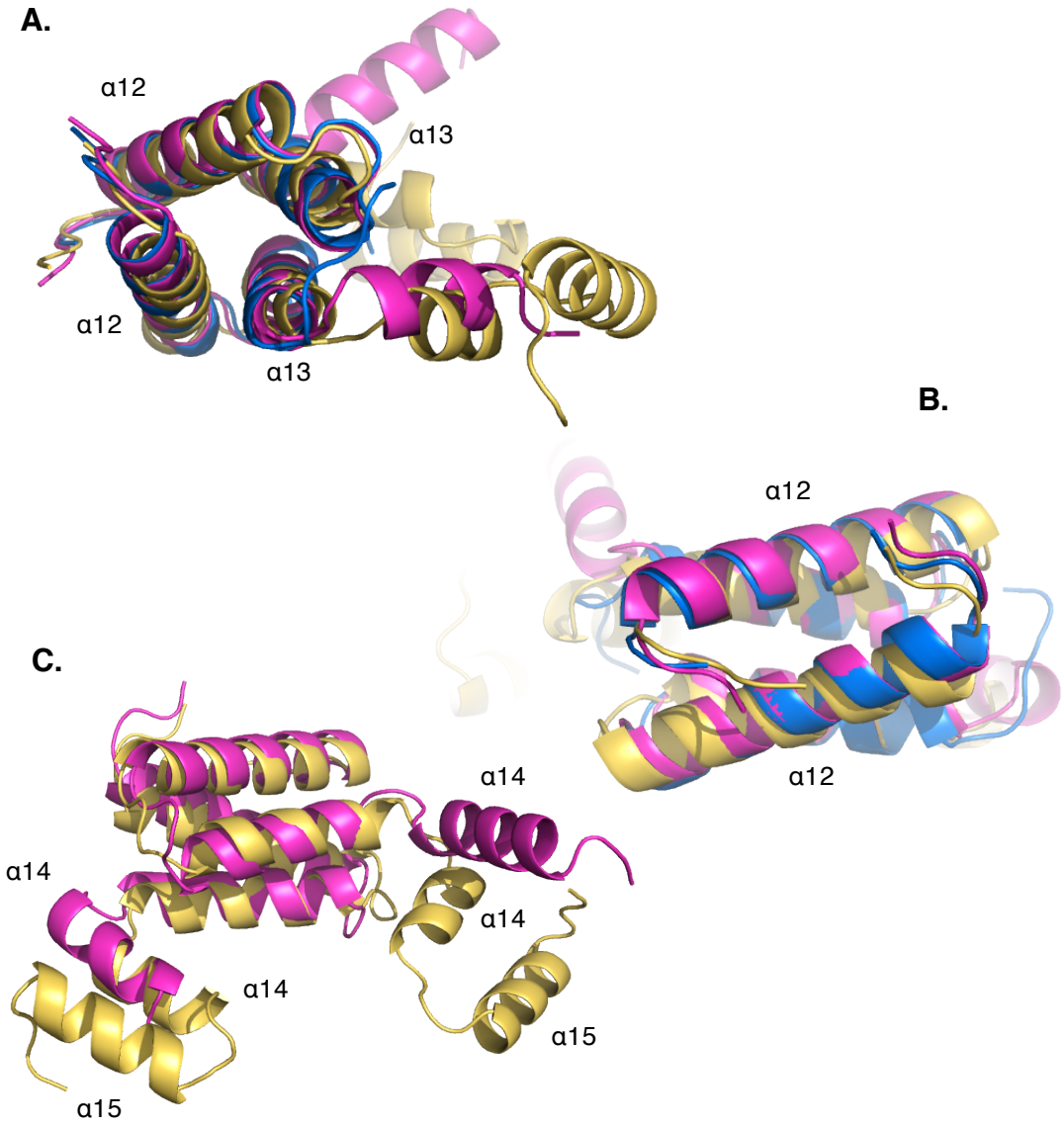


Figure 6.23. $TssA_{CTD}$(303-373) ring dimer structure compared to $TssA_{CTD}$(303-373)-two-helices and $TssA_{CTD}$(303-358)-three-helices structures. $TssA_{CTD}$(303-373)-two-helices – blue, $TssA_{CTD}$(303-358)-three-helices – magenta and $TssA_{CTD}$(303-373) dimer – yellow. **A.** and **B.** Overlay of all three $TssA_{CTD}$ structures shows a pair of α12-α13 helices in the same orientation. Helices α12-α13 interacts with a 2-fold related pair to form a dimer. **C.** Helix α14 of one of the $TssA_{CTD}$(303-358) monomers (pink) has a longer helical region and extends out from the core α12-α13 dimer structure at a different angle than α14 of $TssA_{CTD}$(303-373) ring dimer (yellow). Alignments generated in *Coot* – SSM (Emsley et al., 2010; Krissinel and Henrick, 2004) and images generated in PyMOL version 1.1r1.

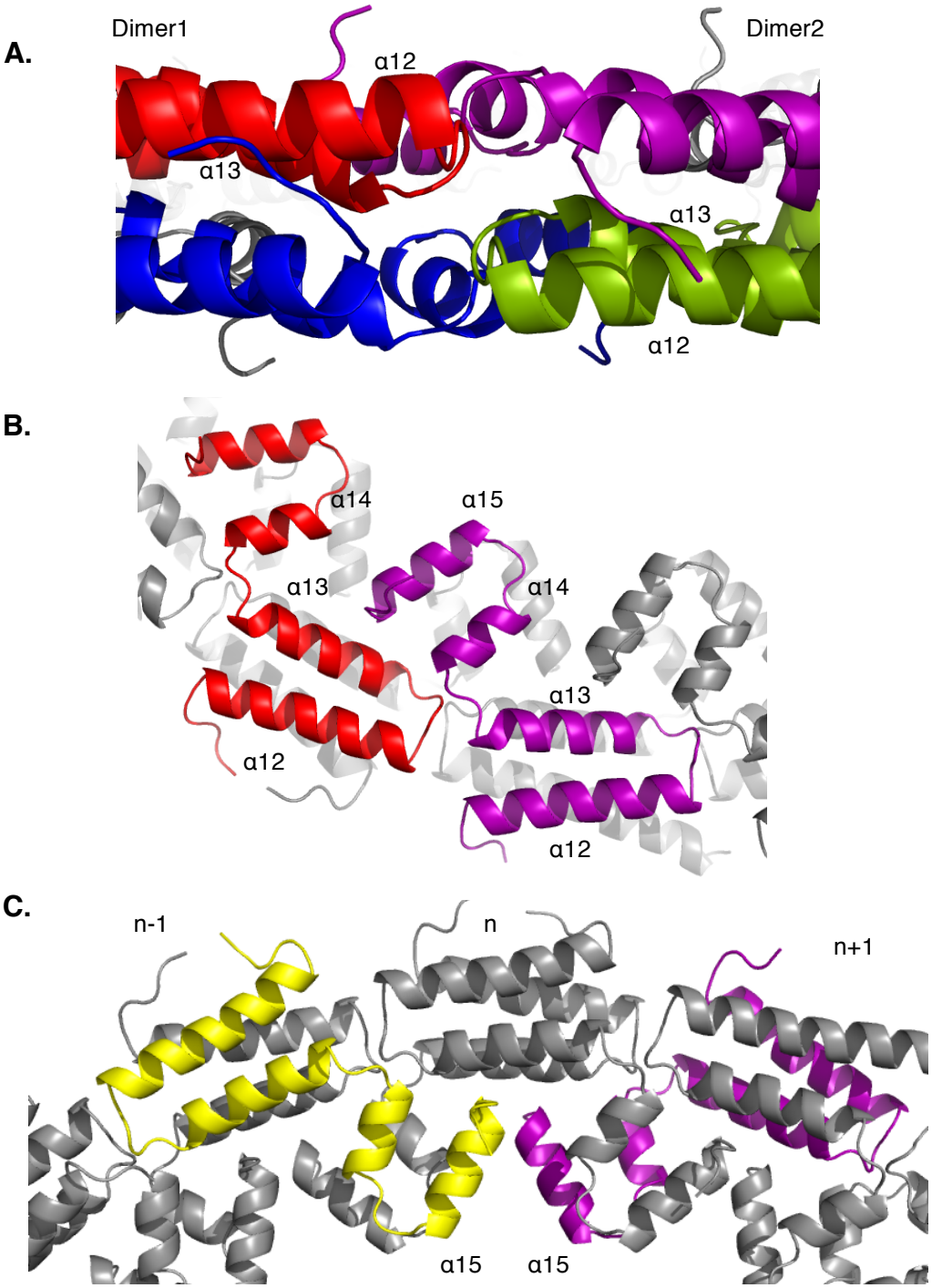


Figure 6.24. Additional $TssA_{CTD}$(303-373) ring interactions
A. Interactions present between the α12-α13 loop of one dimer and the α12-α13 loop of a pseudo-16-fold related dimer. **B.** Interactions present between neighbouring pseudo-16-fold related monomers in one ring layer. **C.** Interactions present between α15 of one monomer of dimer n-1, and α15 of one monomer in the other layer of the ring of dimer n+1. Images generated in PyMOL version 1.1r1.

protein during purification (Dr Svetlana Sedelnikova, personal communication), whose crystals contained only two helices. The pH of the crystallisation solution for the full-length protein (assembled ring) was pH 6.0-8.0 whereas the pH of the crystallisation solution for the truncated protein was pH 4.0. Data collections carried out on other $TssA_{CTD}$(303-373) protein crystals produced a variety of cell dimensions which were compared to the pH of the corresponding crystallisation solution (Table 6.4). The size of the cell may be dependent on the pH of the crystallisation solution, assuming that larger cells contain a fully assembled $TssA_{CTD}$(303-373) ring. The difference in pH may have caused the disassembly or prevented the formation of the ring. The lack of ring assembly may have allowed regions of the protein, which would be inaccessible when in the ring formation, to become accessible to proteases, leading to degradation of the protein.

6.5.8 Formation of the $TssA_{CTD}$(303-373) ring

The formation of the $TssA_{CTD}$(303-373) ring is hypothesised to start with α12-α13 dimer formation. Two monomers of $TssA_{CTD}$(303-373) form a dimer through the α12-α13 interface (Figure 6.25A). Multiple dimers then interact via the α14-α15 helical interaction (Figure 6.25B) to form an interlocking of the dimers around a pseudo-16-fold symmetry axis and ultimate ring formation (Figure 6.25C). The diameter of the full ring is ~185-190 Å, whilst the inner lumen diameter is ~110 Å (Figure 6.26). In the ring, α12 and α13 form the outer section of the ring, whilst α14 and α15 are on the inside of the ring. Therefore the N-terminus of the $TssA_{CTD}$(303-373), where the linker to the $TssA_{NTD}$ would attach, is located on the outside of the ring. Therefore the $TssA_{NTD}$ is most likely present on the periphery of the ring.

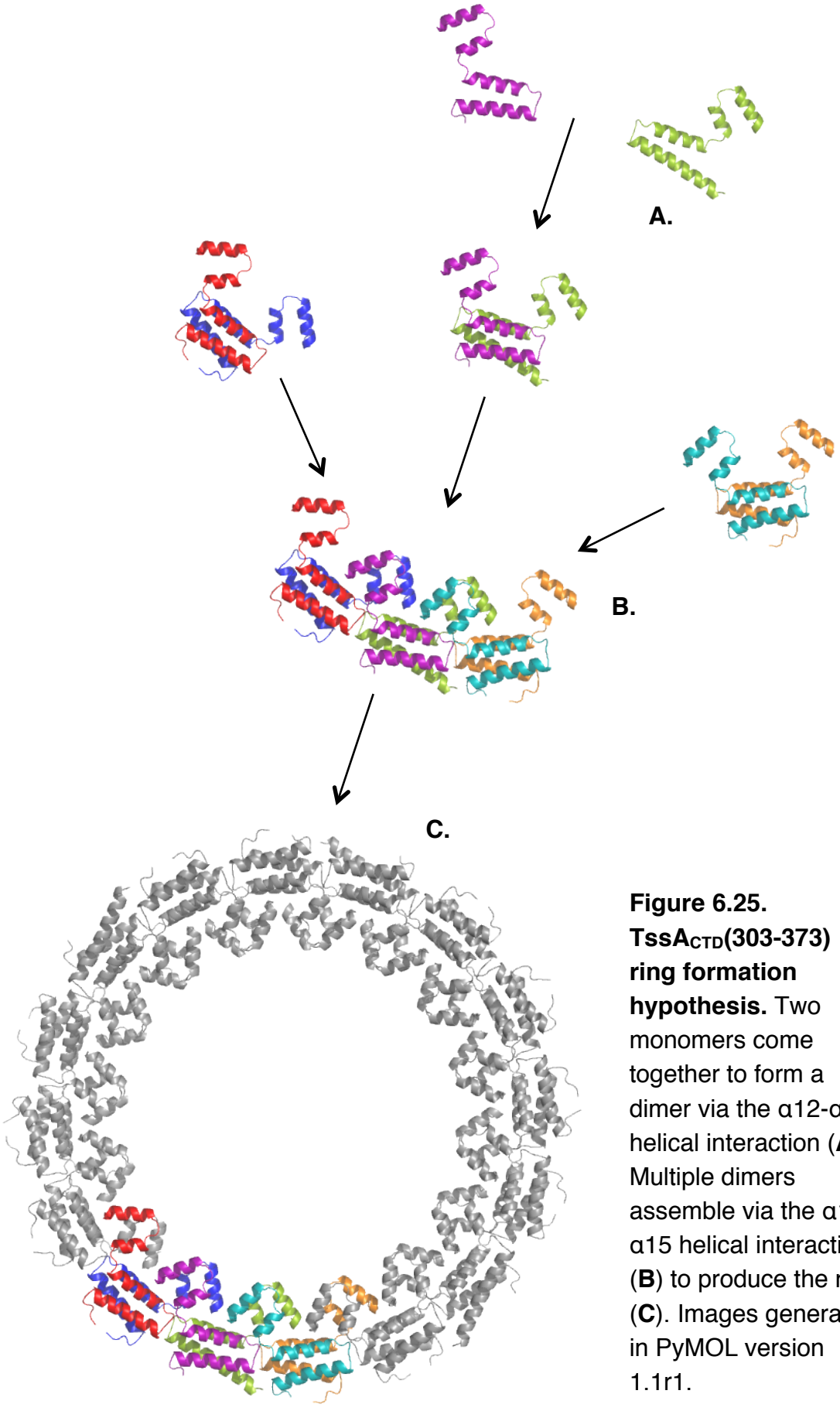


Figure 6.25. $TssA_{CTD}$(303-373) ring formation hypothesis. Two monomers come together to form a dimer via the α12-α13 helical interaction (**A**). Multiple dimers assemble via the α14-α15 helical interaction (**B**) to produce the ring (**C**). Images generated in PyMOL version 1.1r1.

