Genetic, Biochemical and Structural Characterisation of YecA, A Novel Component of the Bacterial Sec Machinery

by

TAMAR CRANFORD-SMITH

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Institute of Microbiology and Infection

School of Biosciences

University of Birmingham

2017

ABSTRACT

The Sec pathway provides a mechanism for the translocation of proteins across or into the cytoplasmic membrane. In bacteria, SecA is a core component of the Sec machinery. YecA has a 20-amino acid sequence at its carboxy-terminus that has high sequence identity to the zinc-binding domain at the carboxy-terminus of SecA. This study provides evidence to show that YecA is a novel component of the Sec machinery of *E. coli*. The *yecA* gene is not essential for the viability of *E. coli* but the deletion of *yecA* interferes with Sec-dependent translocation and the combined deletion of the *yecA* and *secB* genes results in a severely cold-sensitive phenotype. The genetic investigations were supported by biochemical evidence that suggests that YecA improves the translocation-coupled ATPase activity of SecA. Structural investigations suggest that YecA is a monomer in solution. The α-helical domain that forms the main body of YecA is connected via a short linker with limited flexibility to an independent metal-binding domain that has two conformations. The purification of YecA suggested the presence of iron. Biophysical experiments were used to confirm the interaction of the YecA metal-binding domain with iron. This study provides evidence for an additional component of the translocation machinery.

ACKNOWLEDGMENTS

I am most grateful to my supervisor, Dr. Damon Huber, for all his help and guidance. He has provided me with the opportunity to pursue various avenues of research and attend international and national conferences. I thank him for his compassion and understanding. His passion for science remains a source of inspiration.

I am lucky to have had the opportunity of working with brilliant collaborators and I want to thank Prof. Jeff Cole, Prof. Ian Collinson, Dr. Will Allen, Dr. Jack Bryant, Prof. Jeff Green, Dr. Eva Hyde, Dr. Janet Lovett and Anokhi Shah for their support with various aspects of this thesis and my learning. I am particularly grateful to Dr. Tim Knowles for his good-natured and generous contribution to the NMR and SAXS experiments. I am also grateful to Dr. Mohammed Jamshad who is a fount of useful advice and took YecA, with Dr. Tim Knowles, to Geneva for SAXS analysis.

Funding was gratefully received from the Biotechnology and Biological Sciences Research Council (BBSRC) for my studentship. I thank the Microbiology Society for funding my attendance of the SGM in 2016.

T101 is a uniquely wonderful place to work. The six-pack club began as a tongue-in-cheek WhatsApp group and quickly evolved into actual sit-ups. To stalwart and sporadic members alike, thank you! Rachael is a true Ravenclaw. She has never once failed to overcome her shyness to reach out to the newest members of T101 and sort them in to their Harry Potter house. She is a credit to Ravenclaw and I am proud to be her friend.

I have to thank my wonderful friends, Ginny, Eloise, Anna, Flic and Charlotte, who have given me unconditional love and support throughout my time in Birmingham. I have been so touched by the generosity of Bill and Sally Fellows, my not-quite in-laws, who have welcomed me into their home and looked after me like a daughter. Their home has been my sanctuary and I am honoured to be part of their family.

My parents have given me so much, including my name. A male, sporadically hyphenated, excessively long and unpronounceable name. Will I ever meet a new person without having to explain how I came to be named after a river? None of this detracts from the good luck that I had when I was born to parents who prioritised *family* over all else. My father has worked so blisteringly hard, for decades, to provide me with the opportunity to succeed. Through his own actions, he taught me endurance and tenacity and I am grateful for the sacrifices that he made. My mother is the kindest and most devoted person. She is a powerhouse of emotional support and she patiently counselled me through the inevitable lows. Everything that I am, and everything that I have ever achieved, academically and otherwise, I owe to her.

My final vote of thanks is reserved for my partner, Henry. He has been so patient with me throughout my time studying at university and I will never forget it. Henry has empowered me, challenged me and loved me. We have arrived on the cusp of late, in true *Four Weddings and a Funeral* style, to almost everything. Nothing is safe from our collaborative tardiness. But what fun we have when we get there! Falling in love with Henry has changed my life. Together, we have danced until sunrise, cycled up (and down) mountains and attended the most spectacular of cultural and sporting events. Thank you for the journey, Henry. Long may it continue!

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

GENERAL INTRODUCTION

1.1 Prokaryotic cell structure

Prokaryotic cells do not have membrane-bound organelles. For example, instead of a nucleus, prokaryotes have a nucleoid region that contains a circular, double-stranded DNA chromosome. Most cellular processes occur in the cytoplasm. The cytoplasm is bounded by a phospholipid bilayer, which is called the cytoplasmic membrane. Bacteria are divided into Gram-positive and Gram-negative bacteria. Gram-positive bacteria have a thick cell wall of between 20-30 nm of multilayered peptidoglycan. Gram-negative bacteria have a thinner cell wall than Gram-positive bacteria and a single layer of peptidoglycan. Gram-negative bacteria have a periplasmic space that surrounds the cytoplasmic membrane, which is bounded by an outer membrane.

1.2 Gram-negative secretion pathways

Bacteria have evolved several export mechanisms for cell envelope and membrane biogenesis (Figure 1.1). Proteins must be correctly localised to enable cells to function correctly. Approximately a third of proteins do not function within the cytoplasm, where they are synthesised (Driessen and Nouwen, 2008). These proteins can be inserted into membranes, secreted into the periplasm, secreted to the extracellular milieu or injected into other cells. Gram-negative bacteria often use dedicated secretion systems to export virulence proteins or even transport virulence factors directly into the cytoplasm of a target cell. In Gram-negative species, secreted proteins must cross one, two or even three phospholipid membranes before reaching the intended location (Green and Mecsas, 2016).

Specialised protein secretion systems have evolved to transport protein cargo between locations (Green and Mecsas, 2016). Newly synthesised proteins can be inserted into the inner membrane or proteins destined for the periplasm and outer membrane can

be exported through the inner membrane and subsequently inserted into the outer membrane. Some proteins are targeted to the cell surface while others are secreted into the extracellular milieu. The successful secretion of proteins is often required for biofilm formation and other functions such as pathogenicity and nutrient acquisition (Dalbey and Kuhn, 2012). Proteins can be exported directly out of the cell from the cytoplasm by a one-step process or exported by two-step process that involves the export of the protein to the periplasm before its export across the outer membrane. The Sec pathway, which is found in all domains of life, is responsible for the export of the majority of proteins across the inner membrane of bacteria (Pohlschroder *et al.*, 2005; Figure 1.2)

1.2.1 Insertion of proteins into the inner membrane

Substrate proteins of the Sec pathway are translocated through a narrow protein channel, SecYEG, in the cytoplasmic membrane (Osborne *et al.*, 2004). SecYEG, by convention, is known as the Sec translocon. Only unfolded substrate proteins can be translocated through the Sec translocon. YidC, alone or in conjunction with the Sec pathway, facilitates the insertion of some inner membrane proteins (Tsirigotaki *et al.*, 2017). The Sec pathway catalyses the translocation of hydrophilic regions of polypeptide across the membrane and inserts hydrophobic segments into the inner membrane. A subset of proteins is inserted into the inner membrane by YidC (Akopian *et al.*, 2013).

Signal peptidase (Lep) and lactose permease are examples of inner membrane proteins. Lep has two transmembrane segments and a large periplasmic domain at the carboxy-terminus (Dalbey and Kuhn, 2012). Lep is targeted to the inner membrane by the signal recognition particle (SRP) and also requires ATP hydrolysis by SecA, which is unusual (Gierasch, 1989; Dalbey and Kuhn, 2012). Lac permease does not require SecA for insertion into the inner membrane (Dalbey and Kuhn, 2012).

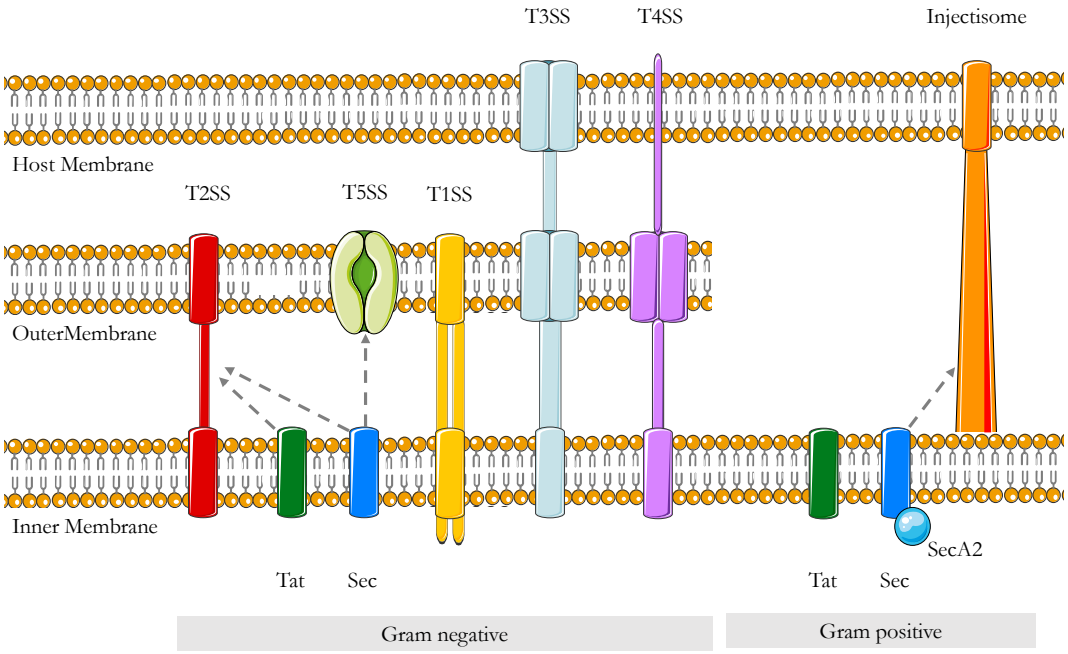


Figure 1.1: Secretion across the bacterial membranes and host membrane

In bacteria, the Sec pathway (blue) mediates the export of most exported proteins. YidC, alone or in complex with SecYEG, facilitates the insertion of inner membrane proteins. The twin-arginine translocation (Tat, dark green) pathway exports a minority of proteins that fold and/or associate with cytoplasmic cofactors in the cytoplasm before crossing the inner membrane. Gram-negative bacteria have specialised export systems. Proteins that function in pathogenesis or nutrient scavenging are transported by the type 1 secretion system (T1SS, yellow). Flagellum proteins are transported by the flagellar type 3 secretion system (T3SS, light blue) and toxins are exported by the pathogenic T3SS. Proteins (e.g. virulence factors) and nucleic acids (e.g. for genetic exchange) are injected into other cells through the type 4 secretion system (T4SS, lilac). Pathogenic effectors are injected into a eukaryotic or bacterial target cell through the type 6 secretion system (T6SS, not drawn). Gram-positive bacteria have three specialised secretion systems. Virulence proteins cross the cytoplasmic membrane through the type 7 secretion system (T7SS, not drawn). Specialised proteins called sortases recognise, cleave and attach proteins that have a carboxy-terminal sorting signal to the membrane. The assembly of pilus structures is dependent on sortases. Virulence factors or genes are injected into other cells by the injectisome (orange). Figure adapted from Green and Mecsas (2016).

Portions of Lep can be inserted into the inner membrane when the Sec pathway is disrupted. Cross-linking studies show that both hydrophobic segments of Lep can be cross-linked to YidC, indicating that Lep inserts into the membrane at a site containing SecYEG and YidC (Samuelson *et al.*, 2000). YidC can also function independently to insert proteins into the inner membrane (Samuelson *et al.*, 2000). This YidC-only pathway is evolutionarily conserved and is also found in mitochondria and chloroplasts but the number of YidC-only substrates is unknown (Dalbey and Kuhn, 2012).

1.2.2 Secretion of proteins into the periplasmic space

Secretion of proteins across the inner membrane into the periplasmic space is dependent on either the Sec or the twin arginine translocation (Tat) pathway. The Sec pathway translocates unfolded proteins through the SecYEG channel in the inner membrane (Driessen and Nouwen, 2008). In contrast to the Sec pathway, the Tat pathway primarily secretes folded proteins (Berks *et al.*, 2005; Figure 1.1), which allows the export of proteins with post-translational modifications, such as redox factors, that form in the cytoplasm (Berks *et al.*, 2005).

In *E. coli*, the TatABC machinery transports folded Tat substrates. It is thought that TatBC is involved in the recognition of the Tat signal peptide while TatA, which forms channel complexes that have variable diameters, can mediate the translocation event (Berks *et al.*, 2005; Dalbey and Kuhn, 2012). One model suggests that the signal sequence binds to the TatBC complex before associating with TatA homooligomer. The substrate is then transported through the pore in the inner membrane that is formed by the TatA homooligomer (Dalbey and Kuhn, 2012).

1.2.3 Secretion of proteins from the periplasm to the outer membrane

There are two pathways by which proteins are sorted and targeted to the outer membrane from the periplasm. The Lol pathway is responsible for the export of lipoproteins, and

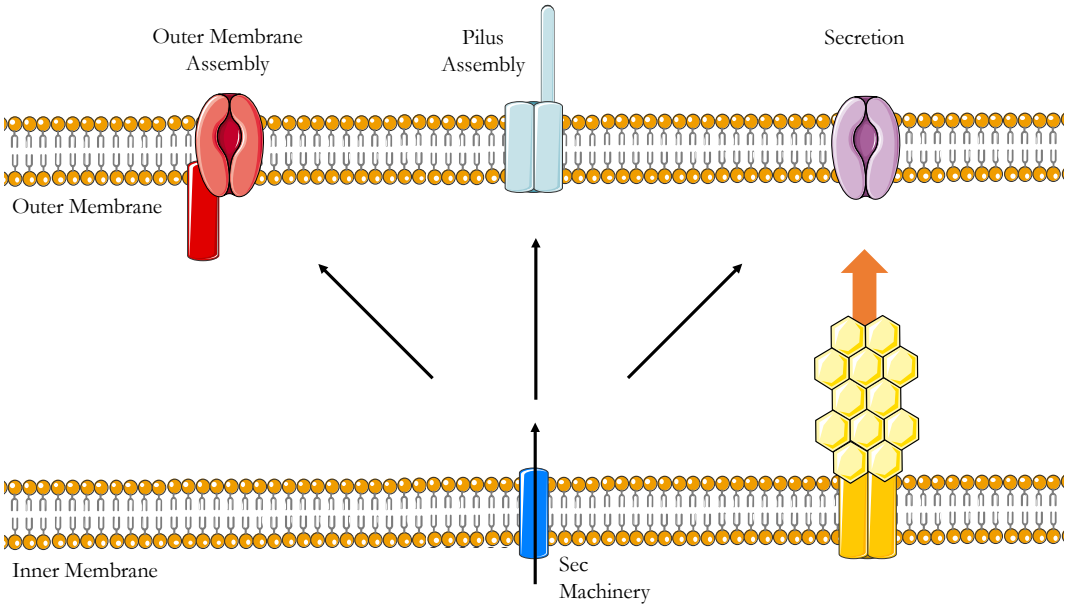


Figure 1.2: The secretion of proteins after translocation by the Sec pathway

The Sec pathway is the primary secretion step for many proteins. Secreted proteins are either inserted into the inner membrane or exported to the periplasm by the Sec machinery, or in some cases by the Tat pathway. After secretion to the periplasm, outer membrane proteins and pili are assembled. In addition, many secreted proteins that have been exported to the periplasm by the Sec machinery are subsequently secreted to the extracellular milieu. Figure adapted from Dalbey and Kuhn (2012).

the β-barrel assembly (Bam) pathway, is responsible for the export of β-barrel proteins.

The Lol pathway is essential for the transport of lipoproteins from the inner to the outer membrane. In *E. coli*, there are five proteins, LolABCDE, that are all essential for viability. Lipoprotein is inserted into the inner membrane by the Sec pathway and modified by diacylglyceride before the cleavage of signal sequence by signal peptidase. ATP hydrolysis by LolCDE releases the lipoprotein from the outer face of the inner membrane. Subsequently, the chaperone, LolA, transports the lipoprotein from the inner to outer membrane. The lipoprotein is transferred from LolA to LolB, the outer membrane receptor, and is then inserted into the outer membrane.

The Bam machinery, which is comprised of BamA and four lipoproteins (BamBCDE) in *E. coli*, can catalyse the outer membrane integration of β-barrel proteins. Outer membrane proteins require periplasmic chaperones to direct them to the outer membrane (Dalbey and Kuhn, 2012; Hussain and Bernstein, 2018). The periplasmic chaperones SurA, Skp, and DegP have all been implicated in the transport of outer membrane proteins to the outer membrane. The structure of the Bam complex has been solved using both X-ray crystallography and cryo-electron microscopy but the mechanism by which it integrates β-barrel proteins into the outer membrane remains poorly understood (Bakelar *et al.*, 2016; Gu *et al.*, 2016; Iadanza *et al.*, 2016; Hussain and Bernstein, 2018). It has been proposed that the rotation of the ring-like structure formed by the periplasmic elements of the Bam complex leads to the opening of an unstable join between the first and last β-strands of the BamA β-barrel, which allows the stepwise insertion of outer membrane proteins into the lipid bilayer (Bakelar *et al.*, 2016; Gu *et al.*, 2016; Iadanza *et al.*, 2016; Hussain and Bernstein, 2018).

1.2.4 Secretion out of the cell

In Gram-negative bacteria, secretion out of the cell can occur via two one-step or two-step secretion. One-step secretion involves export from the cytoplasm into the extracellular milieu or into another cell using a continuous channel that spans the inner membrane, the periplasm, and the outer membrane. Two-step secretion substrates are exported to the periplasm before transportation across the outer membrane. The Sec pathway, or in some instances the Tat pathway, exports two-step secretion proteins to the periplasm (Figures 1.1 and 1.2). The periplasmic intermediate is then exported across the outer membrane in the second step. The two-step systems secrete extracellular proteins that are often involved in pathogenesis (Dalbey and Kuhn, 2012).

Type 1 secretion systems (T1SSs) are one-step secretion systems, which means that substrates are exported across the inner and outer bacterial membranes together (Figure 1.1). T1SSs are comprised of an ABC transporter, a membrane fusion protein and an outer membrane factor. The ABC transporter protein is located in the inner membrane and catalyses the hydrolysis of ATP. The membrane fusion protein crosses the inner membrane and joins the ABC transporter protein to the outer membrane factor of the T1SS. The outer membrane factor forms a pore in the outer membrane. The recognition of substrate proteins by the ABC transporter protein results in the recruitment of the outer membrane factor to the membrane fusion protein, which forms of a channel that spans both the inner and outer membranes (Thomas *et al.*, 2014). Some bacteria have several T1SSs, with each one dedicated to transporting one or a few unfolded substrates (Thomas *et al.*, 2014). T1SSs are generally Sec-independent and proteins that are secreted by this system typically have a signal sequence at the carboxy-terminus that is not cleaved after export.

Type 3 secretions systems (T3SSs) secrete a wide variety of proteinaceous substrates

across both the inner and outer bacterial membranes (Büttner, 2012). Substrates of T3SSs are transported via an "injectisome" from the cytoplasm of the bacterial cell directly to the cytoplasm of a eukaryotic host cell (Figure 1.1). The injectisome is composed of an extracellular needle that is attached to a basal body, which spans the inner membrane, periplasm, and outer membrane of the bacterial cell (Abrusci *et al.*, 2014).

Type 4 secretion systems (T4SSs) secrete substrates into other cells (Figure 1.1). These macromolecular complexes are largely found in Gram-negative bacteria, where they transport protein and DNA across both the inner and outer membranes for insertion into other bacteria or eukaryotic cells (Green and Mecsas, 2016).

Type 2 secretion systems (T2SSs), or the "general secretion pathway", is a two-step secretion pathway (Green and Mecsas, 2016). The first step in this pathway involves the Sec pathway, or in some cases the Tat pathway, to translocate proteins from the cytoplasm to the periplasm (Figures 1.1 and 1.2). The substrates of T2SSs are folded in the periplasm prior to secretion. It is not known how the substrates are recognised in the periplasm but it is unlikely that a linear secretion signal is responsible. The T2SS system has an ATPase, a signal peptidase, and a secretin. ATP hydrolysis provides the energy for export of T2SS substrates across the inner membrane. The formation of a pseudopilus, also stimulated by ATP hydrolysis, in the periplasmic space is thought to push substrates through the outer membrane pore, which is formed by a dodecamer of secretin (Dalbey and Kuhn, 2012).

Type 5 secretion systems (T5SSs) substrates are often called autotransporters because they secrete themselves across the outer membrane (Figure 1.1). Autotransporter proteins are synthesised with an amino-terminal passenger domain, which is secreted across the outer membrane, and a carboxy-terminal domain. After export to the periplasm via the Sec pathway, the carboxy-terminus forms a β-barrel in the outer membrane and the passenger

domain is exposed on the cell surface (Dalbey and Kuhn, 2012; Green and Mecsas, 2016; Leyton *et al.*, 2012).

1.3 The bacterial Sec pathway

The Sec machinery is responsible for the export of unfolded proteins to the periplasm. Most outer membrane lipoproteins, outer membrane β-barrel proteins and surface pili are first translocated across the inner membrane by the Sec system (Dalbey and Kuhn, 2012). All substrate proteins of the Sec pathway are called secretory proteins. Secretory proteins are often synthesised as preproteins, which are also called precursor proteins, that have a mature portion and an amino-terminal signal sequence (Tsirigotaki *et al.*, 2017). The signal sequence is recognised by ribosome-bound components of the Sec machinery as it emerges from the exit channel of the ribosome. The signal sequence is cleaved after translocation and the mature secretory protein folds into its native state.

Most secretory proteins of the Sec pathway are translocated after they have been either fully or mostly synthesised (Randall, 1983; Figure 1.3). This is called posttranslational translocation. The remaining subset of proteins are translocated concurrently with protein synthesis, which is called cotranslational translocation. The translocation of secretory proteins by the cotranslational pathway circumvents one of the major problems for the Sec machinery, which is the maintenance of secretory proteins in a translocation-competent state. The SecYEG pore allows the translocation of unfolded proteins only. Export by the posttranslational translocation pathway is still possible because of several factors that maintain synthesised secretory proteins in a translocation-competent state. Firstly, it is thought that the presence of the amino-terminal signal sequence destabilises the folded protein (Beena *et al.*, 2004), while the reducing environment of the cytoplasm prevents the formation of certain structural features, like disulphide bonds, which form readily in the oxidising environment of the periplasm (Denoncin and Collet, 2013). In

addition, the presence of ribosome-associated or cytoplasmic chaperones maintains pre-proteins in a translocation-competent state. This introduction describes the two mechanisms by which substrates of the Sec pathway are translocated from the cytoplasm.

1.3.1 The SecYEG channel

In bacteria, the channel through which proteins are translocated is known as the SecYEG channel. This channel is a heterotrimer that is comprised of SecY, SecE and SecG (Corey *et al.*, 2016). Bacterial SecY is equivalent to the eukaryotic α-subunit and SecE is equivalent to the eukaryotic γ-subunit. There is high sequence conservation between these bacterial and eukaryotic subunits and both are essential for cell viability. The eukaryotic β-subunit and SecG are not homologous.

In *E. coli*, SecY and SecE are minimally required to support protein translocation (Brundage *et al.*, 1990; Natale *et al.*, 2008). SecY has 10 transmembrane helices. Three cytoplasmic loops of SecY interact with SecA (Mori and Ito, 2006). Studies have suggested that, at the "front " of SecY, there is a lateral opening between two transmembrane helices, which is called the lateral gate. The lateral gate opens to release signal sequences and transmembrane segments from the channel (van den Berg *et al.*, 2004). SecE has 3 TM domains that sit at the "back" of SecY (Kudva *et al.*, 2013). SecE is essential for protein transport because it stabilises SecY and in the absence of SecE, SecY is rapidly degraded by FtsH, a membrane protease (Kihara *et al.*, 1995).

In *E. coli*, the SecG subunit has 2 TM helices. SecG is neither essential for the viability of cells nor protein translocation in *Escherichia coli* (Kudva *et al.*, 2013). SecYEG forms a passive hydrophilic pore through which unfolded secretory proteins can be translocated cotranslationally or posttranslationally (van den Berg *et al.*, 2004; Zimmer *et al.*, 2008; Tsirigotaki *et al.*, 2017). During cotranslational translocation, GTP hydrolysis by the translating ribosome provides the energy for the translocation of emerging secretory

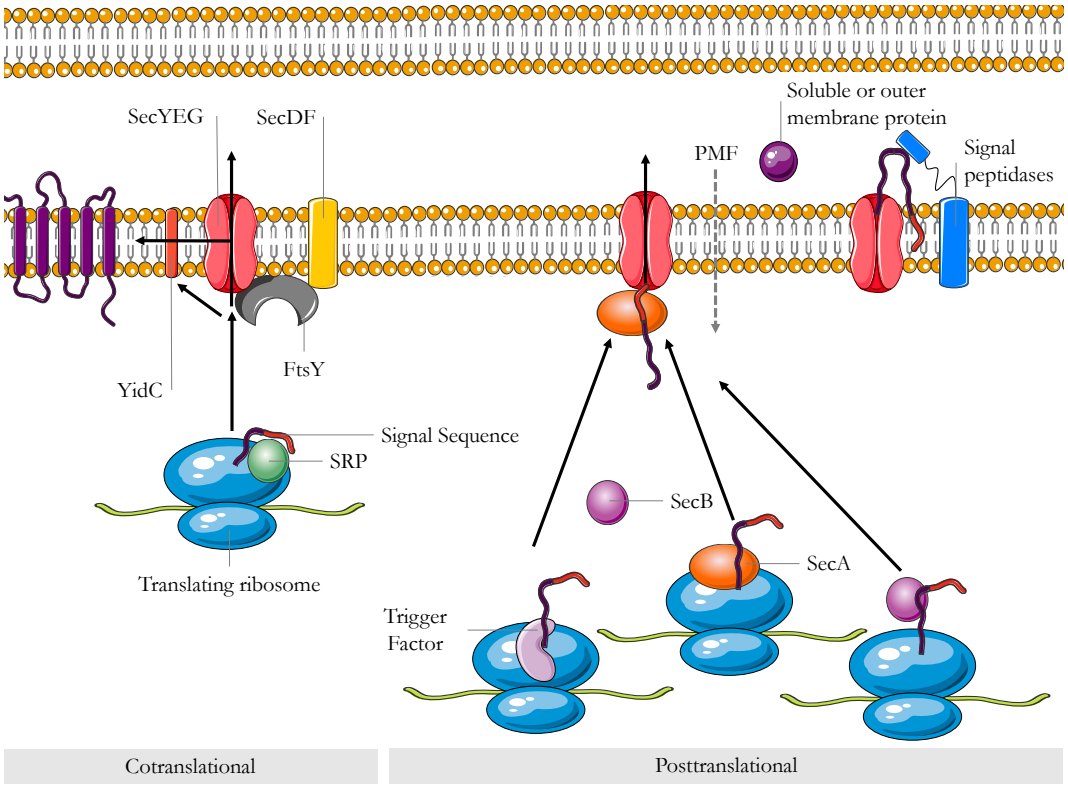


Figure 1.3: The bacterial Sec pathway

Unfolded proteins (dark purple) that contain signal peptides (red) and inner membrane proteins are either cotranslationally recognised and targeted to the transmembrane SecYEG channel (dark pink) by the signal recognition particle (SRP; green) or posttranslationally recognised and targeted by SecA (orange). FtsY (grey) is the membrane receptor of SRP. Trigger factor (lilac) and the ATPase motor SecA bind to preproteins either by docking near the exit channel of the ribosome or in the cytoplasm. SecB (purple) binds to partially or fully synthesised preproteins but does not bind to the ribosome. Preproteins pass through SecYEG and are secreted into the periplasm or into the plasma membrane. The process of posttranslational translocation is powered by repeated cycles of ATP binding and hydrolysis by SecA and the proton motive force (PMF), while cotranslational translocation is powered by the translating ribosome. The auxiliary components SecDF-YajC (yellow) and YidC (dark orange) enhance translocation efficiency. Signal peptidases (bright blue), which are located in the inner membrane, cleave signal peptides.

proteins directly into the SecYEG channel. ATP hydrolysis by SecA and proton motive force provide the energy for the translocation of posttranslationally translocated secretory proteins (Schulze *et al.*, 2014).

The crystal structure of SecYEG from *Methanococcus jannaschii* shows that the SecYEG channel has an hourglass shape with a 5-8 Å ring of aliphatic residues at the narrowest point of the channel, which allows the passage of an unfolded polypeptide (van den Berg *et al.*, 2004). These residues are thought form a hydrophobic seal to prevent the export of proteins with either a secondary or tertiary structure (Park and Rapoport, 2011). The ring is comprised of six isoleucine residues in *E. coli*. The shape of the channel and the lack of charged residues in the cavities on either side of the pore is thought to minimise the contact between the translocating peptide and the SecYEG channel to the hydrophobic ring and increase the efficiency of translocation (van den Berg *et al.*, 2004).

The selectivity of the SecYEG channel is thought to be derived from one of the ten trans-membrane helices that forms a plug domain. The plug locates to the middle of the pore to close the SecYEG channel and maintain the integrity of the cytoplasmic membrane. It is thought that the interaction between the secretory protein and the SecYEG channel displaces the plug domain and opens the channel until translocation is complete, at which point the plug returns to the original, closed conformation (van den Berg *et al.*, 2004). Deletion of the plug does not significantly affect translocation, however the "open" and "closed" states are not as stable without this domain and the permanently open state of SecYEG is toxic in *E. coli* (Harris and Silhavy, 1999; Li *et al.*, 2007; Kudva *et al.*, 2013). Computational modelling indicates that mutations in SecYEG, which are known to allow the export of secretory proteins with defective signal sequences, are likely to destabilise the closed state of the channel. It is possible that the destabilisation of the closed state would facilitate the export of these secretory proteins with defective signal sequences (Smith *et al.*, 2005). Therefore, the SecYEG channel

structure may provide a substrate specificity step in the Sec pathway.

1.3.2 Substrate protein recognition

Secretory proteins are often synthesised as preproteins that have an amino-terminal signal sequence that directs them for export by the Sec pathway (von Heijne, 1985; Figure 1.4). The length of the signal sequence varies but the median length of a signal sequence in *E. coli* is 22 amino acid residues and the minimum length is 15 amino acid residues (Cranford-Smith and Huber, 2018). These signal sequences are cleaved either during or after export (Payne *et al.*, 2012). Signal sequences have a tripartite architecture. The core is comprised of hydrophobic residues that can form an alpha-helical secondary structure. The core is flanked by one or several positively charged, basic residues at the amino terminus and a cleavage site at the carboxy-terminus of the signal sequence. The cleavage site is recognised by the signal peptidase in the periplasm (Figure 1.3).

Most inner membrane proteins contain instructions for their secretion within their hydrophobic, transmembrane domains and do not have a cleavable signal sequence, although some are synthesised with an amino-terminal signal sequence. The subset of inner membrane protein that contain an amino-terminal signal sequence is distinguished from secreted proteins by stop-transfer sequences that anchor the protein in the cytoplasmic membrane (von Heijne, 1994). Stop-transfer sequences are usually 20 amino acid residues long and are hydrophobic (Facey and Kuhn, 2004).

The targeting of secretory substrates to the SecYEG complex in the plasma membrane occurs either cotranslationally or posttranslationally. A greater proportion of preproteins interact with SecA, the motor ATPase, for export. In bacteria, the SRP generally recognises more hydrophobic nascent signal sequences (Lee and Bernstein, 2001; Cranford-Smith and Huber, 2018), which directs the nascent polypeptide for cotrans-

Figure 1.4: General architecture of the Sec signal sequence

The general structure of a signal sequence is conserved in all domains of life (Natale *et al.*, 2008). The signal sequence is located at the amino-terminus of the preprotein and is cleaved from the mature domain after translocation. The median length of a signal sequence in 22 amino acid residues (Cranford-Smith and Huber, 2018). The signal sequence has a tripartite structure that is comprised of a positively charged amino-terminus (N-region, blue), a hydrophobic core (H-region, yellow), and a polar carboxy-terminus (C-region, green) region. The C-region contains the signal peptidase cleavage motif. Apart from the presence of the cleavage motif, Sec signal sequences do not have a conserved amino acid sequence. The tripartite structure of the signal sequence is recognised by the Sec machinery. The H-region typically consists of hydrophobic amino acid residues with a propensity to form an α-helical secondary structure, whereas the positively charged N-region has been implicated in electrostatic interactions with membrane phospholipids. SecA and the signal recognition particle (SRP) recognise the H- and N-regions of the signal sequence.

lational translocation.

1.3.3 Cotranslational translocation

Cotranslational translocation is the translocation of secretory proteins during protein synthesis. The process of cotranslational translocation begins when the emerging secretory protein is recognised by the SRP at the ribosomal exit channel. Subsequently, the ribosome-nascent polypeptide complex (RNC) is delivered to the Sec translocon, which is sometimes supplemented by YidC (Akopian *et al.*, 2013). YidC is a cytoplasmic membrane protein that mediates the insertion of proteins into the cytoplasmic membrane (Samuelson *et al.*, 2000). The SRP delivers the translating ribosome to SecYEG by interacting with the membrane receptor FtsY (Huber *et al.*, 2005; Figure 1.3). FtsY is a peripheral membrane protein that facilitates the insertion of the nascent secretory protein into Sec translocon.

The SRP is a ribonucleoprotein complex that is comprised of Ffh and 4.5S RNA. Ffh is homologous to the 54 kDa, mammalian SRP. Ffh has tripartite domain structure that consists of an α-helical amino-terminal domain, a GTPase domain and a methionine-rich domain. The amino-terminal α-helical domain and the GTPase domain of Ffh both dock at L23 and L29 near the ribosomal exit channel and the methionine-rich domain inserts into the exit channel. The methionine-rich domain is responsible for the recognition of nascent secretory proteins as they are synthesised (Zopf *et al.*, 1990). The emergence of a SRP secretory substrate stabilises the binding of the SRP to the ribosome (Saraogi *et al.*, 2014). Substrate specificity is further enhanced by the recruitment of FtsY.

After the recognition of the nascent secretory proteins by the SRP, the RNC and FtsY form a complex with the SRP (RNC-SRP-FtsY complex) (Shen *et al.*, 2012). FtsY, like Ffh, has an amino terminal domain and a GTPase domain. Ffh and FtsY interact until GTP hydrolysis occurs. GTP hydrolysis occurs when a nascent polypeptide has

only weak affinity for the SRP. As such, GTP hydrolysis causes the dissociation of the RNC-SRP-FtsY complex in the cytoplasm (Shen *et al.*, 2012; Tsirigotaki *et al.*, 2017), which means that only SRP-substrates are targeted to the Sec translocon. The interaction between FtsY, SecYEG and cytoplasmic membrane lipids weakens the affinity of SRP for the RNC. The SRP is released from the RNC by the stimulation of GTP hydrolysis by SecYEG (Saraogi *et al.*, 2014).