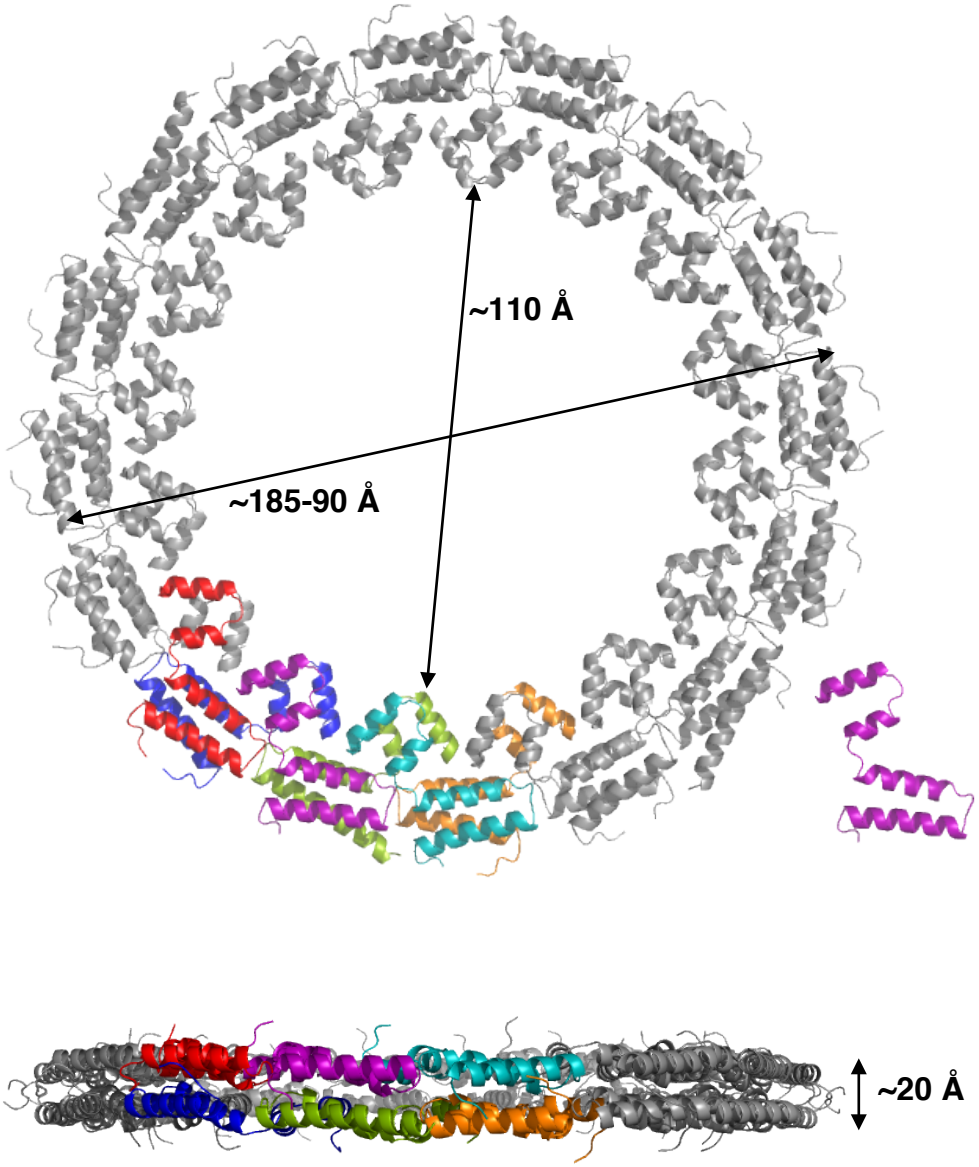


Figure 6.26. $TssA_{CTD}$(303-373) ring dimensions
The ring structure has approximate dimensions of 185-190 Å diameter of the full ring and 110 Å diameter of the central hole, highlighted in the structure (view parallel to the pseudo-16-fold axis). The height of the ring is ~20 Å, highlighted next to the ring, view perpendicular to the pseudo-16-fold axis. A monomer of $TssA_{CTD}$(303-373) (purple) is shown next to the ring. Images and measurements generated in PyMOL version 1.1r1.

6.5.9 Comparison to other $TssA_{CTD}$ structures

*B. cenocep*acia TssA and *P. aeruginosa* TssA belong to the same TssA group, TssA-1, whereas *E. coli* TssA is in a separate group, TssA-2 (Dr. Mark Thomas, personal communication). *E. coli* TssA C-terminus is formed from two stacked hexameric rings creating a dodecamer (Zoued et al., 2016) (Figure 6.27). The structure of *P. aeruginosa* TssA has not yet been determined however it is predicted that the CTD of the phage baseplate protein gp6 is a homolog to the CTD of *P. aeruginosa* TssA (Planamente et al., 2016). *B. cenocepacia* $TssA_{CTD}$ has a significantly different quaternary structure to *E. coli* TssA C-terminus and has no significant structural similarity to the C-terminus of gp6.

All three *B. cenocepacia* $TssA_{CTD}$, *E. coli* TssA C-terminus and *P. aeruginosa* TssA structures form a ring (Planamente et al., 2016; Zoued et al., 2016). The inner diameter of the ring for *B. cenocepacia* $TssA_{CTD}$ and *P. aeruginosa* TssA is similar at ~100 Å (Planamente et al., 2016). However, the *P. aeruginosa* TssA outer diameter is considerably larger at ~260 Å compared to ~185-190 Å (Planamente et al., 2016). The dimensions of *E. coli* TssA C-terminal ring are considerably different with an inner and outer diameter of ~25-48 Å and ~100 Å, respectively (Zoued et al., 2016). The *E. coli* TssA C-terminal lumen has been predicted to expand to accommodate the TssD ring of the tail tube during T6SS assembly (Zoued et al., 2016).

B. cenocepacia $TssA_{CTD}$ ring has 32 monomers with pseudo-D16-fold symmetry, whereas *E. coli* TssA C-terminal ring is a dodecamer with 6-fold symmetry and *P. aeruginosa* is predicted to contain a dodecamer, or double-dodecamer, with the symmetry of the inner section of the visualised TssA ring unknown (Planamente et al., 2016; Zoued et al., 2016). In all three cases the TssA CTD acts as the oligomerisation domain (Planamente et al., 2016; Zoued et al., 2016).

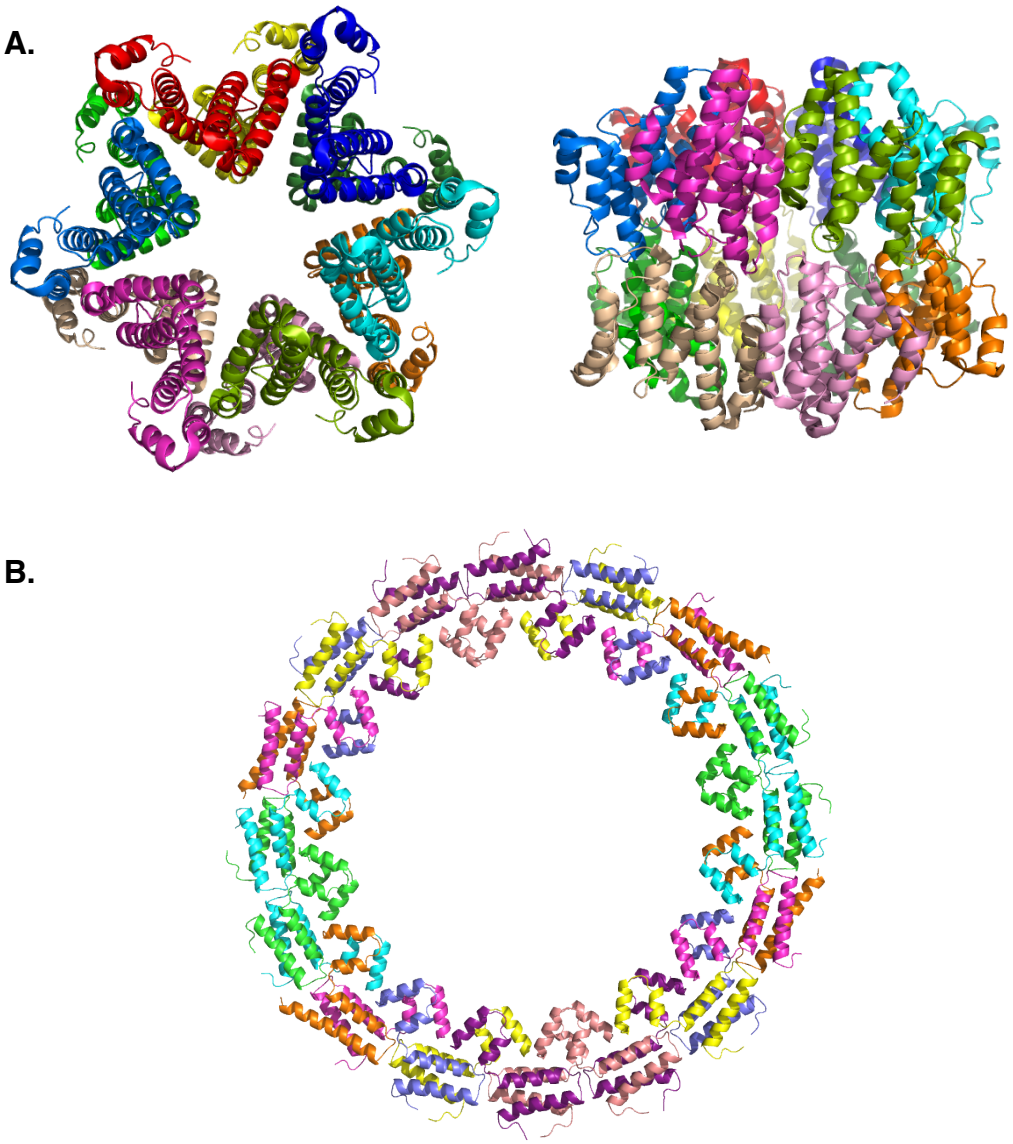


Figure 6.27. Structures of the TssA C-terminal domain from *E. coli* and *B. cenocepacia*. A. *E. coli* C-terminus (PDB:4YO5) (Zoued et al., 2016), viewed parallel to the star shaped lumen and perpendicular to the lumen. **B.** *B. cenocepacia* TssA CTD, viewed parallel to the pseudo-16-fold axis. Images generated in PyMOL version 1.1r1.

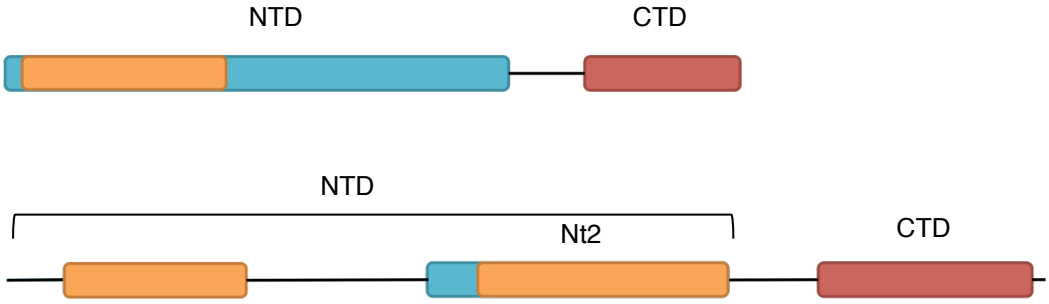
6.6 Conclusions and further work

The structure of *B. cenocepacia* $TssA_{NTD}$(1-255), is the first for the TssA-1 family. The sequence similarity between TssA-1 and TssA-2 (residues 8-133 in TssA-1, *B. cenocepacia* and residues 32-141 in TssA-2, *E. coli*) suggests this region is structurally identical, however the structure of this region in TssA-2 is not known. The structure of *B. cenocepacia* $TssA_{NTD}$(1-255) helices α2-α5 was found to be similar to helices α3-α6 in *E. coli* TssA Nt2 (TssA-2). These observations suggest that TssA-2 NTD is formed from two tandem repeat regions with similar structure, but different sequences (Figure 6.28). Structures of full-length TssA-2 N-terminal regions would allow for further investigation into this hypothesis.

The T6SS components exhibit a mismatch in symmetry between the structures of the different proteins. The membrane complex from *E. coli* contains 5-fold symmetry, the tail tube has 6-fold symmetry and tail sheath has rotated layers of hexametric rings (Durand et al., 2015; Mougous et al., 2006; Kube et al., 2014). TssA from *E. coli* has been shown to have 6-fold symmetry (Zoued et al., 2016), whereas the structure of *B. cenocepacia* TssA has pseudo-D16 symmetry, pseudo symmetry as seen in the crystal lattice. How the mismatch in symmetry is accommodated or utilised within the T6SS has yet to be determined.

The inner diameter of the $TssA_{CTD}$(303-373) ring is ~110 Å and is thus large enough to fit the TssD tail tube inside, outer diameter of ~85 Å (Osipiuk et al., 2011) (Figure 6.29), but not the TssB/TssC tail sheath, whose outer diameter is ~260 Å and inner diameter is ~100 Å (PDB code:3J9G, Kudryashev et al., 2015). The *P. aeruginosa* TssA ring lumen is also predicted to be large enough to fit the TssD ring inside (Planamente et al., 2016). The inside of the *B. cenocepacia* TssA ring shows repeating stripes of negative surface charge evenly distributed around the ring. This suggests that these regions may act as a site of protein:protein interaction, with another protein component of the T6SS

***B. cenocepacia* TssA**



***E. coli* TssA**

Figure 6.28. The predicted structural conservation between TssA family members. A schematic representation of *B. cenocepacia* TssA and *E. coli* TssA, with the protein domains shown as turquoise and red blocks and the interdomain regions shown as lines, except in *E. coli* TssA NTD, where only the region between the NTD and CTD is interdomain linker. The protein regions shown to be similar in sequence and/or structure are coloured orange, leading to the suggestion that all orange regions have a similar structure. *E. coli* TssA structures from Zoued et al., 2016.