Secretory proteins are either translocated across the cytoplasmic membrane or, as is the case for the majority of cotranslationally translocated proteins, they are inserted into the cytoplasmic membrane. YidC is located adjacently to the lateral gate of SecYEG, through which proteins pass for subsequent insertion (Sachelaru *et al.*, 2013; Figure 1.3). Studies have suggested that SecA is recruited to the Sec machinery in the inner membrane to assist with the insertion of membrane proteins (Deiterman *et al.*, 2005; Dalbey and Kuhn, 2012), although the mechanism for this *in vivo* remains unclear because the ribosome and SecA cannot both bind SecYEG *in vitro* (Wu *et al.*, 2011).

1.3.4 Posttranslational translocation

Posttranslational translocation is the translocation of fully or mostly synthesised secretory proteins. The majority of secretory proteins are posttranslationally translocated. In bacteria, all secretory proteins that are exported posttranslationally require SecA, an ATPase, and proton motive force, to provide the energy for efficient translocation (Lill *et al.*, 1989). Historically, it was thought that recognition of these secretory proteins occurred in the cytoplasm, after the protein had been synthesised (Zito and Oliver, 2003; Chatzi *et al.*, 2014).

Subsequent studies indicate that posttranslationally translocated secretory proteins are recognised cotranslationally as they emerge from the ribosomal exit channel by ribosome-bound SecA or Trigger Factor (Oh *et al.*, 2011; Huber *et al.*, 2011; Tsirigotaki *et*

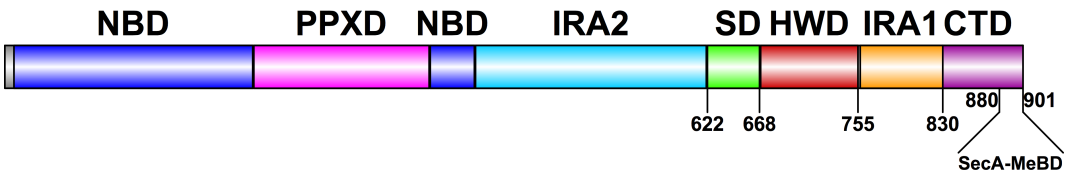
al., 2017). SecA binds L23 near the ribosomal exit channel. Furthermore, SecA interacts with nascent polypeptides earlier in translation than SecB, which is a cytoplasmic chaperone (Huber *et al.*, 2016). The SRP is present in lower concentrations in the cytoplasm than the ribosome, therefore it is likely that substrate recognition by the SRP occurs swiftly to stabilise the interaction between the SRP, the emerging nascent polypeptide and the ribosome (Luirink and Sinning, 2004). The evasion of the SRP by a nascent polypeptide may provide an opportunity for their recognition by ribosome-bound, posttranslational translocation factors, like SecA.

After secretory proteins are synthesised, chaperones, like SecB in *E. coli*, are recruited to the mature portion of the preprotein. SecB functions as a holdase chaperone to maintain the preprotein in an unfolded conformation (Randall *et al.*, 1997; Randall and Hardy, 2002). Subsequently, SecA interacts with the preprotein, which is then translocated through the SecYEG channel (Chatzi *et al.*, 2013, 2014).

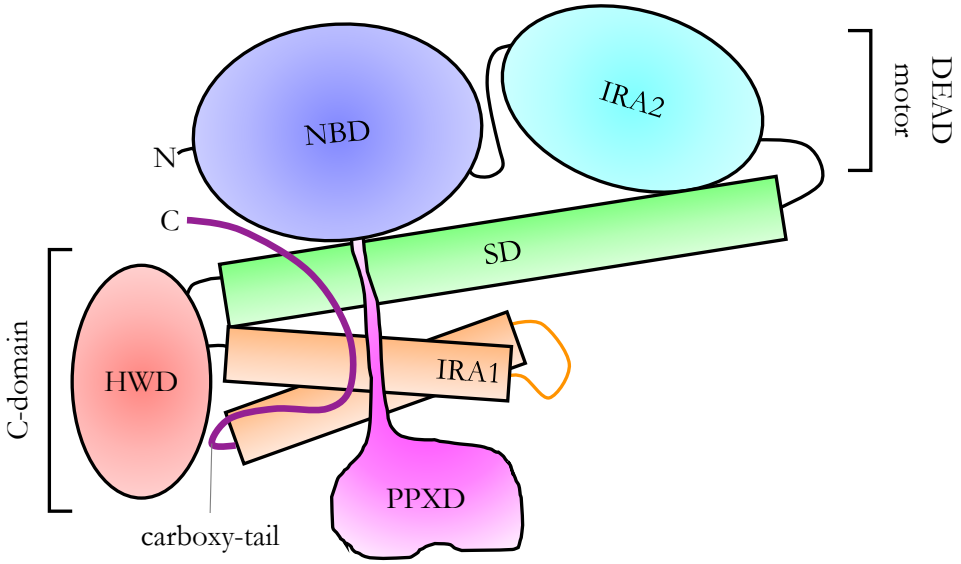
1.3.5 SecA

SecA is a central or "core" component of the Sec pathway in bacteria (Lill *et al.*, 1990). SecA interacts with ATP, SecB, lipids, the ribosome, SecYEG and preproteins (Vrontou and Economou, 2004). SecA cycles between a cytoplasmic form and a peripheral cytoplasmic membrane form (Nakatogawa *et al.*, 2004). The energy for the translocation of preproteins is provided by the hydrolysis of ATP by SecA.

SecA autoregulates the translation of the *secA* mRNA. Immediately upstream of the *secA* gene is the *secM* gene. The genes are co-transcribed as a single mRNA that can form secondary structures (Nakatogawa *et al.*, 2004). The initiation of *secA* translation is prevented through the interaction of SecA with the secondary structure of its own mRNA (Schmidt and Oliver, 1989). SecM is an exported protein and studies have suggested that the defective export of SecM results in the increased expression of *secA*



(a) Domain architecture of SecA



(b) Schematic representation of a monomer of SecA

Figure 1.5: Illustration of the domain organisation of SecA

SecA is comprised of three domains, which are the DEAD motor, the PPXD and the C-domain. 1.5a shows a schematic of the unfolded domain organisation of SecA. The residues indicated in the C-domain refer to the *E. coli* SecA (Papanikou *et al.*, 2007). 1.5 shows a schematic for the tertiary structure of SecA (adapted from Vrontou and Economou, 2004). The DEAD motor contains NDB (blue) and IRA2 (turquoise). The PPXD (magenta) loosely interacts with the helix-turn-helix of IRA1 (orange) in the C-domain. The scaffold domain (green) interacts with the length of the DEAD motor and IRA1. The HWD (red) is flexible. The carboxy-tail (purple) can interact with the PPXD. At the extreme carboxy-terminus of SecA is a 20-amino acid residue domain (SecA-MeBD) that is unresolved crystallographically. Figure 1.5 is adapted from (Vrontou and Economou, 2004).

(Oliver *et al.*, 1998).

The *E. coli* SecA is 102 kDa and has three structural units: the DEAD motor, the polypeptide binding domain (PPXD) and the C-domain (Vrontou and Economou, 2004; Figure 1.5). The amino-terminal DEAD motor is comprised of the nucleotide binding domain (NBD) and the intramolecular regulator of ATPase domain (IRA2), which bind a single molecule of ATP (Ye *et al.*, 2004). The DEAD motor is an ATPase domain that is structurally homologous to the DEAD box helicase superfamily of proteins (Ye *et al.*, 2004). The functional specificity of DEAD box proteins is conferred by the other domains of the protein. ATPase activity of SecA is regulated allosterically by interactions with Mg^{2+} and SecYEG.

The PPXD interacts with the C-domain and binds to secretory proteins. This portion of SecA is sometimes referred to as the substrate specificity domain (Vrontou and Economou, 2004). Studies have suggested that the preprotein binding domain can interact with either the signal sequence or the mature region (Papanikou *et al.*, 2005; Gelis *et al.*, 2007). The C-domain is almost exclusively α-helical and is comprised of four sub-domains. The scaffold domain (SD) is a single α-helix that spans both components of the DEAD motor. The other side of the SD interacts with a second intramolecular regulator of ATPase domain (IRA1). IRA1 is a helix-turn-helix that regulates translocation. The helical wing domain (HWD), which interacts loosely with the rest of SecA, is flexible. During ATP hydrolysis, conformational changes that occur in the DEAD motor are conferred to the HWD (du Plessis *et al.*, 2011). In crystal structures of full-length SecA, the carboxy-tail of SecA is mostly crystallographically unresolved (Vrontou and Economou, 2004). The carboxy-tail of SecA can be further divided into two regions. The more proximal of the two regions can interact with the stem of the PPXD and is known as the flexible linker. The flexible linker could have an autoinhibitory role via the occlusion of the peptide-binding groove (Gelis *et al.*, 2007). The most distal of the

two regions is comprised of 22 amino acid residues at the extreme carboxy-terminus of SecA and contains the only β-sheet secondary structure in the C-domain (Dempsey *et al.*, 2004). This domain is referred to as the SecA-MeBD in this study. Studies have shown that the SecA-MeBD coordinates zinc (Fekkes *et al.*, 1999). Further to this, the SecA-MeBD can bind lipids and SecB (Breukink *et al.*, 1995; Fekkes *et al.*, 1998; Randall *et al.*, 2004).

SecA is dimeric in physiologically relevant conditions *in vitro* and it is thought that SecA forms dimers in the cytoplasm *in vivo* (Woodbury *et al.*, 2012). However, several different, published crystal structures of SecA reveal a number of different dimeric interfaces (Lycklama a Nijeholt and Driessen, 2012). The dimerisation of SecA is affected by temperature, salt concentration, lipid-binding or ligand-binding (Or *et al.*, 2002; Woodbury *et al.*, 2012). The conditions required to form SecA crystals for X-ray crystallography could result in aberrant dimerisation by SecA. Studies of soluble SecA indicate that SecA forms an antiparallel homodimer (Ding *et al.*, 2003; de Keyzer *et al.*, 2005).

The functional role of SecA dimerisation is unclear and remains controversial (Vrontou and Economou, 2004). None of the studies that investigated the state of SecA in response to different conditions included SecYEG (Kusters *et al.*, 2011), therefore a clear understanding of how different components of the Sec machinery influence the state of SecA during translocation remains elusive. Both monomeric and dimeric states of SecA have been suggested to associate with SecYEG. Studies have suggested that SecA is a dimer when actively engaged in preprotein translocation (Ding *et al.*, 2003; Kusters *et al.*, 2011). However, the interaction of SecA with cytoplasmic membranes, substrate protein and SecYEG promotes monomerisation *in vitro*. In addition, a mutant SecA that is preferentially monomeric is functional *in vitro* and *in vivo* (Or *et al.*, 2005) and covalently crosslinked SecA dimers did not have translocation activity *in vitro* (Or and

Rapoport, 2007).

The 22 amino acid residues at the carboxy-terminus of SecA comprise the SecA-MeBD. The SecA-MeBD is present in most bacteria but is dispensable for translocation (Karamanou *et al.*, 2005). The SecA-MeBD is an auto-inhibitor of the peptide binding activity of SecA (Gelis *et al.*, 2007). The *E. coli* SecA-MeBD interacts with the the flat β-sheet of dimeric SecB, although other sites of interaction have been identified (Randall *et al.*, 2004, 2005; Patel *et al.*, 2006). It is thought that the binding of carboxy-terminus of SecB to SecA induces a conformational change in SecA to allow the accommodation of a secretory protein (Patel *et al.*, 2006; Gelis *et al.*, 2007). The interaction between the SecA-MeBD and SecB is stabilised by the presence of zinc (Fekkes *et al.*, 1999), which is an example of how the presence of a metal cofactor can influence the structure and function of a metalloprotein. Four amino acid residues, Cysteine 884 - X_2 - Cysteine 887 - X_8 - Cysteine 896 - Histidine 897 (where X represents any amino acid residue) coordinate the zinc ion, with the assistance of a conserved serine residue at position 889 in *E. coli* (Dempsey *et al.*, 2004). The coordination of zinc by these residues is proposed to fold the SecA-MeBD correctly to allow the interaction with SecB (Dempsey *et al.*, 2004).

1.3.6 The accessory proteins of the Sec pathway

A number of accessory proteins that assist with the translocation of secretory substrates *in vivo* have been identified. The accessory proteins include YajC, YidC, SecD and SecF. YidC is more abundant than SecYEG and it is proposed that a pool of YidC functions independently in the cytoplasmic membrane with one copy of YidC per SecYEG (Sachelaru *et al.*, 2015; Urbanus *et al.*, 2002). Indeed, it has been suggested that YidC enables the translocation of some proteins, without the assistance of the core translocation machinery or ATP *in vitro* (Serek *et al.*, 2004). In *E. coli*, a channel in the cytoplasmic membrane, which is comprised of SecYEG-YidC-SecDF-YajC, is known as the holotranslocon. One copy of each of the seven subunits is present in in

the holotranslocon (Botte *et al.*, 2016). The formation of the holotranslocon *in vitro* stimulates the ATPase activity of SecA and increased membrane insertion efficiency (Schulze *et al.*, 2014; Botte *et al.*, 2016). It is proposed that conformational changes within SecDF couple proton motive force to translocation (Tsukazaki *et al.*, 2013), which is supported by stimulation of translocation through the holotranslocon by proton motive force *in vitro* (Schulze *et al.*, 2014). The ability of the translocation machinery to form multiple different protein channels (YidC alone, dimeric SecYEG, SecYEG, and the holotranslocon) could provide the Sec pathway with the flexibility to respond to specific requirements of different secretory proteins (Botte *et al.*, 2016).

SecD and SecF are found in the cytoplasmic membranes of archaea and bacteria (Eichler, 2003). The two proteins are co-conserved and, in some species, are fused into a single protein (Tseng *et al.*, 1999). Deletion of either *secD* or *secF* results in a cold-sensitive phenotype and the over-expression of both SecD and SecF allows the export of secretory proteins with defective signal sequences (Pogliano and Beckwith, 1994). SecD and SecF stabilise the interaction of SecA with the cytoplasmic membrane, although the effect on translocation is limited (Economou. *et al.*, 1995).

YidC is essential for cell viability and depletion of this protein results in a severe translocation defect (Samuelson *et al.*, 2000). YidC associates with the SecYEG channel (Scotti *et al.*, 2000) at the lateral gate, a position that is displaced by translocating polypeptides, which is proposed to facilitate the insertion of membrane proteins into the lipid phase of the cytoplasmic membrane (Sachelaru *et al.*, 2013). Oxa1p is a homologue of YidC that is found in the mitochondrial membrane, which lacks SecA. Oxa1p is thought to be involved with the insertion of F-ATPase or membrane protein folding. The localisation of YidC at the lateral gate of SecYEG is independent of SecD and SecF.

YajC is an integral membrane protein. YajC interacts with SecD and SecF *in vivo* (Gardel

et al., 1990; Duong and Wickner, 1997). YajC is not essential for cell viability, however overexpression of the *yajC* gene confers ethanol tolerance to *Lactobacillus buchneri* and *E. coli* (Liu *et al.*, 2016). The mechanism by which YajC aids the increased tolerance to ethanol was not investigated, however ethanol stress includes damage to the cell wall and reduced membrane potential.

1.3.7 Cytoplasmic chaperones

Molecular chaperones are proteins that assist with folding and unfolding of other proteins or charging proteins with ligands, for example. A holdase is a specific type of molecular chaperone that prevents the folding of proteins. The posttranslational pathway requires the presence of cytoplasmic chaperones to maintain secretory proteins in a translocation-competent state in the cytoplasm (Randall and Hardy, 2002).

SecB is a cytoplasmic holdase that is responsible for maintaining secretory proteins in a translocation-competent, unfolded state. Early discovery of SecB means that it is the model for Sec chaperones, despite only contributing to the export of approximately 4% of the secreted proteins in *E. coli* (Chatzi *et al.*, 2013; Tsirigotaki *et al.*, 2017). Monomeric SecB is a 17 kDa protein that is found in α-, β- and γ-proteobacteria. Biochemical studies of the *E. coli* SecB and X-ray crystallography of the SecB from *Haemophilus influenzae* both indicate that SecB assembles into a stable dimer of dimers (Topping *et al.*, 2001; Randall and Hardy, 2002). SecB maintains secretory proteins in a translocation-competent, unfolded state. It is suggested that secretory proteins are targeted by SecB to SecA in the cytoplasm. However, as discussed in Section 1.3.5, more recent studies suggest that SecA can cotranslationally interact with nascent secretory proteins before SecB. SecB interacts with nascent proteins after 150 amino acid residues have emerged from the exit channel (Huber *et al.*, 2011). The binding of SecB to an emerging polypeptide is dependent on slow folding rather than the presence of an amino-terminal signal sequence (Bukau *et al.*, 2006). Studies have shown that the signal sequence

decreases the rate of folding in the cytoplasm, which suggests that the signal sequence indirectly contributes to the interaction of SecB with secretory proteins. (Liu *et al.*, 1989).

Studies have suggested that SecB is not a dedicated Sec chaperone (Ullers *et al.*, 2004). The deletion of other chaperones, GroEL, GroES, DnaK or DnaJ caused the overexpression of SecB, which implies that SecB can compensate for the absence of these other chaperones. The deletion of *secB* results in the defective export of some proteins, which can be rescued by the overexpression of DnaK and DnaJ (Wild *et al.*, 1992). Furthermore, SecB prevents the aggregation of insulin B-chains, a non-native substrate, *in vitro* (Pans *et al.*, 2000). The role of the heatshock chaperones, DnaK and DnaJ, in the export of secretory proteins is not specific to proteins that are dependent on SecB for export. For example, deletion of *dnaK* and *dnaJ* results in aberrant export of PhoA (Wild *et al.*, 1992).

Trigger Factor is a chaperone that is proposed to exist as a dimer but binding studies indicate that Trigger Factor binds the ribosome as a monomer near the ribosomal exit channel (Kaiser *et al.*, 2006; Schlünzen *et al.*, 2005). Trigger Factor interacts with substrates with a lower affinity than SecB does, however the location of Trigger Factor at the exit channel means that it is likely to be the first chaperone encountered by emerging polypeptides (Kramer *et al.*, 2002). *E. coli* can tolerate the absence of Trigger Factor. However, the deletion of both *tig*, which encodes Trigger Factor, and *dnaK* genes results in a temperature-sensitive pheonotype, which is the result of protein aggregation (Ullers *et al.*, 2004). It has been proposed that the two chaperones cooperate to facilitate *de novo* folding of proteins *in vivo* (Bukau *et al.*, 2000).

SecA has been predominantly studied as a motor protein that cycles between the cytoplasm and the SecYEG channel in the cytoplasmic membrane. Recently, it has been shown that SecA binds to the ribosome close to the exit channel and interacts with nascent

polypeptides independently of SecB and Trigger Factor (Huber *et al.*, 2011, 2016). A subset of the nascent polypeptides that interact with SecA also interact with the SRP. SecA is implicated in the insertion of a subset of membrane proteins. These results lend support to the hypothesis that a complex network of chaperones recognise, hold and target proteins for translocation (Wild *et al.*, 1992; Randall and Hardy, 2002; Huber *et al.*, 2016).

1.4 Metalloproteins

Up to one half of proteins contain metal cofactors (Bartinkas and Gitlin, 2001; Waldron, 2009). The generic term for a protein that contains a metal cofactor is a metalloprotein. The role of a metal cofactor can be solely structural or it can drive catalysis as well (Waldron, 2009). During the purification of metalloproteins, the metal cofactor can disassociate from the protein, even under non-denaturing conditions, which results in heterogenous populations of metal cofactors (Bollinger Jr., 2010; Cotruvo Jr. and Stubbe, 2012). This illustrates the difficulty for cells in charging the proteins with the correct metals, and for researchers in the characterisation of these metalloproteins. The coordination of a different metal cofactor can result in the mischaracterisation of the function of that protein or enzyme (Waldron, 2009).

The Irving-Williams series refers to the relative stabilities of complexes formed by a metal ion (Irving and Williams, 1953; Waldron, 2009). The stability of high-spin complexes of the divalent ions of first-row transition metals follows the order:

$$Mn^{2+} < Fe^{2+} < Co^{2+} < Ni^{2+} < Cu^{2+} > Zn^{2+}$$

The Irving-Williams series is independent of the ligand. The ionic radii decreases from Mn^{2+} to Cu^{2+} because the additional positive charge is inadequately screened by the additional electron. The discrepancy with copper and zinc occurs because six-coordinate copper complexes are thought to be distorted but more stable than zinc.

Bacteria have evolved a network of pathways and mechanisms to facilitate the acquisition of the correct metal cofactor by the protein. Metal ion chaperones, which are known as metallochaperones, can deliver the metal directly to the metalloprotein. In *E. coli*, for example, SlyD charges hydrogenase enzymes with nickel (Zhang *et al.*, 2005), while IscA coordinates free Fe^{2+} in the cytoplasm for the biogenesis of iron-sulphur clusters (Ding and Clark, 2004). However, cognate metallochaperones have not been identified for the majority of metalloproteins and it is generally considered that most metalloproteins acquire the metal cofactor from the cytoplasmic pool (Tottey *et al.*, 2008; Waldron, 2009). For example, the comparison of MncA and CucA, which can both bind Mn^{2+} and Cu^{2+}, in cyanobacteria suggested that the export of MncA as a quaternary structure by the Tat pathway allows the protein to bind Mn^{2+} in the cytoplasm prior to export. Once MncA is bound to Mn^{2+} exchange with Cu^{2+} was not observed. The concentrations of free ions in the cytoplasm is tightly regulated to limit the contamination of metal-binding sites with the incorrect metal cofactor.

1.5 YecA, a protein with sequence homology to SecA

The current literature suggests that SecA and SecB interact via the SecA-MeBD. However, the SecA-MeBD is not always present in organisms that contain SecB. It was hypothesised that identifying and characterising homologues of the SecA-MeBD may provide insight into the role of this domain *in vivo* (Dr. D. Huber, personal communication). YecA is a protein of unknown function that was identified by searching for similar regions to the carboxy-terminus of the *E. coli* SecA using `BLASTp` (Altschul *et al.* 1990; Table 1.1). Like SecA, this region of YecA is located at the carboxy-terminus of the protein. The 20-amino acid residue domain at the carboxy-terminus is called the YecA-MeBD in this study. 16 amino acid residues of the 22-amino acid sequence of the SecA-MeBD were identical in the YecA-MeBD.

There are several differences between the two carboxy-terminal domains. Firstly, the SecA-MeBD is two amino acid residues longer at the most distal end than YecA. Secondly, the tyrosine residue at position 893 (Y893) of SecA is replaced with a phenylalanine residue at position 214 (F214) of YecA. Structurally, phenylalanine and tyrosine are similar, although phenylalanine is more hydrophobic. In *Rhodobacter capsulatus*, the tyrosine residue of SecA is replaced with a phenylalanine (Fekkes *et al.*, 1997). The histidine residue at position 897 (H897) of SecA is replaced with a cysteine residue at position 219 (C219) in YecA. Histidines are commonly found in the binding sites of proteins and they can coordinate metal ligands (Betts and Russell, 2003). The *E. coli* SecA coordinates a zinc ion via three cysteine residues and the histidine residue (Dempsey *et al.*, 2004). However, H897 is sometimes functionally replaced with a cysteine residue, as in *Bacillus subtilis* and *Staphylococcus carnosus* (Fekkes *et al.*, 1997). The SecA-MeBD has a conserved serine residue at position 888 in *E. coli* that is also found in the YecA-MeBD. The serine residue has been implicated in zinc coordination by the SecA-MeBD although the interaction is not direct (Dempsey *et al.*, 2004). Therefore, the YecA-MeBD is a predicted zinc-binding motif due to the sequence identity to the SecA-MeBD (The UniProt Consortium, 2016).

YecA is a 25 kDa protein that has, in addition to the YecA-MeBD, a 192-amino acid residue domain called UPF0149 (Figure 1.6). UPF0149 is a domain of unknown function. This domain is found in one other protein in *E. coli*, YgfB, which is proposed to interact with RNA polymerase (Malecki *et al.*, 2014). However no function has been attributed to this interaction. The structure of YgfB is mostly α-helical and is structurally similar to the γ-subunit of DNA polymerase III and DnaK, the heat-shock chaperone (Bermand *et al.*, 2000). YgfB has been found to bind within the *wcaC* operon (Ishihama *et al.*, 2016). The genes expressed from this operon are proposed to be involved with acid and heat tolerance (Mao *et al.*, 2001). The binding site of YgfB overlaps with a H-NS binding site,

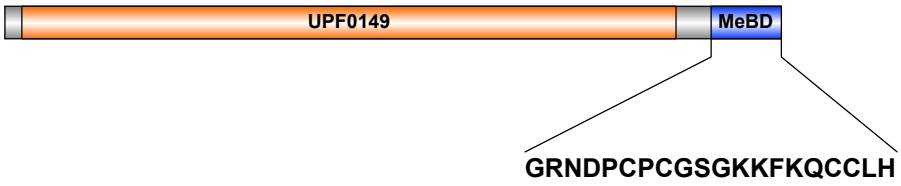


Figure 1.6: The domain structure of YecA

YecA is a 25 kDa protein with two domains. At the amino-terminus is a 192 amino acid residue domain called UPF0149, which is a domain of unknown function. At the carboxy-terminus of YecA is the YecA-MeBD, which is 20-amino acid residues in length. The YecA-MeBD has high sequence similarity to the SecA-MeBD.

Table 1.1: Sequence alignment of the SecA-MeBD and the YecA-MeBD from *Escherichia coli*, strain K12, using BLAST and CLUSTAL O alignment platforms

The YecA-MeBD (residues 201 - 221) is 75% ($\frac{18}{22}$) similar to the SecA-MeBD (880 - 901)

Protein (residues)	Amino acid sequence
SecA (879 - 901)	GRNDPCPCGSGKKYKQCHGRLQ
YecA (201 - 221)	GRNDPCPCGSGKKFKQCCLH
BLAST alignment	GRNDPCPCGSGKK+KQC
CLUSTAL O (1.2.4) multiple sequence alignment	********** :*** :

which could suggest a role for YgfB in antagonising the general gene-silencer H-NS. However, the specific function of YgfB in relation to these interactions remains unclear and YgfB is still classified as a protein of unknown function (The UniProt Consortium, 2016).

Initial studies indicated that YecA could interact with the ribosome *in vitro* (Dr. D. Huber, personal communication) but the function of YecA remains unknown. The sequence homology between the SecA-MeBD and the YecA-MeBD suggested that the two domains had a related function. However, the specific function of the SecA-MeBD is not clear because although the SecA-MeBD interacts with lipids and SecB, a SecA mutant that lacks the most distal 70-amino acid residues is still functional (Karamanou *et al.*, 2005; van Wely *et al.*, 2000). Furthermore, Gram-positive bacteria do not contain a SecB homologue but often contain a SecA that has a SecA-MeBD.

The primary aim of this study was to investigate whether YecA was involved with the translocation of secretory proteins *in vivo* and *in vitro* and to investigate whether the YecA-MeBD was important for the function of YecA. The secondary aim of this study was to investigate the structure of YecA. The final aim of this study was to investigate the cognate metal cofactor of YecA.

Chapter 2

MATERIALS AND METHODS

2.1 Culture media, growth conditions and strains

Lysogeny broth (LB), composed of 10 g/l tryptone, 10 g/l NaCl and 5 g/l yeast extract, was used for most experiments. Overnight cultures were grown in 5 ml LB in 20 ml glass universal tubes. Protein expression cultures were grown in 2 l baffled Erlenmeyer flasks. All other liquid cultures were grown in unbaffled Erlenmeyer flasks. All cultures were incubated in a shaking incubator at 180 rpm.

LB agar was made by adding 1.5% (w/v) agar to LB. Unless otherwise indicated, cultures and plates were grown at 37°C. Where indicated, media was supplemented with 50 μM/ml kanamycin, 200 μM/ml ampicillin or 50 μM/ml ampicillin (chromosomal expression of the resistance cassette).

2.1.1 Dilution plates

Square Petri dishes of 120 mm x 120 mm (Greiner Bio One) were filled with 50 ml LB agar. Overnight cultures were serially diluted 1 in 10 to 10^{-7} in a 96-well plate (Greiner Bio One). From each dilution, 2 μl was pipetted onto the plate with a multichannel pipette.

2.2 Molecular genetics

2.2.1 Preparation of DNA

Plasmid DNA was purified from an overnight culture using the GeneJET Plasmid Miniprep Kit (Thermo Fisher) according to the manufacturer's instructions. For the preparation of genomic DNA, the STRATEC RTP Bacteria DNA Mini kit was used according to the manufacturer's instructions.

2.2.2 Polymerase chain reaction

Primers used in this study are found in Table 2.7. DNA fragments used for cloning and mutagenesis were amplified using Phusion® High-Fidelity DNA Polymerase (New England Biolabs). Unless otherwise specified, reaction mixtures were compiled according to Table 2.1 the thermocycling conditions for QuikChange PCR were as in Table 2.2. PCR products amplified by Phusion® High-Fidelity DNA Polymerase were purified using the GeneJet PCR purification kit (ThermoFisher Scientific).

MyTaq Red Mix (Bioline) was used for colony PCR. A single bacterial colony was picked using a sterile 10 µl pipette tip and resuspended directly in the reaction mixture. The thermocycling conditions included a 10-minute initial denaturation step of 95°C to lyse the cells, and an annealing temperature of 55°C (unless otherwise specified).

2.2.3 Agarose gel electrophoresis

DNA fragments were separated using 1% (w/v) agarose (Bioline) gels in 1X TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA) plus the DNA dye, 0.01% SYBR® Safe (Thermo Fisher). DNA samples were prepared in 6X DNA loading buffer (Bioline) (1 µl loading buffer to 5 µl DNA sample). Samples that were amplified using MyTaq Red mix were not prepared with the loading buffer because the mix included a loading dye. A 1 kb Hyperladder (New England Biolabs) was used as a ladder. Gels were run in 1X TAE buffer at 120 V until the dye front had reached the end. Gels were viewed under UV light (300 nm) in a Gel Doc (Bio-Rad).

2.2.4 Cloning

Purified PCR fragments and plasmids were digested with two restriction enzymes, *BsaI* and *BamHI HF* (Thermo Scientific), in CutSmart® Buffer (New England Biolabs). Digested products were separated by agarose gel electrophoresis and extracted using GeneJet

Table 2.1: QuikChange® mutagenesis reaction mixture

Reagent	Control	Construct
SDW	41 µl	39 µl
5X Phusion HF Buffer	5 µl	5 µl
10 mM dNTPs	2	2
10 µM forward primer	1.25 µl	1.25 µl
10 µM reverse primer	1.25 µl	1.25 µl
Template	1 ng	1 ng
Nuclease free water	to a final volume of 50 µl	
5X Phusion HF Buffer	1 µl	1 µl

Table 2.2: QuikChange® mutagenesis PCR reaction programme

	Cycles	Temperature (°C)	Time (s)
Initial denaturation	1	95	300
Denaturation	20	95	10
Annealing		65	5
Extension		72	300
Final Extension	1	72	120
Hold	-	4	∞

Gel Extraction kit. The vector was treated with Antarctic Phosphatase (New England Biolabs) to prevent vector re-ligation. DNA was quantified using Nanodrop (ThermoFisher Scientific) and the insert was ligated into the vector at a 3:1 ratio with 50 ng of vector DNA, using T4 ligase (New England Biolabs). An overnight temperature gradient to 4°C overnight was used for the ligation. The temperature gradient was achieved by placing the ligation reaction in a 22°C waterbath inside a sealed Styrofoam™ box, which was then incubated at 4°C. Subsequently, the salts were removed from the ligation reaction by incubating ligation reaction with an equal volume 7.5 M NH_4AOc and two volumes of ice-cold isopropanol. The mixture was vortexed and incubated on ice for 20 minutes. The mixture was centrifuged for 15 minutes at 12000 *g* at 4°C. The supernatant was discarded and the pellet was dried for 5 minutes in the Concentrator 5301 (Eppendorf) before 10 µl sterile, deionised water was added. Subsequently, 5 µl of the ligation mixtures were transformed into DH5α *E. coli* (unless otherwise stated) and plated onto LB agar supplemented with the appropriate antibiotics.

2.3 Bacterial transformation

Chemical transformation

An overnight culture of the desired strain was subcultured 1:1000 into 5 ml LB and grown to OD_{600} 0.1 - 0.2. 100 µl of culture was incubated on ice with TSS (5 g 10% Polyethylene glycol, 2.5 ml 5% Dimethyl Sulfoxide, 1 ml 20mM $MgCl_2$ and made up to 50 ml sterile LB) and 3 µl plasmid DNA in a 1.5 ml microcentrifuge tube for 5 minutes. The mixture was heat-shocked for 1 minute in a 42°C ThermoMixer (Eppendorf) and returned to ice for 5 minutes. 1 ml LB was added to the mixture and the cultures were incubated for 1 hour at 37°C (unless otherwise stated) for 1 hour. The cultures were pelleted at 6000 *g* for 5 minutes. The supernatant was decanted and the cell pellet was resuspended in 100 µl LB. The resuspended pellet was plated on LB agar containing the appropriate antibiotic and incubated overnight at 37°C (unless otherwise stated).

Transformation by electroporation

All reagents, microcentrifuge tubes and electroporation cuvettes were kept on ice throughout. All centrifugation steps were done at 6000 *g* for 5 minutes at 4°C in 1.5 ml microcentrifuge tubes unless otherwise stated. 1 ml of an overnight culture of the desired strain was centrifuged. The LB was removed completely and the cell pellet was resuspended in 100 µl ice-cold H_2O. The mixture was centrifuged. This step was repeated three times: twice more with 100 µl ice-cold H_2O and once with 100 µl ice-cold 10% glycerol. The glycerol supernatant was removed and the pellet was resuspended in 50 µl 10% glycerol.

30 µl electrocompetent cells were mixed with 50 ng plasmid DNA in a 1 mm electroporation cuvette (Geneflow). The mixture was electroporated at 1750 V and 1 ml LB was added immediately. The cells were mixed with the LB by pipetting and the mixture was transferred to a 1.5 ml microcentrifuge tube for recovery for 1 hour at 37°C (unless otherwise stated). The cultures were pelleted at 6000 *g* for 5 minutes at room temperature. The supernatant was decanted and the cell pellet was resuspended in 100 µl LB. The resuspended pellet was plated on LB agar containing the appropriate antibiotic and incubated overnight at 37°C (unless otherwise stated).

2.4 Strain construction

2.4.1 P1 transduction

P1 transduction, first described in 1972 by Miller (1972), was used to transduce mutations from the Keio collection (Baba *et al.*, 2006) to new strains of *E. coli*.

To make the P1 lysate, 5 ml LB was inoculated with a single colony of the donor strain using a sterile loop. The cultures were grown at 37°C to OD_{600} 0.1. 50 µl $CaCl_2$ and 2 µl

P1 virus stock was added. The cultures were grown at 37°C in a shaking incubator (180 rpm) until lysis occurred. Using a Pasteur pipette, 5 drops of $CHCl_3$ were added to the culture. 10 µl 10% w/v SDS was also added. The culture was vortexed and left on the bench for 5 minutes to lyse. The top 2 ml of the culture was centrifuged for 10 minutes at 6000 *g*. The top 1 ml of the supernatant was stored at 4°C.