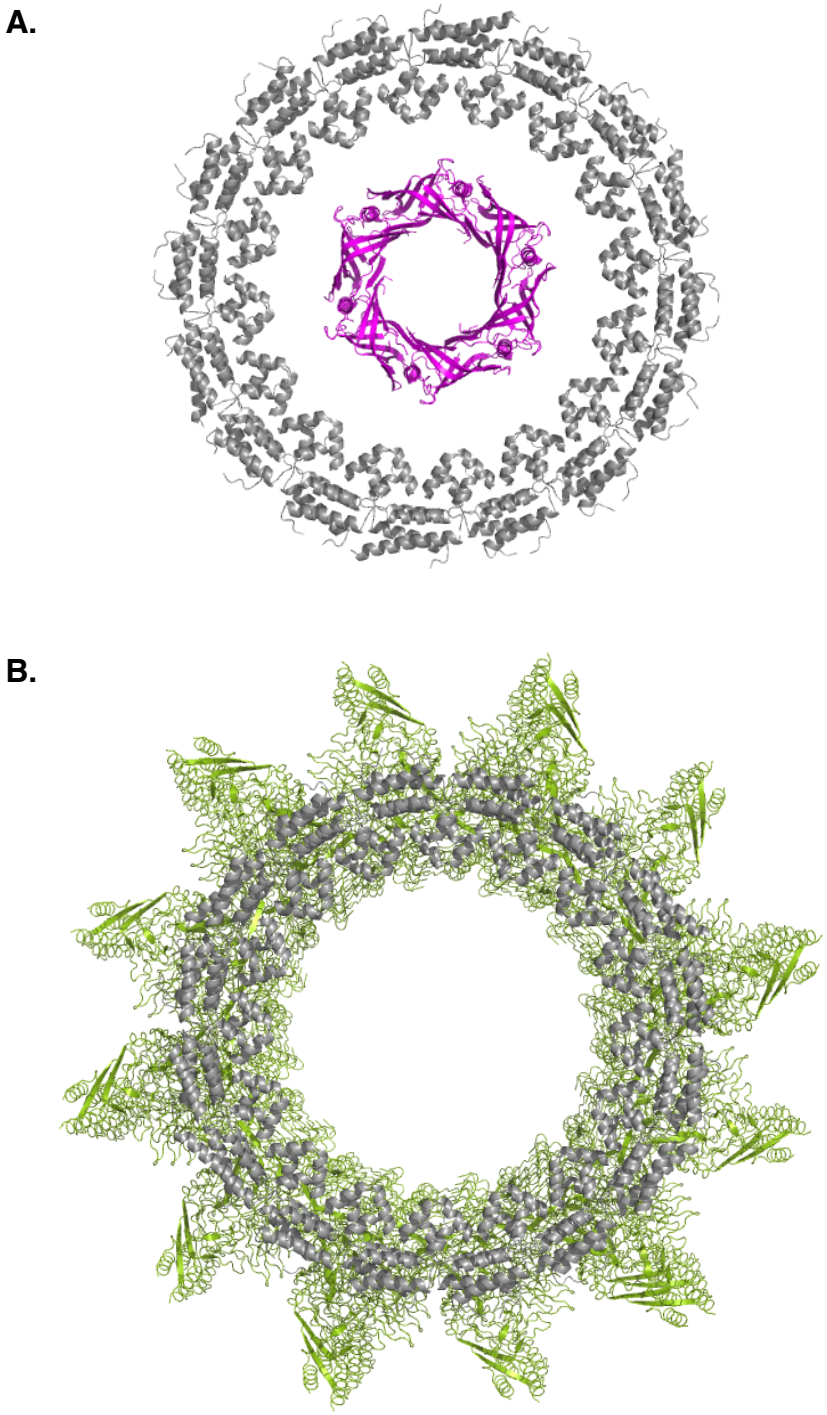


Figure 6.29. $TssA_{CTD}$(303-373) ring structure modelled with TssD and TssB/TssC. **A.** $TssA_{CTD}$(303-373) shown in grey, with Hcp3, in pink, from *P. aeruginosa* (PDB:3HE1) (Osipiuk et al., 2011) equivalent to TssD. Hcp3 fits inside the $TssA_{CTD}$(303-373) ring. **B.** $TssA_{CTD}$(303-373) shown in grey, with VipA/VipB sheath from *V. cholerae* (PDB:3J9G) (Kudryashev et al., 2015), equivalent to TssB/TssC. $TssA_{CTD}$(303-373) fits onto the VipA/VipB sheath. Images generated in PyMOL version 1.1r1.

inserting into the ring (Figure 6.30). The outer diameter of the $TssA_{CTD}$(303-373) ring is ~185-190 Å which means it could fit onto the end of TssB/TssC tail sheath, (Figure 6.29) this was also described by Planamente et al., 2016 for TssA from *P. aeruginosa*. The outside of the ring has repeating sections of positive charge, which again may be involved in protein-protein interactions (Figure 6.30). *E. coli* TssA has been predicted to exist in an open and closed state, with the open state lumen large enough to accommodate the tail tube (Zoued et al., 2016). However the *B. cenocepacia* $TssA_{CTD}$(303-373) ring structure has a lumen large enough to accommodate the tail tube, suggesting it exists only in one state or the structure has been trapped in the open conformation of the ring, and it too can exist in a closed state, which has yet to be visualised. *B. cenocepacia* TssA and *E. coli* TssA have different symmetry within the CTD ring, however the NTD have a similar architecture.

These observations suggest that *B. cenocepacia* TssA could sit at one end of the tail sheath, and allow the tail tube to pass through it (Figure 6.30). This is similar to the function predicted for the open form of *E. coli* TssA and also *P. aeruginosa* TssA (Planamente et al., 2016; Zoued et al., 2016), however the role of *B. cenocepacia* TssA in relation to the baseplate and other T6SS proteins remains to be defined. Further structures are required to determine how TssA interacts with other T6SS components and what its function within the T6SS might be.

The elucidation of the structure of the first domain of the TssA-2 NTD would indicate whether this section is conserved in sequence and structure. Structural analysis of the TssA in complex with other T6SS core components might give some indication as to its function within the system, specifically with other putative baseplate proteins TssE, TssF, TssG and TssK.

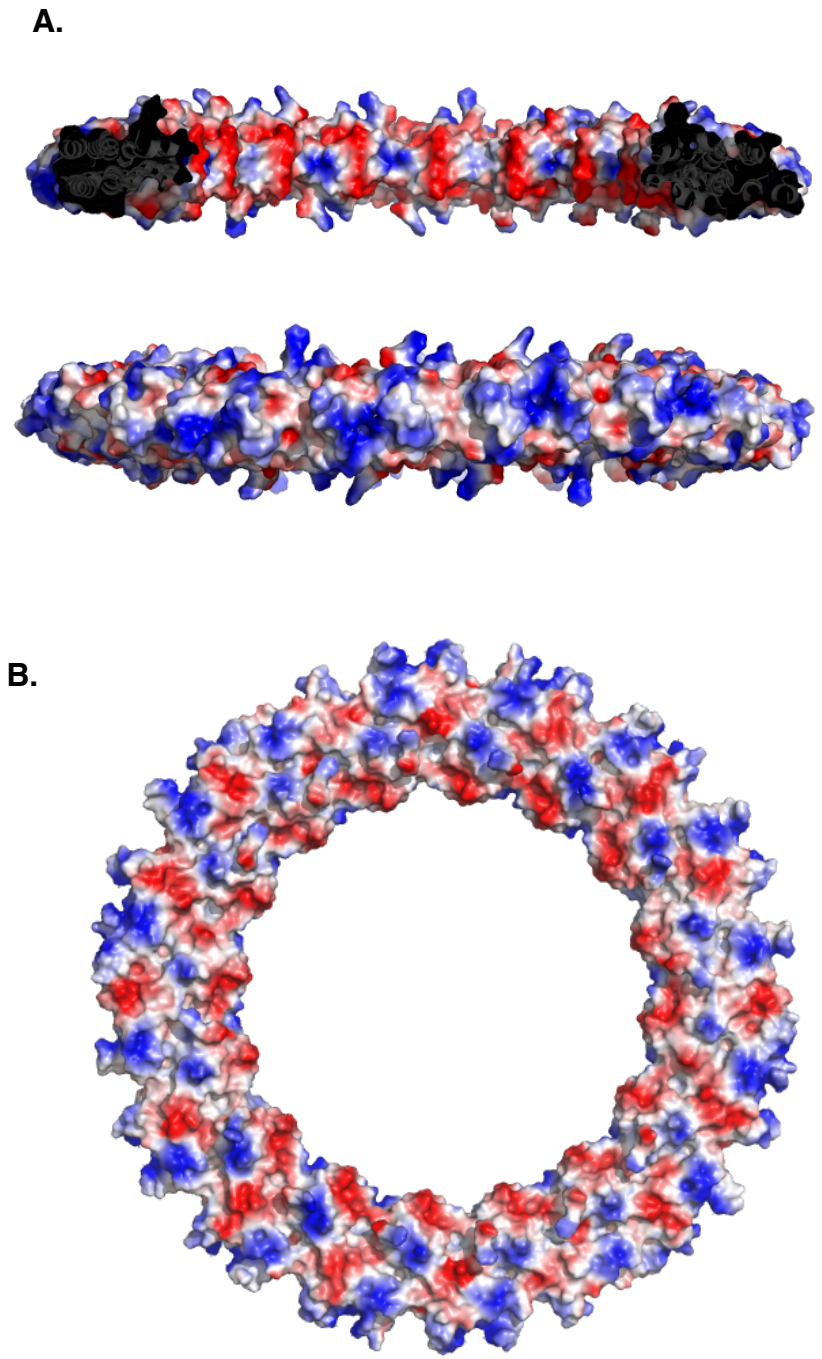


Figure 6.30. Surface charge of $TssA_{CTD}$(303-373) ring.
Electrostatic surface of $TssA_{CTD}$(303-373) ring was generated in PyMOL version 1.1r1. **A**. The ring is slabbed to view the internal surface of the ring. Stripes of negative charge (red) are seen on the internal face of the ring, with positive charge (blue) stripes on the external surface of the ring. View is perpendicular to the pseudo-16-fold symmetry axis. **B.** Surface charge of the ring, view is parallel to the pseudo-16-fold symmetry axis. Images generated in PyMOL version 1.1r1.

Bibliography

Abbink, F.C., Orendi, J.M., and de Beaufort, A.J. (2001). Mother-to-child transmission of *Burkholderia pseudomallei*. New England Journal of Medicine *344*, 1171-1172.

Aksyuk, A.A., Leiman, P.G., Shneider, M.M., Mesyanzhinov, V.V., and Rossmann, M.G. (2009). The Structure of Gene Product 6 of Bacteriophage T4, the Hinge-Pin of the Baseplate. Structure *17*, 800-808.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). BASIC LOCAL ALIGNMENT SEARCH TOOL. Journal of Molecular Biology *215*, 403-410.

Angus, A.A., Agapakis, C.M., Fong, S., Yerrapragada, S., Estrada-de los Santos, P., Yang, P., Song, N., Kano, S., Caballero-Mellado, J., de Faria, S.M.*, et al.* (2014). Plant-Associated Symbiotic *Burkholderia* Species Lack Hallmark Strategies Required in Mammalian Pathogenesis. PLOS ONE *9*, e83779.

Aschtgen, M.S., Bernard, C.S., De Bentzmann, S., Lloubès, R., and Cascales, E. (2008). SciN Is an Outer Membrane Lipoprotein Required for Type VI Secretion in Enteroaggregative *Escherichia coli*. Journal of Bacteriology *190*, 7523-7531.

Attree, O., and Attree, I. (2001). A second type III secretion system in *Burkholderia pseudomallei*: who is the real culprit? Microbiology-Sgm *147*, 3197-3199.

Aubert, D.F., Flannagan, R.S., and Valvano, M.A. (2008). A novel sensor kinase-response regulator hybrid controls biofilm formation and type VI secretion system activity in *Burkholderia cenocepacia*. Infection and Immunity *76*, 1979-1991.

Bagos, P.G., Nikolaou, E.P., Liakopoulos, T.D., and Tsirigos, K.D. (2010). Combined prediction of Tat and Sec signal peptides with hidden Markov models. Bioinformatics *26*, 2811-2817.

Baker, A., Tahani, D., Gardiner, C., Bristow, K.L., Greenhill, A.R., and Warner, J. (2011). Groundwater Seeps Facilitate Exposure to *Burkholderia pseudomallei*. Applied and Environmental Microbiology *77*, 7243-7246.

Barman, P., Kaur, R., and Kumar, K. (2013). Clinically lesser known entity in India: A Report of two cases of Melioidosis. Indian journal of critical care medicine : peer-reviewed, official publication of Indian Society of Critical Care Medicine *17*, 46-48.

Baskaran, N., Kandpal, R.P., Bhargava, A.K., Glynn, M.W., Bale, A. and Weissman, S.M. (1996). Uniform amplification of a mixture of deoxyribonucleic acids with varying GC content. Genome methods *6*, 633-638.

Basler, M. (2015). Type VI secretion system: secretion by a contractile nanomachine. Philosophical Transactions of the Royal Society B *370*, 20150021.

Basler, M., Pilhofer, M., Henderson, G.P., Jensen, G.J., and Mekalanos, J.J. (2012). Type VI secretion requires a dynamic contractile phage tail-like structure. Nature *483*, 182-186.

Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G.W. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallographica Section D-Biological Crystallography *67*, 271-281.

Baugh, L., Gallagher, L.A., Patrapuvich, R., Clifton, M.C., Gardberg, A.S., Edwards, T.E., Armour, B., Begley, D.W., Dieterich, S.H., Dranow, D.M., *et al.* (2013). Combining Functional and Structural Genomics to Sample the Essential *Burkholderia* Structome. Plos One *8*, e53851.

Benjdia, A., Subramanian, S., Leprince, J., Vaudry, H., Johnson, M.K., and Berteau, O. (2008). Anaerobic sulfatase-maturating enzymes, first dual substrate radical S-adenosylmethionine enzymes. Journal of Biological Chemistry *283*, 17815-17826.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Research *28*, 235-242.

Berman, H., Henrick, K., and Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. Nature Structural Biology *10*, 980-980.

Bertani, G. (2004). Lysogeny at Mid-Twentieth Century: P1, P2, and Other Experimental Systems. Journal of Bacteriology *186*, 595-600

Bönemann, G., Pietrosiuk, A., Diemand, A., Zentgraf, H., and Mogk, A. (2009). Remodelling of VipA/VipB tubules by ClpV-mediated threading is crucial for type VI protein secretion. EMBO Journal *28*, 315-325.

Bornberg-Bauer, E., Rivals, E., and Vingron, M. (1998). Computational approaches to identify leucine zippers. Nucleic Acids Research *26*, 2740-2746.

Boyer, F., Fichant, G., Berthod, J., Vandenbrouck, Y., and Attree, I. (2009). Dissecting the bacterial type VI secretion system by a genome wide *in silico* analysis: what can be learned from available microbial genomic resources? BMC Genomics *10*, 104.

Bradford, M.M. (1976). Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. Analytical Biochemistry *72*, 248-254.

Brent, A.J., Matthews, P.C., Dance, D.A., Pitt, T.L., and Handy, R. (2007). Misdiagnosing melioidosis. Emerging Infectious Diseases *13*, 349-351.

Brett, P.J., Deshazer, D., and Woods, D.E. (1997). Characterization of *Burkholderia pseudomallei* and *Burkholderia pseudomallei*-like strains. Epidemiology and Infection *118*, 137-148.

Brett, P.J., DeShazer, D., and Woods, D.E. (1998). *Burkholderia thailandensis* sp. nov., a *Burkholderia pseudomallei*-like species. International Journal of Systematic Bacteriology *48*, 317-320.

Brunet, Y.R., Espinosa, L., Harchouni, S., Mignot, T., and Cascales, E. (2013). Imaging Type VI Secretion-Mediated Bacterial Killing. Cell Reports *3*, 36-41.

Brunet, Y.R., Hénin, J., Celia, H., and Cascales, E. (2014). Type VI secretion and bacteriophage tail tubes share a common assembly pathway. EMBO Reports *15*, 315-321.

Brunet, Y.R., Zoued, A., Boyer, F., Douzi, B., and Cascales, E. (2015). The Type VI Secretion TssEFGK-VgrG Phage-Like Baseplate Is Recruited to the TssJLM Membrane Complex via Multiple Contacts and Serves As Assembly Platform for Tail Tube/Sheath Polymerization. PLOS Genetics *11*, e1005545.

Buchan, D.W.A., Minneci, F., Nugent, T.C.O., Bryson, K., and Jones, D.T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Research *41*, W349-W357.

Buetow, L., Flatau, G., Chiu, K., Boquet, P., and Ghosh, P. (2001). Structure of the Rho-activating domain of *Escherichia coli* cytotoxic necrotizing factor 1. Nature Structural Biology *8*, 584-588.

Burtnick, M.N., Brett, P.J., Harding, S.V., Ngugi, S.A., Ribot, W.J., Chantratita, N., Scorpio, A., Milne, T.S., Dean, R.E., Fritz, D.L.*, et al.* (2011). The Cluster 1 Type VI Secretion System Is a Major Virulence Determinant in *Burkholderia pseudomallei*. Infection and Immunity *79*, 1512-1525.

Caballero-Mellado, J., Martínez-Aguilar, L., Paredes-Valdez, G., and Estrada-de los Santos, P. (2004). *Burkholderia unamae* sp nov., an N_2-fixing rhizospheric and endophytic species. International Journal of Systematic and Evolutionary Microbiology *54*, 1165-1172.

Carr, S.A., Hemling, M.E., Bean, M.F., and Roberts, G.D. (1991). Integration of mass-spectrometry in analytical biotechnology. Analytical Chemistry *63*, 2802-2824.

Casey, W.T., Spink, N., Cia, F., Collins, C., Romano, M., Berisio, R., Bancroft, G.J., and McClean, S. (2016). Identification of an OmpW homologue in *Burkholderia pseudomallei*, a protective vaccine antigen against melioidosis. Vaccine *34*, 2616-2621.

Centers for Disease Control (CDC), Department of Health and Human Services (HHS). (2012). Possession, use, and transfer of select agents and toxins; biennial review. Final rule. Federal Register. *77*, 61083-115.

Champion, O.L., Gourlay, L.J., Scott, A.E., Lassaux, P., Conejero, L., Perletti, L., Hemsley, C., Prior, J., Bancroft, G., Bolognesi, M.*, et al.* (2016). Immunisation with proteins expressed during chronic murine melioidosis provides enhanced protection against disease. Vaccine *34*, 1665-1671.

Chan, H.P., and Yip, H.S. (2015). Mediastinal Lymphadenopathy: Melioidosis Mimicking Tuberculosis. Tropical medicine and health *43*, 93-94.

Chantratita, N., Wuthiekanun, V., Boonbumrung, K., Tiyawisutsri, R., Vesaratchavest, M., Limmathurotsakul, D., Chierakul, W., Wongratanacheewin, S., Pukritiyakamee, S., White, N.J.*, et al.* (2007). Biological relevance of colony morphology and phenotypic switching by *Burkholderia pseudomallei*. Journal of Bacteriology *189*, 807-817.

Chantratita, N., Tandhavanant, S., Wikraiphat, C., Trunck, L.A., Rholl, D.A., Thanwisai, A., Saiprom, N., Limmathurotsakul, D., Korbsrisate, S., Day, N.P.J.*, et al.* (2012). Proteomic analysis of colony morphology variants of *Burkholderia pseudomallei* defines a role for the arginine deiminase system in bacterial survival. Journal of Proteomics *75*, 1031-1042.

Chaowagul, W., Chierakul, W., Simpson, A.J., Short, J.M., Stepniewska, K., Maharjan, B., Rajchanuvong, A., Busarawong, D., Limmathurotsakul, D., Cheng, A.C.*, et al.* (2005). Open-label randomized trial of oral trimethoprim-sulfamethoxazole, doxycycline, and chloramphenicol compared with trimethoprim-sulfamethoxazole and doxycycline for maintenance therapy of melioidosis. Antimicrobial Agents and Chemotherapy *49*, 4020-4025.

Chen, Y.S., Chen, S.C., Kao, C.M., and Chen, Y.L. (2003). Effects of soil pH, temperature and water content on the growth of *Burkholderia pseudomallei*. Folia Microbiologica *48*, 253-256.

Chen, V.B., Arendall, W.B., III, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallographica Section D-Biological Crystallography *66*, 12-21.

Chen, Y., Wong, J., Sun, G.W., Liu, Y., Tan, G.-Y.G., and Gan, Y.-H. (2011). Regulation of Type VI Secretion System during *Burkholderia pseudomallei* Infection. Infection and Immunity *79*, 3064-3073.

Cheng, A.C., and Currie, B.J. (2005). Melioidosis: Epidemiology, pathophysiology, and management. Clinical Microbiology Reviews *18*, 383-416.

Cheng, A.C., O'Brien, M., Freeman, K., Lum, G., and Currie, B.J. (2006). Indirect hemagglutination assay in patients with melioidosis in Northern

Australia. American Journal of Tropical Medicine and Hygiene *74*, 330-334.

Cheng, A.C., Limmathurotsakul, D., Chierakul, W., Getchalarat, N., Wuthiekanun, V., Stephens, D.P., Day, N.P.J., White, N.J., Chaowagul, W., Currie, B.J.*, et al.* (2007). Randomized controlled trial of granulocyte colony-stimulating factor for the treatment of severe sepsis due to melioidosis in Thailand. Clinical Infectious Diseases *45*, 308-314.

Cheng, A.C., Chierakul, W., Chaowagul, W., Chetchotisakd, P., Limmathurotsakul, D., Dance, D.A.B., Peacock, S.J., and Currie, B.J. (2008). Short report: Consensus guidelines for dosing of amoxicillin-clavulanate in melioidosis. American Journal of Tropical Medicine and Hygiene *78*, 208-209.

Chierakul, W., Winothai, W., Wattanawaitunechai, C., Wuthiekanun, V., Rugtaengan, T., Rattanalertnavee, J., Jitpratoom, P., Chaowagul, W., Singhasivanon, P., White, N.J.*, et al.* (2005). Melioidosis in 6 tsunami survivors in Southern Thailand. Clinical Infectious Diseases *41*, 982-990.

Chiswell, B., Lovelace, L.L., Brannen, C., Ortlund, E.A., Lebioda, L., and Sodetz, J.A. (2007). Structural features of the ligand binding site on human complement protein C8 gamma: A member of the lipocalin family. Biochimica Et Biophysica Acta-Proteins and Proteomics *1774*, 637-644.

Choh, L.C., Ong, G.H., Vellasamy, K.M., Kalaiselvam, K., Kang, W.T., Al-Maleki, A.R., Mariappan, V., and Vadivelu, J. (2013). *Burkholderia* vaccines: are we moving forward? Frontiers in Cellular and Infection Microbiology *3*, 5.

Chusri, S., Hortiwakul, T., Charoenmak, B., and Silpapojakul, K. (2012). Outcomes of Patients with Melioidosis Treated with Cotrimoxazole Alone for Eradication Therapy. American Journal of Tropical Medicine and Hygiene *87*, 927-932.

Clemens, D.L., Ge, P., Lee, B.Y., Horwitz, M.A., and Zhou, Z.H. (2015). Atomic Structure of T6SS Reveals Interlaced Array Essential to Function. Cell *160*, 940-951.

Collaborative Computational Project, Number 4 (1994). The CCP4 suite: programs for protein crystallography. Acta crystallographica Section D, Biological crystallography *50*, 760-763.

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

Cowtan, K. (1994). DM: an automated procedure for phase improvement by density modification. Joint CCP4 and ESF-EACBM newsletter on protein crystallography.

Cowtan, K. (2008). Fitting molecular fragments into electron density. Acta Crystallographica Section D-Biological Crystallography *64*, 83-89.

Cowtan, K. (2010). Recent developments in classical density modification. Acta Crystallographica Section D-Biological Crystallography *66*, 470-478.

Cruz-Migoni, A., Hautbergue, G.M., Artymiuk, P.J., Baker, P.J., Bokori-Brown, M., Chang, C.-T., Dickman, M.J., Essex-Lopresti, A., Harding, S.V., Mahadi, N.M.*, et al.* (2011). A *Burkholderia pseudomallei* Toxin Inhibits Helicase Activity of Translation Factor eIF4A. Science *334*, 821-824.

Currie, B.J., Fisher, D.A., Anstey, N.M., and Jacups, S.P. (2000a). Melioidosis: acute and chronic disease, relapse and re-activation. Transactions of the Royal Society of Tropical Medicine and Hygiene *94*, 301-304.

Currie, B.J., Fisher, D.A., Howard, D.M., Burrow, J.N.C., Lo, D., Selva-Nayagam, S., Anstey, N.M., Huffam, S.E., Snelling, P.L., Marks, P.J.*, et al.* (2000b). Endemic melioidosis in tropical northern Australia: A 10-year prospective study and review of the literature. Clinical Infectious Diseases *31*, 981-986.

Currie, B.J., Fisher, D.A., Howard, D.M., Burrow, J.N.C., Selvanayagam, S., Snelling, P.L., Anstey, N.M., and Mayo, M.J. (2000c). The epidemiology of melioidosis in Australia and Papua New Guinea. Acta Tropica *74*, 121-127.

Currie, B.J., and Jacups, S.P. (2003). Intensity of rainfall and severity of melioidosis, Australia. Emerging Infectious Diseases *9*, 1538-1542.

Currie, B.J., Jacups, S.P., Cheng, A.C., Fisher, D.A., Anstey, N.M., Huffam, S.E., and Krause, V.L. (2004). Melioidosis epidemiology and risk factors from a prospective whole-population study in northern Australia. Tropical Medicine & International Health *9*, 1167-1174.

Currie, B.J., Ward, L., and Cheng, A.C. (2010). The Epidemiology and Clinical Spectrum of Melioidosis: 540 Cases from the 20 Year Darwin Prospective Study. Plos Neglected Tropical Diseases *4*, e900.

Dance, D.A.B., Wuthiekanun, V., Chaowagul, W., and White, N.J. (1989). The antimicrobial susceptibility of *Pseudomonas pseudomallei* - Emergence of resistance *in vitro* and during treatment. Journal of Antimicrobial Chemotherapy *24*, 295-309.

Das, A.K., Cohen, P.T.W., and Barford, D. (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. Embo Journal *17*, 1192-1199.

Das, S., and Chaudhuri, K. (2003). Identification of a unique IAHP (IcmF Associated Homologous Proteins) cluster in *Vibrio cholerae* and other proteobacteria through In silico analysis. In Silico Biology *3*, 287-300.

Datta, S., Mori, Y., Takagi, K., Kawaguchi, K., Chen, Z.W., Okajima, T., Kuroda, S., Ikeda, T., Kano, K., Tanizawa, K.*, et al.* (2001). Structure of a quinohemoprotein amine dehydrogenase with an uncommon redox cofactor and highly unusual crosslinking. Proceedings of the National Academy of Sciences of the United States of America *98*, 14268-14273.

Day, M. (2012). Structural studies on protein targets from the pathogenic bacterium *Burkholderia pseudomallei*. (PhD thesis. University of Sheffield).

Delano, W.L. The PyMOL Molecular Graphics, Version 1.1r.1 Schrödinger LLC. DeLano Scientific San Carlos, CA, USA (2002).

Depoorter, E., Bull, M.J., Peeters, C., Coenye, T., Vandamme, P., and Mahenthiralingam, E. (2016). *Burkholderia*: an update on taxonomy and biotechnological potential as antibiotic producers. Applied Microbiology and Biotechnology *100*, 5215-5229.

Diepold, A. and Armitage, J.P. (2015). Type III secretion systems: the bacterial flagellum and the injectisome. Philosophical Transactions of the Royal Society B *370*, 20150020.

Doker, T.J., Quinn, C.L., Salehi, E.D., Sherwood, J.J., Benoit, T.J., Elrod, M.G., Gee, J.E., Shadomy, S.V., Bower, W.A., Hoffmaster, A.R.*, et al.* (2014). Case Report: Fatal *Burkholderia pseudomallei* Infection Initially Reported as a *Bacillus* Species, Ohio, 2013. American Journal of Tropical Medicine and Hygiene *91*, 743-746.

Durand, E., Cambillau, C., Cascales, E., and Journet, L. (2014). VgrG, Tae, TIe, and beyond: the versatile arsenal of Type VI secretion effectors. Trends in Microbiology *22*, 498-507.