The recipient strain was grown overnight in LB. 200 µl of the overnight culture plus 50 µl $CaCl_2$ was incubated for 20 minutes at 37°C without shaking in the presence of either 0 µl , 1 µl 2 µl, 5 µl or 10 µl of P1 lysate in a sterile 25 ml universal. Subsequently, 0.1 ml 1 M $Na_3C_6H_5O_7$ and 0.5 ml LB were added and the cultures were incubated for 1 hour. The culture was pelleted and the supernatant was discarded. The cell pellet was resuspended in 50 µl LB and plated on LB agar supplemented with the appropriate antibiotic selection and 20mM Na-Citrate. A phage only control was also plated by spotting 20 µl of P1 lysate onto LB agar.

2.4.2 λInCH

To stably integrate a single copy of a gene at the λ-locus, λInCH was used according to a protocol described by Boyd *et al.* (2000). DHB6521 was transformed by electroporation and re-streaked to purify to colonies. An overnight culture was grown at 30°C. The overnight culture was diluted 1:100 in LB supplemented with 2 mM $MgSO_4$ and grown at 30°C (180 rpm) to mid-log phase. The culture was put in a 42°C water-bath for 15 minutes and then in a 37°C incubator until lysis begins. 100 µl chloroform was added to the cultures and vortexed. The cultures were allowed to lyse on the bench at room temperature for 5 minutes. 2 ml of lysate was centrifuged at 6000 *g* for 5 minutes to clarify the lysate. The top 1 ml was aliquoted and stored at 4°C.

An overnight of the recipient strain was grown in LB supplemented with 2 mM $MgSO_4$. In 4 separate microcentrifuge tubes 100 µl overnight culture was mixed with 1 µl, 2 µl, 5

μl, 10 μl and incubated at 30°C for 15 minutes without shaking. 1 ml LB supplemented with 200 mM $Na_3C_6H_5O_7$ was added to each of the mixtures and then incubated at 30°C (180 rpm) for 1 h. The cultures were pelleted at 6000 *g* for 5 minutes and resuspended in 100 μl LB. The resuspended cells were plated on LB agar supplemented with the appropriate antibiotic at the correct concentration. The plates were incubated at 30°C overnight.

To avoid double lysogens, colonies from the lowest dilution plate were selected. The colonies were re-streaked on the appropriate antibiotic at 30°C, kanamycin at 30°C and on the appropriate antibiotic at 42°C. Correct lysogens were resistant to the appropriate antibiotic, sensitive to kanamycin and temperature sensitive.

An overnight culture of the correct lysogen was grown to cure the strain of the λlytic phage. A sterile loop was used to inoculate a LB agar plate supplemented with the appropriate antibiotic in tight, but not over-lapping, oscillations, and incubated overnight at 42°C. Single colonies from the lowest dilution were re-streaked and tested for resistance to the appropriate antibiotic and growth at 42°C.

2.5 Protein analysis

2.5.1 Buffers for protein purification

Buffer 1	10 mM [HEPES]-KOH [pH 7.4], 100 mM KOAc, 10 mM $Mg(OAc)_2$
Lysis buffer	10 mM [HEPES]-KOH [pH 7.4], 100 mM KOAc, 10 mM $Mg(OAc)_2$) plus 10 mg Dnase I (Thermo Scientific) and 1 mM phenylmethylsulfonyl fluoride (PMSF; Thermo Scientific)

Elution buffer	10 mM [HEPES]-KOH [pH 7.4], 100 mM KOAc, 10 mM $Mg(OAc)_2$, 0.5 mM imidazole
High Salt buffer	10 mM [HEPES]-KOH [pH 7.4], 1 M KOAc, 10 mM Mg(OAc)
Protease buffer	10 mM [HEPES]-KOH [pH 7.4], 100 mM KOAc, 10 mM Mg(OAc), 5 mM β-mercaptoethanol

2.5.2 SDS-PAGE analysis

Protein samples were analysed on 15% resolving and 6% stacking gels made in 0.75 mm casts (BioRad). 10 ml of resolving buffer contained 5 ml 30% (w/v) acrylamide (Protogel), 2.6 ml 1.5 M Tris-HCl [pH 8.8], 0.1 ml 10% (w/v) sodium dodecyl sulphate (SDS), 0.1 ml 10% (w/v) ammonium persulfate (APS) and 4 µl tetramethylethylenediamine (TEMED). 10 ml of stacking buffer contained 1.7 ml 30% (w/v) acrylamide (Protogel), 2.5 ml 0.5 M Tris-HCl [pH 6.8], 0.1 ml 10% (w/v) SDS, 0.1 ml 10% (w/v) APS and 1 µl TEMED.

Protein samples were prepared in 2X Laemmli sample buffer (Sigma Aldrich) at a 1:1 ratio and boiled, unless otherwise indicated. Gels were run at 100 V until the loading dye reached the bottom of the gel. Gels were stained with 1.25% (w/v) R-250 Coomassie Brilliant Blue (Thermo Scientific) in 50% ethanol.

2.5.3 Western blotting

Samples for Western blotting were separated by SDS-PAGE. The proteins were transferred to a nitrocellulose membrane using either the iBlot 2 Dry Blotting System (Thermo Fisher) or by using a Western Sandwich cassette with 1x Transfer Buffer (25 mM Tris, 190 mM glycine, 20% ethanol). Membranes were blocked in 5% (w/v) Milk Buffer (per l: 50 g skimmed milk powder (Tesco), 2.42 g Tris base, 8 g NaCl, pH 8.4) for 1 h at 4°C on an orbital shaker. The membrane was washed for 15 minutes at 4°C with TBST (per litre: 1 ml Tween-20, 2.42 g Tris base, 8 g NaCl, pH 8.4) three times. The membrane was then in-

cubated with the primary antibody (diluted in TBST) overnight on the orbital shaker at 4°C. The membrane was washed three times with TBST for 15 minutes and subsequently incubated with HRP-linked α-rabbit antibody (BioRad) for 1 hour and washed a further three times with TBS. The western blot was developed using ECL Prime Western Blotting Detection Reagent (Amersham), and visualised using a Bio-Rad Chemi XR.

2.5.4 Protein expression

For protein production, the desired gene was cloned into either pCA528 or pCA597, which are cloning vectors that encode either His_6-SUMO or Strep(II)-SUMO tag at the respective amino-termini. The plasmids were transformed by electroporation into electrocompetent *E. coli* strain BL21 DE3. Transformants were incubated overnight in LB supplemented with kanamycin and subcultured 1:1000 into 1 l of LB (protein expression for antibody production or biochemical assays) or 1 l of M9 media containing $^{15}NH_4Cl$ (preparation of single-labelled samples for NMR studies) of 1 l of M9 media containing $^{15}NH_4Cl$ and $^{1}H_7$-$^{13}C_6$-glucose (2 g/l) (preparation of double-labelled samples for NMR studies). Cultures were grown to an OD_{600} of 0.8 at 37°C, at which point the temperature was dropped to 18°C. After 30 minutes at 18°C protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) overnight.

2.6 Protein purification

2.6.1 Cell lysis

Protein expression was done as described in Section 2.5.4. IPTG-induced cells were harvested by centrifugation at 16000 *g* for 30 minutes at 4°C. The cell pellet was resuspended in Lysis Buffer (10 mM [HEPES]-KOH [pH 7.4], 100 mM KOAc, 10 mM $Mg(OAc)_2$) plus 10 mg Dnase I (Thermo Scientific) and 1 mM phenylmethylsulfonyl fluoride (PMSF; Thermo Scientific) and lysed in a C3 cell disruptor until lysis was visually apparent. Un-

broken cells and membranes were removed from the lysate by centrifuging the raw lysate at 17400 *g* for 20 minutes at 4°C. The lysate was further clarified by injecting the supernatant over a 0.45 μM filter unit (Millex-HP), where indicated.

2.6.2 Nickel affinity chromatography and cleavage of the affinity tag

Protino Ni-TED Resin

The clarified lysate was incubated for 1 h with 1 g of Protino Ni-TED Resin (Macherey-Nagel) at 4°C for 1 hour on a roller. The resin was washed five times with High Salt Buffer (10 mM [HEPES]-KOH [pH 7.4], 1 M KOAc, 10 mM $Mg(OAc)_2$) for 15 minutes and five times with Buffer 1. The sample was incubated with 100 μl 4.3 μM SUMO protease overnight at 4°C on the Protino Ni-TED Resin in protease buffer (10 mM [HEPES]-KOH [pH 7.4], 100 mM KOAc, 10 mM Mg(OAc), 5 mM β-mercaptoethanol). The sample and the SUMO protease were eluted from the column with Buffer 1 + 0.5 mM imidazole. The flow through, washes and elutions were analysed by SDS-PAGE.

1 ml HisTrap column

The clarified lysate was then bound to a 1 ml HisTrap FT column (GE Healthcare) overnight using a peristaltic pump at 4°C. The clarified lysate was stirred using a magnetic stirrer and stir bar. After binding the column was washed with Buffer 1 (15x column volume) and High Salt Buffer(5x column volume). Protein elutions in 15x 1 ml fractions in elution buffer (Buffer 1 + 500 mM imidazole). The flow through, washes and elutions were analysed by SDS-PAGE.

The sample was dialysed against Buffer 1 + 5 mM β-mercaptoethanol at 4°C overnight to remove the imidazole and incubated with His_6-Ulp1 to cleave the amino-terminal tag using SnakeSkin™ Dialysis Tubing (10K MWCO) (ThermoFischer Scientific). The dialysis was set up at a 1:500 ml sample:buffer ratio. Where indicated, the tag and His_6-Ulp1

was separated from the cleaved protein by a negative nickel affinity purification step: the sample was passed over the column and the flow through was collected. His_6-Ulp1, the cleaved tag and the uncleaved recombinant protein was eluted with Buffer 1 + 500 mM imidazole. The resulting fractions were analysed fractions were analysed by SDS-PAGE.

2.6.3 Anion exchange chromatography

The sample was applied to a 1 ml RESOURCE Q column (GE Healthcare) the sample was purified using a linear gradient of 100 mM KOAc to 1 M KOAc on the ÅKTA pure (GE Healthcare). A flow rate of 0.5 ml/min was used. The elution volume was collected in fractions of 500 µl were analysed by SDS-PAGE. Protein-containing fractions were detected automatically by the ÅKTA pure system using that absorbances at 280 nm and 254 nm. These fractions were analysed by SDS-PAGE and Coomassie staining. The fractions were pooled and concentrated using a Vivaspin 500 concentrator (GE Healthcare, MWCO 10000) indicated.

2.6.4 Size exclusion chromatography

A Superdex 75 10/300 GL (GE Healthcare; S75 10/300) was used as a final purification step. The column was pre-equilibrated in Buffer 1 + 1 mM TCEP. A flow rate of 1 ml/min was used. The elution volume was collected in 5 ml fractions. Final elutions were analysed by SDS-PAGE. Protein-containing fractions were pooled and concentrated using a Vivaspin 500 concentrator (GE Healthcare, MWCO 10000) where indicated. Protein was stored at -80°C after being snap-frozen in liquid nitrogen.

2.6.5 Determination of protein concentration

Protein concentrations were determined spectrophotometrically at 280 nm. The extinction coefficient was determined using the online platform ProtParam (Artimo *et al.*, 2012) as 44710 M^{-1} for YecA and 44460 M^{-1} for YecA201.

2.7 Generation of polyclonal α-YecA

Full-length YecA was purified as described in Section 2.5.4. The protein was sent to Eurogentec for the Anti-protein 87-day package. This package raises antibodies to the protein of interest by injecting 2 rabbits with 200 μg YecA on days 0, 14, 28 and 56. The serum was collected on day 87. Azide was added to the final serum at a final concentration of 0.02% (w/v) to prevent microbial contamination. The final serum was aliquoted and used to detect YecA and YecA201 by Western blot (1:1500 dilution)

2.8 β-galactosidase assay

2.8.1 Data collection

The β-galactosidase activity was determined according to a protocol described by Miller (1972) with a few adjustments. Cultures were grown in M63 media supplemented with 0.2% maltose to an OD_{600} of between 0.4 and 0.5 at 37°C. 1 mM IPTG was added to the cultures. After 2 h the cells were lysed. For cell lysis, 0.5 ml cell culture was added to 0.5 ml of Z-buffer (60mM $Na_2HPO_4{\cdot}7H_2O$, 40mM $NaH_2PO_4{\cdot}H_2O$, 10 mM KCl, 1 mM $MgSO_4{\cdot}7H_2O$, 1 mM β-mercaptoethanol, pH 7.0), 5 μl SDS, 10 μl chloroform and vortexed for one minute. After incubating the cell lysate for 5 minutes in a 28°C waterbath, the assay was initiated by adding 200 μl of 4 mg/ml *o*-nitrophenyl β-D-galactopyranoside to each reaction. When a visible yellow colour was produced the reaction was terminated by the addition of 500 μl of 1 M Na_2CO_3 and the time was recorded. Subsequently, the samples were centrifuged at 13000 *g* for 5 minutes to pellet the cell debris and the top 1 ml of the reaction solution was extracted and A_{420} and A_{550} values were determined using a UV spectrophotometer.

2.8.2 Data analysis

The absorbance data A_{420}, A_{550} and OD_{600} were collated and β-galactosidase specific activities in Miller units (a.u.) were calculated according to the formula:

$$\text{Miller Units} = \frac{1000\ (A_{420} - 1.75 \times A_{550})}{T \times V \times OD_{600}}$$

where T refers to the time between the initiation of the reaction and its termination and V is the volume of cell culture.

Data were analysed using one-way ANOVA and a Tukey *post hoc* test. A p value of <0.05 was considered to be significant.

2.9 Cold sensitivity assay

Strains that did not contain a plasmid were streaked on LB agar. Strains that contained a multi-copy plasmid were streaked on LB agar supplemented with ampicillin. The strains were grown for 24 h at 37°C and re-streaked in duplicate. One plate was incubated at 37°C and the other was grown at 25°C. Suppression was measured after colony formation at 25°C after 36 h. Plates were imaged using a Gel Doc (BioRad) or iPhone 5s (Apple).

2.10 Cold-sensitive suppressor frequency assay

Overnight cultures were grown at 37°C and serially diluted in duplicate, as described above in Section 2.1.1. One plate was incubated at 37°C and the other at 25°C and grown for 36 h. The suppressor frequency was calculated by dividing the number of suppressors at 25°C by the total number of colonies at 37°C.

2.11 Pull-down assay

2.11.1 Protein expression

The desired gene was cloned into pCA597, which has a amino-terminal Strep(II)-SUMO tag, and the plasmid was transformed into *E. coli* strain BL21 DE3 by electroporation. Transformants were streaked on LB supplemented with kanamycin. A single colony was used to inoculate 25 ml LB supplemented with kanamycin in a 250 ml Erlenmeyer flask. Cells were grown to late log phase at 37°C until an OD_{600} of 1.00 was achieved. The flasks were moved to 30°C for 30 minutes before protein expression was induced for 2 hours with 1 mM IPTG.

2.11.2 Protein purification

The induced cells were pelleted at 16000 *g* for 5 minutes at 4°C. The pelleted cells were resuspended in Lysis Buffer (10 mM [HEPES]-KOH [pH 7.4], 100 mM KOAc, 10 mM $Mg(OAc)_2$) plus 10 mg Dnase I (Thermo Scientific) and 1 mM phenylmethylsulfonyl fluoride (PMSF; Thermo Scientific). Lysis was achieved by sonication at 4°C using the Biodisruptor® Plus (diagenode). Cells were sonicated for two sets of 10 x 30 seconds on the highest setting. The cell debris was pelleted at 16000 *g* for 5 minutes at 4°C. 50 µl of Strep(II)-Tactin® Sepharose® 50% suspension (iba) was pipetted into a microcentrifuge tube using a pipette tip with a cut tip and centrifuged at 4000 *g* for 5 minutes. The supernatant was discarded carefully and the resin was resuspended in 1 ml Buffer 1. The wash step was repeated three times. On the third wash, the resin was resuspended in 50 µl Buffer 1. The clarified lysate and the Strep(II)-Tactin® Sepharose® 50% suspension were incubated together in a microcentrifuge tube for 5 minutes at 4°C on an orbital shaker. The resin was pelleted at 4000 *g* at 4°C for 1 minute and the supernatant was discarded. The resin was resuspended in 1 ml Buffer 1 and incubated at 4°C for 1 minute on the orbital shaker. The resin was pelleted at 4000 *g* at 4°C for 1 minute and the supernatant was

discarded. The resin was resuspended in 1 ml High Salt Buffer (10 mM [HEPES]-KOH [pH 7.4], 1 M KOAc, 10 mM $Mg(OAc)_2$) and incubated at 4°C for 1 minute on the orbital shaker. The resin was pelleted at 4000 *g* for 1 minute at the supernatant was discarded. The resin was dried using the Concentrator 5301 (Eppendorf) at 60°C. The dried resin was resuspended in 50 µl 2X Laemmli sample buffer (Sigma Aldrich) and analysed by SDS-PAGE and Coomassie staining.

2.11.3 Mass Spectrometry

Coomassie stained protein samples were excised from the SDS-PAGE gel using a razor blade and stored at -20°C in a microcentrifuge tube that had been sealed with parafilm until analysis. Samples were submitted for analysis by MALDI mass spectrometry (Dr. Jinglei Yu, Functional Genomics, University of Birmingham). Peptides were referenced to *E. coli* K12.

2.11.4 Analysis of mass spectrometry data

The results of the mass spectrometry were analysed manually. Each protein identified by the mass spectrometry was given a score that was the sum of the ion scores of all peptides that were identified, which gave an indication of the prevalence of the protein in the sample analysed. The score was determined by Dr. Jinglei Yu. A minimum score of 40 was applied to each of the biological replicates. If the protein was present in both biological replicates after the initial scoring filter was applied, then the interaction was considered to be positive.

2.12 Citrate synthase aggregation assay

The citrate synthase assay was done as described in Shah *et al.* (2016), with a few differences. Porcine heart citrate synthase (Sigma) at a concentration of 200 nM (monomer concentration) was mixed in 1 ml of assay buffer (10 mM [HEPES]-KOH [pH 7.4], 100

mM KOAc, 10 mM $Mg(OAc)_2$) with the protein of interest at a 1:1 molar ratio in micro-centrifuge tubes. The reaction mixture was incubated in a preheated ThermoMixer® C (Eppendorf) at 50°C. After 30 minutes, the samples were transferred to 1 ml cuvettes and the light scattering was measured at 320 nm. Data were analysed using one-way ANOVA. A p value of <0.05 was considered to be significant.

2.13 ATPase assay

For the ATPase assay, the protocol described by Robson *et al.* (2009b) was used. proOmpA-myc was used as the model substrate. Reactions of 30 μM SecA, 4.6 μM SecYEG-reconstituted proteoliposomes, 20 mM phosphoenolpyruvate, 20 μM NADH, 9-14 Units of Pyruvate Kinase/Lactic Dehydrogenase and 30 nM SecAN95 (monomer concentration) in TKM buffer (20 mM [Tris]-HCl [pH 7.5], 50 mM KCl, 2 mM $MgCl_2$ were incubated at 25°C with 10 μM of the protein of interest. 1mM ATP and 0.7 μM proOmpA-myc were added sequentially to each reaction and mixed by pipetting.

ATP consumption was coupled to NADH depletion using a pyruvate kinase/lactate dehydrogenase regenerating system in the presence of excess phosphoenolpyruvate at 25°C. The absorbance at 340 nm followed using a Lambda 25 spectrophotometer (PerkinElmer Life Sciences). Rates were determined by manually fitting the change in NADH concentration to a straight line, and the slopes (in $\Delta A_{390}.min^{-1}$) divided by the concentration of SecAN95 and the molar extinction coefficient of NADH (6220 M^{-1} at 340 nm) to give (M ATP).(M SecA)$^{-1}$.min^{-1}. Significance was determined using single-factor ANOVA. A p value of <0.05 was considered to be significant.

2.14 Translocation assay

Translocation efficiencies were determined using an *in vitro* translocation assay (Gold *et al.*, 2007; Robson *et al.*, 2009b; Allen *et al.*, 2016). proOmpA-myc was used as the model

substrate. Reactions of 30 μM SecA, 4.6 μM SecYEG-reconstituted proteoliposomes, 20 mM phosphoenolpyruvate, 20 μM NADH, 9-14 Units of Pyruvate Kinase/Lactic Dehydrogenase and 30 nM SecAN95 (monomer concentration) in TKM buffer (20 mM [Tris]-HCl [pH 7.5], 50 mM KCl, 2 mM $MgCl_2$ were incubated at 25°C with 10 μM of the protein of interest. 1mM ATP and 0.7 μM proOmpA-myc were added sequentially to each reaction and mixed by pipetting. After a 30 min *in vitro* translocation reaction at 25°C, all material that had not been translocated was degraded with protease K (Sigma Aldrich), and the translocated material quantified by western blotting against the carboxy-terminus myc-tag of pro-OmpA. Translocation bands were quantified using an Odyssey imaging system (LICOR).

2.15 One-dimensional thin-layer chromatography

To determine the lipid content of the protein, one-dimensional thin-layer chromatography was used according to a protocol described by Isom (2017). Thin-layer chromatography (TLC) plates (Silica gel 60 -Merck Millipore) were used. The plates were 10 cm long and a width that allowed 1 cm between each sample, plus 1 cm at each edge. 5 μl glass microcapillary tubes (Sigma Aldrich) were used to load 10 μl of the sample, control and each of the standards 1 cm from the bottom of the TLC plate. A solvent system of 65:25:4 chloramphenecol:methanol:water was used to separate phospholipids by the polarity of the head group. The TLC plate was placed in the solvent system, with the samples at the bottom, until the solvent front was approximately 1 cm from the top. The TLC plate was dried for 30 minutes. The TLC plate was stained with phosphomolybdic acid and heated with a hair-dryer to visualise the lipid species.

2.16 Circular dichrosism

2.16.1 Data collection

Protein was diluted in 10 mM potassium phosphate buffer, pH 7.2 (Sigma Aldrich). The spectra were acquired using a Jasco 1.5 spectropolarimeter. Data were collected between 200 and 260 nm with a wavelength step of 0.2 nm. After each spectrum, the temperature was increased incrementally from 15.55CC to 87.4CC.

2.16.2 Data analysis

Data were analysed using the online K2D software platform (Andrade *et al.*, 1993).

2.17 UV-vis spectroscopy

Apo-protein was prepared by incubating the purified protein for 1 hour with 0.1 mM EDTA to chelate divalent cations. The EDTA was removed by dialysing the protein into Buffer 1 (10 mM [HEPES]-NaOH [pH 7.4], 100 mM KOAc, 10 mM $Mg(OAc)_2$ using SnakeSkin™ Dialysis Tubing (10K MWCO) (ThermoFischer Scientific) overnight at 4°C. A Cary 50 Bio UV-Vis spectrophotometer and Hellma 10 mm cuvettes were used for scanning UV-vis spectroscopy experiments. The temperature was maintained at 22°C. A buffer only sample was measured in triplicate. The average buffer only UV-vis spectrum was subtracted from the experimental data. All samples were measured in triplicate.

2.18 Inductively coupled plasma optical emission spectrometry

Protein at a final concentration of 50 mg/l in Buffer 1 plus 2% v/v nitric acid (Sigma-Aldrich) was used. The nickel affinity column was cleaned with 10 ml of 0.1 M EDTA

(Thermo Fischer Scientific). Subsequently, nitric acid was added to the EDTA solution to a final concentration of 2% v/v. $ZnSO_4$ (Sigma-Aldrich), $FeSO_4$ (Sigma-Aldrich) and $MnSO_4$ (Sigma-Aldrich) were each prepared at 18 nM, 180 nM and 1.8 μM in sterile, deionised H_2O containing 2% v/v nitric acid.

The samples were submitted to Dr. Maria Thompson (School of Geography, Earth and Environmental Sciences, University of Birmingham) and analysed using a Varian 725-ES ICP-OES instrument.

The peak wavelengths for each element were determined manually. The emission intensity of the standards at the respective wavelength for each element was used to derive the calibration curves. The gradient and y-intercepts of each calibration curve were used to calculate the concentration of each element from its emission intensity at the appropriate wavelength using the formula.

$$\text{Concentration} = \frac{\varepsilon - m}{c} \tag{2.1}$$

where ε was the emission intensity, m was the gradient and c was the y-intercept. The peak wavelength for Fe was 238.204 nm. The peak wavelength was 257.61 nm for Mn and for Zn the peak wavelength was 206.1888 nm.

2.19 Electronparamagnetic resonance (EPR) spectroscopy

$FeCl_3$ (Sigma Aldrich) was analysed in the presence and absence of an equimolar concentration of apo-protein by Dr. Janet Lovett and Anokhi Shah (University of St. Andrews) on a Bruker Elexsys E580 spectrometer with an ER 5106QT-2w cylindrical resonator operating at 34 GHz (Q-band) at 10 K, which was maintained by a cryogen free variable temperature cryostat (from Cryogenic Limited). Quartz tubes with 3 mm outer diameter were used. Both samples were analysed in Buffer 1 containing 30% v/v glycerol. Exper-

iments were done with field sweeps from 1000 G to 17000 G with 4000 points using a $\frac{\pi}{2-\pi}$ Hahn echo sequence where the π pulse length was 32 ns and the time between pulses was 400 ns. The power level was determined by observing the maximum echo. The shot repetition time was set at 100 μs with 50 shots per point, which was sufficient for the iron, though caused some saturation of the Mn^{2+} contaminant peak. The resultant echo-detected field swept profiles were plotted taking account any video gain differences and numbers of averages. The results were normalised to the maximal EPR intensity by Dr. Janet Lovett. Using the formula

$$hv = gB_0\beta, \tag{2.2}$$

where h is Planck's constant, v is the frequency (in this case, at the Q band, 34 GHz), B_0 is the magnetic field and β is the Bohr magneton, the g-factor for the protein-bound iron cofactor was determined.

2.20 Small Angle X-ray scattering (SAXS)

2.20.1 Data collection

SAXS data were collected by Dr. Tim Knowles and Dr. Mohammed Jamshad on the BM28 beamline at ESRF, Grenoble using an in-line S200 10/300 GL size exclusion chromatography column (GE Healthcare) to remove aggregates. The sample was run at a concentration of 10 mg/ml and data were collected under continuous flow every second.

2.20.2 Data analysis

Two-dimensional scattering plots were averaged automatically at the beamline. PRIMUS software was used to analyse the scattering pattern, subtract the buffer background, generate Guinier plots and generate Kratky plots (Konarev *et al.*, 2003). Kratky plots were used to verify that the protein was correctly folded. The slope of the Guinier plot was

used to calculate the radius of gyration (R_g). SAXS MoW2 was used to determine the molecular weight of the protein from the scattering data (Fischer *et al.*, 2010). Standard settings in CORAL (Petoukhov *et al.*, 2012) were used to model the SAXS data onto the known crystal structures of YgfB (PDB ID: 1IZM, Galkin *et al.*, 2004) and the SecA-MeBD (PDB ID: 1TM6, Matousek and Alexandrescu, 2004). At the amino-terminus of 1IZM, 10 dummy glycine residues were modelled. At the carboxy-terminus of 1IZM a flexible linker of 19 dummy glycine residues was modelled to link with 1TM6. The 1IZM structure was fixed and the 1TM6 was free in solution. The spatial step was 5 Å and the angular step was 20°. This was done in collaboration with Dr. Tim Knowles.

2.21 Phyre2 modelling

The FASTA protein sequence was uploaded to the online Phyre2 platform (Konarev *et al.*, 2003). Phyre2 builds a structural model of a protein using the structures of known proteins with sequential homology as a template. Templates with a confidence of >90% were considered to provide accurate models of the core of the protein (2-4Å rmsd from native, true structure) (Konarev *et al.*, 2003). Templates with <90% confidence were considered in the absence of an alternative region. 8 residues were modelled *ab initio* by Phyre2.

2.22 X-ray crystallography

After gel filtration, YecA was concentrated using a Vivaspin 500 concentrator (GE Healthcare, MWCO 10000), to 20 mg/ml. The following 96 condition screens were used to test for crystal formation: JCSG+, ProPlex, Morpheus HT-96, Midas HT-96 (Molecular Dimensions). Sitting drops were set up at a 300:300 nl protein:condition ratio manually, with a 30 µl reservoir in each condition. The crystal trays were checked at regular intervals for crystal formation.

2.23 NMR Spectroscopy

2.23.1 $^{1}H^{15}N$-HSQC NMR spectroscopy

$^{1}H^{15}N$-labelled apo-protein was prepared by treating with 2 mM EDTA at 4°C, as indicated. The EDTA was removed by dialysis using SnakeSkin™ Dialysis Tubing (10K MWCO) (ThermoFischer Scientific) into 20 mM [MES]-NaOH [pH 6.5] 10 mM NaCl. $^{1}H^{15}N$-YecA was analysed at a concentration of 1.1 mM in 20 mM [MES]-NaOH [pH 6.5] 10 mM NaCl. $^{1}H^{15}N$-YecA201 was analysed at a concentration of 0.84 mM in 20 mM [MES]-NaOH [pH 6.5] 10 mM NaCl using a 5 mm Shigemi tube (Shigemi Inc.).

For all experiments the temperature was set to 298 K and a Bruker 900 MHz spectrometer that was equipped with a 4-channel AVANCE III HD console and a 5mm TCI z-PFG cryogenic probe with enhanced 13C, 15N, 1H, 2H sensitivity was used. The spectral width was set to 14 ppm in the ^{1}H direction and 30 ppm in the ^{15}N direction for all experiments. The centre of the spectrum was 4.698 ppm in the ^{1}H direction and 118 in the ^{15}N direction. Prior to Fourier Transformation, the ^{1}H dimension was solvent filtered then both dimensions were apodised with a standard cosine bell function. Each dimension was then zero filled, the ^{1}H dimension to 256 (t1) and the ^{15}N dimension to 2048 (t2) complex points. The spectra were then phased and polynomial baseline corrected in all dimensions to produce the spectra shown.

2.23.2 NMR Spectroscopy for backbone assignments

$^{15}N^{13}C$-labelled protein was prepared in 20 mM [MES]-NaOH [pH 6.0], 10 mM NaCl in a 5 mm Shigemi tube (Shigemi Inc.). YecA was used at a concentration of 0.5 mM using a 5 mm Shigemi tube (Shigemi Inc.). The ^{1}H, ^{15}N, and ^{13}C resonances of the YecA backbone were assigned using triple resonance experiments. BEST TROSY versions of HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, HNCO, HN(CA)CO were used

for backbone assignments (Clubb *et al.*, 1992; Grzesiek and Bax, 1992; Lescop *et al.*, 2007; Kay *et al.*, 1992; Kay and G. Y. Xu, 1992; Salzmann *et al.*, 1998, 1999; Schanda *et al.*, 2006). The spectral width was set to 14 ppm in the ^{1}H direction and 30 ppm in the ^{15}N direction for all experiments. The centre of the spectrum was 4.698 ppm in the ^{1}H direction and 118 in the ^{15}N direction. For the HN(CO)CACB and HNCACB experiments the spectral width was 76 ppm in the ^{13}C direction with 32% and 25% non-uniform sampling with 64 scans per increment, respectively. For the HNCA and the HN(CO)CA experiments the spectral width was 30 ppm and the centre of the spectrum set to 55.9 ppm. Non-uniform sampling was set to 20 and 15% respectively, with 32 scans per increment. The non-uniform sampling for the HNCA was 15% and for the HN(CO)CA the non-uniform sampling was 20% with 32 scans per experiment. HN(CA)CO and HNCO 16 ppm 176.2 centre. The HN(CO)CACB experiment had 25% non-uniform sampling. Non-uniform sampling was used for all backbone experiments using a Bruker 900 MHz spectrometer that was equipped with a 4-channel AVANCE III HD console and a 5mm TCI z-PFG cryogenic probe with enhanced 13C, 15N, 1H, 2H sensitivity. The temperature was set at 298 K for all experiments.

Data were processed using QMDD software and nmrPipe, the spectra were analysed in nmrDraw and Sparky.

Table 2.4: *E. coli* cloning and expression strains

Strain	Genotype and Description	Reference or Source
BL21 DE3	B F- *dcm ompT hsdS*(r_B- m_B-) gal [$malB^+$]$_{K12}$(λ^S)	Lab Stock
	E. coli DE3 T7 express, protein expression strain	
DH5α	F^- endA1 *glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG* Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, *hsdR17*(r_K- m_K+), λ–	Lab Stock
	Transforms with high efficiency	
MG1655	F^-, $lambda^-$, rph-1	Lab Stock
DHB6501	K12 F^-, λ^-, λ^s Δ*lac(MS265), mel-1, nalA2, supF58* (=*suIII, tyrT58*)	Boyd *et al.* (2000)
DHB6521	DHB6501 λInCh1 (Kan^r)	Boyd *et al.* (2000)

Table 2.5: *E. coli* experimental strains

Strain	Genotype and Description	Reference or Source
N48	N48 [MM18 (MC4100 Φ(*malE'-'lacZ*)) *recA1*]	Bassford Jr. *et al.* (1979)
TCS103	DRH509 *yecA*::kan	This study
TCS138	TCS103 att_λ::*yecA*	This study
TCS139	TCS138 att_λ::*yecA201*	This study
DRH901	MG1655 *yecA*::kan	D. Huber
DRH901	MG1655 Δ*yecA*	D. Huber
DRH902	MG1655 *secB::kan*	D. Huber
DRH903	MG1655 Δ*yecA secB::kan*	D. Huber

Table 2.6: Plasmids used in this study

Plasmid	Description	Source
pCA597	Kanamycin-resistance, empty plasmid vector that contains a T7-inducible promoter. Cloned proteins will have a Strep(II)-SUMO tag at the amino-terminus	Andréasson *et al.* (2008)
pCA528	Kanamycin-resistance, empty plasmid vector that contains a T7-inducible promoter. Cloned proteins will have a His_6-SUMO tag at the amino-terminus	Andréasson *et al.* (2008)
pTrc99a	Ampicillin-resistant, bacterial expression vector with inducible *lacI* promoter	Invitrogen
pCS070	pCA528-His_6-SUMO-*yecA*	This study
pCS071	pCA528-His_6-SUMO-*yecA201*	This study
pDH963	pCA597-Strep(II)-SUMO-*yecA*	D. Huber
pCS163	pCA528-His_6-SUMO-*yecA201*	This study

Table 2.7: Primers used in this study

Primer	Sequence (5′to 3′) and Description
YecA_for	`CCA-GTG-GGT-CTC-AGG-TGG-TAT-GAA-AAC-GGG-ACC-GTT-AAA-CG` Forward primer for the amplification of *yecA*
YecA_rev	`CGC-GGA-TCC-TTA-ATG-CAG-GCA-GCA-CTG-CTT-AAA-TTT-C` Reverse primer for the amplification of *yecA*
YecA201_for	`GCA-GAA-AAA-CGG-TAA-GAA-TTC-GAG-CTC` Forward primer to delete the YecA-MeBD by QuikChange® mutagenesis
YecA201_rev	`GAG-CTC-GAA-TTC-TTA-CCG-TTT-CTC-TTC` Reverse primer to delete the YecA-MeBD by QuikChange® mutagenesis
Strep_YecA201_for	`AAG-CAG-AAG-AGA-AAT-AAG-GAT-CCG-AAT-TCG-AGC` Forward primer to delete the YecA-MeBD by QuikChange® mutagenesis
Strep_YecA201_rev	`AAT-TCG-GAT-CCT-TAT-TTC-TCT-TCT-GCT-TTG-ATC-G` Reverse primer to delete the YecA-MeBD by QuikChange® mutagenesis

Chapter 3

THE ROLE OF YECA IN SEC-DEPENDENT PROTEIN EXPORT IN *ESCHERICHIA COLI*

3.1 Introduction

YecA is a protein of unknown function that has a carboxy-terminal domain, the YecA-MeBD. The YecA-MeBD has sequence similarity to the SecA-MeBD. It was hypothesised that these two domains may be functionally homologous. SecA interacts with SecB via the SecA-MeBD (Patel *et al.*, 2006), an interaction that is stabilised when the SecA-MeBD coordinates zinc (Fekkes *et al.*, 1999). The SecA-MeBD and SecB are not fully co-conserved (Fekkes *et al.*, 1997). Therefore, the SecA-MeBD may have other binding partners. Due to the role of SecA in the secretion of proteins, YecA was investigated using techniques that were used to identify and derive the function of Sec proteins.

One genetic strategy for investigating Sec-dependent translocation *in vivo* is to use gene fusions. In the 1970s, gene fusions that were composed of the amino terminus of a secreted protein and the carboxy-terminus of β-galactosidase (LacZ) were used to investigate how proteins were localised to the periplasm of *E. coli* (Beckwith, 2013). Maltose-binding protein, encoded by the *malE* gene, is a periplasmic protein and *malE′-′lacZ* is an example of a gene fusion that was used to investigate how proteins were localised to the correct subcellular compartment. Initially, it was shown that high expression of MalE′-′LacZ was toxic due to the jamming of the cytoplasmic membrane channel (Bassford Jr. *et al.*, 1979). Other fusion proteins were used to identify mutations in the signal sequence that caused the fusion protein to avoid recognition by the Sec machinery and remain in the cytoplasm (Bassford Jr. *et al.*, 1979). β-galactosidase is only active when folded in the cytoplasm and subsequent studies screened for defects in the Sec pathway that yielded increased β-galactosidase activity. Using this method *secA* (Oliver and Beckwith, 1981, 1982), *secB* (Kumamoto and Beckwith, 1983) and *secDF* (Gardel *et al.*, 1987, 1990) were identified.

Genetic interactions, such as synthetic lethality or suppression, can suggest that there is a functional relationship between the gene products. For example, the relationship between *secA* and *secB* was investigated by constructing double mutants (Kumamoto and Beckwith, 1983). The accumulation of secretory proteins in the cytoplasm was more severe in the double mutant strains compared with the single mutant strain. The subsequent discovery of a direct and functional interaction between SecA and SecB underlines the role of genetic interaction studies in understanding biological systems (Hartl *et al.*, 1990).

Biochemical techniques have also been invaluable for characterising the Sec machinery. For example, the *in vitro* translocation system that was described by Lill *et al.* (1989) can directly measure the translocation-coupled hydrolysis of ATP. Initially, radioisotope-labelled ATP was used but the development of a coupled regeneration system allowed spectroscopic measurement of ATP hydrolysis. This system has been used to investigate the mechanism of protein export by investigating the effect of SecB (Wang *et al.*, 2008; Mao *et al.*, 2009), lipids, SecYEG and secretory proteins (Robson *et al.*, 2009b) on the translocation-coupled ATPase activity of SecA.

In this chapter, a combination of *in vivo* assays and biochemical experiments were used to investigate whether YecA is an auxiliary Sec protein. Initially, the effect of YecA on protein translocation was studied using a reporter gene, *malE'-'lacZ*. The genetic interaction between *secB* and *yecA* was then investigated. Further interaction partners of YecA were investigated using a pull-down assay. Subsequently, YecA was characterised using an *in vitro* assay designed to investigate whether YecA was a chaperone, and then an *in vitro* translocation system was used. This chapter describes the first data to suggest that YecA is involved with the translocation of secretory proteins in *E. coli*.

3.2 Results

3.2.1 Construction of isogenic strains to investigate the role of YecA in posttranslational translocation

The *yecA* gene spans the positions 1990954 to 1991619 on the *E. coli* K12 MG1655 genome (NC_000913.3). N48 is a K12 derivative of *E. coli* strain MC4100 that has a chromosomal *malE'-'lacZ* gene fusion. To investigate the function of YecA, a kanamycin resistance cassette was introduced at the *yecA* locus of N48 by P1 transduction, thereby creating strain TCS103. Subsequently, using λInCh, two strains were constructed that encoded *yecA* (denoted TCS138) or *yecA201* (denoted TCS139) at the λ locus. The expression of the *yecA* and *yecA201* genes was under IPTG control. This resulted in four isogenic strains that were verified by whole genome sequencing using Illumina (Microbes NG) and used in the following assay.

3.2.2 The effect of YecA and YecA201 on posttranslational translocation using a β-galactosidase assay

To investigate whether YecA had a role in the Sec pathway, *E. coli* strain MC4100 that contained a gene fusion between the *lacZ* gene, which encoded the cytoplasmic enzyme β-galactosidase (LacZ), and the *malE* gene, which encoded the periplasmic maltose binding protein (MalE, MBP), was used (Bassford Jr. *et al.*, 1979). MBP is secreted to the periplasm by posttranslational translocation. The activity of β-galactosidase can be detected colourimetrically either through the hydrolysis of the substrate 5-bromo-4-chloro-3-indolyl-D-galactopyranoside, which produces an insoluble blue precipitate, or by the hydrolysis of ortho-nitrophenyl-β-galactoside, which is colourless, to ortho-nitrophenol, which is yellow (Matsumara and Rowe, 2005). When LacZ is fused to the carboxy-terminus of MalE, β-galactosidase activity is low because the fusion protein is targeted for export and LacZ is unable to fold (Dwyer *et al.*, 2014). A

mutation that decreases the efficiency of translocation of the fusion protein results in the increased accumulation of MalE′-′LacZ in the cytoplasm and a concurrent increase in β-galactosidase activity compared with the parent strain. This assay was used to determine the effect of YecA and the YecA-MeBD on posttranslational translocation.

In initial experiments, the effect of the presence or absence of YecA on the β-galactosidase activity of the fusion protein MalE′-′LacZ was determined (Figure 3.1). N48 (the parent strain), TCS103 (*yecA::kan*), and TCS138 (*yecA::kan* λ-*yecA*) were grown in M63 minimal medium supplemented with 0.2% maltose to mid-exponential phase and induced with 1 mM IPTG for 2 hours. To investigate the importance of the YecA-MeBD for activity of YecA in relation to the export of the MalE′-′LacZ fusion protein, strain TCS139 (*yecA::kan* λ-*yecA201*) was assayed for β-galactosidase activity under the same conditions. The cultures were harvested, lysed and assayed for β-galactosidase activity. Data from three biological replicates, which were measured on separate days, were collated and normalised to the β-galactosidase activity of TCS103. N48 had the lowest relative β-galactosidase activity at 0.0186 ± 0.0142. The relative β-galactosidase activity of TCS103 was 1.00 ± 0.563 units. The deletion of the *yecA* gene resulted in nearly a 50-fold increase in the relative β-galactosidase activity compared with N48, the parent strain. The difference between the relative β-galactosidase activities of the parent strain and TCS103 indicated that the mutation of *yecA* caused the accumulation of MalE′-′LacZ in the cytoplasm.

The relative β-galactosidase activity of strain TCS138 was 0.522 ± 0.164. The relative β-galactosidase activity of strain TCS139 was 0.116 ± 0.0264. These results suggested that expression of YecA from the λ locus decreased the accumulation of MalE′-′LacZ in the cytoplasm compared with the *yecA* deletion mutant. However, the relative β-galactosidase activity was lower when YecA201 was expressed compared with the full-length protein. When YecA201 was expressed, the β-galactosidase activity was similar to that of the parent

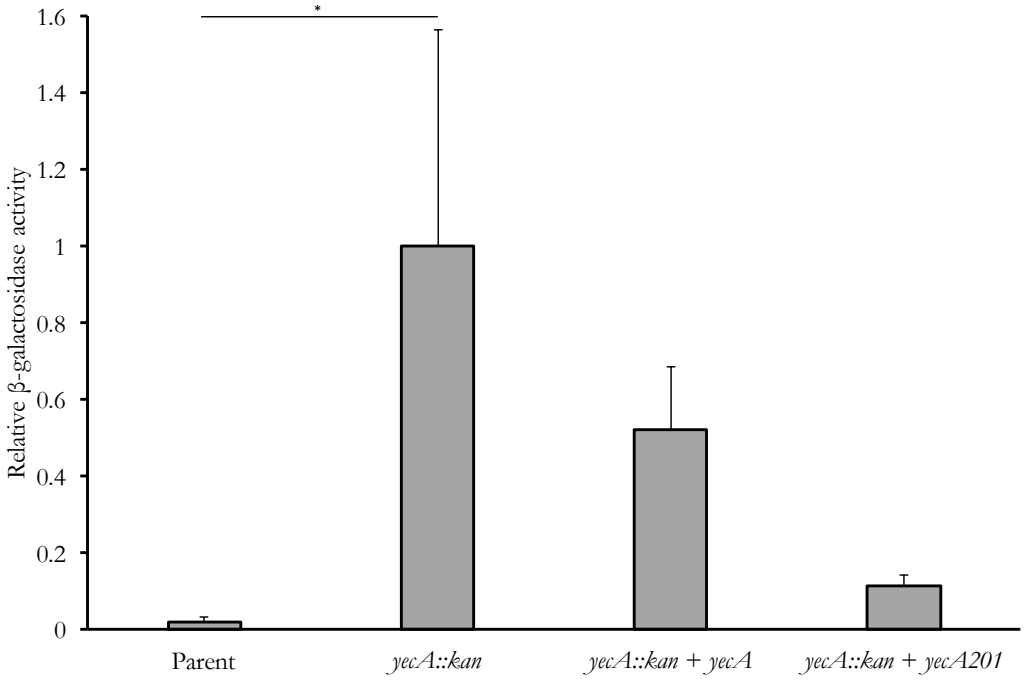


Figure 3.1: β-galactosidase assay to show the effect of YecA and YecA201 on the targeting of MalE′-′LacZ to the cytoplasmic membrane.

Strains were grown in M63 minimal medium supplemented with 0.2% maltose to induce *malE* expression. At mid-log phase the cells were induced with 1 mM IPTG for 2 h. 0.5 ml of culture was added to 0.5 ml Z-Buffer and lysed using 0.1% SDS and chloroform. Subsequently, the β-galactosidase activity was assayed at 28°C by the addition of ortho-Nitrophenyl-β-galactoside to the cell lysate. The hydrolysis of ortho-Nitrophenyl-β-galactoside to ortho-Nitrophenyl-β-galactose results in a colourimetric change from colourless to yellow. When a yellow colour had developed the reaction was stopped. To decrease the contribution of cell debris to the light scattering, the samples were pelleted and the colour change of the supernatant was measured spectrophotometrically at 420 nm. The contribution of cell debris to the light scattering was measured spectrophotometrically at 550 nm. The β-galactosidase activity was calculated using equation 2.8.2. Results are the mean of three independent biological replicates that were undertaken on separate days. Data are expressed as relative to the mean of the *yecA* deletion strain and analysed by one-way ANOVA (p = 0.0375). The Tukey *post hoc* test identifed a significant difference between the parent strain and the *yecA::kan* mutant strain p = 0.0434, indicated by an asterisk) The error bars represent the standard error.

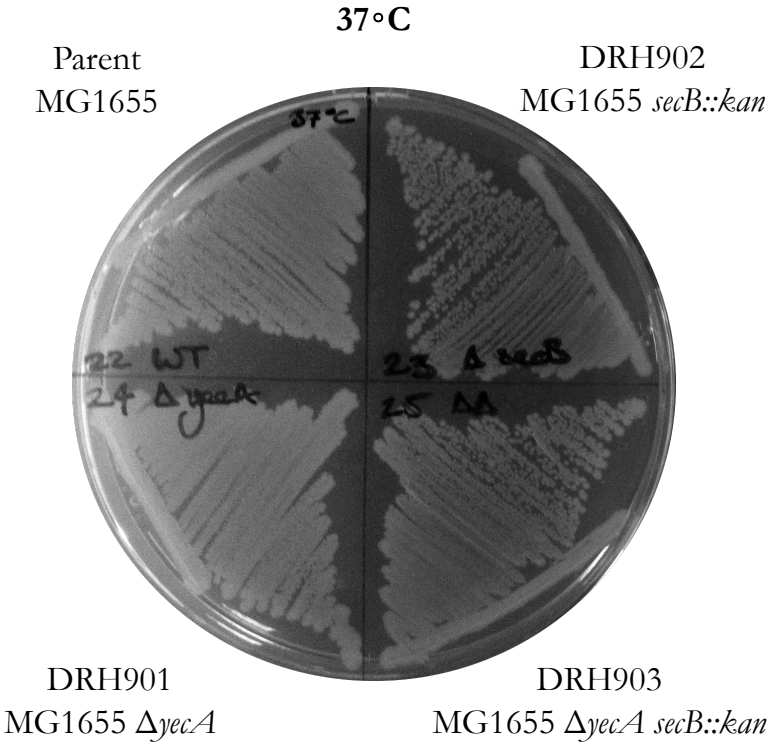
strain, which suggested that the truncated protein was either more active or more stable than the full-length protein *in vivo*.

3.2.3 The genetic interaction between *secB* and *yecA*

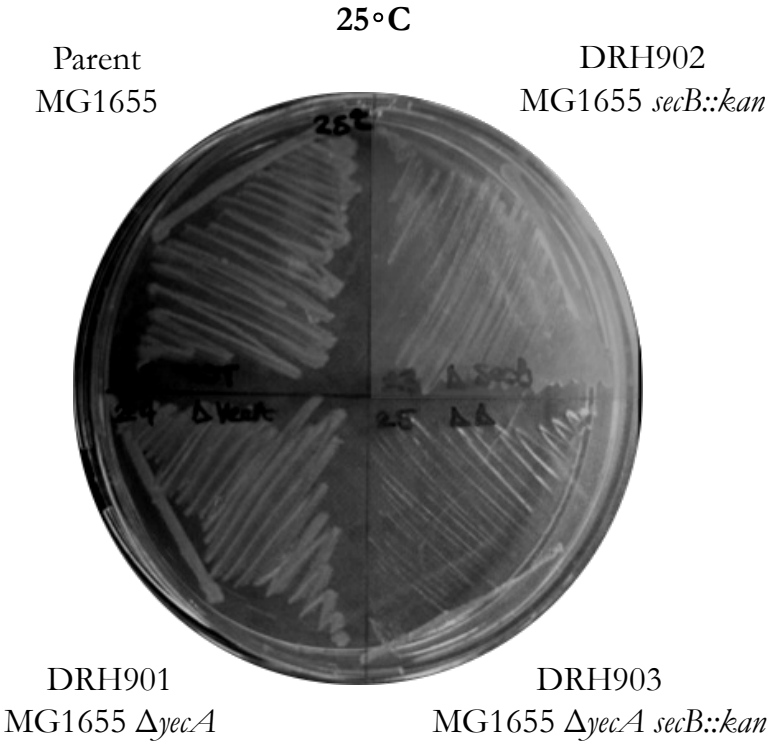
The effect of deleting *yecA* in *E. coli* strain MG1655 was investigated. Defects in protein secretion have been linked to a cold sensitive (CS) phenotype (Pogliano and Beckwith, 1993). For example, a *secB::kan* mutant is mildly CS (Ullers *et al.*, 2007). SecB is a chaperone that maintains preproteins in an unfolded state in the cytoplasm for post-translational translocation. MalE is an example of a protein that is SecB-dependent for export to the periplasm. The result of the β-galactosidase assay described in Section 3.2.2 indicated that YecA had an effect on the targeting and export of MalE to the periplasm. Therefore, to investigate whether the deletion of *yecA* resulted in a similar CS phenotype, the growth of *E. coli* strains MG1655, DRH901 (MG1655 Δ*yecA*), DRH902 (MG1655 *secB::kan*) and DRH903 (MG1655 Δ*yecA secB::kan*) were compared at 37°C and 25°C (Figure 3.2). The growth of DRH901 at 37°C was not different from MG1655 or DRH902 (Figure 3.2a). DRH902, the *secB::kan* mutant, had in a mild growth defect at 25°C compared with the parent strain, MG1655 (Figure 3.2b). However, DRH901 grew similarly to MG1655 at 25°C (Figure 3.2b). Therefore the deletion of *yecA* did not result in a CS phenotype.

DRH903, the double deletion strain, was used to investigate whether the deletion of *yecA* had an effect on the growth defect of DRH902, the *secB::kan* mutant strain. Growth of DRH903 at 37°C was similar to the parent strain (MG1655), DRH901 and DRH902. At 25°C, growth of DRH903 was severely CS compared with DRH902, the *secB::kan* mutant. After 24 h, there was no observed growth of DRH903 (Figure 3.2b).

The growth of DRH903 resulted in the selection of suppressor mutations at 25°C. To determine the frequency of suppressor mutations, overnight cultures of the double



(a) The effect of deleting *yecA* on growth at 37°C



(b) The effect of deleting *yecA* on growth at 25°C

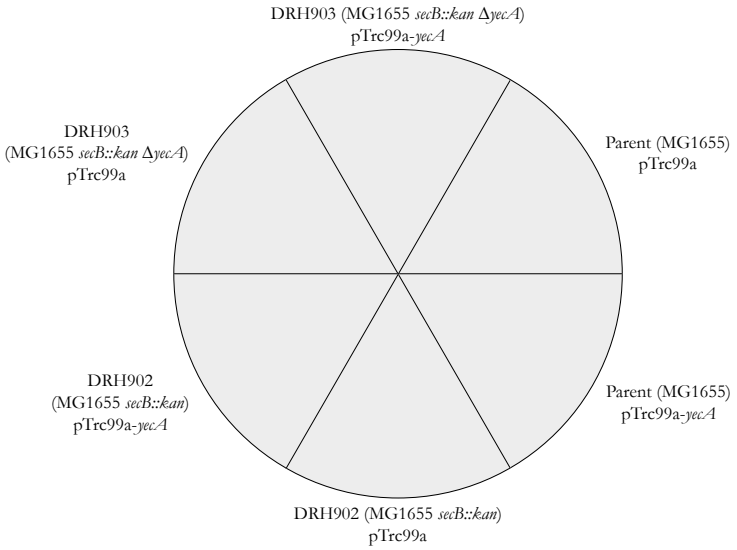
Figure 3.2: The effect of deleting *yecA* on growth

Strains were streaked on LB agar at 37°C and grown for 24 h (3.2a) or 25°C and grown for 36 h (3.2b). The figure is representative of six, independent biological replicates.

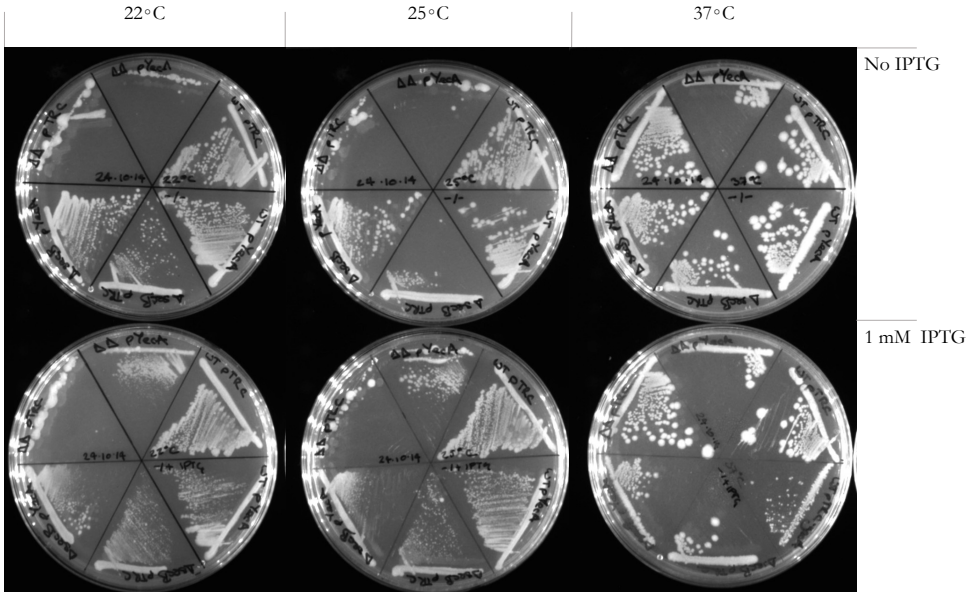
deletion strain were serially diluted to 10^{-7}, plated on LB agar and grown at 25°C and 37°C. The suppressor mutation frequency was determined by dividing the number of mutants with the total number of colonies, which resulted in an apparent suppressor frequency of 10^{-2} ($n = 3$).

To investigate whether expression of *yecA* from a plasmid could complement the mild CS growth defect of the DRH902 and the severe CS phenotype of DRH903, the strains MG1655, DRH902 and DRH903 were transformed with pTrc99a or pTrc99a-*yecA* and grown on LB agar in the presence or absence of 1 mM IPTG at three different temperatures: 37°C, 25°C and 22°C. The plasmid pTrc99a had no effect on the growth of any strain in the presence or absence of IPTG at 37°C, 25°C and 22°C. The growth of MG1655 pTrc99a-*yecA* was unaffected by the presence or absence of IPTG at 25 and 22°C compared with the growth of MG1655 pTrc99a, however a mild growth defect of MG1655 pTrc99a-*yecA* was observed in the presence of IPTG at 37°C compared with the growth of MG1655 pTrc99a (Figure 3.3). This result suggested that the over-expression of *yecA* was mildly toxic at 37°C.

The growth phenotype of DRH902 pTrc99a was the same as DRH902 pTrc99a-*yecA* at each temperature in the presence of IPTG. However, in the absence of IPTG, the growth phenotype of DRH902 pTrc99a-*yecA* strain was similar to that of MG1655 pTrc99a at both 25°C and 22°C. This result indicated that the basal level of YecA expression from the pTrc99a-*yecA* plasmid was sufficient to suppress the mild CS phenotype caused by the deletion of *secB*. The deletion of both *secB* and *yecA* caused a severe CS phenotype that was not suppressed in the absence of IPTG by the basal level of expression of *yecA* from pTrc99a-*yecA*. In the presence of IPTG the production of YecA from pTrc99a-*yecA* suppressed the CS phenotype of the double deletion strain at 25°C. At 22°C, the expression of *yecA* from pTrc99a-*yecA* partially suppressed the CS growth defect of the double deletion strain. These results suggested that there was a genetic interaction between *secB*



(a) Diagram to show the layout of strains



(b) Cold sensitivity assay to show the inducible suppression of the CS phenotype by *yecA*

Figure 3.3: The expression of *yecA* affects the CS phenotype of a *secB::kan* strain

The strains were streaked from a single colony onto LB agar supplemented with ampicillin and grown for for 24 hours in the presence or absence of IPTG at either 22°C, 25°C or 37°C. Plates are representative of three independent experiments. The plate map in 3.3a shows the strain name, the genotype and the plasmid for each segment. 3.3b shows the growth of these strains at 22°C, 25°C or 37°C in the presence or absence of 1 mM IPTG.

and *yecA* and that the suppression of the CS phenotype was affected by the expression level of YecA.

3.2.4 Identification of the interaction partners of YecA and YecA201 using a pull-down assay

The genetic analysis suggested a role for YecA in Sec-dependent translocation. It was reasoned that if YecA was an auxiliary Sec protein, then proteins from Sec machinery, exported proteins or the ribosome would co-purify with YecA. A pull-down assay was used to analyse the protein-protein interactions of YecA *in vivo*.

To investigate the effect of deleting the YecA-MeBD on the protein-protein interactions of YecA, pCA597-*yecA201* was first constructed using primers `Strep_YecA201_for` and `Strep_YecA201_rev` and pCA597-*yecA* (pDH963) as a DNA template. Following the PCR reaction the mixture was incubated with *DpnI* to digest the template DNA. Subsequently, the template DNA was incubated with *BsaHI HF* to linearise the DNA. The linear DNA was analysed by agarose gel electrophoresis. Figure 3.4 compares the linearised template DNA (Lane 1) with the PCR product DNA (Lane 2). There was a small decrease in the molecular weight of the PCR product compared with the template DNA. The difference in size between the two samples is 60 base pairs, therefore the difference was expected to be small. The PCR product was transformed into *E. coli* strain BL21 DE3 by electroporation and the deletion of the YecA-MeBD was confirmed by sequencing. The resulting plasmid was denoted pCS163 and the strain was denoted TCS163.

Three plasmid constructs were used for the pull-down assay (Figure 3.5). pCA597 contained a Strep(II)-SUMO construct. pCA597-*yecA* contained Strep(II)-SUMO-*yecA* and pCA597-yecA201 contained Strep(II)-SUMO-*yecA201*. Each plasmid was transformed into *E. coli* strain BL21 DE3 and grown to mid-log phase and protein

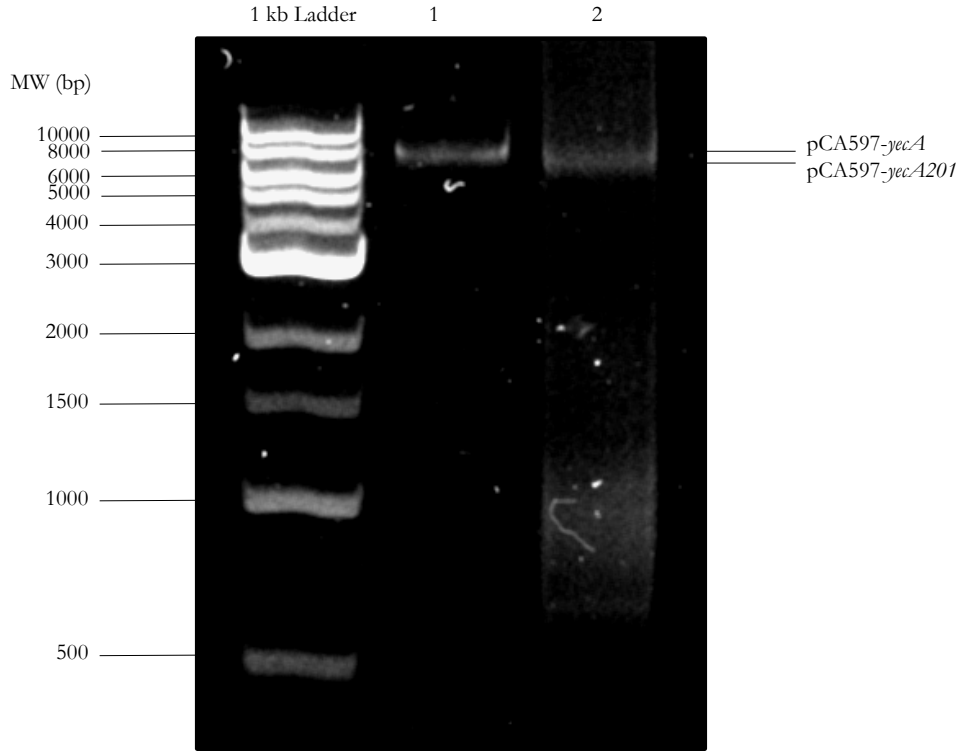


Figure 3.4: Agarose gel electrophoresis of restriction digest products to screen for the construction of pCA528-*yecA201*

Following PCR amplification of pCA597-*yecA* to delete 60 base pairs at the 5′ end of *yecA*, the resulting PCR product was incubated with *DpnI* to digest the template plasmid DNA. The PCR product and the template were digested with *Bam HI HF* analysed by 1% agarose gel electrophoresis. Lane 1 shows the linear template DNA, pCA597-*yecA* and Lane 2 shows the PCR product, which was confirmed by sequencing to be pCA597-*yecA201*.

expression was induced by the addition of 1 mM IPTG for 2 hours. The cells were pelleted and lysed by sonication. The lysate was centrifuged to pellet the cell debris and then the clarified lysate was incubated with Strep(II)-Tactin beads. The beads were washed once with Buffer 1 and washed once with High Salt Buffer to remove proteins that were interacting non-specifically. After the washes, the supernatant was discarded. The beads were dried in a speed vacuum and resuspended in Laemmli buffer. The pull-down samples were analysed by SDS-PAGE and Coomassie staining.

A representative Coomassie gel of the pull-down is presented in Figure 3.6. The control, Strep(II)-SUMO, resolved at approximately 22 kDa. Strep(II)-SUMO-YecA resolved at 45 kDa and Strep(II)-SUMO-YecA201 resolved at 43 kDa. The full-length protein co-purified with several protein bands with masses of between 75 and 180 kDa. YecA201 co-purified with several proteins between 63 and 245 kDa. The brackets in Figure 3.6 indicate the region for each sample that was excised from the gel and analysed by mass spectrometry.

The pull down was repeated six times and the mass spectrometry was repeated twice. The results of the mass spectrometry were analysed manually. Each protein identified by the mass spectrometry was given a score that was the sum of the ion scores of all peptides that were identified, which gave an indication of the prevalence of the protein in the sample analysed. A minimum score of 40 was applied to each of the biological replicates. If the protein was present in both biological replicates after the initial scoring filter was applied, then the interaction was considered to be positive. Using this filter, Strep(II)-SUMO did not have any positive protein-protein interactions. YecA201 interacted with CdsA, an RNA DEAD helicase. The full-length protein interacted with CdsA, EF-G, PNPase and DnaK.

The RNA helicase CdsA (also known as DeaD) co-purified with both YecA and

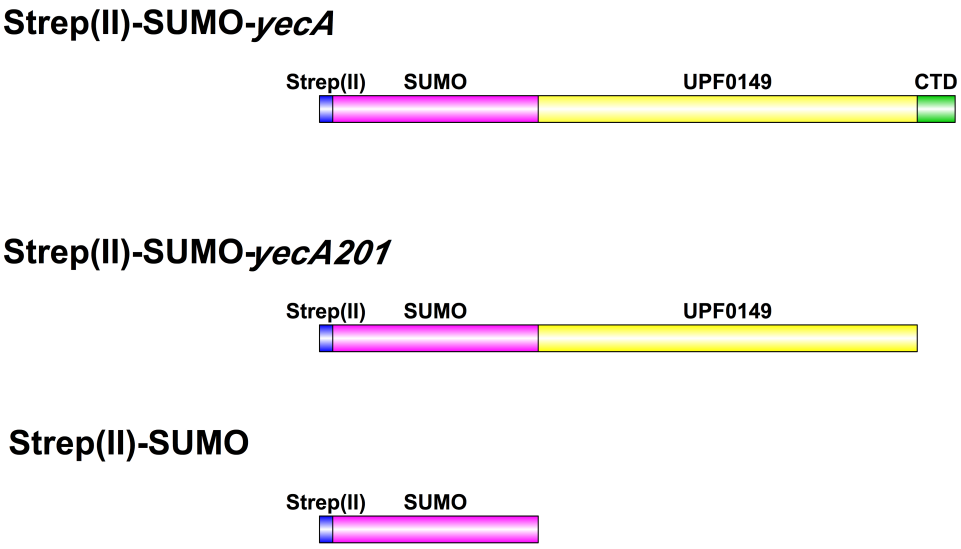


Figure 3.5: Design of the constructs of recombinant proteins that were used in the pull-down assay.

YecA has two domains, UPF0149 and the MeBD. YecA201 lacks only the MeBD. Both YecA and YecA201 were expressed with an amino-terminal Strep(II)-SUMO tag. The Strep(II)-SUMO tag was also expressed as a control. This figure was prepared with the DOG2.0 software (Ren *et al.*, 2009).

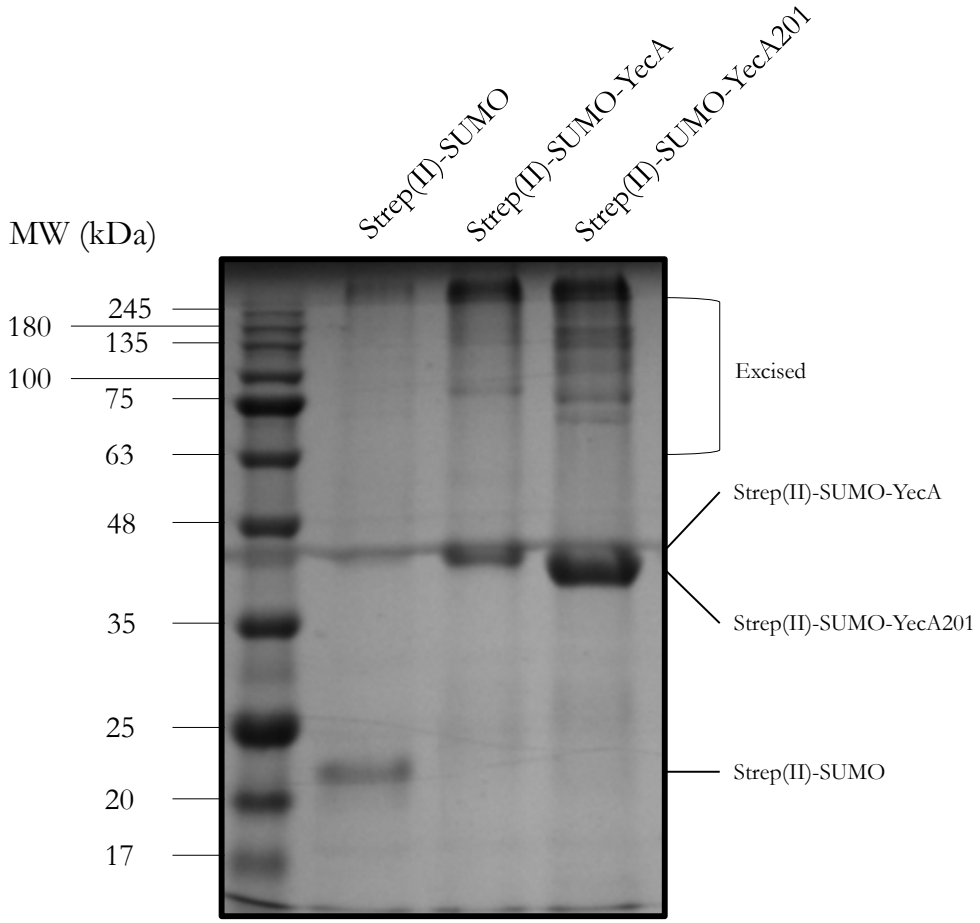


Figure 3.6: Representative SDS-PAGE analysis of the protein-protein interactions of YecA and YecA201

The protein-protein interactions of YecA was analysed using a pull-down assay. Expression of Strep(II)-SUMO, Strep(II)-SUMO-YecA and Strep(II)-SUMO-YecA201 was induced with 1 mM IPTG. Induced cultures were pelleted, lysed and clarified by centrifugation. The clarified lysate was incubated with Strep(II)-Tactin beads for 5 minutes at 4°C. The resin was washed with Buffer 1 and High Salt Buffer, each for 1 minute. The resin was pelleted, dried in a centrifugal evaporator and resuspended in Laemmli sample buffer. 10 µl of each sample was loaded on a 15% SDS-PAGE gel. The gel is representative of four independent experiments. The brackets indicate the excised bands that were analysed by mass spectrometry.