Durand, E., Nguyen, V.S., Zoued, A., Logger, L., Péhau-Arnaudet, G., Aschtgen, M.S., Spinelli, S., Desmyter, A., Bardiaux, B., Dujeancourt, A.*, et al.* (2015). Biogenesis and structure of a type VI secretion membrane core complex. Nature *523*, 555-560.

Echols, N., Grosse-Kunstleve, R.W., Afonine, P.V., Bunkoczi, G., Chen, V.B., Headd, J.J., McCoy, A.J., Moriarty, N.W., Read, R.J., Richardson, D.C.*, et al.* (2012). Graphical tools for macromolecular crystallography in PHENIX. Journal of Applied Crystallography *45*, 581-586.

Egan, A.M., and Gordon, D.L. (1996). *Burkholderia pseudomallei* activates complement and is ingested but not killed by polymorphonuclear leukocytes. Infection and Immunity *64*, 4952-4959.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallographica Section D-Biological Crystallography *66*, 486-501.

English, G., Byron, O., Cianfanelli, F.R., Prescott, A.R., and Coulthurst, S.J. (2014). Biochemical analysis of TssK, a core component of the bacterial Type VI secretion system, reveals distinct oligomeric states of TssK and identifies a TssK-TssFG subcomplex. Biochemical Journal *461*, 291-304.

Estrada-de los Santos, P., Bustillos-Cristales, R., and Caballero-Mellado, J. (2001). *Burkholderia*, a genus rich in plant-associated nitrogen fixers with wide environmental and geographic distribution. Applied and Environmental Microbiology *67*, 2790-2798.

Estrada-de los Santos, P., Vinuesa, P., Martínez-Aguilar, L., Hirsch, A.M., and Caballero-Mellado, J. (2013). Phylogenetic Analysis of *Burkholderia* Species by Multilocus Sequence Analysis. Current Microbiology *67*, 51-60.

Evans, G., and Pettifer, R.F. (2001). CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. Journal of Applied Crystallography *34*, 82-86.

Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallographica Section D-Biological Crystallography *62*, 72-82.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A.*, et al.* (2016). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Research *44*, D279-D285.

Flatau, G., Lemichez, E., Gauthier, M., Chardin, P., Paris, S., Fiorentini, C., and Boquet, P. (1997). Toxin-induced activation of the G protein p21 Rho by deamidation of glutamine. Nature *387*, 729-733.

Flower, D.R. (1993a). Structural relationship of streptavidin to the calycin protein superfamily. Febs Letters *333*, 99-102.

Flower, D.R. (1996). The lipocalin protein family: Structure and function. Biochemical Journal *318*, 1-14.

Flower, D.R., North, A.C.T., and Attwood, T.K. (1991). Mouse oncogene protein 24p3 is a member of the lipocalin protein family. Biochemical and Biophysical Research Communications *180*, 69-74.

Flower, D.R., North, A.C.T., and Attwood, T.K. (1993b). Structure and sequence relationships in the lipocalins and related proteins. Protein Science *2*, 753-761.

Flower, D.R., Sansom, C.E., Beck, M.E., and Attwood, T.K. (1995). The first prokaryotic lipocalins. Trends in Biochemical Sciences *20*, 498-499.

French, C.T., Toesca, I.J., Wu, T.-H., Teslaa, T., Beaty, S.M., Wong, W., Liu, M., Schröder, I., Chiou, P.-Y., Teitell, M.A.*, et al.* (2011). Dissection of the *Burkholderia* intracellular life cycle using a photothermal nanoblade.

Proceedings of the National Academy of Sciences of the United States of America *108*, 12095-12100.

Galyov, E.E., Brett, P.J., and DeShazer, D. (2010). Molecular Insights into *Burkholderia pseudomallei* and *Burkholderia mallei* Pathogenesis. Annual Review of Microbiology, Vol 64, 2010 *64*, 495-517.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I, Appel, R.D. and Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Research *31*, 3784-3788.

Gasteiger, E., Hoogland, G., Gattiker, A., Duvand, S., Wilkins, M.R., Appel, R.D. and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook, John M. Walker, ed. (Totowa, NJ: Humana Press Inc.), pp. 571-607

Gattiker, A., Bienvenut, W.V., Bairoch, A. and Gasteiger, E. (2002). FindPept, a tool to identify unmatched masses in peptide mass fingerprinting protein identification. Proteomics *2*, 1435-1444.

Gauthier, Y.P., Hagen, R.M., Brochier, G.S., Neubauer, H., Splettstoesser, W.D., Finke, E.J., and Vidal, D.R. (2001). Study on the pathophysiology of experimental *Burkholderia pseudomallei* infection in mice. Fems Immunology and Medical Microbiology *30*, 53-63.

Geer, L.Y., Domrachev, M., Lipman, D.J., and Bryant, S.H. (2002). CDART: Protein homology by domain architecture. Genome Research *12*, 1619-1623.

Glass, M.B., Steigerwalt, A.G., Jordan, J.G., Wilkins, P.P., and Gee, J.E. (2006). *Burkholderia oklahomensis* sp nov., a *Burkholderia pseudomallei*-like species formerly known as the Oklahoma strain of *Pseudomonas pseudomallei*. International Journal of Systematic and Evolutionary Microbiology *56*, 2171-2176.

Godovac-Zimmermann, J. (1988). The structural motif of β-lactoglobulin and retinol-binding protein - a basic framework for binding and transport of small hydrophobic molecules?. Trends in Biochemical Sciences *13*, 64-66.

Goldman, P.J., Grove, T.L., Sites, L.A., McLaughlin, M.I., Booker, S.J., and Drennan, C.L. (2013). X-ray structure of an AdoMet radical activase

reveals an anaerobic solution for formylglycine posttranslational modification. Proceedings of the National Academy of Sciences of the United States of America *110*, 8519-8524.

Gorrec, F. (2009). The MORPHEUS protein crystallization screen. Journal of Applied Crystallography *42*, 1035-1042.

Green, R.N., and Tuffnell, P.G. (1968). Laboratory acquired melioidosis. American Journal of Medicine *44*, 599-605.

Guang-Han, O., Leang-Chung, C., Vellasamy, K.M., Mariappan, V., Li-Yen, C., and Vadivelu, J. (2016). Experimental Phage Therapy for *Burkholderia pseudomallei* Infection. Plos One *11*, e0158213.

Guentas, L., Gensous, S., Cavaloc, Y., Ducousso, M., Amir, H., de Ledenon, B.D.G., Moulin, L., and Jourand, P. (2016). *Burkholderia novacaledonica* sp nov and *B. ultramafica* sp nov isolated from roots of *Costularia* spp. pioneer plants of ultramafic soils in New Caledonia. Systematic and Applied Microbiology *39*, 151-159.

Gutierrez, M.G., Pfeffer, T.L., and Warawa, J.M. (2015). Type 3 Secretion System Cluster 3 Is a Critical Virulence Determinant for Lung-Specific Melioidosis. Plos Neglected Tropical Diseases *9*, e3441.

Hachani, A., Lossi, N.S., Hamilton, A., Jones, C., Bleves, S., Albesa-Jove, D., and Filloux, A. (2011). Type VI Secretion System in *Pseudomonas aeruginosa* SECRETION AND MULTIMERIZATION OF VgrG PROTEINS. Journal of Biological Chemistry *286*, 12317-12327.

Harley, V.S., Dance, D.A.B., Tovey, G., and Drasar, B.S. (1994). Interaction of *Pseudomonas pseudomallei* with macrophages. Biochemical Society Transactions *22*, S88-S88.

Harris, P.N.A., Ketheesan, N., Owens, L., and Norton, R.E. (2009). Clinical Features That Affect Indirect-Hemagglutination-Assay Responses to *Burkholderia pseudomallei*. Clinical and Vaccine Immunology *16*, 924-930.

Hirano, T., Kinoshita, N., Morikawa, K., and Yanagida, M. (1990). Snap helix with knob and hole - essential repeats in s-pombe nuclear-protein nuc2+. Cell *60*, 319-328.

Hoffmaster, A.R., AuCoin, D., Baccam, P., Baggett, H.C., Baird, R., Bhengsri, S., Blaney, D.D., Brett, P.J., Brooks, T.J.G., Brown, K.A.*, et al.* (2015). Melioidosis diagnostic workshop, 2013. Emerging infectious diseases *21*, e141045.

Holden, M.T.G., Titball, R.W., Peacock, S.J., Cerdeño-Tárraga, A.M., Atkins, T., Crossman, L.C., Pitt, T., Churcher, C., Mungall, K., Bentley, S.D.*, et al.* (2004). Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. Proceedings of the National Academy of Sciences of the United States of America *101*, 14240-14245.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Research *38*, W545-W549.

Hood, R.D., Singh, P., Hsu, F., Güvener, T., Carl, M.A., Trinidad, R.R.S., Silverman, J.M., Ohlson, B.B., Hicks, K.G., Plemel, R.L.*, et al.* (2010). A Type VI Secretion System of *Pseudomonas aeruginosa* Targets, a Toxin to Bacteria. Cell Host & Microbe *7*, 25-37.

Hoppe, I., Brenneke, B., Rohde, M., Kreft, A., Haussler, S., Reganzerowski, A., and Steinmetz, I. (1999). Characterization of a murine model of melioidosis: Comparison of different strains of mice. Infection and Immunity *67*, 2891-2900.

Incardona, M.-F., Bourenkov, G.P., Levik, K., Pieritz, R.A., Popov, A.N., and Svensson, O. (2009). EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. Journal of Synchrotron Radiation *16*, 872-879.

James, G.L., Delaney, B., Ward, L., Freeman, K., Mayo, M., and Currie, B.J. (2013). Surprisingly Low Seroprevalence of *Burkholderia pseudomallei* in Exposed Healthy Adults in the Darwin Region of Tropical Australia Where Melioidosis Is Highly Endemic. Clinical and Vaccine Immunology *20*, 759-760.

Jeong, Y., Kim, J., Kim, S., Kang, Y., Nagamatsu, T., and Hwang, I. (2003). Toxoflavin produced by *Burkholderia glumae* causing rice grain rot is responsible for inducing bacterial wilt in many field crops. Plant Disease *87*, 890-895.

Jitprasutwit, S., Ong, C., Juntawieng, N., Ooi, W.F., Hemsley, C.M., Vattanaviboon, P., Titball, R.W., Tan, P., and Korbsrisate, S. (2014).

Transcriptional profiles of *Burkholderia pseudomallei* reveal the direct and indirect roles of Sigma E under oxidative stress conditions. BMC Genomics *15*, 787.

Jones, A.L., Beveridge, T.J., and Woods, D.E. (1996). Intracellular survival of *Burkholderia pseudomallei*. Infection and Immunity *64*, 782-790.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. Journal of Molecular Biology *293*, 195-202.

Kabsch, W. (2010). XDS. Acta Crystallographica Section D-Biological Crystallography *66*, 125-132.

Kanamaru, S. (2009). Structural similarity of tailed phages and pathogenic bacterial secretion systems. Proceedings of the National Academy of Sciences of the United States of America *106*, 4067-4068.

Kanamaru, S., Leiman, P.G., Kostyuchenko, V.A., Chipman, P.R., Mesyanzhinov, V.V., Arisaka, F., and Rossmann, M.G. (2002). Structure of the cell-puncturing device of bacteriophage T4. Nature *415*, 553-557.

Kantardjieff, K.A., and Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. Protein Science *12*, 1865-1871.

Karpenahalli, M.R., Lupas, A.N., and Soeding, J. (2007). TPRpred: a tool for prediction of TPR-, PPR- and SELI-like repeats from protein sequences. Bmc Bioinformatics *8*, 2.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols *10*, 845-858.

Kespichayawattana, W., Rattanachetkul, S., Wanun, T., Utaisincharoen, P., and Sirisinha, S. (2000). *Burkholderia pseudomallei* induces cell fusion and actin-associated membrane protrusion: a possible mechanism for cell-to-cell spreading. Infection and Immunity *68*, 5377-5384.

Kikuchi, Y., Meng, X.Y., and Fukatsu, T. (2005). Gut symbiotic bacteria of the genus *Burkholderia* in the broad-headed bugs *Riptortus clavatus* and

Leptocorisa chinensis (Heteroptera : Alydidae). Applied and Environmental Microbiology *71*, 4035-4043.

Koh, S.F., Tay, S.T., and Puthucheary, S.D. (2013). Colonial morphotypes and biofilm forming ability of *Burkholderia pseudomallei*. Tropical Biomedicine *30*, 428-433.

Koponen, M.A., Zlock, D., Palmer, D.L., and Merlin, T.L. (1991). Melioidosis - forgotten, but not gone. Archives of Internal Medicine *151*, 605-608.

Korbsrisate, S., Vanaporn, M., Kerdsuk, P., Kespichayawattana, W., Vattanaviboon, P., Kiatpapan, P., and Lertmemongkolchai, G. (2005). The *Burkholderia pseudomallei* RpoE (AlgU) operon is involved in environmental stress tolerance and biofilm formation. FEMS Microbiology Letters *252*, 243-249.

Kostyuchenko, V.A., Leiman, P.G., Chipman, P.R., Kanamaru, S., van Raaij, M.J., Arisaka, F., Mesyanzhinov, V.V., and Rossmann, M.G. (2003). Three-dimensional structure of bacteriophage T4 baseplate. Nature Structural Biology *10*, 688-693.

Kostyuchenko, V.A., Chipman, P.R., Leiman, P.G., Arisaka, F., Mesyanzhinov, V. V. and Rossmann, M.G. (2005). The tail structure of bacteriophage T4 and its mechanism of contraction. Nature Structural and Molecular Biology *12*, 810-813.

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallographica Section D-Biological Crystallography *60*, 2256-2268.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. Journal of Molecular Biology *372*, 774-797.

Kube, S., Kapitein, N., Zimniak, T., Herzog, F., Mogk, A., and Wendler, P. (2014). Structure of the VipA/B Type VI Secretion Complex Suggests a Contraction-State-Specific Recycling Mechanism. Cell Reports *8*, 20-30.

Kudryashev, M., Wang, R.Y.R., Brackmann, M., Scherer, S., Maier, T., Baker, D., DiMaio, F., Stahlberg, H., Egelman, E.H., and Basler, M.

(2015). Structure of the Type VI Secretion System Contractile Sheath. Cell *160*, 952-962.

Laemmli, U.K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature *227*, 680-685.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947-2948.