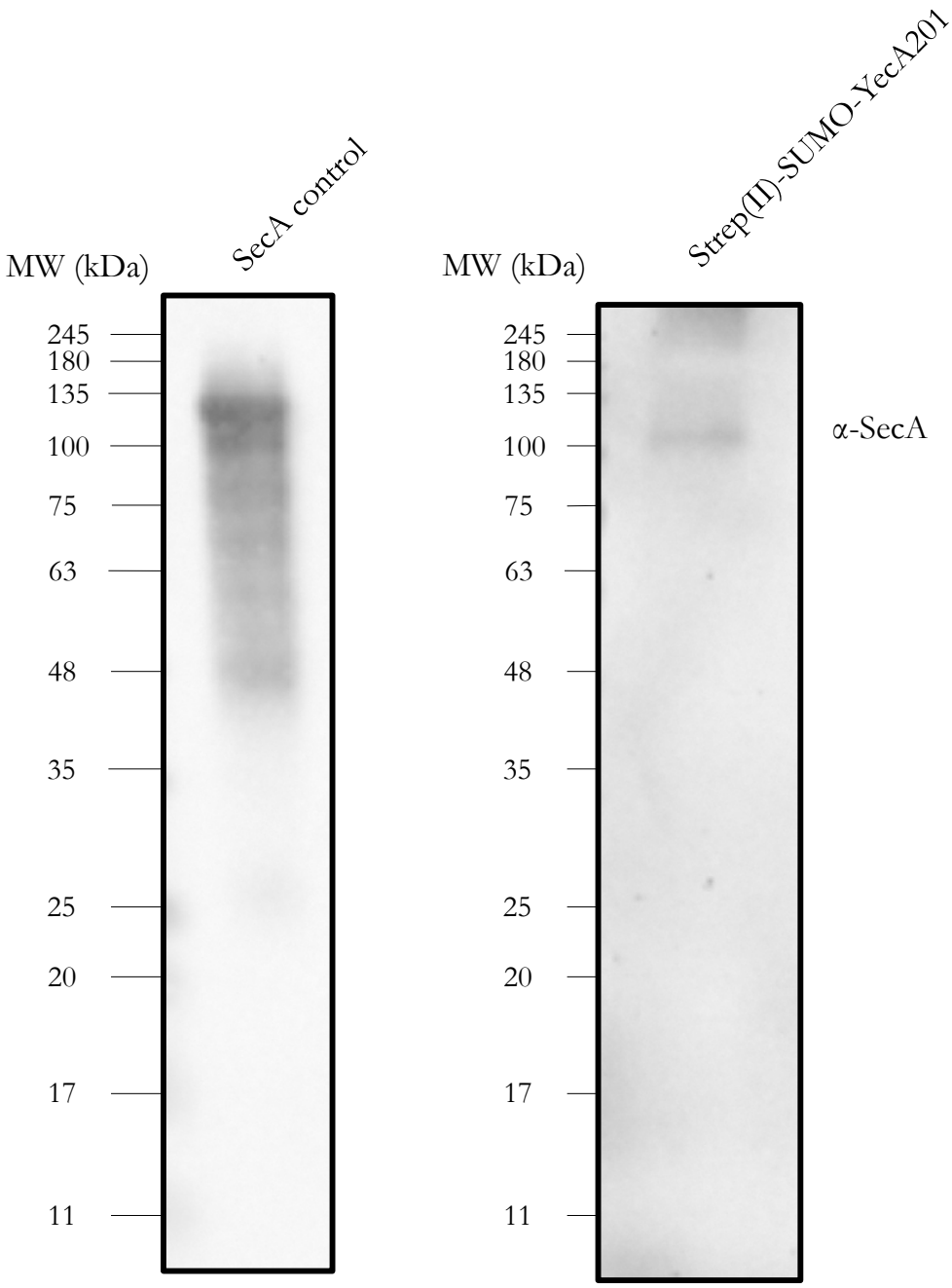


Figure 3.7: Western blot analysis of the interaction of SecA with Strep(II)-SUMO-YecA201

The pull down samples were analysed for the presence of SecA using α-SecA antibody. 2μl of 0.8 mM SecA was used as a control. The lane on the left shows the purified SecA and the lane on the right shows the Strep(II)-SUMO-YecA201 pull down. SecA was resolved at approximately 100 kDa. SecA was not detected in the Strep(II)-SUMO-YecA sample or the Strep(II)-SUMO control sample.

YecA201. The identification of CdsA in this assay was in agreement with other, more stringent purifications that found that CdsA co-purified with YecA201 (Section 3.2.6). YecA also co-purified with PNPase, an exoribonuclease (Briani *et al.*, 2016).

DnaK co-purified with YecA. DnaK is a member of the HSP70 (Heat Shock Protein of 70 kDa) family of ATP-dependent molecular chaperones (Bukau *et al.*, 2006). DnaK is a multifuctional chaperone that, in *E.coli*, assists with protein folding and disaggregation and protein targeting and translocation through biological membranes. A *dnaK* deletion confers a cold-sensitive phenotype. The deletion of both *dnaK* and *tig* causes severe protein aggregation and a severe growth defect that is suppressed by the overexpression of SecB (Ullers *et al.*, 2007). As such, DnaK forms part of the network of chaperones that indirectly enable the translocation of secretory proteins.

EF-G catalyses the movement of mRNA from the A to P to E sites in the ribosome, together with the cognate tRNA molecule via the hydrolysis of GTP to GDP (Slasi *et al.*, 2015) and catalyses the recycling of the ribosome after the termination of translation (Zhang *et al.*, 43). An *in vitro* study found that EF-G can promote the folding of two model substrates, citrate synthase and α-glucosidese, and that the GTPase activity of EF-G stimulates the disaggregation of citrate synthase. EF-G is one of the most abundant proteins in the cell (McLennan *et al.*, 2013).

The mass spectrometry also suggested that YecA and YecA201 interacted with SecA however the score for one of the replicates for both YecA and YecA201 was below the minimum. In addition, SecA was identified in the control using the same medium stringency filter. This interaction was investigated by Western blot, which indicated the presence of SecA in the YecA201 sample, but not in either the control or the YecA sample (Figure 3.7).

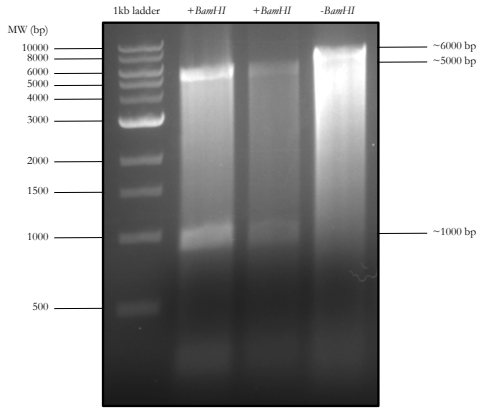
3.2.5 Construction of pCA528-*yecA* and pCA528-*yecA201*

To investigate the function of YecA *in vitro*, a plasmid that contained a gene fusion between His_6-SUMO and *yecA* was constructed. The *yecA* gene sequence was amplified by PCR using `YecA_for` and `YecA_rev primers` primers, purified and digested with *BsaI* and *BamHI HF*. The PCR-amplified *yecA* was ligated into pCA528, which expresses a T7-promoted His_6-SUMO fusion tag, that had been digested with *BsaI* and *BamHI HF*, and transformed into *E. coli* strain DH5α. The resulting colonies were screened before sequencing by digesting with *BamHI HF* (Figure 3.8a). The plasmid construct is designated pCA528-*yecA*. The plasmid was purified and transformed into *E. coli* strain BL21 DE3 pCA528-*yecA* to create strain TCS070.

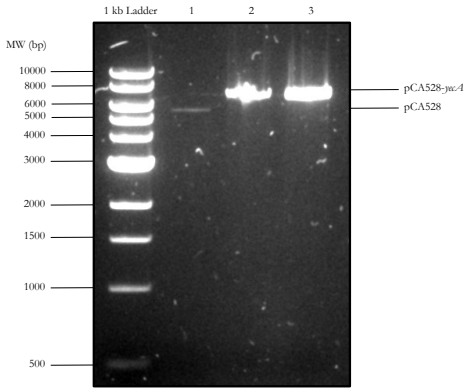
To investigate the role of the YecA-MeBD, a truncated variant of YecA that lacked the distal 20-amino acid residues was constructed using pCA597-*yecA* as a template. Residue 202 was replaced with a stop codon using the primers `YecA201_for` and `YecA201_rev`. A diagnostic restriction digest was used to identify the loss of the *BamHI* site in the construct. Plasmids that had lost the *BamHI* site were sequenced using the primer `T7-FP`. The plasmid was transformed into *E. coli* strain BL21 DE3 by electroporation to create the strain TCS071.

3.2.6 Purification of YecA and YecA201 for *in vitro* assays

To purify YecA and YecA201, the His_6-SUMO-tagged fusion proteins were expressed and purified as described in Section 2.5.4 and 2.6.2. Figures 3.9 and 3.10 show the Coomassie-stained SDS-PAGE gels of the 0.5 ml imidazole elution fractions from the nickel-affinity column. His_6-SUMO-YecA resolved at approximately 37 kDa, which approximately corresponds to the molecular weight of the fusion protein. His_6-SUMO-YecA201 also resolved at 37 kDa. Both His_6-SUMO-YecA and His_6-SUMO-YecA201 co-purified with contaminating proteins. A co-purifying protein that resolved at 75 kDa



(a) Agarose gel electrophoresis of *Bam HI HF*-restriction digested pCA528 and pCA528-*yecA*



(b) Agarose gel electrophoresis of restriction digest products to screen for the construction of pCA528-*yecA201*

Figure 3.8: Agarose gel electrophoresis to show the construction of pCA528-*yecA* and pCA528-*yecA201*

3.8a shows the screen for the construction of pCA528-*yecA201*. `YecA_for` and `YecA_rev primers` were used to amplify *yecA* from the chromosome. The resulting PCR product was purified and digested with *BsaI* and *BamHI HF* at 37°C. The PCR-amplified *yecA* was ligated into pCA528 at 4°C overnight, which expresses a T7-promoted His_6-SUMO fusion tag that had also been digested with *BsaI* and *BamHI HF*, and transformed into *E. coli* strain BL21 DE3 by electroporation and plated on LB agar supplemented with kanamycin. The resulting colonies were screened before sequencing by digesting with *BamHI HF*.3.8b shows the agarose gel electrophoresis of the pCA528-*yecA* following QuikChange® mutagenesis to introduce a stop codon at position 202 of the amino acid sequence of YecA using `YecA201_for` and `YecA201_rev` primers. The resulting PCR product was incubated with *DpnI* to digest the template plasmid DNA and transformed into *E. coli* strain DH5α. The plasmid DNA was isolated and incubated with *Bam HI HF* and *NcoI* restriction digest enzymes. The template DNA in Lane 1 had both restriction sites because two DNA fragments of 1000 base pairs and 5000 base pairs were resolved by agarose gel electrophoresis. The plasmid resolved in Lane 2 also had both restriction sites. Lane 3 suggested the successful insertion of the stop codon because a single band at 6000 base pairs was resolved, which indicated that the *Bam HI HF* site had been lost. This was confirmed by sequencing.

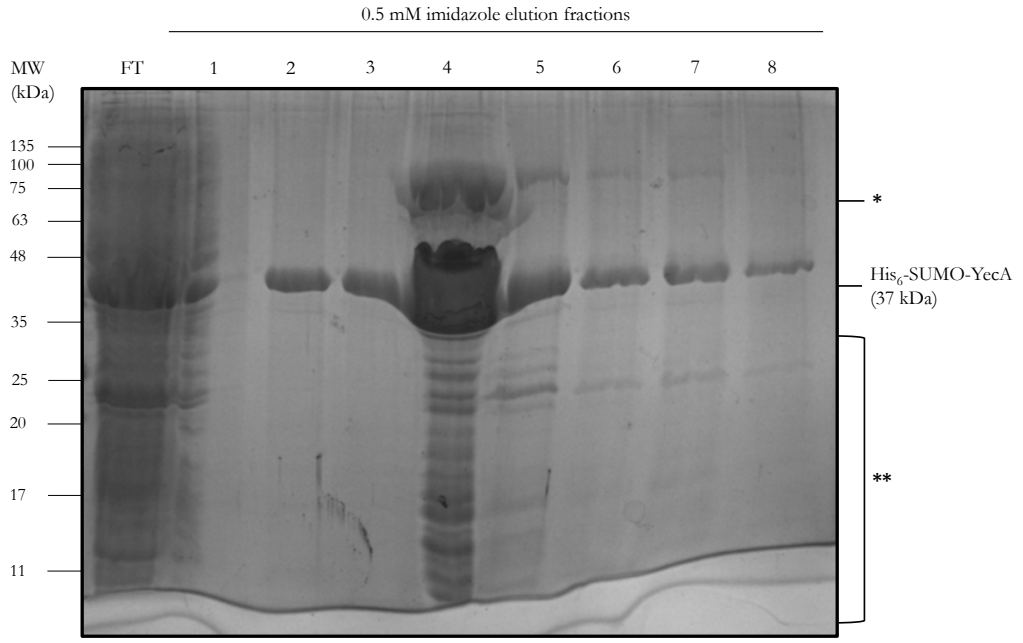


Figure 3.9: SDS-PAGE analysis of protein-containing fractions after nickel-affinity chromatography of YecA

YecA201 was expressed as a recombinant protein with an amino-terminal His_6-SUMO tag in *E. coli* strain BL21 DE3. The cultures were grown at 37°C in LB supplemented with kanamycin until late log phase. The temperature was dropped to 18°C and after 30 minutes at this temperature protein expression was induced by the addition of 1 mM IPTG. The cells were induced overnight. The cultures were centrifuged to harvest the cells. The cell pellet was resuspended in Buffer 1 supplemented with PMSF (to prevent protein degredation) and Dnase I (to degrade DNA) and lysed by high pressure homogenisation. The lysate was centrifuged to pellet un-lysed cells and cell debris. The supernatant passed through a 0.45 μM filter to further clarify the lysate. The clarified lysate was applied to a nickel-affinity column overnight at 4°C. The column was washed with five column volumes of High Salt Buffer and with five column volumes of Buffer 1. The protein was eluted from the column with 0.5 mM imidazole. After elution from the column, 10 μl of the flow through and each protein-containing imidazole elution was mixed with 10 μl Laemmli buffer and analysed by SDS-PAGE and Coomassie staining. * indicates a protein with a molecular weight of approximately 73 kDa that co-purified with YecA, which is possibly CdsA and ** indicates co-purifying proteins of below 25 kDa that are possibly ribosomal proteins.

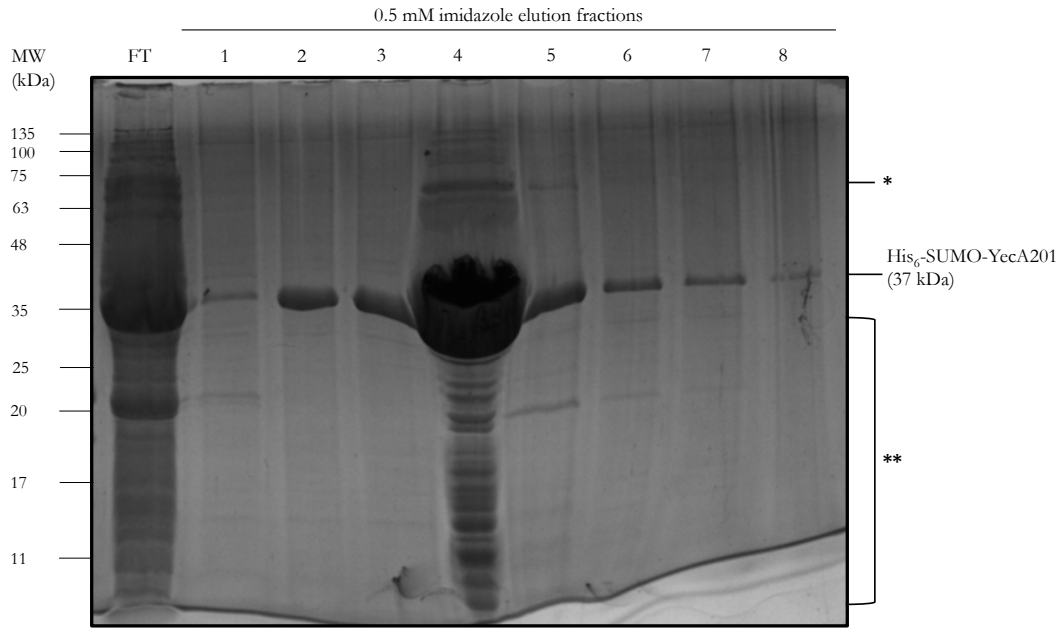


Figure 3.10: SDS-PAGE analysis of protein-containing fractions after nickel-affinity chromatography of YecA201

YecA201 was expressed as a recombinant protein with an amino-terminal His_6-SUMO tag in *E. coli* strain BL21 DE3. The cultures were grown at 37°C in LB supplemented with kanamycin until late log phase. The temperature was dropped to 18°C and after 30 minutes at this temperature protein expression was induced by the addition of 1 mM IPTG. The cells were induced overnight. The cultures were centrifuged to harvest the cells. The cell pellet was resuspended in Buffer 1 supplemented with PMSF (to prevent protein degredation) and Dnase I (to degrade DNA) and lysed by high pressure homogenisation. The lysate was centrifuged to pellet un-lysed cells and cell debris. The supernatant passed through a 0.45 μM filter to further clarify the lysate. The clarified lysate was applied to a nickel-affinity column overnight at 4°C at 0.5 ml/min. The column was washed with five column volumes of High Salt Buffer and with five column volumes of Buffer 1. The protein was eluted from the column with 0.5 mM imidazole. After elution from the column, 10 μl of the flow through and each protein-containing imidazole elution was mixed with 10 μl Laemmli buffer and analysed by SDS-PAGE and Coomassie staining. * indicates a protein with a molecular weight of approximately 73 kDa that co-purified with YecA, which is possibly CdsA and ** indicates co-purifying proteins of below 25 kDa that are possibly ribosomal proteins.

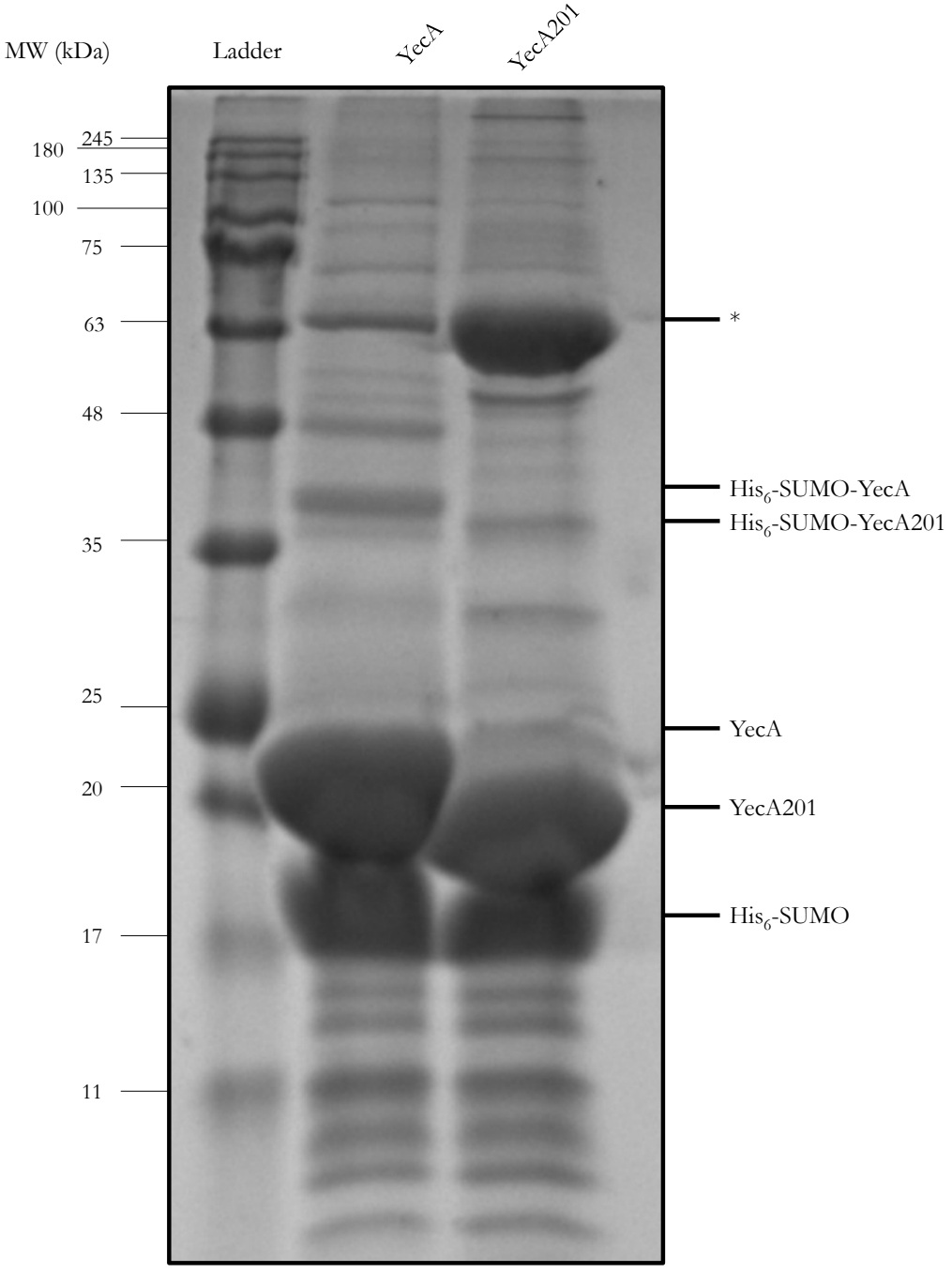
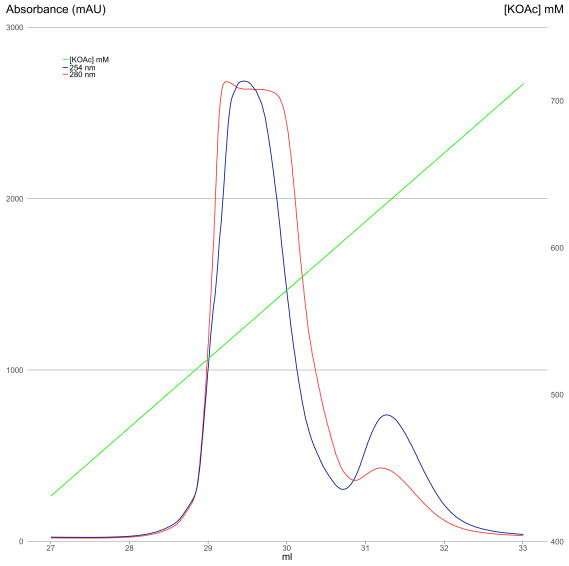
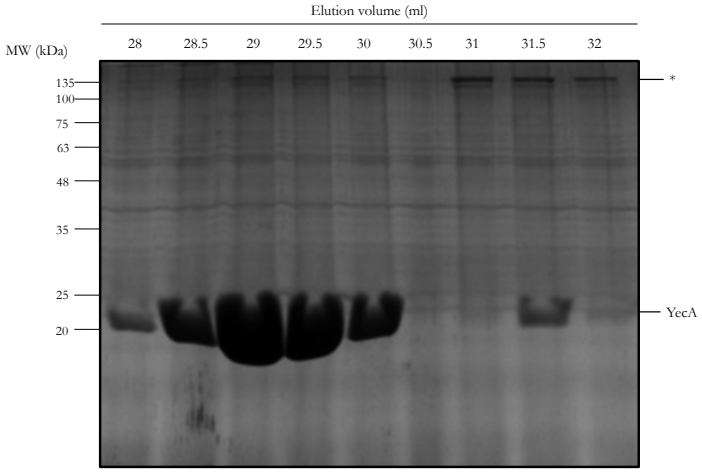


Figure 3.11: SDS-PAGE to analyse the cleavage of the His_6-SUMO tag from YecA and YecA201 by Ulp1

After nickel-affinity chromatography, the protein was incubated with His_6-Ulp1 at 4°C overnight in the presence of 5 mM β-mercaptoethanol. Ulp1 hydrolyses the peptide bond at the carboxy-terminus of SUMO, which cleaves the amino-terminal tag from the recombinant protein. * indicates an unknown co-purifying protein with a molecular weight of 63 kDa.



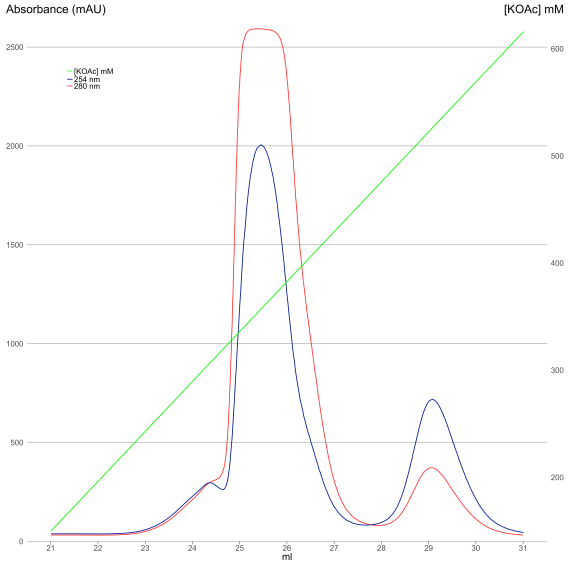
(a) Spectrophotometric analysis of the elution of YecA from the anion exchange column



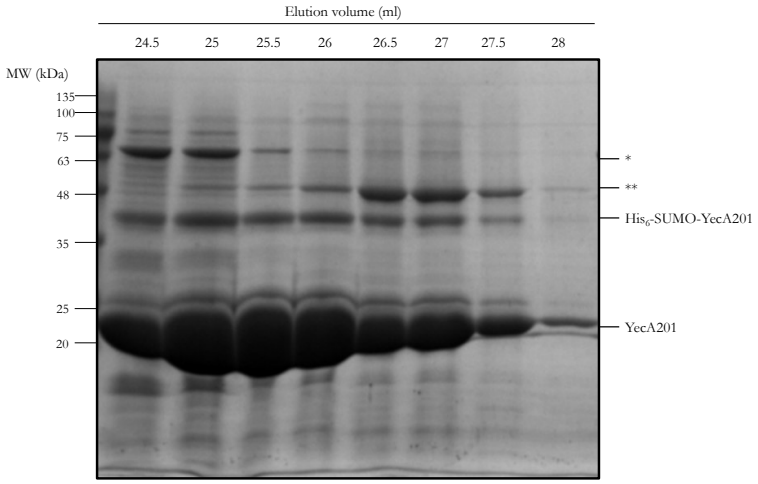
(b) SDS-PAGE to analyse anion exchange chromatography of YecA

Figure 3.12: Anion exchange chromatography of YecA

After the amino-terminal tag had been cleaved, YecA was applied to an anion exchange column. The sample was applied manually to an anion exchange chromatography column that had be pre-equilibrated with Buffer 1. The flow rate was 1 ml/min. 3.12a shows the elution trace of the anion exchange chromatography of YecA201. Right hand-side *y*-axis: The concentration of KOAc was increased from 100 mM to 1 M (green). Left hand-side *y*-axis: The elution of YecA from the anion exchange column was measured spectrophotometrically at 280 nm (red) and 254 nm (blue). The elutions were collected in 0.5 ml fractions for analysis by SDS-PAGE. 3.12b shows the SDS-PAGE analysis of protein-containing fractions. 10 µl from every other fraction between 18.5 ml and 24.5 ml was mixed with 10 µl Laemmli buffer and analysed by SDS-PAGE and Coomassie staining. * indicates an unknown co-purifying protein with a molecular weight of 135 kDa.



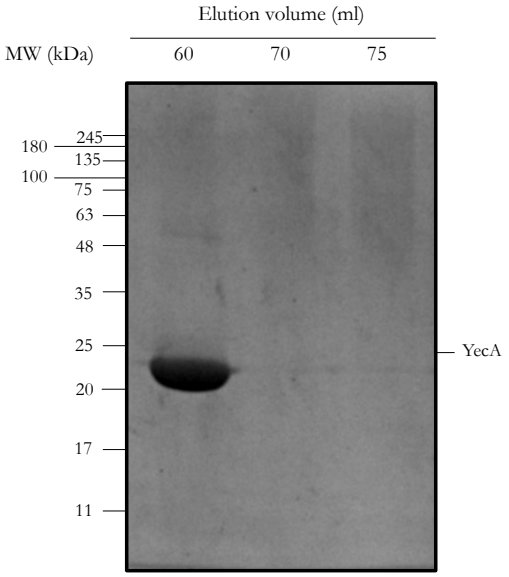
(a) Spectrophotometric analysis of the elution of YecA201 from the anion exchange column



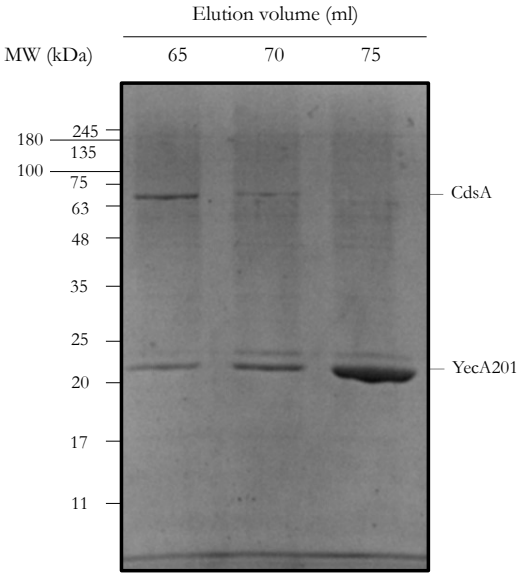
(b) SDS-PAGE to analyse anion exchange chromatography of YecA201

Figure 3.13: Anion exchange chromatography of YecA201

After the amino-terminal tag had been cleaved, YecA201 was applied to an anion exchange column. The sample was applied manually to an anion exchange chromatography column that had be pre-equilibrated with Buffer 1. The flow rate was 1 ml/min. 3.13a shows the elution trace of the anion exchange chromatography of YecA201. Right hand-side *y*-axis: The concentration of KOAc was increased from 100 mM to 1 M (green). Left hand-side *y*-axis: The elution of YecA201 from the anion exchange column was measured spectrophotometrically at 280 nm (red) and 254 nm (blue). The elutions were collected in 0.5 ml fractions for analysis by SDS-PAGE. 3.13b shows the SDS-PAGE analysis of protein-containing fractions. 10 µl from every other fraction between 28 ml and 28 ml was mixed with 10 µl Laemmli buffer and analysed by SDS-PAGE and Coomassie staining. * indicates an unknown co-purifying protein with a molecular weight of 63 kDa and ** indicates an unknown co-purifying protein with a molecular weight of 48 kDa.



(a) SDS-PAGE analysis of protein-containing fractions after gel filtration of YecA



(b) SDS-PAGE analysis of protein-containing fractions after gel filtration of YecA201

Figure 3.14: SDS-PAGE analysis of YecA and YecA201 after gel filtration

Each protein was applied to an S75 10/300 GL column that had been pre-equilibrated with Buffer 1 + 1 mM TCEP. YecA and YecA201 were concentrated and injected onto the column. The flow rate was 0.5 ml/min and the elutions were collected in 5 ml fractions. 10 µl from each protein-containing fraction was mixed with 10 µl Laemmli buffer and analysed by SDS-PAGE and Coomassie staining. 3.14a shows the Coomassie-stained SDS-PAGE gel of the three protein-containing fractions of YecA after gel filtration chromatography. YecA did not co-purify with any other protein bands. 3.14b shows the Coomassie-stained SDS-PAGE gel of the three protein-containing fractions of YecA201 after gel filtration chromatography. YecA201 co-purified with another protein. The co-purifying band in lane 1 (elution volume 65 ml) was excised and analysed by mass spectrometry and identified as CdsA, a RNA DEAD helicase with a score of 3789 and peptide coverage of 81%.

co-purified with both YecA and YecA201, as well as several bands of between 25 and 11 kDa, which were possibly ribosomal proteins.

Fractions 2-8 for both YecA and YecA201 were pooled and incubated with SUMO protease (Ulp1) overnight at 4°C in the presence of 5 mM β-mercaptoethanol to prevent the oxidation of cysteine residues. Figure 3.11 shows that the SUMO protease had cleaved approximately 90% of the amino-terminal tag from YecA and YecA201. To remove contaminating proteins, cleaved YecA was purified using anion exchange chromatography. Figure 3.12a shows the elution trace of YecA from the anion exchange column. The elution volume was collected in 0.5 ml fractions and fractions 28.5 to 32 ml contained protein, according to the elution trace. These fractions were analysed by SDS-PAGE and Coomassie staining (Figure 3.12b), which showed that YecA was nearly pure. Fractions 28 to 30 ml were pooled and concentrated. Fractions 31 to 32 ml did not contain much YecA but an unidentified protein (marked with an asterisk) was resolved in these fractions at a molecular weight of approximately 135 kDa.

YecA201 was also purified by anion exchange chromatography. Figure 3.13a shows the elution trace of YecA201 from the anion exchange column. The elution volume was collected in 0.5 ml fractions, as before, and fractions 24.5 to 28 ml contained protein, according to the elution trace. These fractions were analysed by SDS-PAGE and Coomassie staining (Figure 3.13b), which showed that YecA201 had co-purified with several contaminating proteins. The prominent bands are indicated on the gel. One of these bands was the correct molecular weight to be the His_6-SUMO-YecA201. A second band of 75 kDa and a third band of 48 kDa also co-purified with YecA201.

Size exclusion chromatography was used as a final purification step for both YecA and YecA201 to remove co-purifying proteins or aggregates. Final elutions were analysed by SDS-PAGE, which showed that YecA purified without any visible, contaminating pro-

teins. The size exclusion chromatography of YecA201 is shown in Figure 3.14a. A 75 kDa protein co-purified with YecA201 in fractions 65 and 70 ml. The band in the 65 ml fraction was excised and analysed by mass spectrometry. The co-purifying protein was identified as DeaD, the RNA helicase that is also known as CdsA. The fractions that did not contain CdsA were concentrated and used for *in vitro* assays.