Larsen, E., Smith, J.J., Norton, R., and Corkeron, M. (2013). Survival, Sublethal Injury, and Recovery of Environmental *Burkholderia pseudomallei* in Soil Subjected to Desiccation. Applied and Environmental Microbiology *79*, 2424-2427.

Lee, S-H., Chong, C-E., Lim, B-S., Chai, S-J., Sam, K-K., Mohamed, R. and Nathan, S. (2007). *Burkholderia pseudomallei* animal and human isolates from Malaysia exhibit different phenotypic characteristics. Diagnostic Microbiology and Infectious Disease *58*, 263-270.

Leiman, P.G., Chipman, P.R., Kostyuchenko, V.A., Mesyanzhinov, V.V. and Rossmann, M.G. (2004). Three-Dimensional Rearrangement of Proteins in the Tail of Bacteriophage T4 on Infection of Its Host. Cell *118*, 419-429.

Leiman, P.G., Basler, M., Ramagopal, U.A., Bonanno, J.B., Sauder, J.M., Pukatzki, S., Burley, S.K., Almo, S.C., and Mekalanos, J.J. (2009). Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. Proceedings of the National Academy of Sciences of the United States of America *106*, 4154-4159.

Leiman, P.G., Arisaka, F., van Raaij, M.J., Kostyuchenko, V.A., Aksyuk, A.A., Kanamaru, S., and Rossmann, M.G. (2010). Morphogenesis of the T4 tail and tail fibers. Virology Journal *7*, 355.

Lengwehasatit, I., Nuchtas, A., Tungpradabkul, S., Sirisinha, S., and Utaisincharoen, P. (2008). Involvement of *B. pseudomallei* RpoS in apoptotic cell death in mouse macrophages. Microbial Pathogenesis *44*, 238-245.

Li, W., Ma, C., Guan, R., Xu, Y., Tomchick, D.R. and Rizo, J. (2011). The Crystal Structure of a Munc13 C-terminal Module Exhibits a Remarkable Similarity to Vesicle Tethering Factors. Structure *19*, 1443-1445.

Lim, Y.W., Baik, K.S., Han, S.K., Kim, S.B., and Bae, K.S. (2003). *Burkholderia sordidicola* sp nov., isolated from the white-rot fungus *Phanerochaete sordida*. International Journal of Systematic and Evolutionary Microbiology *53*, 1631-1636.

Limmathurotsakul, D., and Peacock, S.J. (2011). Melioidosis: a clinical overview. British Medical Bulletin *99*, 125-139.

Limmathurotsakul, D., Chaowagul, W., Chierakul, W., Stepniewska, K., Maharjan, B., Wuthiekanun, V., White, N.J., Day, N.P.J., and Peacock, S.J. (2006). Risk factors for recurrent melioidosis in northeast Thailand. Clinical Infectious Diseases *43*, 979-986.

Limmathurotsakul, D., Jamsen, K., Arayawichanont, A., Simpson, J.A., White, L.J., Lee, S.J., Wuthiekanun, V., Chantratita, N., Cheng, A., Day, N.P.J*., et al.* (2010a). Defining the True Sensitivity of Culture for the Diagnosis of Melioidosis Using Bayesian Latent Class Models. Plos One *5*, e12485.

Limmathurotsakul, D., Wongratanacheewin, S., Teerawattanasook, N., Wongsuvan, G., Chaisuksant, S., Chetchotisakd, P., Chaowagul, W., Day, N.P.J., and Peacock, S.J. (2010b). Increasing Incidence of Human Melioidosis in Northeast Thailand. American Journal of Tropical Medicine and Hygiene *82*, 1113-1117.

Limmathurotsakul, D., Kanoksil, M., Wuthiekanun, V., Kitphati, R., deStavola, B., Day, N.P.J., and Peacock, S.J. (2013). Activities of Daily Living Associated with Acquisition of Melioidosis in Northeast Thailand: A Matched Case-Control Study. Plos Neglected Tropical Diseases *7*, e2072.

Limmathurotsakul, D., Wongsuvan, G., Aanensen, D., Ngamwilai, S., Saiprom, N., Rongkard, P., Thaipadungpanit, J., Kanoksil, M., Chantratita, N., Day, N.P.J*., et al.* (2014). Melioidosis Caused by *Burkholderia pseudomallei* in Drinking Water, Thailand, 2012. Emerging Infectious Diseases *20*, 265-268.

Limmathurotsakul, D., Funnell, S.G.P., Torres, A.G., Morici, L.A., Brett, P.J., Dunachie, S., Atkins, T., Altmann, D.M., Bancroft, G., Peacock, S.J., *et al.* (2015). Consensus on the development of vaccines against naturally acquired melioidosis. Emerging infectious diseases *21*, e141480.

Limmathurotsakul, D., Golding, N., Dance, D.A.B., Messina, J.P., Pigott, D.M., Moyes, C.L., Rolim, D.B., Bertherat, E., Day, N.P.J., Peacock, S.J. and Hay, S.I. (2016). Predicted global distribution of *Burkholderia pseudomallei* and burden of melioidosis. Nature Microbiology *1*, 15008

Losada, L., Ronning, C.M., DeShazer, D., Woods, D., Fedorova, N., Kim, H.S., Shabalina, S.A., Pearson, T.R., Brinkac, L., Tan, P., *et al.* (2010). Continuing Evolution of *Burkholderia mallei* Through Genome Reduction and Large-Scale Rearrangements. Genome Biology and Evolution *2*, 102-116.

Lumbiganon, P., and Viengnondha, S. (1995). Clinical manifestations of melioidosis in children. Pediatric Infectious Disease Journal *14*, 136-140.

Lunelli, M., Lokareddy, R.K., Zychlinsky, A., and Kolbe, M. (2009). IpaB-IpgC interaction defines binding motif for type III secretion translocator. Proceedings of the National Academy of Sciences of the United States of America *106*, 9661-9666.

MacIntyre, D.L., Miyata, S.T., Kitaoka, M., and Pukatzki, S. (2010). The *Vibrio cholerae* type VI secretion system displays antimicrobial properties. Proceedings of the National Academy of Sciences of the United States of America *107*, 19520-19524.

Maharjan, B., Chantratita, N., Vesaratchavest, M., Cheng, A., Wuthiekanun, V., Chierakul, W., Chaowagul, W., Day, N.P.J., and Peacock, S.J. (2005). Recurrent melioidosis in patients in northeast Thailand is frequently due to reinfection rather than relapse. Journal of Clinical Microbiology *43*, 6032-6034.

Majerczyk, C.D., Brittnacher, M.J., Jacobs, M.A., Armour, C.D., Radey, M.C., Bunt, R., Hayden, H.S., Bydalek, R., and Greenberg, E.P. (2014a). Cross-Species Comparison of the *Burkholderia pseudomallei*, *Burkholderia thailandensis*, and *Burkholderia mallei* Quorum-Sensing Regulons. Journal of Bacteriology *196*, 3862-3871.

Majerczyk, C., Brittnacher, M., Jacobs, M., Armour, C.D., Radey, M., Schneider, E., Phattarasokul, S., Bunt, R., and Greenberg, E.P. (2014b). Global analysis of the *Burkholderia thailandensis* Quorum Sensing-Controlled Regulon. Journal of Bacteriology 196, 1412-1424.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S.N., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I.*, et al.* (2015). CDD: NCBI's conserved domain database. Nucleic Acids Research *43*, D222-D226.

Matthews, B.W. (1968). Solvent content of protein crystals. Journal of Molecular Biology *33*, 491-497.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. Journal of Applied Crystallography *40*, 658-674.

McGuinness, E.T. (1973). Estimation of protein size, weight, and asymmetry by gel chromatography. Journal of Chemical Education *50*, 826-830.

McLeod, C., Morris, P.S., Bauert, P.A., Kilburn, C.J., Ward, L.M., Baird, R.W., and Currie, B.J. (2015). Clinical Presentation and Medical Management of Melioidosis in Children: A 24-Year Prospective Study in the Northern Territory of Australia and Review of the Literature. Clinical Infectious Diseases *60*, 21-26.

Meir, A., Helppolainen, S.H., Podoly, E., Nordlund, H.R., Hytonen, V.P., Maatta, J.A., Wilchek, M., Bayer, E.A., Kulomaa, M.S., and Livnah, O. (2009). Crystal Structure of Rhizavidin: Insights into the Enigmatic High-Affinity Interaction of an Innate Biotin-Binding Protein Dimer. Journal of Molecular Biology *386*, 379-390.

Meumann, E.M., Cheng, A.C., Ward, L., and Currie, B.J. (2012). Clinical Features and Epidemiology of Melioidosis Pneumonia: Results From a 21-Year Study and Review of the Literature. Clinical Infectious Diseases *54*, 362-369.

Miller, J.H. (1972). Experiments in Molecular Genetics. (Cold Springs Harbor New York: Cold Spring Harbour Laboratory Press)

Miller, W.R., Pannell, L., Cravitz, L., Tanner, W.A., and Ingalls, M.S. (1948a). Studies on certain biological characteristics of *Malleomyces mallei* and *Malleomyces pseudomallei* .1. Morphology, cultivation, viability, and isolation from contaminated specimens. Journal of Bacteriology *55*, 115-126.

Miller, W.R., Pannell, L., and Ingalls, M.S. (1948b). Experimental chemotherapy in glanders and melioidosis. American Journal of Hygiene *47*, 205-213.

Mitchell, A., Chang, H.Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S.*, et al.* (2015). The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Research *43*, D213-D221.

Morse, L.P., Moller, C.-C.B., Harvey, E., Ward, L., Cheng, A.C., Carson, P.J., and Currie, B.J. (2009). Prostatic Abscess Due to *Burkholderia pseudomallei*: 81 Cases From a 19-Year Prospective Melioidosis Study. Journal of Urology *182*, 542-547.

Mougous, J.D., Cuff, M.E., Raunser, S., Shen, A., Zhou, M., Gifford, C.A., Goodman, A.L., Joachimiak, G., Ordoñez, C.L., Lory, S.*, et al.* (2006). A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. Science *312*, 1526-1530.

Moule, M.G., Hemsley, C.M., Steet, Q., Guerra-Assunçäo, J.A., Lim, J., Sarkar-Tyson, M., Clark, T.G., Tan, P.B.O., Titball, R.W., Cuccui, J., et al., (2014). Genome-Wide Saturation Mutagenesis of *Burkholderia pseudomallei* K96243 Predicts Essential Genes and Novel Targets for Antimicrobial Development.

Moule, M.G., Spink, N., Willcocks, S., Lim, J., Guerra-Assunçäo, J.A., Cia, F., Champion, O.L., Senior, N.J., Atkins, H.S., Clark, T.*, et al.* (2016). Characterization of New Virulence Factors Involved in the Intracellular Growth and Survival of *Burkholderia pseudomallei*. Infection and Immunity *84*, 701-710.

Moulin, L., Munive, A., Dreyfus, B., and Boivin-Masson, C. (2001). Nodulation of legumes by members of the β-subclass of Proteobacteria. Nature *411*, 948-950.

Muangman, S., Korbsrisate, S., Muangsombut, V., Srinon, V., Adler, N.L., Schroeder, G.N., Frankel, G., and Galyov, E.E. (2011). BopC is a type III secreted effector protein of *Burkholderia pseudomallei*. FEMS Microbiology Letters *323*, 75-82.

Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Verezemska, O., Isbandi, M., Thomas, A.D., Ali, R., Sharma, K., Kyrpides, N.C*., et al.* (2016). Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. Nucleic acids research. doi: 10.1093/nar/gkw992.

Murdoch, S.L., Trunk, K., English, G., Fritsch, M.J., Pourkarimi, E., and Coulthurst, S.J. (2011). The Opportunistic Pathogen *Serratia marcescens* Utilizes Type VI Secretion To Target Bacterial Competitors. Journal of Bacteriology *193*, 6057-6069.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallographica Section D-Biological Crystallography *53*, 240-255.

Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallographica Section D-Biological Crystallography *67*, 355-367.

Nandakumar, R., Shahjahan, A.K.M., Yuan, X.L., Dickstein, E.R., Groth, D.E., Clark, C.A., Cartwright, R.D., and Rush, M.C. (2009). *Burkholderia glumae* and *B. gladioli* Cause Bacterial Panicle Blight in Rice in the Southern United States. Plant Disease *93*, 896-905.

NCBI Resource Coordinators. (2016). Database resources of the National Center for Biotechnology Information. Nucleic Acids Research *44*, D7-D19.

Ngauy, V., Lemeshev, Y., Sadkowski, L., and Crawford, G. (2005). Cutaneous melioidosis in a man who was taken as a prisoner of war by the Japanese during World War II. Journal of Clinical Microbiology *43*, 970-972.

Nieves, W., Asakrah, S., Qazi, O., Brown, K.A., Kurtz, J., AuCoin, D.P., McLachlan, J.B., Roy, C.J., and Morici, L.A. (2011). A naturally derived outer-membrane vesicle vaccine protects against lethal pulmonary *Burkholderia pseudomallei* infection. Vaccine *29*, 8381-8389.

North, A.C.T. (1989). 3-dimensional arrangement of conserved amino-acid residues in a superfamily of specific ligand-binding proteins. International Journal of Biological Macromolecules *11*, 56-58.

Notredame, C., Higgins, D.G. and Heringa, J. (2000). T-Coffee: A Novel Method for Fast and Accurate Multiple Sequence Alignment. Journal of Molecular Biology *302*, 205-217.

Ooi, W.F., Ong, C., Nandi, T., Kreisberg, J.F., Chua, H.H., Sun, G.W., Chen, Y.H., Mueller, C., Conejero, L., Eshaghi, M.*, et al.* (2013). The Condition-Dependent Transcriptional Landscape of *Burkholderia pseudomallei*. Plos Genetics *9*, e1003795.

Osipiuk, J., Xu, X., Cui, H., Savchenko, A., Edwards, A., and Joachimiak, A. (2011). Crystal structure of secretory protein Hcp3 from *Pseudomonas aeruginosa*. Journal of Structural and Functional Genomics *12*, 21-26.

Osiriphun, Y., Wongtrakoongate, P., Sanongkiet, S., Suriyaphol, P., Thongboonkerd, V., and Tungpradabkul, S. (2009). Identification and Characterization of RpoS Regulon and RpoS-Dependent Promoters in *Burkholderia pseudomallei*. Journal of Proteome Research *8*, 3118-3131.