3.2.7 Interaction of YecA with a model substrate

An *in vitro* chaperone assay was used to investigate whether YecA could function as a chaperone *in vitro*. The genetic analysis of *yecA* suggested that *yecA* and *secB* could be functionally similar. SecB is a molecular chaperone that has been shown to interact with secretory substrates as well as other, more general, polypeptides in the cytoplasm (Randall and Hardy, 2002). It is possible to investigate the chaperone activity of a protein by measuring its ability to prevent aggregation of porcine heart citrate synthase (citrate synthase) at 50°C (Shah *et al.*, 2016). The assay used investigates whether a chaperone has holdase activity by preventing aggregation. Citrate synthase was incubated at 50°C in the presence or absence of YecA and the light scattering at 320 nm was used to measure the aggregation of citrate synthase. The normalised light scattering of 200 nM citrate synthase in the absence of YecA was 1.0 ± 0.3, which indicated that the protein had aggregated. In the presence of 0.1 μM YecA, the normalised light scattering of was 0.7 ±0.2. In the presence of 1 μM YecA, the normalised light scattering was 0.4 ±0.2. Therefore, the decreased aggregation of citrate synthase could be caused by the presence of YecA.

To control for the aggregation of YecA, 1 μM YecA was incubated at 50°C. The normalised light scattering of the control was 0.2 ± 0.1, which suggested that YecA was aggregating at 50°C. To correct for this observation, 0.2 was subtracted from the light scattering of 1 μM YecA + 0.2 μM citrate synthase reaction, which gave a normalised light scattering of 0.01 ±0.04. Statistical significance between the corrected value and the citrate synthase

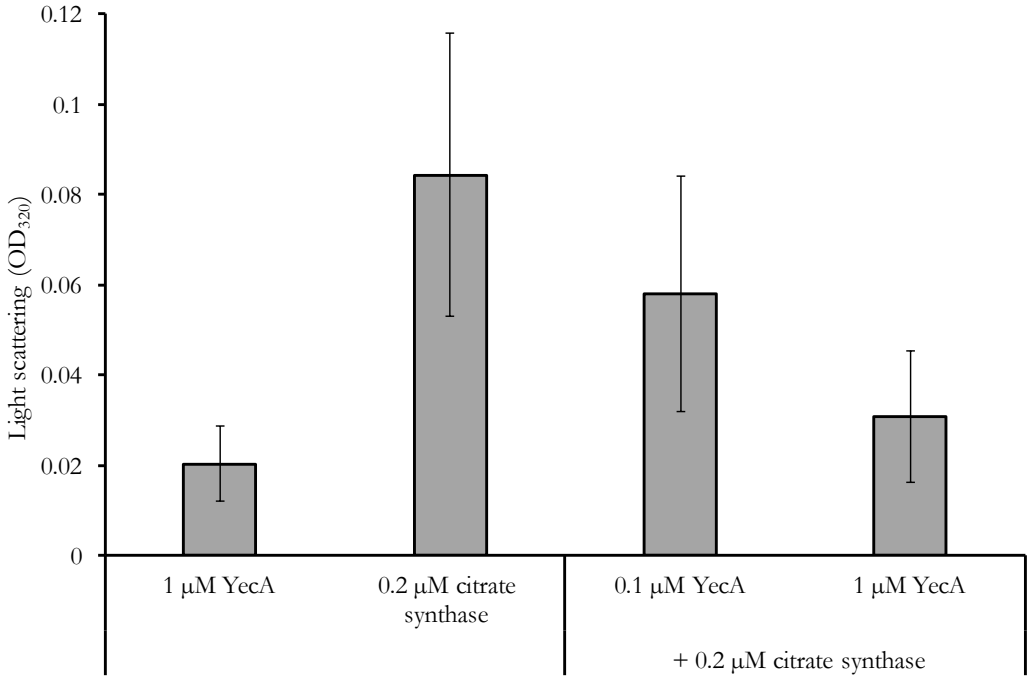


Figure 3.15: The effect of YecA on the aggregation of citrate synthase at 50°C

The ability of YecA to function as a holdase was assessed using a citrate synthase aggregation assay. Citrate synthase, at 200 nM, was incubated in either the presence or absence of YecA at 50°C for 30 minutes. YecA was used at either 0.1 μM or 1 μM, which was an equimolar concentration or a 10-fold excess of the citrate synthase dimer. The aggregation of citrate synthase was assayed at the endpoint by measuring the light scattering at 320 nm. The data represent the mean of three independent experiments normalised to the mean aggregation of the 0.2 μM citrate synthase control. The error bars represent the standard error. All samples were were analysed using one-way ANOVA with a p-value of 0.05 considered to be significant. $p = 0.0961$ which suggested that YecA did not significantly decrease the aggregation of citrate synthase at 50°C.

control was determined using a two-sample T-test assuming equal variances ($p = 0.05$). This result suggested that the YecA could be a holdase.

3.2.8 Lipid analysis by thin-layer chromatography of YecA

The 70-amino acid residues at the carboxy-terminus of SecA is required for the stable interaction of SecA with phospholipids in the inner membrane (Breukink *et al.*, 1995; Eichler and Wickner, 1997; Vrontou and Economou, 2004). To investigate whether YecA binds to phospholipids, YecA was purified and lipid content was analysed by thin-layer chromatography. Lipids were extracted from 2 ml of 1 mg/l YecA by the addition of 3 ml of 2:1 chloroform:methanol. To control for the cross-contamination of samples with lipid, a buffer-only sample was also treated with the chloroform:methanol solution. The samples were centrifuged to separate the aqueous phase (upper) from the organic phase (lower). The aqueous phase was discarded and the organic phase was analysed by one-dimentional thin-layer chromatography (TLC) in a 65:25:4 chloroform:methanol:water solvent system. Phosphatidylethanolamine (PE), phosphatidylglycerol (PG) and cardiolipin (CL) were used as standards. The TLC plate was stained with phosphomolybdic acid to show that YecA did not co-purify with any lipid species (Figure 3.16) and no lipid species were visible in the buffer only control. It was not possible to conclude that YecA interacted with lipids from this result.

3.2.9 The effect of YecA and YecA201 on the ATPase activity of SecA *in vitro*

SecB stimulates the ATPase activity of SecA *in vitro* (Mao *et al.*, 2009). To investigate whether YecA had a similar effect on the ATPase activity of SecA, an *in vitro* translocation system was used (Schulze *et al.*, 2014). The reaction consisted of liposomes that had been reconstituted with SecYEG, SecAΔ95 (a variant of SecA that lacked the most distal 95 amino acid residues), ATP, proOmpA-myc, pyruvate kinase, phosphoenolpyruvate,

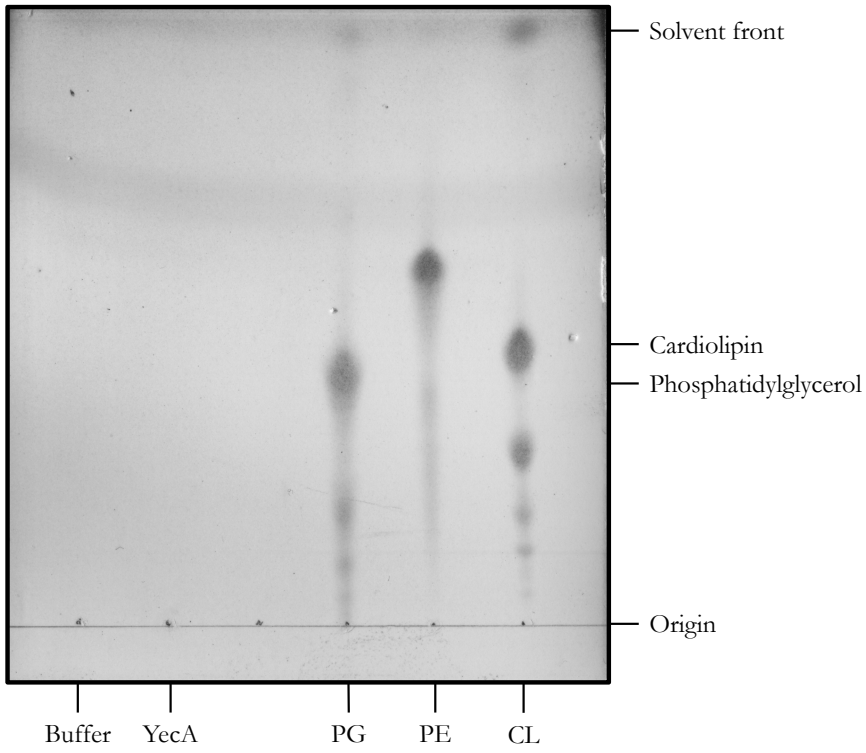


Figure 3.16: One-dimentional thin-layer chromatography to analyse binding of YecA to phospholipids

Any lipids that were associated with YecA were extracted with a 2:1 chloroform:methanol solution. The extracted lipids and PE, PG and CL standards were separated in a 65:35:4 chloroform:methanol:water solvent system. A buffer only sample was included to control for disassociation of YecA during the purification. The TLC plate was stained with phosphomolybdic acid.

lactate dehydrogenase and NADH (Figure 3.17). YecA and YecA201 were purified as described in Section 3.2.6. The *in vitro* ATPase assay measures the rate of ATP hydrolysis by SecAΔ95 through a coupled regeneration system. ATP was hydrolysed to ADP and regenerated using phosphoenolpyruvate and pyruvate kinase. The pyruvate kinase was reduced by NADH to lactate, a reaction that was catalysed by lactate dehydrogenase. NADH was not regenerated. Therefore, as the assay progressed, the concentration of NADH decreased. The oxidation of NADH was measured spectrophotometrically at 340 nm.

The carboxy-tail of SecA, which includes the SecA-MeBD, occludes binding of the signal peptide to the peptide-binding groove (Gelis *et al.*, 2007). The ATPase activity of SecAΔ95 increases in the presence of preprotein (Robson *et al.*, 2009a; Corey *et al.*, 2016) and has decreased variability of ATPase activity compared with full-length YecA (Robson *et al.*, 2009a; Corey *et al.*, 2016). The preprotein proOmpA-myc was added to each reaction to stimulate the ATPase activity of SecAΔ95. ProOmpA-myc was used as a substrate because several other studies have used proOmpA, an outer membrane protein, to investigate Sec-dependent translocation *in vitro* and *in vivo* (Geller and Green, 1989; Robson *et al.*, 2009a; Corey *et al.*, 2016; Lee and Bernstein, 2001).

The positive control included all components of the *in vitro* translocation system, except YecA or YecA201. The first negative control lacked only ATP from the reaction to control for background oxidation of NADH to NAD^+. The second negative control included ATP and YecA but lacked SecYEG and SecA from the *in vitro* translocation system. This control was included to investigate the ATPase activity of YecA. The effect of YecA and YecA201 on the ATPase activity of SecAΔ95 was measured before and after addition of proOmpA-myc. The rate of ATP hydrolysis, as measured by the change in NADH concentration, was normalised to the positive control after proOmpA-myc had been added for each of the three independent replicates (Figure 3.18).

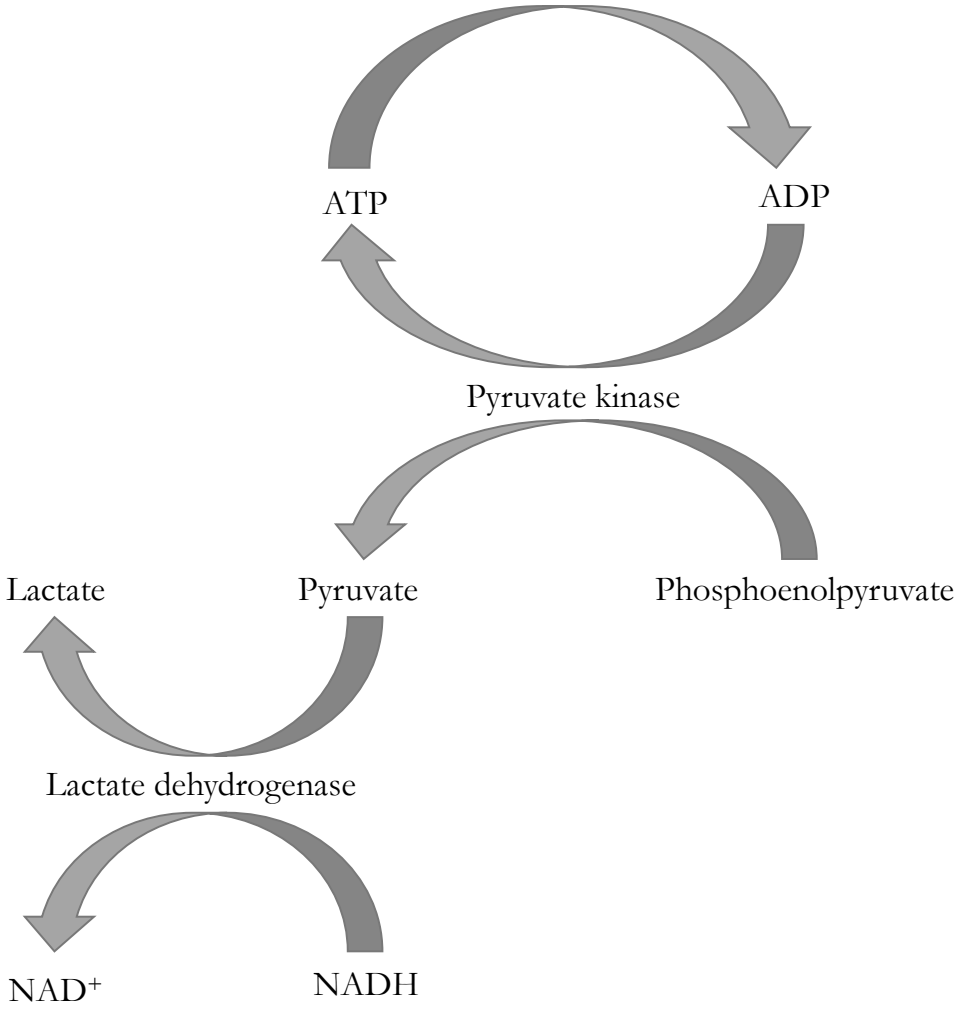


Figure 3.17: Illustration of the coupled regeneration system of the *in vitro* ATPase assay

ATP hydrolysis to ADP by SecA is coupled to the conversion of pyruvate to phosphoenolpyruvate by pyruvate kinase. Pyruvate is reduced to lactate by lactate dehydrogenase by the oxidation of NADH to NAD^+. The depletion of NADH is measured spectrophotometrically at a wavelength of 390 nm.

Before proOmpA-myc was added to the system, the rate of ATP hydrolysis was measured. For the control that lacked ATP, the rate of ATP hydrolysis was 0.03 ± 0.01 min^{-1}. The normalised rate of ATP hydrolysis for the control that lacked SecYEG and SecAΔ95 but contained YecA (minus SecYEGA + YecA) was 0.01 ± 0.006 min^{-1}. The normalised rate of the control reaction that contained all components of the system except YecA or YecA201 (minus YecA) was 0.07 ± 0.009 min^{-1}. The rate of ATP hydrolysis when YecA was included in the reaction was 0.08 ± 0.009 min^{-1}. When YecA201 was included in the reaction, the normalised rate of ATP hydrolysis was 0.07 ± 0.002 min^{-1}. One-way ANOVA analysis of the data indicated that there was a significant difference between the reactions (p = 0.002). A *post hoc* Tukey test showed that neither YecA nor YecA201 affected the rate of ATP hydrolysis compared with the reaction in the absence of YecA (p = 0.9 and p = 0.9, respectively). Therefore, it is not possible to conclude from this data that YecA increased ATPase activity of SecA in the absence of preprotein *in vitro*.

The rate of ATP hydrolysis was measured after the addition of the preprotein to investigate whether YecA or YecA201 had an effect on the ATPase activity of SecAΔ95 *in vitro* in the presence of proOmpA-myc. The rate of ATP hydrolysis by SecAΔ95 in the control reaction that lacked only YecA was normalised to 1 for each replicate. The normalised rate of ATP hydrolysis in the absence of ATP was 0.02 ± 0.01 min^{-1}. The normalised rate of ATP hydrolysis was 0.02 ± 0.005 min^{-1} for the control that lacked SecYEG and SecAΔ95 but contained YecA. The two negative controls both had a similar rate of ATP hydrolysis in the presence or absence of proOmpA-myc. The latter control indicated that YecA did not have *in vitro* ATPase activity.

When proOmpA-myc was added to the system that contained YecA, the normalised rate of ATP hydrolysis by SecAΔ95 was 1.4 ± 0.1 min^{-1}, which is nearly a 40% increase compared with the minus YecA control. When proOmpA-myc was added to the reaction

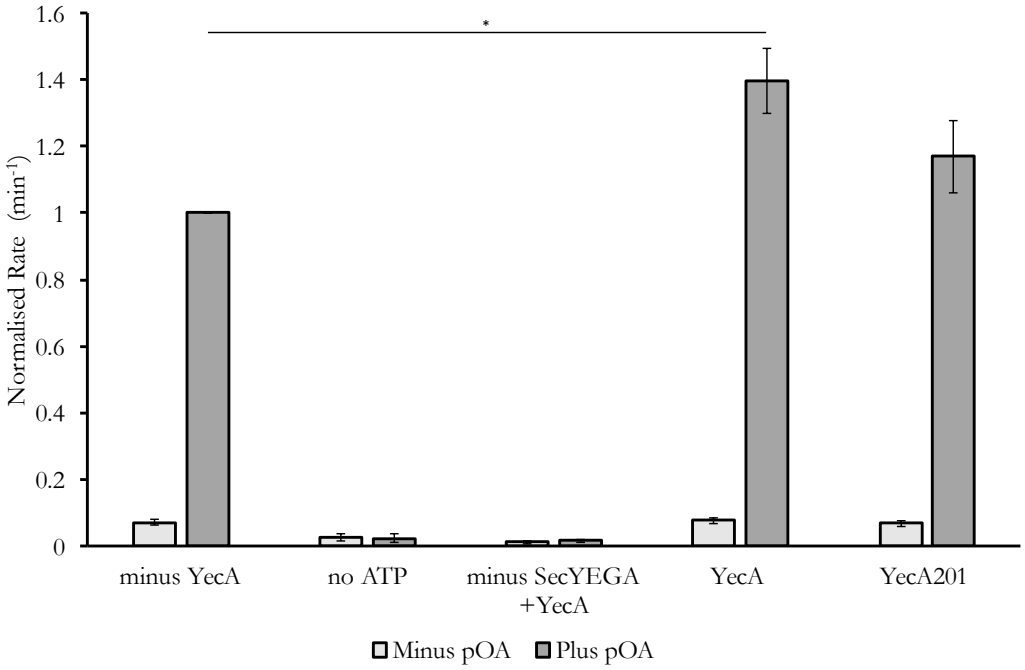


Figure 3.18: Relative ATPase activity of SecAΔ95 in the presence or absence of YecA or YecA201 *in vitro*

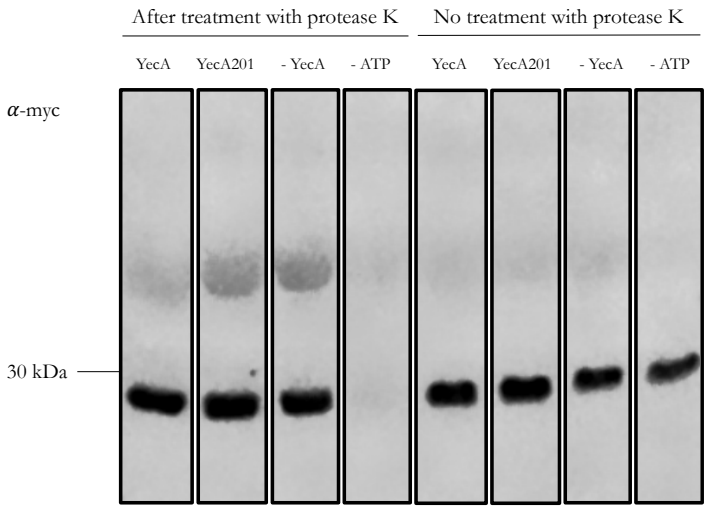
The effect of YecA or YecA201 on the ATPase activity of SecAΔ95 was measured using a coupled regeneration reaction that measures the hydrolysis of ATP via the decrease in NADH over time spectrophotometrically. The rate of ATP hydrolysis by 0.3 μM SecAN95 in the presence of 0.3 μM proteoliposomes and 10 μM YecA or YecA201 was measured before and after the addition of 1 μM proOmpA-myc. The rate was normalised within each replicate to the positive control, which was the reaction that contained all of the reagents except YecA. The error bars are the standard error of the normalised data. Statistical analysis indicated that there was a significant difference between the rate of ATPase hydrolysis after proOmpA-myc had been added to the reactions ($p = 5.64 \times 10^{-8}$). Further analysis using a *post hoc* Tukey investigate revealed that there was a significant difference when YecA was included in the assay compared with the minus YecA control ($p = 0.01$, indicated by an asterisk). Inclusion of YecA201 also caused a significant increase in the rate of ATP hydrolysis compared the rate of ATP hydrolysis in the absence of YecA ($p = 0.01$). There was no statistical difference between YecA and YecA201 on the normalised rate of ATP hydrolysis ($p = 0.7$). $n = 3$.

that contained the full system and YecA201, the normalised rate of ATP hydrolysis was 1.2 ± 0.1 min^{-1}. After proOmpA-myc had been added, the inclusion of YecA201 in the reaction resulted in a 30% increase in the rate of ATP hydrolysis compared with the minus YecA control. Both YecA and YecA201 stimulated the ATPase activity of SecAΔ95, however the activity of the full-length protein was higher than that of YecA201.

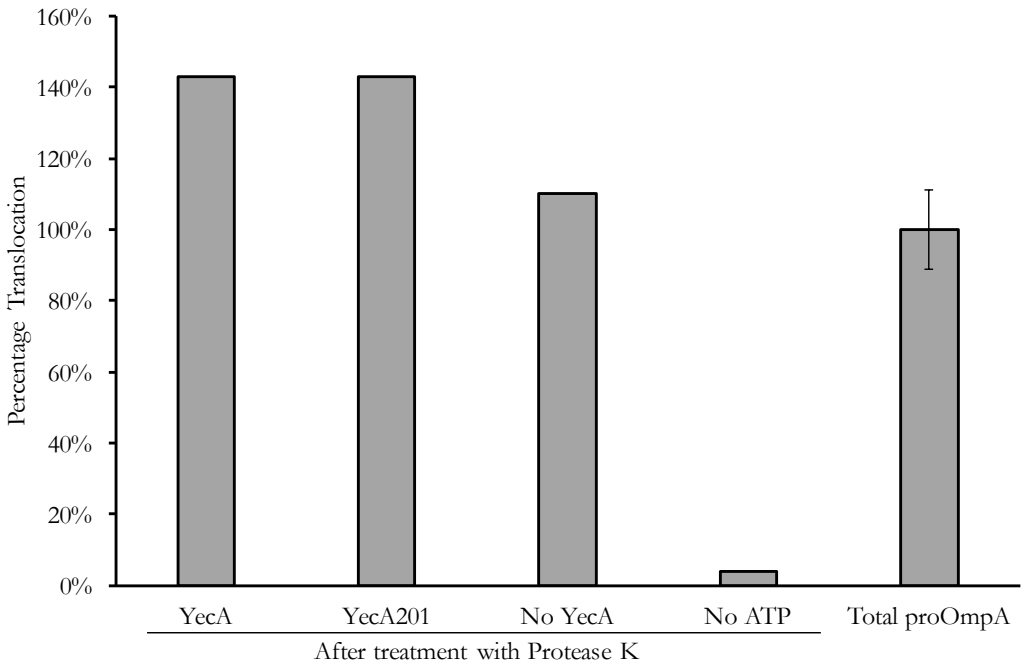
3.2.10 The effect of YecA and YecA201 on the rate of translocation *in vitro*

The assay described in Section 3.2.9 indicated that YecA and YecA201 had increased the ATPase activity of SecAΔ95. The ATPase activity of SecA is required for translocation of preproteins through the SecYEG membrane channel *in vivo*. In the *in vitro* translocation assay, successfully translocated proOmpA is sequestered inside the SecYEG proteoliposomes. A protein protection digest assay was used to investigate the effect of YecA on the translocation of proOmpA. This assay results in the digestion of all proteinaceous matter that was not translocated into the proteoliposomes. After treatment with protease K, the liposomes were analysed by SDS-PAGE. The *in vitro* translocation efficiency of each reaction was determined by immunoblotting against the carboxy-terminal myc-tag of proOmpA-myc (Figure 3.19). After the reaction that lacked ATP was incubated with protease K, no proOmpA-myc was detected by immunoblot. All reactions successfully coupled ATP hydrolysis with the translocation of proOmpA-myc (Figure 3.19).

The total mean proOmpA-myc (before protease K digestion) of 10% of each sample was 0.1 ±0.03. This result shows the error associated with this assay. Although there was a small increase in the translocation activity of YecA (0.2) and YecA201 (0.2) compared the translocation activity in the absence of YecA (0.1), it was reasoned that this was within the error of this experiment and the translocation assay was repeated only once.



(a) Western blot to investigate the translocation activity of SecAΔ95



(b) Quantified translocation activity of SecA Δ95

Figure 3.19: Translocation activity of SecAΔ95 in the presence and absence of YecA or YecA201

The translocation activity of SecAΔ95 was investigated through the detection of proOmpA-myc by α-myc immunoblotting. 3.19a shows the Western blot of samples that had been incubated with protease K. The samples were treated with protease K to digest the proOmpA-myc that had not been translocated into the proteoliposomes. 10% of the sample was not treated with protease K and represents the total input of proOmpA-myc. $n = 1$. 3.19b shows the quantified translocation activity (a.u.) of the *in vitro* translocation system. The signals shown in 3.19a were quantified using ImageJ software. The mean quantified signal of 10% each sample that was not treated with protease K represents the total input of proOmpA-myc. The error bar is the standard error of the total input controls. $n = 4$. The detection of proOmpA-myc after the samples had been treated with protease K represented the translocation activity of each sample. $n = 1$.

3.3 Discussion

In this chapter, three novel observations were described that together suggest that YecA, a protein of unknown function, is involved in Sec-dependent translocation of secretory proteins in *E. coli*. Firstly, it was shown that YecA was involved with the export of a fusion protein that has been used to identify components of the Sec machinery. Secondly, the genetic interaction between *secB*, which encodes a cytoplasmic chaperone, and *yecA* was identified. Thirdly, YecA interacted with SecA to increase the translocation-coupled ATPase activity of SecA. It is possible that YecA functions as a cytoplasmic chaperone of secretory substrates.

The first aim of this chapter was to investigate the effect of deleting *yecA* on the Sec-dependent translocation of MalE′-′LacZ *in vivo*. The β-galactosidase assay showed that the deletion of *yecA* resulted in a secretion defect. It was not possible to fully complement the deletion of *yecA* by expressing *yecA* from another locus, however the expression of *yecA201* did alleviate the observed secretion defect. There are several possible explanations for this. The first explanation is that YecA201 is more stable *in vivo* than YecA, which could be investigated using a Western blot to visualise the relative concentrations of each protein. It is also possible that YecA is required at a specific subcellular concentration that was not achieved under the experimental conditions of this assay. The CS assay suggested that, under similar conditions to the β-galactosidase assay, the over-expression of *yecA* caused a mild growth defect. This growth defect was not observed when expression of *yecA* was not induced. In the future, the β-galactosidase assay will be repeated with different levels of induced expression of *yecA*. The result of the β-galactosidase assay described in this chapter could also indicate that YecA201 is more active than YecA. YecA interacts with the ribosome and this interaction is dependent on the YecA-MeBD (Dr. D. Huber, personal communication). It is possible that a proportion of YecA that interacted with the ribosome was diverted from interacting

with the fusion protein, which resulted in a higher β-galactosidase activity compared with YecA201.

The connection between YecA and SecB was investigated in this chapter. SecB is a cytoplasmic chaperone that maintains preproteins in an unfolded state for posttranslational translocation and MalE is dependent on SecB for export (Collier *et al.*, 1988). To investigate this, single gene deletions of *yecA* and *secB* in *E. coli* strain MG1655 were compared with the parent strain and a double deletion strain that lacked both *yecA* and *secB*. The enhanced CS phenotype of the double mutant compared with the *secB* deletion strain indicated that *yecA* and *secB* could be functionally redundant. Studies have shown that, like *yecA* deletion strains, increased basal β-galactosidase activity is observed in *secB* deletion strains that express *malE'-'lacZ* (Kumamoto and Beckwith, 1983) because SecB increases the concentration of preprotein at the translocation site by interacting with full-length SecA. A SecB-induced conformational change increases the affinity of SecA for unfolded secretory proteins, and it is this interaction that increases the rate of ATP hydrolysis by SecA (Miller *et al.*, 2002; Gelis *et al.*, 2007; Mao *et al.*, 2009). In the *in vitro* system described in this chapter to investigate the rate of ATP hydrolysis by SecA, SecAΔ95 and not full-length SecA was used. Inclusion of SecB in this *in vitro* system resulted in a decrease in the ATPase activity of SecAΔ95 (Prof. I. Collinson and Dr. W. Allen, personal communication). It is possible that the decrease in ATP hydrolysis was caused by the inability of SecB to bind SecAΔ95. Therefore, it is possible that the mechanism by which YecA increased the ATPase activity of SecA involved a direct interaction between YecA and SecAΔ95. Furthermore, the pull down assay suggested that YecA co-purified with SecA. In the future, an immunoprecipitation assay will be used to confirm these data.

The observed phenotypes described in this chapter could be linked to the stimulation of the ATPase activity of SecA by YecA and YecA201. ATP hydrolysis by SecA provides

the energy for the translocation of secretory proteins through SecYEG. YecA was shown to increase the ATPase activity of SecAΔ95 *in vitro*. However, it was not clear whether the increased ATPase activity was the result of a direct interaction between SecAΔ95 and YecA. The increase in ATPase activity of SecAΔ95 in the presence of YecA could be caused by YecA functioning as a chaperone to effectively increase the preprotein concentration at the SecYEG channel.

The protein-protein interactions of YecA that were identified using the pull-down assay and the mass spectrometry analysis indicated that YecA interacted with CdsA. The expression of *cdsA* is induced by cold-shock (Jones *et al.*, 1996). Like SecA, CdsA is an ATP-dependent DEAD-box protein. While SecA hydrolyses ATP to drive translocation of secretory proteins, CdsA is a RNA helicase that is thought to assist with several processes that include the initiation of translation, cold-shock induced gene regulation, mRNA decay and biogenesis of the large ribosomal subunit (Charollais *et al.*, 2004). Deletion of the *cdsA* gene in *E. coli* causes a cold-sensitive growth phenotype due to aberrant biogenesis of the large ribosomal subunit (Charollais *et al.*, 2004). CdsA interacts with RNase E, which is known to associate with PNPase and the DEAD-box protein RhlB to form an "RNA degradosome" (Regonesi *et al.*, 2006). RhlB and RNase E were both identified by mass spectrometry as a co-purifying with YecA and YecA201. However, the parameters of the manual filter meant that these proteins were not considered to have a positive interaction with either YecA or YecA201. PNPase was, however, positively identified as interacting with YecA. In the future, the interaction of YecA with the RNA degradosome will be investigated.

The possibility that YecA was a chaperone was investigated *in vitro* using a non-native substrate, citrate synthase. YecA decreased the aggregation of citrate synthase by approximately 50%, although this result was not statistically significant. The experimental conditions that were used during this assay were observed to cause the aggregation of

YecA. The aggregation of secretory substrate proOmpA has be used to investigate the chaperone activity of SecB at room temperature (Lecker *et al.*, 1990). In the future, a different model substrate, such as the precursor protein proOmpA, will be used to investigate the holdase activity of YecA.

It not is possible to conclude whether YecA interacted with lipids from the investigation presented in this chapter. YecA was purified to homogeneity prior to the TLC assay. Subsequent discussion revealed that similar levels of purification can disrupt the interaction between SecA and lipids (Dr. M. Jamshad, personal communication). SecA interacts with lipids and SecB via the SecA-MeBD. To investigate whether YecA interacts with lipids, the TLC assay could be repeated with a less stringently purified YecA sample. If YecA interacts with lipids, the role of the YecA-MeBD could be investigated by repeating the TLC assay with YecA201. If YecA does not interact with lipid, the MeBDs of YecA and SecA could be swapped to investigate whether the SecA requires the specific amino acid residues in the SecA-MeBD to interact with lipid.

The role of the YecA-MeBD was investigated by comparing the full-length protein with YecA201, a truncated variant of YecA that lacked the most distal 20-amino acid residues. The expression of *yecA201* complemented the effect of deleting *yecA* on the export of MalE′-′LacZ *in vivo*, unlike the expression of *yecA*, which only partially complemented the deletion. The partial complementation by *yecA201* of this phenotype could be caused by differential expression of *yecA* compared with the parent strain. It is possible that YecA was required at a specific cellular concentration, or under a particular environmental condition that was not achieved under the experimental conditions that were used. The MeBD of SecA can prevent binding of signal peptides to the peptide-binding groove of SecA (Gelis *et al.*, 2007). It is possible that the YecA-MeBD has an auto-inhibitory role, like the SecA-MeBD. Therefore, it is possible that the partial complementation of the *yecA* deletion by the full-length protein could be explained by the inhibition of YecA by

its MeBD. In the next chapter the investigation of the structure of YecA to analyse both the YecA-MeBD and the full-length protein is described.

Chapter 4

STRUCTURAL ANALYSIS OF YECA

4.1 Introduction

The results presented in Chapter 3 suggested a role for YecA in the Sec pathway of *E. coli*. These investigations were predicated on the hypothesis that the YecA-MeBD and the SecA-MeBD were functionally homologous. The structure of YecA was investigated to gain insight into the molecular mechanism of the protein. Structures of sequence homologues of both domains of YecA have been resolved, however it is not known how the two domains interact to form the tertiary structure of YecA.

UPF0149 is an 192-amino acid residue domain that is situated at the amino terminus of YecA. In *E. coli*, the only other protein that contains UPF0149 is YgfB, which is a 20.3 kDa protein of unknown function. A 1.95 Å crystal structure of the dimer of YgfB from *H. influenzae* has been resolved (Galkin *et al.*, 2004). The monomer has a two structural regions formed of four and three α-helices, respectively. The carboxy-terminal 3-helix bundle is similar to the "lid" of the substrate binding domain of DnaK, a chaperone (Galkin *et al.*, 2004). The sequence similarity between YgfB and the 3-helix bundle of DnaK is low. The two α-helices of YgfB fold to form a novel fold that could be formed by YecA but the function of this fold has not been investigated. UPF0149 remains a domain of unknown function, however studies have suggested a role for YgfB in DNA-binding and the regulation of gene expression in *E. coli* (Ishihama *et al.*, 2016).