Pallen, M.J., Francis, M.S., and Futterer, K. (2003). Tetratricopeptide-like repeats in type-III-secretion chaperones and regulators. FEMS Microbiology Letters *223*, 53-60.

Pape, T. and Shneider, T.R. (2004). *HKL2MAP*: a graphical user interface for macromolecular phasing with *SHELX* programs. Journal of Applied Crystallography *37*, 843-844.

Parte, A.C. (2014). LPSN-list of prokaryotic names with standing in nomenclature. Nucleic Acids Research *42*, D613-D616.

Pautsch, A., and Schulz, G.E. (2000). High-resolution structure of the OmpA membrane domain. Journal of Molecular Biology *298*, 273-28

Payne, G.W., Vandamme, P., Morgan, S.H., LiPuma, J.J., Coenye, T., Weightman, A.J., Jones, T.H., and Mahenthiralingam, E. (2005). Development of a *recA* gene-based identification approach for the the entire *Burkholderia* genus. Applied and Environmental Microbiology *71*, 3917-3927.

Peabody, C. R., Chung, Y.J., Yen, M-R, Vidal-Ingigliardi, D., Pugsley, A.P. and Saier Jr, M.H. (2003). Type II protein secretion and its relationship to bacterial type IV pili and archael flagella. Microbiology *149*, 3051-3072.

Peacock, S.J., Limmathurotsakul, D., Lubell, Y., Koh, G.C.K.W., White, L.J., Day, N.P.J., and Titball, R.W. (2012). Melioidosis Vaccines: A Systematic Review and Appraisal of the Potential to Exploit Biodefense Vaccines for Public Health Purposes. Plos Neglected Tropical Diseases *6*, e1488.

Pell, L.G., Kanelis, V., Donaldson, L.W., Howell, P.L., and Davidson, A.R. (2009). The phage λ major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. Proceedings of the National Academy of Sciences of the United States of America *106*, 4160-4165.

Pervaiz, S., and Brew, K. (1985). Homology of beta-lactoglobulin, serum retinol-binding protein, and protein hc. Science *228*, 335-337.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods *8*, 785-786.

Pevsner, J., Reed, R.R., Feinstein, P.G., and Snyder, S.H. (1988). Molecular-cloning of odorant-binding protein - member of a ligand carrier family. Science *241*, 336-339.

Pietrosiuk, A., Lenherr, E.D., Falk, S., Bönemann, G., Kopp, J., Zentgraf, H., Sinning, I., and Mogk, A. (2011). Molecular Basis for the Unique Role of the AAA^+ Chaperone ClpV in Type VI Protein Secretion. Journal of Biological Chemistry *286*, 30010-30021.

Pitman, M.C., Luck, T., Marshall, C.S., Anstey, N.M., Ward, L., and Currie, B.J. (2015). Intravenous Therapy Duration and Outcomes in Melioidosis: A New Treatment Paradigm. Plos Neglected Tropical Diseases *9*, e0003586.

Pitt, T.L., Trakulsomboon, S., and Dance, D.A.B. (2007). Recurrent melioidosis: Possible role of infection with multiple strains of *Burkholderia pseudomallei*. Journal of Clinical Microbiology *45*, 680-681.

Planamente, S., Salih, O., Manoli, E., Albesa-Jové, D., Freemont, P.S., and Filloux, A. (2016). TssA forms a gp6-like ring attached to the type VI secretion sheath. EMBO Journal *35*, 1613-1627.

Price, E.P., Sarovich, D.S., Mayo, M., Tuanyok, A., Drees, K.P., Kaestli, M., Beckstrom-Sternberg, S.M., Babic-Sternberg, J.S., Kidd, T.J., Bell, S.C*., et al.* (2013). Within-Host Evolution of *Burkholderia pseudomallei* over a Twelve-Year Chronic Carriage Infection. Mbio *4*, e00388-13.

Pruksachartvuthi, S., Aswapokee, N., and Thankerngpol, K. (1990). Survival of *Pseudomonas pseudomallei* in human phagocytes. Journal of Medical Microbiology *31*, 109-114.

Pukatzki, S., Ma, A.T., Sturtevant, D., Krastins, B., Sarracino, D., Nelson, W.C., Heidelberg, J.F., and Mekalanos, J.J. (2006). Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. Proceedings of the National Academy of Sciences of the United States of America *103*, 1528-1533.

Pukatzki, S., Ma, A.T., Revel, A.T., Sturtevant, D., and Mekalanos, J.J. (2007). Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. Proceedings of the National Academy of Sciences of the United States of America *104*, 15508-15513.

Pumpuang, A., Chantratita, N., Wikraiphat, C., Saiprom, N., Day, N.P.J., Peacock, S.J., and Wuthiekanun, V. (2011). Survival of *Burkholderia pseudomallei* in distilled water for 16 years. Transactions of the Royal Society of Tropical Medicine and Hygiene *105*, 598-600.

Rainbow, L., Hart, C.A., and Winstanley, C. (2002). Distribution of type III secretion gene clusters in *Burkholderia pseudomallei*, *B. thailandensis* and *B. mallei*. Journal of Medical Microbiology *51*, 374-384.

Ralph, A., McBride, J., and Currie, B.J. (2004). Transmission of *Burkholderia pseudomallei* via breast milk in northern Australia. Pediatric Infectious Disease Journal *23*, 1169-1171.

Robertson, J., Levy, A., Sagripanti, J.-L., and Inglis, T.J.J. (2010). The Survival of *Burkholderia pseudomallei* in Liquid Media. American Journal of Tropical Medicine and Hygiene *82*, 88-94.

Rotz, L.D., Khan, A.S., Lillibridge, S.R., Ostroff, S.M., and Hughes, J.M. (2002). Public health assessment of potential biological terrorism agents. Emerging Infectious Diseases *8*, 225-230.

Royer, W.E., Jr., Sharma, H., Strand, K., Knapp, J.E., and Bhyravbhatla, B. (2006). Lumbricus erythrocruorin at 3.5 angstrom resolution: Architecture of a megadalton respiratory complex. Structure *14*, 1167-1177.

Ruggiero Bachega, J.F., Maluf, F.V., Andi, B., Pereira, H.D.M., Carazzollea, M.F., Orville, A.M., Tabak, M., Brandao-Neto, J., Garratt, R.C., and Reboredo, E.H. (2015). The structure of the giant haemoglobin from *Glossoscolex paulistus*. Acta Crystallographica Section D-Structural Biology *71*, 1257-1271.

Rutten, L., Geurtsen, J., Lambert, W., Smolenaers, J.J.M., Bonvin, A.M., De Haan, A., Van der Ley, P., Egmond, M.R., Gros, P., and Tommassen, J. (2006). Crystal structure and catalytic mechanism of the LPS 3-O-deacylase PagL from *Pseudomonas aeruginosa*. Proceedings of the National Academy of Sciences of the United States of America *103*, 7071-7076.

Saiprom, N., Amornchai, P., Wuthiekanun, V., Day, N.P.J., Limmathurotsakul, D., Peacock, S.J., and Chantratita, N. (2015). Trimethoprim/sulfamethoxazole resistance in clinical isolates of *Burkholderia pseudomallei* from Thailand. International Journal of Antimicrobial Agents *45*, 557-559.

Salomon, D., Kinch, L.N., Trudgian, D.C., Guo, X., Klimko, J.A., Grishin, N.V., Mirzaei, H., and Orth, K. (2014). Marker for type VI secretion system effectors. Proceedings of the National Academy of Sciences of the United States of America *111*, 9271-9276.

Santos, A.V., Dillon, R.J., Dillon, V.M., Reynolds, S.E., and Samuels, R.I. (2004). Ocurrence of the antibiotic producing bacterium *Burkholderia* sp in colonies of the leaf-cutting ant Atta sexdens rubropilosa. Fems Microbiology Letters *239*, 319-323.

Sarkar-Tyson, M., Smither, S.J., Harding, S.V., Atkins, T.P., and Titball, R.W. (2009). Protective efficacy of heat-inactivated *B. thailandensis, B. mallei* or *B. pseudomallei* against experimental melioidosis and glanders. Vaccine *27*, 4447-4451.

Sarovich, D.S., Price, E.P., Limmathurotsakul, D., Cook, J.M., Von Schulze, A.T., Wolken, S.R., Keim, P., Peacock, S.J., and Pearson, T. (2012). Development of ceftazidime resistance in an acute *Burkholderia pseudomallei* infection. Infection and drug resistance *5*, 129-132.

Sarovich, D.S., Ward, L., Price, E.P., Mayo, M., Pitman, M.C., Baird, R.W., and Currie, B.J. (2014). Recurrent Melioidosis in the Darwin Prospective Melioidosis Study: Improving Therapies Mean that Relapse Cases Are Now Rare. Journal of Clinical Microbiology *52*, 650-653.

Sauter, N.K., Grosse-Kunstleve, R.W., and Adams, P.D. (2004). Robust indexing for automatic data collection. Journal of Applied Crystallography *37*, 399-409.

Sawana, A., Adeolu, M., and Gupta, R.S. (2014). Molecular signatures and phylogenomic analysis of the genus *Burkholderia*: proposal for division of this genus into the emended genus *Burkholderia* containing pathogenic organisms and a new genus *Paraburkholderia* gen. nov harboring environmental species. Frontiers in Genetics *5*, 429.

Schell, M.A., Ulrich, R.L., Ribot, W.J., Brueggemann, E.E., Hines, H.B., Chen, D., Lipscomb, L., Kim, H.S., Mrázek, J., Nierman, W.C*., et al.* (2007). Type VI secretion is a major virulence determinant in *Burkholderia mallei*. Molecular Microbiology *64*, 1466-1485.

Schmid, A.B., Lagleder, S., Graewert, M.A., Roehl, A., Hagn, F., Wandinger, S.K., Cox, M.B., Demmer, O., Richter, K., Groll, M*., et al.* (2012). The architecture of functional modules in the Hsp90 co-chaperone Sti1/Hop. Embo Journal *31*, 1506-1517.

Schwarz, S., West, T.E., Boyer, F., Chiang, W.-C., Carl, M.A., Hood, R.D., Rohmer, L., Tolker-Nielsen, T., Skerrett, S.J., and Mougous, J.D. (2010). Burkholderia Type VI Secretion Systems Have Distinct Roles in Eukaryotic and Bacterial Cell Interactions. Plos Pathogens *6*, e1001068.

Schwarz, S., Singh, P., Robertson, J.D., LeRoux, M., Skerrett, S.J., Goodlett, D.R., West, T.E., and Mougous, J.D. (2014). VgrG-5 Is a *Burkholderia* Type VI Secretion System-Exported Protein Required for Multinucleated Giant Cell Formation and Virulence. Infection and Immunity *82*, 1445-1452.

Sessitsch, A., Coenye, T., Sturz, A.V., Vandamme, P., Barka, E.A., Salles, J.F., Van Elsas, J.D., Faure, D., Reiter, B., Glick, B.R.*, et al.* (2005). *Burkholderia phytofirmans* sp. nov., a novel plant-associated bacterium with plant-beneficial properties. International Journal of Systematic and Evolutionary Microbiology *55*, 1187-1192.

Shalom, G., Shaw, J.G., and Thomas, M.S. (2007). *In vivo* expression technology identifies a type VI secretion system locus in *Burkholderia pseudomallei* that is induced upon invasion of macrophages. Microbiology-Sgm *153*, 2689-2699.

Sheahan, K.L., Cordero, C.L., and Satchell, K.J.F. (2004). Identification of a domain within the multifunctional *Vibrio cholerae* RTX toxin that covalently cross-links actin. Proceedings of the National Academy of Sciences of the United States of America *101*, 9798-9803.

Sheldrick, G.M. (2008). A short history of *SHELX*. Acta Crystallographica Section A *64*, 112-122.

Shneider, M.M., Buth, S.A., Ho, B.T., Basler, M., Mekalanos, J.J., and Leiman, P.G. (2013). PAAR-repeat proteins sharpen and diversify the type VI secretion system spike. Nature *500*, 350-353.

Shumilin, I.A., Cymborowski, M., Chertihin, O., Jha, K.N., Herr, J.C., Lesley, S.A., Joachimiak, A., and Minor, W. (2012). Identification of Unknown Protein Function Using Metabolite Cocktail Screening. Structure *20*, 1715-1725.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D. and Higgins, D.G. (2011). Fast scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology *7*, 539.

Sikorski, R.S., Boguski, M.S., Goebl, M., and Hieter, P. (1990). A repeating amino-acid motif in cdc23 defines a family of proteins and a new relationship among genes required for mitosis and rna-synthesis. Cell *60*, 307-317.

Silva, E.B., and Dow, S.W. (2013). Development of *Burkholderia mallei* and *pseudomallei* vaccines. Frontiers in Cellular and Infection Microbiology *3,* 10.

Silverman, J.M., Agnello, D.M., Zheng, H., Andrews, B.T., Li, M., Catalano, C.E., Gonen, T., and Mougous, J.D. (2013). Haemolysin Coregulated Protein Is an Exported Receptor and Chaperone of Type VI Secretion Substrates. Molecular Cell *51*, 584-593.

Sonnhammer, E.L., von Heijne, G., and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. Proceedings International Conference on Intelligent Systems for Molecular Biology *6*, 175-182.

Stevens, M.P., Wood, M.W., Taylor, L.A., Monaghan, P., Hawes, P., Jones, P.W., Wallis, T.S., and Galyov, E.E. (2002). An Inv/Mxi-Spa-like type III protein secretion system in *Burkholderia pseudomallei* modulates intracellular behaviour of the pathogen. Molecular Microbiology *46*, 649-659.

Stevens, M.P., Friebel, A., Taylor, L.A., Wood, M.W., Brown, P.J., Hardt, W.D., and Galyov, E.E. (2003). A *Burkholderia pseudomallei* type III secreted protein, BopE, facilitates bacterial invasion of epithelial cells and exhibits guanine nucleotide exchange factor activity. Journal of Bacteriology *185*, 4992-4996.

Stevens, M.P., Haque, A., Atkins, T., Hill, J., Wood, M.W., Easton, A., Nelson, M., Underwood-Fowler, C., Titball, R.W., Bancroft, G.J.*, et al.* (2004). Attenuated virulence and protective efficacy of a *Burkholderia pseudomallei* bsa type III secretion mutant in murine models of melioidosis. Microbiology-Sgm *150*, 2669-2676.