The SecA-MeBD binds to lipid and SecB (Breukink *et al.*, 1995; Zhou and Xu, 2003). When bound to zinc the SecA-MeBD forms a ββα structural motif that is similar to the classical zinc finger motif (Zhou and Xu, 2003; Dempsey *et al.*, 2004; Matousek and Alexandrescu, 2004). A single zinc ion is coordinated via three cysteine residues and a histidine residue, which is sometimes replaced by a fourth cysteine residue (Zhou and Xu, 2003). The interaction between SecA and SecB is stabilised by the presence of zinc (Fekkes *et al.*, 1997), possibly because the coordination of zinc creates a positively charged

interface on the SecA-MeBD (Zhou and Xu, 2003). The positively charge residues and the zinc-binding residues are conserved between the SecA-MeBD and the YecA-MeBD.

To investigate the structure of YecA, several techniques were used. These techniques (X-ray crystallography, NMR spectroscopy, small-angle X-ray scattering and circular dichroism) are described below.

X-ray crystallography

The objective of X-ray crystallography is to resolve the three-dimensional structure of the desired complex from its crystal (Smyth and Martin, 2000). To form crystals, protein is encouraged to come out of solution by vapour diffusion (Figure 4.1). For example, the SecA-polypeptide-binding domain from different organisms has been crystallised in two different conformations, with the PPXD close to the HWD, or rotated away from the HWD (Hunt *et al.*, 2002; Osborne *et al.*, 2004). The crystal structure of a single SecA bound to a single SecY indicates that the PPXD rotates further away from the HWD to create an interface between the PPXD and the nucleotide binding domain 2 to interact with SecY (Zimmer *et al.*, 2008). The observed conformational changes provide the foundations for further mechanistic studies. Structural information can provide the basis for focussed downstream research, therefore X-ray crystallography has a role in many areas of biology.

NMR spectroscopy

NMR spectroscopy is an analytical technique that is used to determine the molecular structure of proteins in solution. After the molecular structure has been determined, interactions and other physical properties can be ascertained through conformational changes. NMR allows the study of proteins in solution, unlike X-ray crystallography, and as such has been used to study structure and molecular dynamics in conditions that

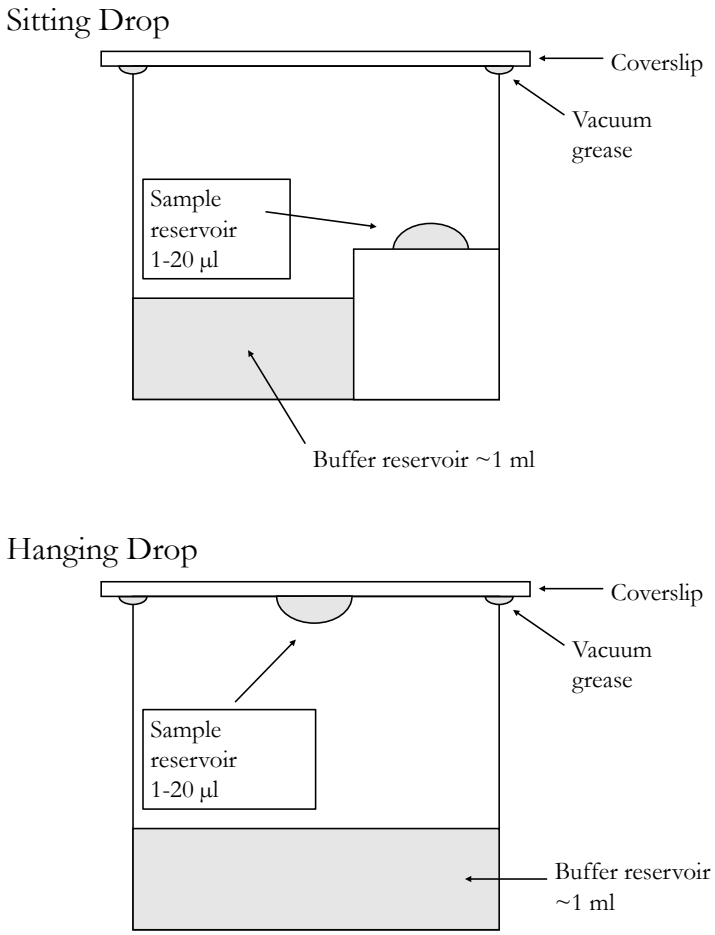


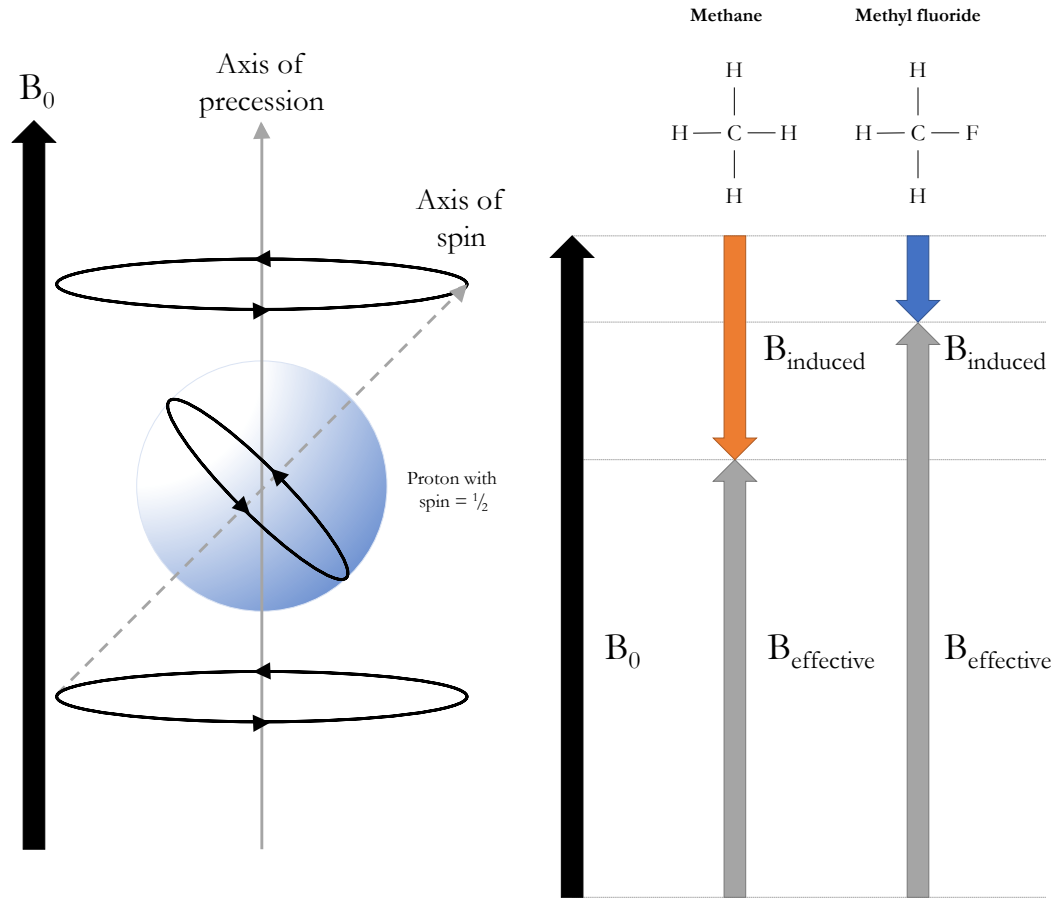
Figure 4.1: Sitting drop and hanging drop methods of protein crystallisation.

Vapour diffusion is used to encourage protein to come out of solution and crystallise. There are two types of vapour diffusion: sitting drop and hanging drop. In both set-ups, a sample reservoir containing purified protein, buffer and precipitant is placed within a sealed well with a reservoir of similar buffers. Water moves from the sample reservoir and collects in the buffer reservoir by vapour diffusion over time. This promotes protein crystallisation because the protein and precipitant concentrations both increase slowly. The only difference between the two methods of vapour diffusion is the orientation of the sample reservoir to the buffer reservoir.

approximate the physiological environment. Full-length SecA remains unresolved by X-ray crystallography due to the flexibility of the carboxy-tail. NMR spectroscopy has revealed that the carboxy-tail interacts with the PPXD in the main body of the protein (Keramisanou *et al.*, 2006; Gelis *et al.*, 2007), which could regulate the interaction of SecA with polypeptides in the cytoplasm.

Protons, neutrons and electrons are sometimes referred to in the literature as quantum mechanically subatomic particles. Collectively, these particles have nuclear spin (Keeler, 2002). A nucleus with spin is a charged particle that rotates about a given axis, which creates a magnetic field. The overall spin of the nucleus is determined by the number and ratio of protons and neutrons. If a nucleus has an even number of protons and an even number of neutrons then the paired spins often cancel each other out such that the spin of the nucleus is equal to zero. If the number of neutrons plus the number of protons is odd then the nucleus has a half integer spin, e.g. $\frac{1}{2}$, $\frac{3}{2}$, $\frac{5}{2}$, $\frac{7}{2}$. A nucleus has an integer spin if the number of neutrons and protons are both odd. The magnetic field around a nucleus with spin is the nuclear magnetic moment. NMR spectroscopy applies a magnetic field (B_0) to the sample. The nuclear magnetic moment of the nucleus causes the nucleus to align with or against the external magnetic field. The rotational axis of a nucleus cannot be exactly parallel or antiparallel to the external magnetic field but must precess at an angle (Figure 4.2a). This precession is known as the Larmor precession. The frequency of the Larmor precession is proportional to the strength of the external magnetic field: the greater the strength of the external magnetic field, the greater the frequency.

In an example compound with two populations of hydrogen atoms there will be two precessional frequencies. A little more than half of the protons align with B_0 and the rest align against B_0. Electromagnetic radiation in the radio frequency range is applied to the sample. The frequency that is equivalent to the precessional frequencies causes the corresponding hydrogen populations to absorb radiation and "flip" so that both are



(a) The origin of an NMR signal

(b) The effective magnetic field

Figure 4.2: The effect of an external magnetic field on a sample with spin

4.2a shows that when a magnetic field (B_0) is applied to a nucleus with spin, the nucleus precesses at an angle. 4.2b compares the effective magnetic field strength experienced by protons in methane and methyl fluoride. When an external magnetic field B_0 is applied to methane, the electrons that belong to the methyl carbon create their own magnetic field that opposes B_0. The magnetic field that is created by the electrons is known as the induced magnetic field. The induced magnetic field shields the protons of the methane from B_0 such that they experience an effective magnetic field that is B_0 less the induced magnetic field. In the case of methyl fluoride, the fluorine has a greater electromagnetic strength than carbon. This means that the electrons that belong to the carbon are pulled towards the fluorine and the protons are not shielded to the same extent as the protons in methane. Therefore the effective magnetic field experienced by the protons in methyl fluoride is greater than the effective magnetic field experienced by the protons in methane. As a result, the chemical shift of the protons of methyl fluoride is greater than the chemical shift of the protons of methane. Figure adapted from Soderberg (2016).

aligned against B_0. The absorbance of the radiation, and the intensity of the absorbance, is recorded by the NMR instrument. The frequency information is converted to chemical shift to standardise recordings of samples across different NMR instruments with different field strengths.

When a compound is placed inside a magnetic field (B_0), non-equivalent protons will resonate with a different frequency. The chemical shift of a proton is determined primarily by the neighbouring electronic environment. Methane (CH_4) has four equivalent protons. When B_0 is applied to methane the electrons of the carbon atom circulate and create an induced magnetic field that opposes B_0. The induced magnetic field reduces the strength of the magnetic field that the four protons experience. This is known as local diamagnetic shielding and results in a lower resonance frequency of the protons (Soderberg, 2016). Methyl fluoride (CH_3F) is an example of a compound with three protons that experience a deshielding effect because the fluorine is more electronegative than carbon so the carbon's electrons are pulled towards the fluorine and away from the hydrogens. This means that the three protons of methyl fluoride experience a larger effective magnetic field strength than methane (Figure 4.2b). This means that the chemical shift of the protons of methyl fluoride is greater than the chemical shift of the protons of methane (Soderberg, 2016). The strength of the deshielding effect of electronegative moiety is correlated with the proximity to the proton, *i.e.* the closer the proton to the electronegative moiety, the stronger the deshielding effect and the greater the chemical shift will be.

Heteronuclear Single Quantum Coherence (HSQC) is a simple 2D NMR spectrum that maps cross peaks of correlating ^{1}H and heteronuclei (for example ^{15}N and ^{13}C). The ^{1}H-^{15}N-HSQC spectrum of a protein will show peaks with slightly different frequencies depending on the amino acid composition and the folding of a protein. In the case of a folded protein, the peaks are usually well dispersed in the NH region (6-12 ppm) whereas

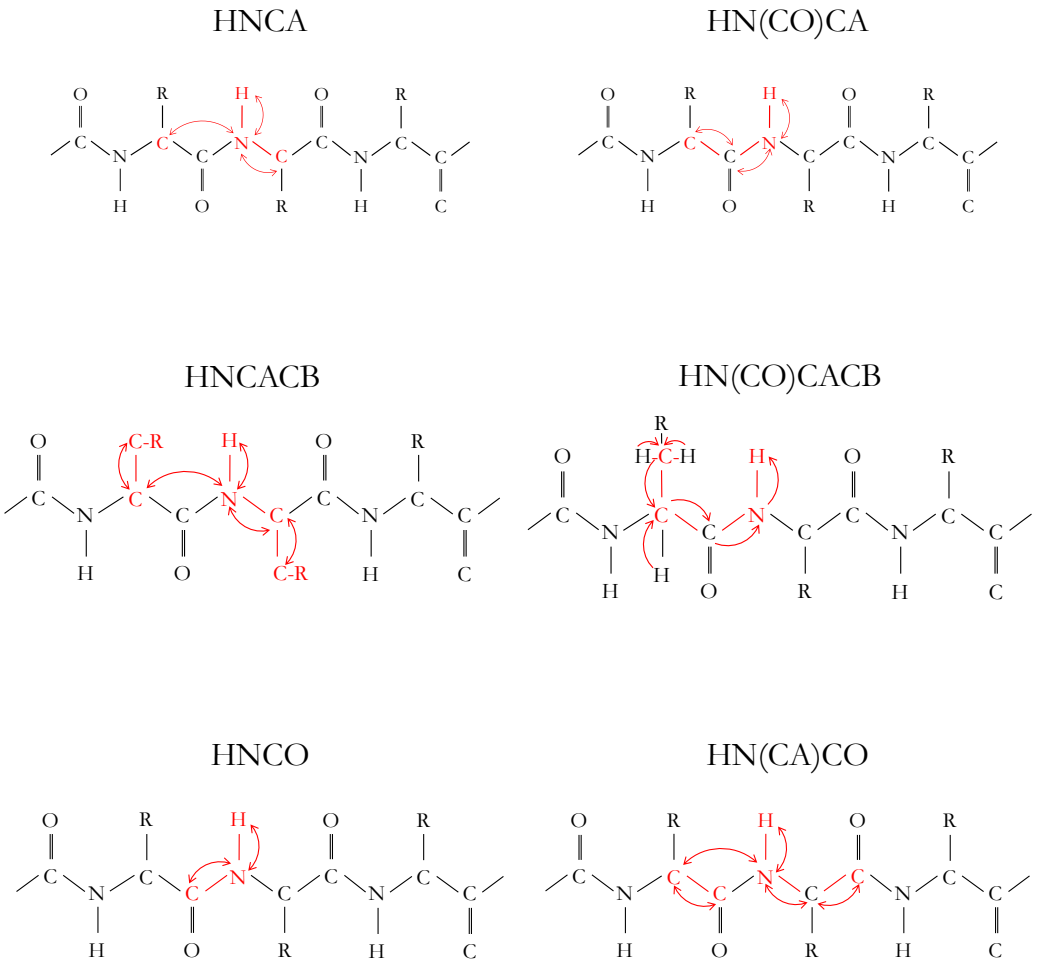


Figure 4.3: Diagram to show the NMR experiments used for the assignment of YecA

The red arrows show the direction of magnetisation transfer. For each experiment, the chemical shift information that can be determined from each experiment is indicated by the red atoms. HNCA provides information about the amide proton of a residue, the CA of the same residue and the CA of the neighbouring residue. HN(CO)CA provides information about the the amide proton of a residue and both the carbon atoms of the preceding residue. HNCACB correlates the chemical shifts of the amide bond of a residue and the CA and the CB of both that residue and the neighbouring residue. HN(CO)CACB provides information about the amide proton and the CA, CB and the CO of the neighbouring residue, as well as providing information about the CO of its residue. HNCO correlates the amide proton with the CO of the neighbouring residue. HN(CA)CO correlates the amide bond with the CO (and the CA) of the same residue and the neighbouring residue. Figure adapted from Knowles (2005).

in an unfolded protein resonance peaks are poorly dispersed. This is due to the high sensitivity of the chemical shift to the chemical environment. Each ^{1}H-^{15}N cross peak in the ^{1}H-^{15}N-HSQC spectrum represents a resonance peak from a single N-H. Therefore, in an ^{1}H-^{15}N-HSQC, the number of residues of the protein should be roughly equal to the number of peaks. Proline lacks an NH-backbone moiety so is not observed in an ^{1}H-^{15}N-HSQC experiment. ^{1}H-^{15}N-HSQC spectra also contain signals from the NH_2 groups of the side chains of asparagine and glutamine and of the aromatic HN protons of tryptophan. The ^{1}H-^{15}N-HSQC experiment is a useful tool in detecting and studying interactions with ligands, such as other proteins or drugs. When a protein binds a ligand, the chemical shifts of the co-ordinating and nearby residues are perturbed (Williamson, 2013).

Isotope-labelling of a protein is necessary for the assignment of the backbone amino acid sequence. The isotopes used are ^{15}N and ^{13}C and the double isotope-labelled proteins can be investigated using three-dimensional assignment experiments. The three dimensional experiments together correlate the amide proton with the CA and the CB and the CO for a given residue and the adjacent residue, depending on the experiment. By sequentially assigning adjacent amino acid residues to the resonances of these chemical shifts it is possible to assign the backbone of the protein (Knowles, 2005). There are several experiments that are required to achieve this, which are described below (Figure 4.3).

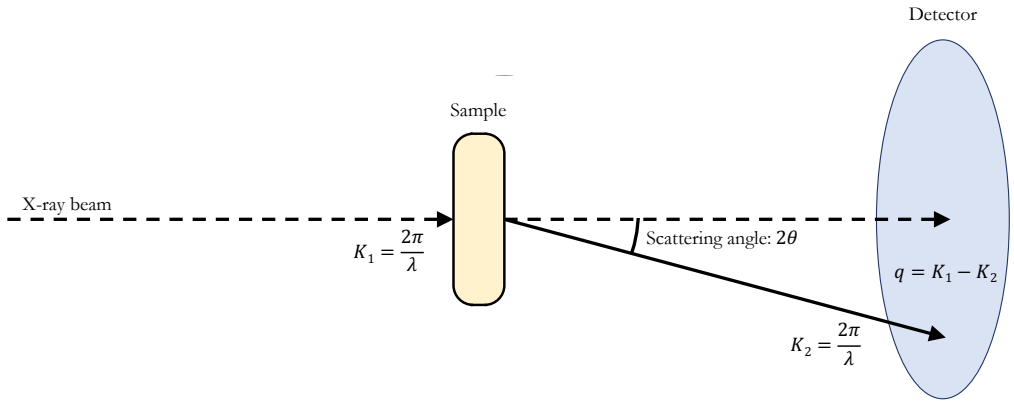
The HNCA experiment correlates the resonances of the amide proton, the nitrogen and the CA for a residue and CA for the adjacent residue can also been derived from this experiment. The partner experiment, HN(CO)CA experiment correlates the amide proton and the nitrogen of a residue, with the adjacent residue. Therefore, the HN(CO)CA experiment allows the distinction between the two CA carbons.

The HNCACB experiment correlates amide proton, the nitrogen, the CA and CB chemical shifts for residue. The HNCACB experiment can be used to observe and correlate the resonances of the CA and CB of residue. The ^{13}C dimension of the HNCACB spectrum has additional CB chemical shifts compared with the HNCA spectrum but is otherwise similar to the HNCA experiment. The HN(CO)CACB correlates amide proton and nitrogen chemical shifts for a residue, with those of the CA and CB for the adjacent residue. The spectrum of the HN(CO)CACB spectrum is therefore similar to the HN(CO)CACB spectrum but in the ^{13}C dimension there are additional CB chemical shifts.

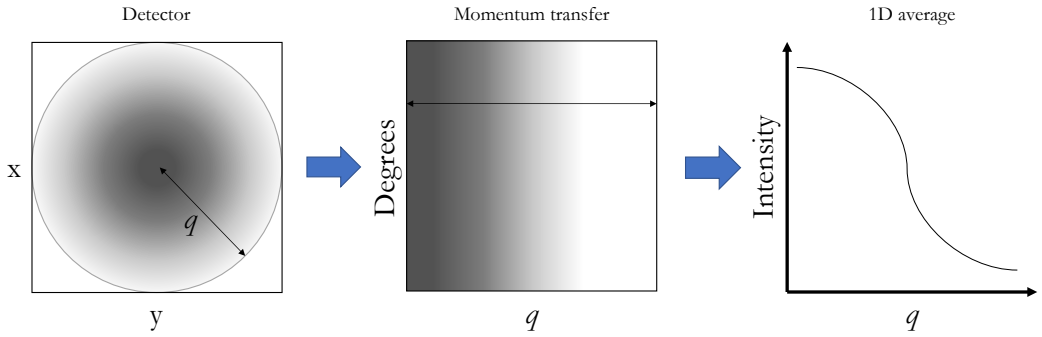
The HNCO and HN(CA)CO experiments both correlate the amide proton and nitrogen with that of the CO. The HNCO experiment correlates the amide proton and nitrogen with that of the CO of the adjacent residue and the HN(CA)CO experiment correlates the chemical shifts of the amide proton and nitrogen of a residue with the CO of the same residue with the adjacent residue.

Small-angle X-ray scattering

Small-angle X-ray scattering (SAXS) is a low resolution technique that can be used to study the overall shape and conformation of a biological macromolecules in solution. A protein sample in solution is illuminated with an X-ray beam. The sample scatters scattered X-ray intensity is measured by a detector (Figure 4.4a). The scattering pattern is related to the overall shape and size of the particles (Figure 4.4b). The scattering intensity, I is represented as a function of momentum transfer q, or the distance from zero that an X-ray is scattered. Using SAXS data, it is possible to model the structure of a protein using a program called CORAL (Petoukhov *et al.*, 2012). CORAL uses template structures and dummy residues to model multidomain proteins. Template domains are moved by increments and scattering intensity profiles are generated computationally. The model



(a) Schematic of a SAXS experiment



(b) The pattern of the scattering is related to the shape and size of the analysed particles

Figure 4.4: Overview of the SAXS experiment

During a SAXS experiment, the sample is illuminated by a monochromatic X-ray beam and the intensity of the scattered X-rays (K_2) is recorded by an X-ray detector. The scattering pattern of the pure solvent (K_1) is collected as well and subtracted from the sample solution scattering leaving only the signal from the particles of interest. The resulting scattering pattern is related to the overall shape and size of the particles under investigation ($q = K_1 - K_2$).

structure of the protein is generated from these computationally-generated SAXS profiles.

Circular dichroism

Circular dichroism spectroscopy (CD) measures the difference in absorption between left-hand and right-hand circularly polarised light by a chiral molecule (Kelly *et al.*, 2005). Chiral molecules are mirror-image isomers and all amino acids, apart from glycine, have chirality. The far-UV range (160 - 260 nm) is often used to investigate the secondary structure of proteins because the peptide bond, the side chains and the disulphide bonds absorb light in this region (Kelly *et al.*, 2005). One of the strengths of CD is that it is can be to investigate structural changes of a protein in response to denaturing conditions, pH, mutations in the amino acid sequence or even binding of a ligand.

In this chapter, a combination of structural techniques were used to study the structure of YecA. The data that is presented in this chapter suggests that YecA is a monomeric protein that has two distinct domains, which do not interact directly. In addition, using NMR spectroscopy, two conformations of the YecA-MeBD were identified. While the structure of YecA was not fully resolved, insight into the thermostability and the orientation of the two domains was gained, this chapter provides the basis for further structural analysis of YecA.

4.2 Results

4.2.1 A prediction model of the structure of YecA

Initially, the structure of YecA was investigated using Phyre2. Phyre2 is an online platform that builds a structural model of a protein using sequence alignments to proteins with known structures (Kelley *et al.*, 2015). The amino acid sequence of the *E.*

coli YecA (UniProt ID: P0AD05) was submitted to Phyre2. 90% of YecA was modelled with 90% confidence. The template structures that were used to model YecA are listed in Table 4.1. Phyre2 predicted that 71% of YecA would be α-helical and the remaining portion would be random coil. Figure 4.5 shows the predicted structure of YecA. The model has 7-αhelices that form the UPF0149 domain and a random coil that links the main body of the protein with the YecA-MeBD.

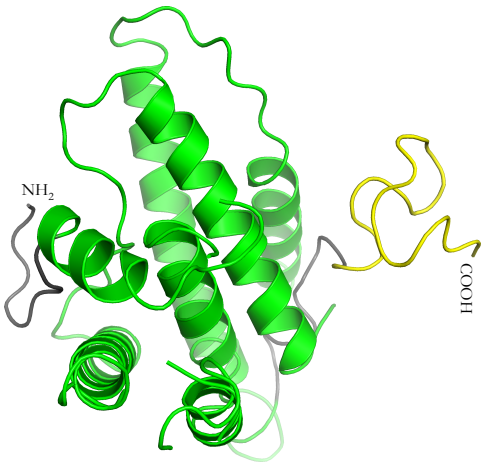
The templates that were used to model YecA are indicated in Table 4.1. The UPF0149 domain was modelled to a confidence of >90% using protein data bank structures 4GYT (Michalska *et al.*, 2012) and 1IZM (Galkin *et al.*, 2004). The YecA-MeBD was modelled to a confidence of >90% using protein data bank structures 2I9W (Joint Center for Structural Genomics, 2006), 1SX0 (Dempsey *et al.*, 2004), 1SX1 (Dempsey *et al.*, 2004), 1OZB (Zhou and Xu, 2003) and 1TM6 (Matousek and Alexandrescu, 2004). Residues 186 to 200 were modelled to a confidence of >80% using the protein data bank structure 1NNJ (Pereira de Jesus *et al.*, 2005).

Phyre2 aligned residues 69 - 154 of YecA to residues 721 - 880 of SecA from *Thermatoga maritima* (14% identity; PDB ID: 3JUX; Zimmer and Rapoport, 2009) and residues 51 - 154 of YecA to residues 703 - 825 of SecA from *Mycobacterium tuberculosis* (12% identity; 1NKT; Sharma *et al.*, 2003) and residues 8 - 154 of YecA to residues 790 - 928 SecA from *Thermus thermophilus* (13% identity; PDB ID: 2IPC; Vassylyev *et al.*, 2006) and residues 8 - 154 of YecA to residues 632 -770 of SecA from *B. subtilis* (15% identity; PDB ID: 1TF5; Osborne *et al.*, 2004). This region of SecA forms part of the SD and HWD. Phyre aligned residues 12 to 154 of YecA to residues 628 - 808 of SecA from *B. subtilis* (13% identity; PDB ID: 3DIN; Zimmer *et al.*, 2008), which corresponds to the flexible linker. The carboxy-terminal domain of SecA is exclusively α-helical. The sequence identity between these templates and the UPF0149 of YecA is between 12 and 15% but the confidence of homology for all alignments was >90%. This result suggested that the UPF0149 domain

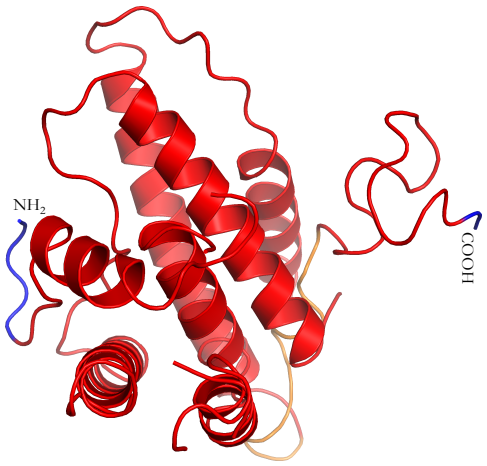
Table 4.1: Phyre2 sequence alignment of high confidence proteins to YecA

The amino acid sequence of *E. coli* YecA was submitted to Phyre2 for structural modelling. ⋆ indicates a template that was used to model the structure of YecA. The confidence is an indication that the YecA and the template sequence are homologous.

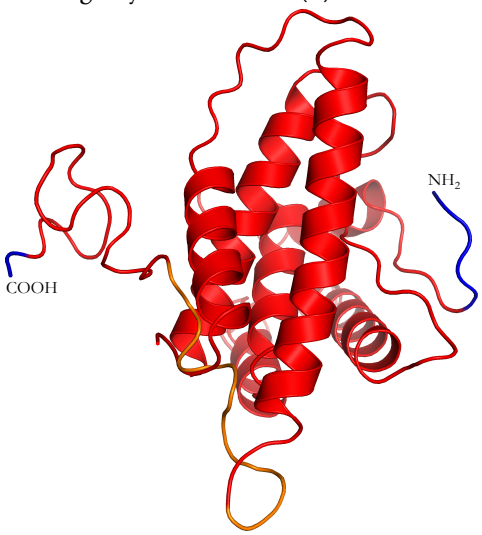
Template ID	Protein	Alignment to region of YecA	Confidence (%)	Reference
4GYTB⋆	Lpg0076	6 - 185	99.9	(Michalska *et al.*, 2012)
1IZMa⋆	YgfB-like	10 - 183	99.9	(Galkin *et al.*, 2004)
2I9WA	Psyc_2064	199-219	99.6	(Joint Center for Structural Genomics, 2006)
1SX0A⋆	SecA	200 - 219	99.4	(Dempsey *et al.*, 2004)
SX1A⋆	SecA	200 - 219	99.4	(Dempsey *et al.*, 2004)
1OZBI	SecA	201 - 219	99.3	(Zhou and Xu, 2003)
1OZBi	SecA	201 - 219	99.3	(Zhou and Xu, 2003)
1OZBJ	SecA	201 - 219	99.3	(Zhou and Xu, 2003)
1TM6A⋆	SecA	201 - 219	99.3	(Matousek and Alexandrescu, 2004)
1TM6B⋆	SecA	201 - 219	99.3	(Matousek and Alexandrescu, 2004)
2I9A3	uPA	148 - 211	97.7	(Lubkowski and Barinka, 2006)
2IPCB	SecA	8 - 154	95.4	(Vassylyev *et al.*, 2006)
3DINB	SecA	12 -154	95.3	(Zimmer *et al.*, 2008)
1TF5A2	SecA	8 - 154	95.3	(Osborne *et al.*, 2004)
1NKTA2	SecA	51 - 154	93.5	(Sharma *et al.*, 2003)
3JUXA	SecA	69 - 154	92.8	(Zimmer and Rapoport, 2009)
1NNJA⋆	MutM	148-211	84.5	(Pereira de Jesus *et al.*, 2005)



(a) Predicted model of YecA using Phyre2



(b) Predicted model of YecA using Phyre2



(c) Predicted model of YecA using Phyre2

Figure 4.5: Phyre 2 structural model to show the predicted orientation of the two domains of YecA

The amino acid sequence of YecA was submitted to Phyre 2 (Kelley *et al.*, 2015) to predict a structural model using homologous sequences with known structures as templates. 8 templates were used to predict the structure of YecA (Table 4.1). In Subfigure 4.5a UPF0149 (11 - 189) is shown in green and the YecA-MeBD (201 - 221) is shown in yellow and other residues (1 - 10; 190 - 200) are shown in grey. In Subfigures 4.5c and 4.5b red indicates that the structure was modelled with confidence of >90%, orange indicates that the structure was modelled with confidence of >80% and blue indicates that the structure was modelled *ab initio* and is unreliable (Kelley *et al.*, 2015). 8 residues were modelled *ab initio*. 90% of YecA was modelled with confidence of >90%. Subfigures 4.5c and 4.5b are 180° rotations of the predicted model of YecA. Subfigure 4.5a shows the predicted model of YecA in the same orientation as Subfigure 4.5b.

of YecA might be structurally similar to the C-domain of SecA.

4.2.2 Purification of full-length YecA for X-ray crystallography

To obtain a crystal structure of YecA, full-length YecA was expressed and purified for X-ray crystallography screen trials. The *yecA* gene sequence was amplified by PCR using `yecA_for` and `yecA_rev` and ligated into pCA528 using *BsaI* and *BamHI* restriction sites, and transformed into *E. coli* strain BL21 DE3. Bacteria were cultured as described in Section 2.5.4. The cells were lysed using high pressure homogenisation. The lysate was centrifuged to remove cell debris and subsequently incubated with Protino Ni-TED resin, which chelates the His_6 moiety of the amino-terminal tag of YecA. The recombinant YecA bound to the resin, although a small fraction was eluted by the wash steps (Figure 4.6). His_6-Ulp1 (SUMO protease) was added to the resin to cleave the amino-terminal tag from YecA. Ulp1 is a protease that recognises the tertiary structure of SUMO. Subsequently, the proteins were eluted using imidazole. The protein-containing fractions were analysed by SDS-PAGE, which showed that cleaved YecA, the His_6-SUMO tagged YecA and the cleaved tag all eluted together (Figure 4.6). SDS-PAGE analysis of these fractions also indicated the presence of contaminating proteins. Fractions E2 to E7 were pooled and concentrated for further purification.

Anion exchange chromatography was used to separate proteins according to their net surface charge. YecA eluted when the concentration of KOAc was between 600 and 700 mM (Figure 4.7a). Analysis of the elution fractions indicated that YecA co-purified with the His_6-SUMO tagged YecA (Figure 4.7b). Therefore, size exclusion chromatography was used to separate YecA from the recombinant variant of YecA. The gel filtration column was pre-equilibrated with Buffer 1 + 1 mM TCEP and the protein sample was loaded manually. The protein eluted in 5 ml fractions between 60 and 75 ml (Figure 4.8. The fractions were pooled after analysis by SDS-PAGE and concentrated to 10 mg/ml for crystallography screen trials.