Suarez, G., Sierra, J.C., Erova, T.E., Sha, J., Horneman, A.J., and Chopra, A.K. (2010). A Type VI Secretion System Effector Protein, VgrG1, from *Aeromonas hydrophila* That Induces Host Cell Toxicity by ADP Ribosylation of Actin. Journal of Bacteriology *192*, 155-168.

Subsin, B., Thomas, M.S., Katzenmeier, G., Shaw, J.G., Tungpradabkul, S., and Kunakorn, M. (2003). Role of the stationary growth phase sigma factor RpoS of *Burkholderia pseudomallei* in response to physiological stress conditions. Journal of Bacteriology *185*, 7008-7014.

Suputtamongkol, Y., Hall, A.J., Dance, D.A.B., Chaowagul, W., Rajchanuvong, A., Smith, M.D., and White, N.J. (1994). The epidemiology of melioidosis in Ubon-Ratchatani, Northeast Thailand. International Journal of Epidemiology *23*, 1082-1090.

Suputtamongkol, Y., Chaowagul, W., Chetchotisakd, P., Lertpatanasuwun, N., Intaranongpai, S., Ruchutrakool, T., Budhsarawong, D., Mootsikapun, P., Wuthiekanun, V., Teerawatasook, N.*, et al.* (1999). Risk factors for melioidosis and bacteremic melioidosis. Clinical Infectious Diseases *29*, 408-413.

Suwannasaen, D., Mahawantung, J., Chaowagul, W., Limmathurotsakul, D., Felgner, P.L., Davies, H., Bancroft, G.J., Titball, R.W., and Lertmemongkolchai, G. (2011). Human Immune Responses to *Burkholderia pseudomallei* Characterized by Protein Microarray Analysis. Journal of Infectious Diseases *203*, 1002-1011.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P.*, et al.* (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Research *43*, D447-D452.

Tandhavanant, S., Thanwisai, A., Limmathurotsakul, D., Korbsrisate, S., Day, N.P.J., Peacock, S.J., and Chantratita, N. (2010). Effect of colony morphology variation of *Burkholderia pseudomallei* on intracellular survival and resistance to antimicrobial environments in human macrophages in vitro. BMC Microbiology *10*, 303.

Taylor, N.M.I., Prokhorov, N.S., Guerrero-Ferreira, R.C., Shneider, M.M., Browning, C., Goldie, K.N., Stahlberg, H., and Leiman, P.G. (2016). Structure of the T4 baseplate and its function in triggering sheath contraction. Nature *533*, 346-352.

Thaipadungpanit, J., Chierakul, W., Pattanaporkrattana, W., Phoodaeng, A., Wongsuvan, G., Huntrakun, V., Amornchai, P., Chatchen, S., Kitphati, R., Wuthiekanun, V.*, et al.* (2014). *Burkholderia pseudomallei* in Water Supplies, Southern Thailand. Emerging Infectious Diseases *20*, 1947-1949.

Thibault, F.M., Hernandez, E., Vidal, D.R., Girardet, M., and Cavallo, J.D. (2004). Antibiotic susceptibility of 65 isolates of *Burkholderia pseudomallei* and *Burkholderia mallei* to 35 antimicrobial agents. Journal of Antimicrobial Chemotherapy *54*, 1134-1138.

Thongboonkerd, V., Vanaporn, M., Songtawee, N., Kanlaya, R., Sinchaikul, S., Chen, S.-T., Easton, A., Chu, K., Bancroft, G.J., and Korbsrisate, S. (2007). Altered proteome in *Burkholderia pseudomallei*

rpoE operon knockout mutant: Insights into mechanisms of rpoE operon in stress tolerance, survival, and virulence. Journal of Proteome Research *6*, 1334-1341.

Tong, S.D., Yang, S., Lu, Z.Z., and He, W.H. (1996). Laboratory investigation of ecological factors influencing the environmental presence of *Burkholderia pseudomallei*. Microbiology and Immunology *40*, 451-453.

Tuanyok, A., Kim, H.S., Nierman, W.C., Yu, Y., Dunbar, J., Moore, R.A., Baker, P., Tom, M., Ling, J.M.L., and Woods, D.E. (2005). Genome-wide expression analysis of iron regulation in *Burkholderia pseudomallei* and *Burkholderia mallei* using DNA microarrays. FEMS Microbiology Letters *252*, 327-335.

Ura, H., Furuya, N., Iiyama, K., Hidaka, M., Tsuchiya, K., and Matsuyama, N. (2006). *Burkholderia gladioli* associated with symptoms of bacterial grain rot and leaf-sheath browning of rice plants. Journal of General Plant Pathology *72*, 98-103.

Utaisincharoen, P., Arjcharoen, S., Limposuwan, K., Tungpradabkul, S., and Sirisinha, S. (2006). *Burkholderia pseudomallei* RpoS regulates multinucleated giant cell formation and inducible nitric oxide synthase expression in mouse macrophage cell line (RAW 264.7). Microbial Pathogenesis *40*, 184-189.

Vanaporn, M., Vattanaviboon, P., Thongboonkerd, V., and Korbsrisate, S. (2008). The *rpoE* operon regulates heat stress response in *Burkholderia pseudomallei*. FEMS Microbiology Letters *284*, 191-196.

Vandeputte-Rutten, L., Bos, M.P., Tommassen, J., and Gros, P. (2003). Crystal structure of Neisserial surface protein A (NspA), a conserved outer membrane protein with vaccine potential. Journal of Biological Chemistry *278*, 24825-24830.

Vander Broek, C.W., Chalmers, K.J., Stevens, M.P., and Stevens, J.M. (2015). Quantitative Proteomic Analysis of *Burkholderia pseudomallei* Bsa Type III Secretion System Effectors Using Hypersecreting Mutants. Molecular & Cellular Proteomics *14*, 905-916.

Vanlaere, E., LiPuma, J.J., Baldwin, A., Henry, D., De Brandt, E., Mahenthiralingam, E., Speert, D., Dowson, C., and Vandamme, P. (2008). *Burkholderia latens* sp. nov., *Burkholderia diffusa* sp. nov.,

Burkholderia arboris sp. nov., *Burkholderia seminalis* sp. nov and *Burkholderia metallica* sp. nov., novel species within the *Burkholderia cepacia* complex. International Journal of Systematic and Evolutionary Microbiology *58*, 1580-1590.

Velapatiño, B., Limmathurotsakul, D., Peacock, S.J., and Speert, D.P. (2012). Identification of differentially expressed proteins from *Burkholderia pseudomallei* isolated during primary and relapsing melioidosis. Microbes and Infection *14*, 335-340.

Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007). Automated structure solution with autoSHARP. Methods in molecular biology (Clifton, NJ) *364*, 215-230.

Vorobiev, S., Neely, H., Chen, Y., Seetharaman, J., Patel, P., Xiao, R., Ciccosanti, C., Maglaqui, M., Everett, J.K., Nair, R., Acton, T.B., Rost, B., Montelione, G.T., Tong, L. and Hunt, J.F. Northeast Structrual Genomics Consortium. Crystal Structure of the C-terminal part of the TPR repeat-containing protein Q11TI6_CYTH3 from *Cytophaga hutchinsonii*.

Warawa, J., and Woods, D.E. (2005). Type III secretion system cluster 3 is required for maximal virulence of *Burkholderia pseudomallei* in a hamster infection model. FEMS Microbiology Letters *242*, 101-108.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189-1191.

Wetmore, P.W., and Gochenour, W.S. (1956). Comparative studies of the genus *malleomyces* and selected *pseudomonas* species .1. Morphological and cultural characteristics. Journal of Bacteriology *72*, 79-89.

Whitlock, G.C., Estes, D.M., and Torres, A.G. (2007). Glanders: off to the races with *Burkholderia mallei*. Fems Microbiology Letters *277*, 115-122.

Whitmore, A. (1913). An account of a glanders-like disease occurring in rangoon. Journal of Hygiene *13*, 1-35.

Whitmore, A. and Krishnaswami, C.S. (1912). An account of the discovery of a hitherto undescribed infective disease occuring among the population of Ragoon. Indian Medical Gazette *47*, 262-267.

Wiersinga, W.J., van der Poll, T., White, N.J., Day, N.P., and Peacock, S.J. (2006). Melioidosis: insights into the pathogenicity of *Burkholderia pseudomallei*. Nature Reviews Microbiology *4*, 272-282.

Wiersinga, W.J., Currie, B.J., and Peacock, S.J. (2012). Melioidosis. New England Journal of Medicine *367*, 1035-1044.

Willcocks, S.J., Denman, C.C., Atkins, H.S., and Wren, B.W. (2016). Intracellular replication of the well-armed pathogen *Burkholderia pseudomallei*. Current Opinion in Microbiology *29*, 94-103.

Williams, S.G., Varcoe, L.T., Attridge, S.R., and Manning, P.A. (1996). *Vibrio cholerae* Hcp, a secreted protein coregulated with HlyA. Infection and Immunity *64*, 283-289.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A.*, et al.* (2011). Overview of the CCP4 suite and current developments. Acta Crystallographica Section D-Biological Crystallography *67*, 235-242.

Winsor, G.L., Khaira, B., Van Rossum, T., Lo, R., Whiteside, M.D., and Brinkman, F.S.L. (2008). The Burkholderia Genome Database: facilitating flexible queries and comparative analyses. Bioinformatics *24*, 2803-2804.

Winstanley, C., Hales, B.A., and Hart, C.A. (1999). Evidence for the presence in *Burkholderia pseudomallei* of a type III secretion system-associated gene cluster. Journal of Medical Microbiology *48*, 649-656.

Winter, G. (2010). xia2: an expert system for macromolecular crystallography data reduction. Journal of Applied Crystallography *43*, 186-190.

Winter, G., and McAuley, K.E. (2011). Automated data collection for macromolecular crystallography. Methods *55*, 81-93.

Wongtrakoongate, P., Mongkoldhumrongkul, N., Chaijan, S., Kamchonwongpaisan, S., and Tungpradabkul, S. (2007). Comparative proteomic profiles and the potential markers between *Burkholderia pseudomallei* and *Burkholderia thailandensis*. Molecular and Cellular Probes *21*, 81-91.

Wongtrakoongate, P., Tumapa, S., and Tungpradabkul, S. (2012). Regulation of a quorum sensing system by stationary phase sigma factor RpoS and their co-regulation of target genes in *Burkholderia pseudomallei*. Microbiology and Immunology *56*, 281-294.

Wuthiekanun, V., Dance, D.A.B., Wattanagoon, Y., Supputtamongkol, Y., Chaowagul, W., and White, N.J. (1990). The use of selective media for the isolation of *Pseudomonas pseudomallei* in clinical-practice. Journal of Medical Microbiology *33*, 121-126.

Wuthiekanun, V., Smith, M.D., Dance, D.A.B., Walsh, A.L., Pitt, T.L., and White, N.J. (1996). Biochemical characteristics of clinical and environmental isolates of *Burkholderia pseudomallei*. Journal of Medical Microbiology *45*, 408-412.

Wuthiekanun, V., Suputtamongkol, Y., Simpson, A.J.H., Kanaphun, P., and White, N.J. (2001). Value of throat swab in diagnosis of melioidosis. Journal of Clinical Microbiology *39*, 3801-3802.

Wuthiekanun, V., Cheng, A.C., Chierakul, W., Amornchai, P., Limmathurotsakul, D., Chaowagul, W., Simpson, A.J.H., Short, J.M., Wongsuvan, G., Maharjan, B.*, et al.* (2005). Trimethoprim/sulfamethoxazole resistance in clinical isolates of *Burkholderia pseudomallei*. Journal of Antimicrobial Chemotherapy *55*, 1029-1031.

Wuthiekanun, V., Chierakul, W., Langa, S., Chaowagul, W., Panpitpat, C., Saipan, P., Thoujaikong, T., Day, N.P., and Peacock, S.J. (2006). Short report: Development of antibodies to *Burkholderia pseudomallei* during childhood in melioidosis-endemic northeast Thailand. American Journal of Tropical Medicine and Hygiene *74*, 1074-1075.

Yabuuchi, E., Kosako, Y., Oyaizu, H., Yano, I., Hotta, H., Hashimoto, Y., Ezaki, T., and Arakawa, M. (1992). Proposal of *Burkholderia* gen.nov. and transfer of 7 species of the genus pseudomonas homology group-ii to the new genus, with the type species *Burkholderia cepacia* (Palleroni and Holmes 1981) comb.nov. Microbiology and Immunology *36*, 1251-1275.

Yoder-Himes, D.R., Chain, P.S.G., Zhu, Y., Wurtzel, O., Rubin, E.M., Tiedje, J.M., and Sorek, R. (2009). Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing.

Proceedings of the National Academy of Sciences of the United States of America *106*, 3976-3981.

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.*, et al.* (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics *26*, 1608-1615.

Zanotti, G., Folli, C., Cendron, L., Alfieri, B., Nishida, S.K., Gliubich, F., Pasquato, N., Negro, A., and Berni, R. (2008). Structural and mutational analyses of protein-protein interactions between transthyretin and retinol-binding protein. Febs Journal *275*, 5841-5854.

Zhang, Z., Sauter, N.K., van den Bedem, H., Snell, G., and Deacon, A.M. (2006). Automated diffraction image analysis and spot searching for high-throughput crystal screening. Journal of Applied Crystallography *39*, 112-119.

Zheng, J., and Leung, K.Y. (2007). Dissection of a type VI secretion system in *Edwardsiella tarda*. Molecular Microbiology *66*, 1192-1206.

Zoued, A., Durand, E., Bebeacua, C., Brunet, Y.R., Douzi, B., Cambillau, C., Cascales, E., and Journet, L. (2013). TssK Is a Trimeric Cytoplasmic Protein Interacting with Components of Both Phage-like and Membrane Anchoring Complexes of the Type VI Secretion System. Journal of Biological Chemistry *288*, 27031-27041.

Zoued, A., Durand, E., Brunet, Y.R., Spinelli, S., Douzi, B., Guzzo, M., Flaugnatti, N., Legrand, P., Journet, L., Fronzes, R.*, et al.* (2016). Priming and polymerization of a bacterial contractile tail structure. Nature *531*, 59-63.

Zueter, A., Yean, C.Y., Abumarzouq, M., Rahman, Z.A., Deris, Z.Z., and Harun, A. (2016). The epidemiology and clinical spectrum of melioidosis in a teaching hospital in a North-Eastern state of Malaysia: a fifteen-year review. Bmc Infectious Diseases *16*, 333.