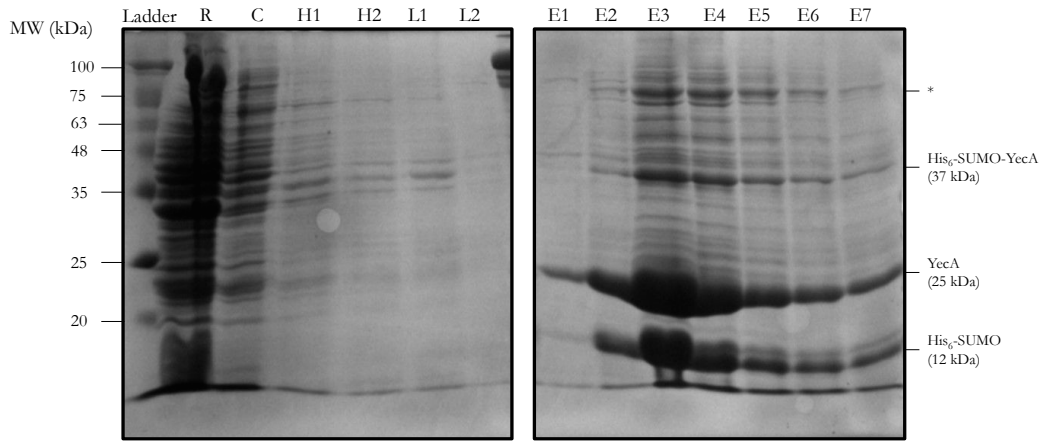
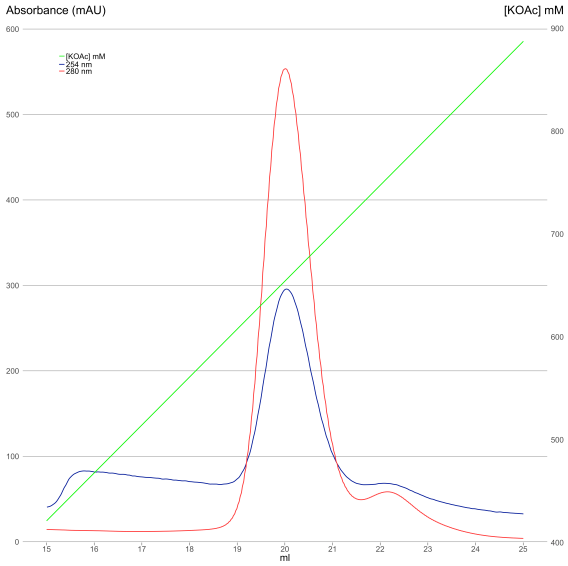
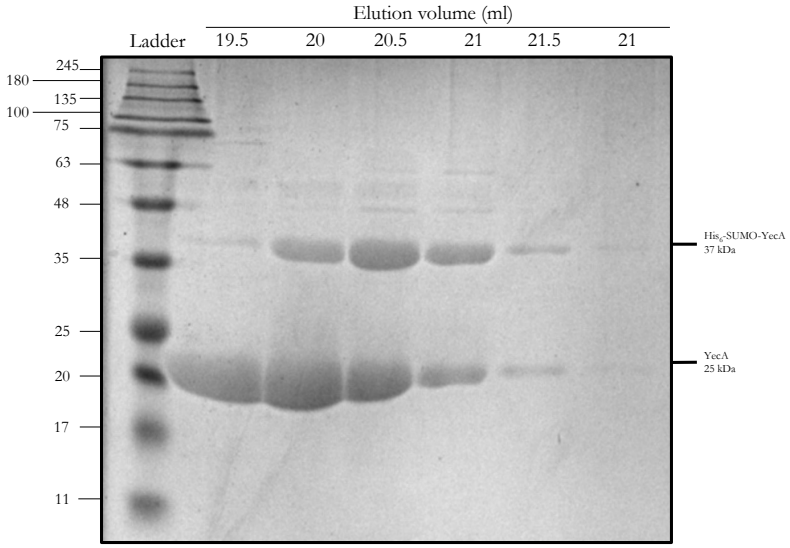


Figure 4.6: SDS-PAGE of YecA expression (*E. coli* strain BL21 DE3 pCA528-His_6-SUMO-YecA) and purification using Protino Ni-TED resin.

His_6-SUMO-YecA was expressed from a plasmid in *E. coli* strain BL21 DE3 that was grown in LB supplemented with kanamycin. The IPTG-induced cells were harvested and lysed. The lysate was clarified by centrifugation and filtration before incubation with Protino resin. The lysate was equilibrated with the lysate for 1 hour and subsequently washed with Buffer 1 (LW). Subsequently, the resin was washed with five column volumes of High Salt Buffer (HW). YecA was eluted in 1 ml fractions with 0.5 mM imidazole. The washes and imidazole elutions were analysed by SDS-PAGE and Coomassie staining. R: raw lysate, L: clarified lysate, H1-2: washes with Buffer 1 + 900 mM KOAc, L1-2: washes with Buffer 1, E1-7: 1 ml elutions with 500 mM imidazole from the Protino resin after incubation with SUMO protease. * indicates a protein of 75 kDa that copurified with YecA, which is possibly CdsA.



(a) Anion Exchange Chromatography Elution Profile of YecA for Crystallography Trials



(b) SDS-PAGE analysis of anion exchange chromatography elutions.

Figure 4.7: Anion exchange chromatography of YecA for crystal screens

After the amino-terminal tag had been cleaved, YecA was applied to an anion exchange column. The sample was applied manually to an anion exchange chromatography column that had be pre-equilibrated with Buffer 1. The flow rate was 1 ml/min. 4.7a shows the elution trace of the anion exchange chromatography of YecA201. Right hand-side *y*-axis: The concentration of KOAc was increased from 100 mM to 1 M (green). Left hand-side *y*-axis: The elution of YecA from the anion exchange column was measured spectrophotometrically at 280 nm (red) and 254 nm (blue). The elutions were collected in 0.5 ml fractions for analysis by SDS-PAGE. 4.7b shows the SDS-PAGE analysis of protein-containing fractions. 10 µl from each fraction between 19 and 21.5 ml was mixed with 10 µl Laemmli buffer and analysed by SDS-PAGE and Coomassie staining.

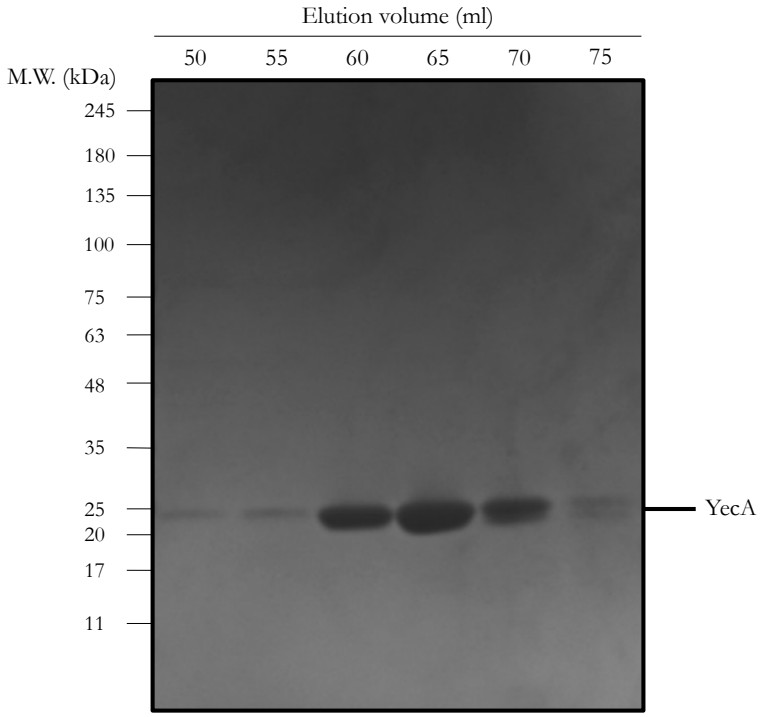


Figure 4.8: Size exclusion chromatography of YecA for crystal screens

Each protein was applied to an S75 10/300 GL column that had been pre-equilibrated with Buffer 1 + 1 mM TCEP. YecA and YecA201 were concentrated and injected onto the column. The flow rate was 0.5 ml/min and the elutions were collected in 5 ml fractions. 10 µl from each protein-containing fraction was mixed with 10 µl Laemmli buffer and analysed by SDS-PAGE and Coomassie staining. YecA resolved at approximately 25 kDa, which corresponds to its molecular weight. No contaminating bands or degradation products were resolved by SDS-PAGE. Fractions 60 to 70 ml were pooled and concentrated to 10 mg/ml for crystal screen trials.

4.2.3 X-ray Crystallography of full-length YecA

Purified YecA was screened in crystal screen matrices designed for soluble protein crystallisation. The commercial screens that were used were JCSG+, ProPlex, Morpheus HT-96, and Midas HT-96. Each screen consisted of 96 conditions. YecA was used at a concentration of 10 mg/ml. The protein sample was diluted in condition buffer at a ratio of 1:1 to form a sitting drop of 600 nl in 96-well crystallisation plates. After 1 month no crystals were observed.

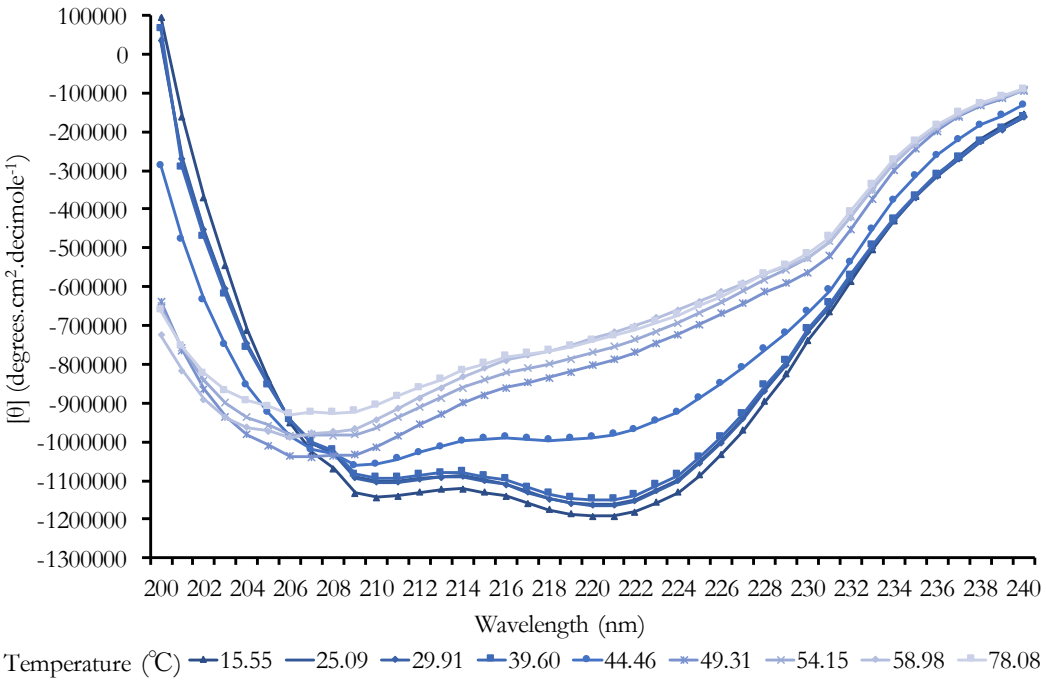
A brown-red precipitate was observed in several conditions in the JCSG+ screen: 0.1 MCAPS, pH 10.5, 40% v/v MPD (E5), 0.2M zinc acetate dihydrate, 0.0.5 mM imidazole, pH 8.0, 20% v/v PEG 3000 (E6) and 0.2 M zinc acetate dihydrate, 0.1 M sodium cacodylate, pH 6.5, 10% v/v 2-propanol (E7). A similarly brown-red precipitate was observed in 0.1 M magnesium acetate, 0.1 M MOPS, pH 7.5, 12% w/v PEG 8000 (E8), pH 7.0, 1.3 M Na/K hydrogen phosphate (G9), 0.1 M Na HEPES, pH 7.5, 1 M sodium acetate (G11), 0.1 M Na HEPES, pH 7.0, 1.5 M ammonium sulphate (F12) in the ProPlex screen. YecA is not predicted to bind any chemicals that would result in crystals of this colour and so it was thought that these precipitates were contaminants. Further analysis of these conditions was not pursued.

4.2.4 Analysis of the secondary structure of YecA as a function of temperature by circular dichroism

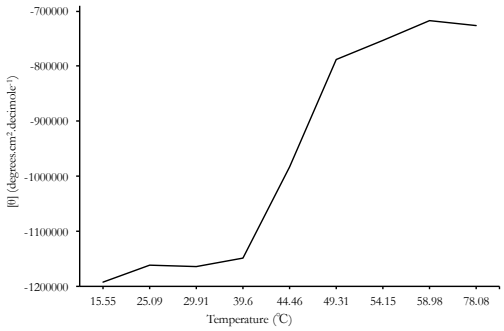
Circular dichroism spectroscopy (CD) can be used analyse changes to the secondary structure of protein in response to changes in temperature. NMR spectroscopy can be used to investigate the structure of macromolecules in solution. However, the temperature used for NMR spectroscopy experiments, which is 298 K (24.85°C), can affect protein stability (Kozak *et al.*, 2016). CD was used to investigate the thermal stability of YecA prior to NMR spectroscopy, and to gain an insight into the secondary structure of folded YecA.

Purified YecA was dialysed into 10 mM potassium phosphate buffer (pH 7.4) and analysed by CD. A buffer only control was also analysed. The buffer only control was subtracted from the CD spectra of YecA at each temperature and analysed by K2D (Andrade *et al.*, 1993; Whitmore and Wallace, 2004, 2008). K2D is an online resource that estimates the secondary structure content of a protein from the raw CD data. The buffer-corrected, K2D back-calculated CD spectra of YecA as a function of temperature is shown in Figure 4.9 (Andrade *et al.*, 1993). The NRMSD is a measure of the difference between the experimental ellipticities and the ellipticities of the back-calculated spectra for the derived structure (Table 4.3). NRMSD values ≤0.05 were considered to be a good fit. Table 4.3 shows that the back-calculated spectra for temperatures up to 58.98°C had NRMSD values that were ≤0.05. The back-calculated spectra of YecA at 58.98 and 78.08°C did not have such a good fit, according to the NRMSD value, which could be due to the aggregation of YecA at the higher temperatures.

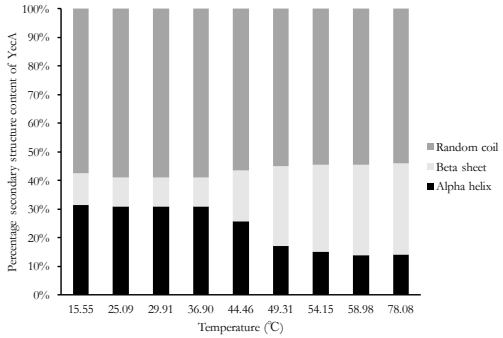
The CD spectra of YecA for the temperatures 15.55, 25.09, 29.91 and 39.6°C were similar with negative ellipticity peaks at 221 nm and 210 nm (Figure 4.9a). Analysis of these CD spectra by K2d indicated that the secondary structure of YecA was stable to 39.6°C because there was little change in the proportion of α-helical, β-sheet or random coil (Figure 4.9c). Figure 4.9a shows that the intensity of the negative ellipticity peak at 221 nm was decreased when the temperature was increased to 44.46°C. Figure 4.9b shows that between 39.6 and 49.31°C there a large decrease in the molar ellipticity at 221 nm, which indicated protein unfolding. The observed decrease in the negative ellipticity peak at 221 nm suggested that the α-helical portion of the YecA molecule was either decreased or destabilised or both as the temperature was increased to 44.46°C. Analysis of the CD spectrum of YecA at 44.46°C by K2D indicated that the proportion of α-helical structure had decreased and the β-sheet secondary structure had increased compared with the lower temperatures (Figure 4.9c). Figure 4.9 shows that the negative ellipticity



(a) Buffer-corrected and back-calculated CD spectra of 0.8 μM YecA in 10 mM potassium phosphate buffer as a function of temperature using K2D



(b) The change in mean residue molar ellipticity, $[\theta]$ (degrees.cm^2.decimole^{-1}), at 221 nm



(c) The proportion of the secondary structures of YecA, as calculated by K2D

Figure 4.9: Circular dichroism to analyse the thermostability of YecA

Purified YecA was dialysed into 10 mM potassium phosphate buffer and analysed by CD at different temperatures. The spectra were analysed by K2D (Andrade *et al.*, 1993) and the back-calculated results are shown in Subfigure 4.9a. The *y*-axis shows intensity as the mean residue molar ellipticity, $[\theta]$ (degrees.cm^2.decimole^{-1}). Subfigure 4.9b shows the change in molar ellipticity, $[\theta]$ (degrees.cm^2.decimole^{-1}), at 221 nm of the K2D analysis as a function of temperature. Subfigure 4.9c shows an analysis of the CD spectra of YecA as a function of temperature by K2D. The percentage of α-helix secondary structure is shown in black, while the percentage of β-sheet structure is shown in dark grey. The percentage content of random coil is shown in grey.

Table 4.3: Calculated NRMSD parameters of YecA CD spectra (Micsonai *et al.*, 2015)

Temperature (°C)	NRMSD (3 s.f.)
15.55	0.0199
25.09	0.0215
29.91	0.0180
36.90	0.0253
44.46	0.0224
49.31	0.0333
54.15	0.429
58.98	0.637
78.08	0.597

peak at 221 nm was most decreased in intensity for the four highest temperatures. The CD spectra of YecA at 49.31, 54.15, 58.98 and 78.08°C have lost most of the secondary structure of YecA compared with the protein at lower temperatures. The spectra of YecA at the four highest temperatures were similar.

The α-helical secondary structure of YecA was stable between 15..55°C and 39.6°C. The melting temperature of YecA was determined as 46.9°C and the results of the CD spectroscopy suggested that YecA was likely to be stable at 298 K (24.85°C) for analysis by NMR spectroscopy.

4.2.5 Preparation of ^{1}H-^{15}N-YecA

NMR spectroscopy was used to investigate the structure of YecA. To analyse the structure of a protein by NMR it is often necessary to isotopically label the protein with two isotopes, ^{13}C and ^{15}N, by growing the strain that expresses the gene of interest in minimal medium supplemented with $^{15}NH_4Cl$ and ^{13}C-glucose. ^{13}C-glucose is expensive and, at 25 kDa, YecA is close to the molecular weight limit of NMR spectroscopy. Therefore ^{1}H-^{15}N-YecA was purified for analysis by ^{1}H-^{15}N-HSQC to test the suitability of this protein for structural analysis by NMR spectroscopy (McIntosh and Dahlquist, 1990).

Initially, YecA was produced as a recombinant protein with a His_6-SUMO tag at the amino-terminus in *E. coli* strain BL21 DE3. The strain was grown in minimal media that was supplemented with $^{15}NH_4Cl$. The first purification step required the application of the cell lysate to a nickel-affinity column. The recombinant YecA bound to the column via the amino-terminal His_6-SUMO tag. YecA was eluted from the column using imidazole, which competes with the His_6 tag for binding to the nickel column. The imidazole elutions were analysed by SDS-PAGE, which showed that the recombinant YecA protein eluted from the nickel column with many other proteins (Figure 4.10).

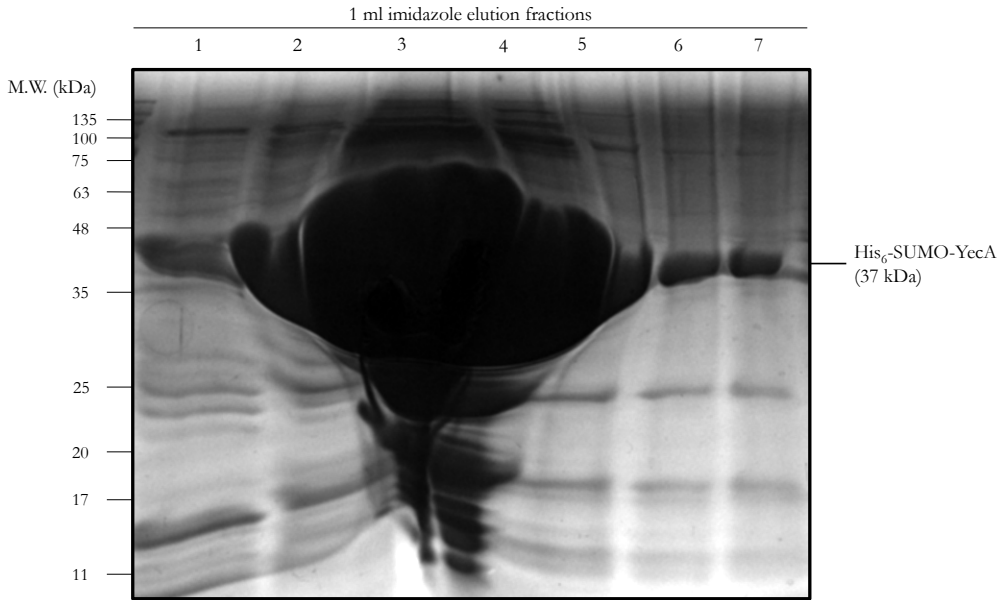
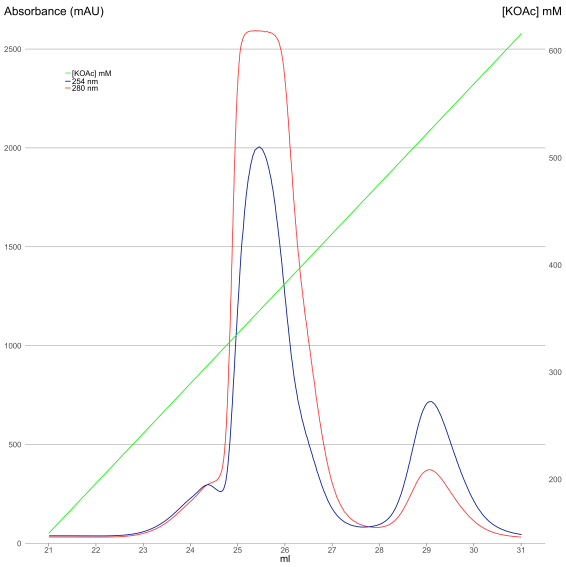
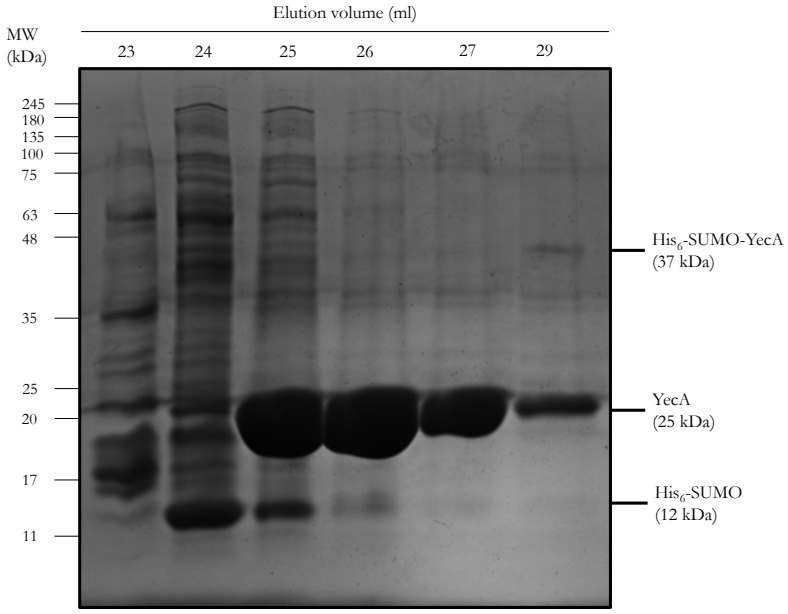


Figure 4.10: SDS-PAGE analysis of the nickel-affinity chromatography of ^{1}H-^{15}-N-His_6-SUMO-YecA

His_6-SUMO-YecA was expressed in *E. coli* strain BL21 DE3 in M9 media that was supplemented with $^{15}NH_4Cl$ and kanamycin. The IPTG-induced cells were harvested and lysed. The lysate was clarified by centrifugation and filtration before application to a nickel affinity column. The lysate was equilibrated with the lysate for 1 hour and subsequently washed with Buffer 1. YecA was eluted in 1 ml fractions with 0.5 mM imidazole. The imidazole elutions were analysed by SDS-PAGE and Coomassie staining. A significant amount of His_6-SUMO-YecA eluted in imidazole elution fractions 2-4 and these fractions were pooled and concentrated incubation with SUMO protease to remove the His_6-SUMO tag.



(a) UV-vis analysis of the purification of ^{1}H-^{15}N-YecA after anion exchange chromatography



(b) Anion exchange chromatography elution trace of ^{1}H-^{15}N-YecA

Figure 4.11: Anion exchange chromatography of ^{1}H-^{15}N-YecA for NMR spectroscopy

To purify cleaved YecA after the incubation with SUMO protease, ^{1}H-^{15}N-YecA was concentrated and applied to an anion exchange column. 4.11a: shows the elution trace of YecA. The anion exchange column was pre-equilibrated with Buffer 1, which contains 100 mM KOAc, and the protein sample was injected on to the column manually. A salt gradient from 100 mM KOAc to 1 M KOAc was used to elute the proteins according to their net charge and is shown in green (right *y*-axis).The absorbance (left *y*-axis) at 280 nm is in red and the absorbance at 254 nm is in blue. ^{1}H-^{15}N-YecA was eluted in 500 µl fractions from the anion exchange column between 23 and 28 ml when the concentration of KOAc in buffer 1 was 680 mM. 4.11b: shows the SDS-PAGE analysis of every other protein-containing fraction. Fraction 23 and 24 contained a small amount of YecA and several contaminating bands. Most of YecA eluted between 25 and 27 ml and these fractions were pooled for further purification.

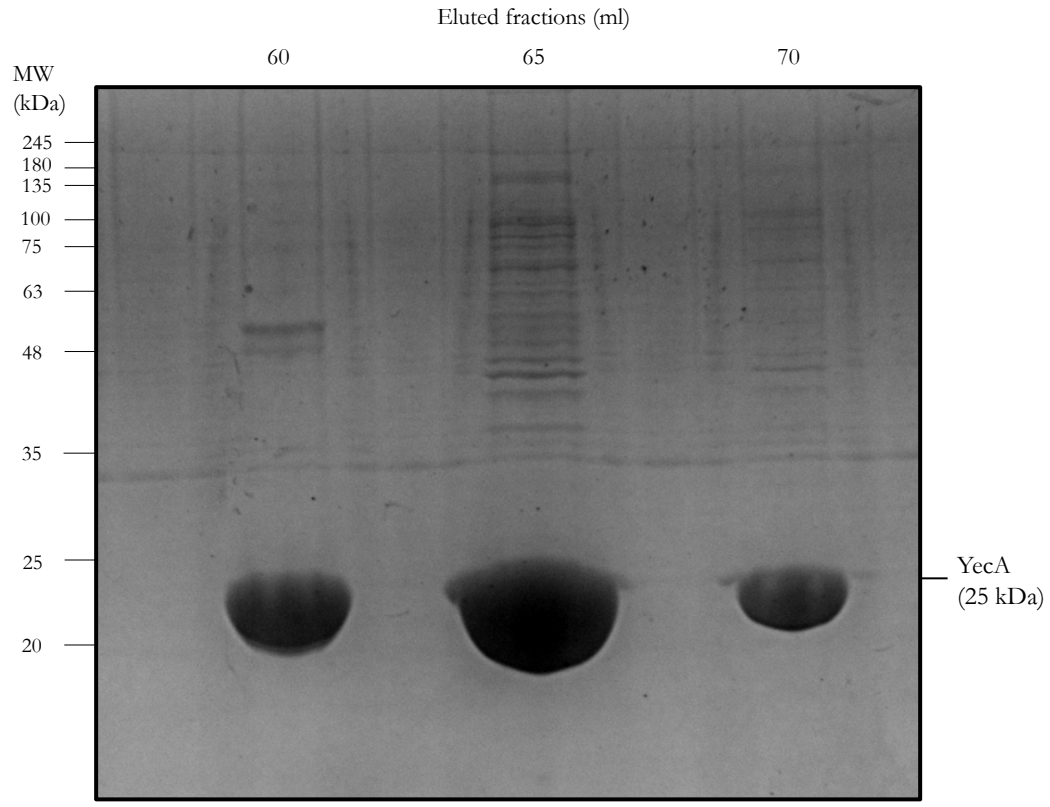


Figure 4.12: SDS-PAGE analysis of the size exclusion chromatography of ^{1}H-^{15}N-YecA

Size exclusion chromatography was used as a final purification step to remove co-purifying proteins and remove aggregates prior to NMR spectroscopy experiments. ^{1}H-^{15}N-YecA applied manually to a S75 10/300 GL column that had been pre-equilibrated with Buffer 1 supplemented with 1 mM TCEP. The flow rate was 0.5 ml/min and the elutions were collected in 5 ml fractions. 10 µl from each protein-containing fraction was mixed with 10 µl Laemmli buffer and analysed by SDS-PAGE and Coomassie staining. 10 µl of each protein-containing fraction was analysed for purity by SDS-PAGE. A final concentration of 1.1 mM ^{13}C-^{15}N-YecA was achieved.

The nickel-affinity column is normally blue due to the presence of nickel. After the application of His_6-SUMO-YecA to the column, the blue was replaced with a yellow-brown colour that remained after the elution of the protein (Figure 4.13). The crystallography screen had suggested the potential presence of a metal cofactor due to the brown-red precipitates formed in some of the crystal screen conditions. The colours associated with YecA suggested that this metal cofactor could be iron.

The amino-terminal tag was cleaved by incubating the pooled protein fractions with SUMO protease (His_6-Ulp1). The cleaved YecA was separated from the tagged YecA and the protease by applying the sample to a nickel-affinity column. The cleaved YecA cannot not bind to the column and flows through the column. The flow through was applied to an anion exchange chromatography column. YecA eluted from the column as the concentration of KOAc increased from 300 to 450 mM (Figure 4.11a) and these fractions were analysed by SDS-PAGE analysis, which showed that most of YecA had eluted between 23 -27 ml but several contaminating proteins were also resolved (Figure 4.11b). The protein-containing fractions were pooled and applied to a size-exclusion chromatography column. Figure 4.12 shows that YecA eluted between 60 and 70 ml with a small amount of contaminating protein that only resolved after 48 h in Coomassie stain. This result shows that YecA was >90% pure for analysis by ^{1}H-^{15}N-HSQC NMR spectroscopy experiments. Subsequently, the protein was concentrated to 1.1 mM in 20 mM [MES]-NaOH [pH 6.5] 10 mM NaCl.

4.2.6 NMR ^{1}H-^{15}N-HSQC of YecA

NMR ^{1}H-^{15}N-HSQC experiments were used to assess whether YecA was folded and to asses the quality of the spectrum prior to pursuing more expensive techniques. The ^{1}H-^{15}N-HSQC spectrum of 1.1 mM YecA showed that YecA was folded because the resonances are well-dispersed (Figure 4.14). 170 resonance peaks were identified. YecA con-

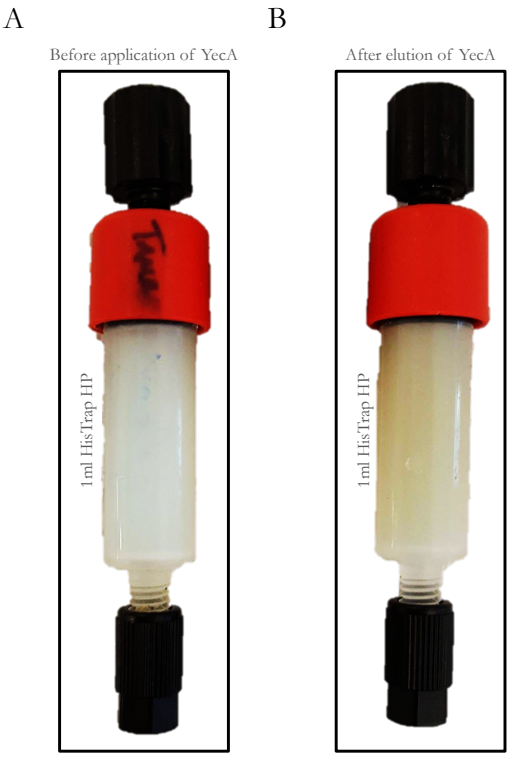


Figure 4.13: The 1 ml HisTrap column before and after nickel affinity chromatography of YecA

His_6-SUMO-YecA was purified initially by nickel affinity chromatography. Panel A: 1 ml HisTrap column before application of YecA. Panel B: 1 ml HisTrap column after YecA was eluted from the column with 0.5 mM imidazole.

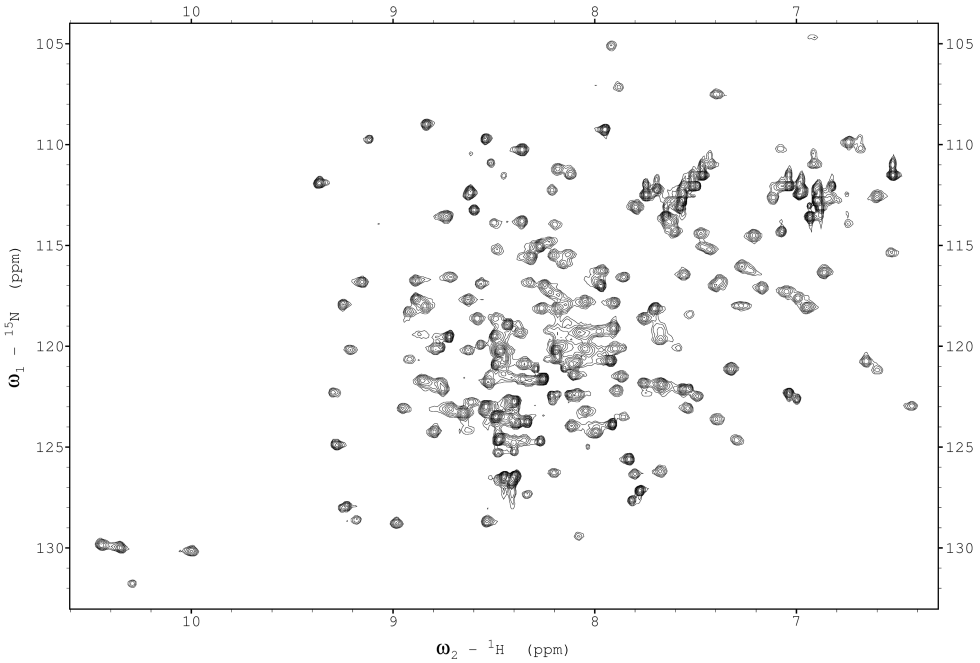


Figure 4.14: ^{1}H-^{15}N-HSQC spectrum of YecA

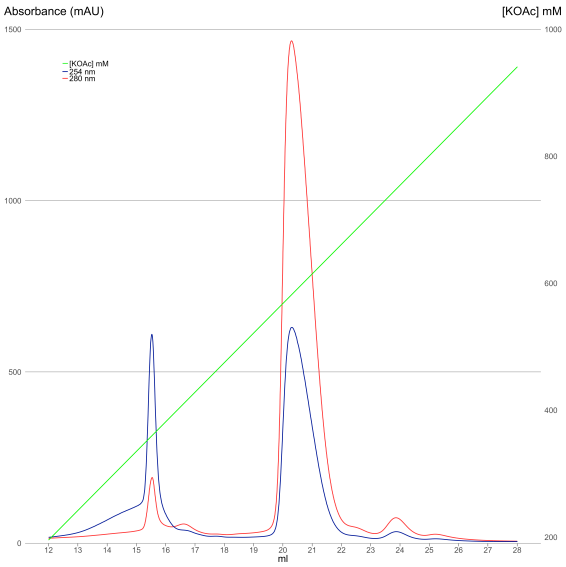
^{1}H-^{15}N-YecA was dialysed into 20 mM [MES]-NaOH [pH 6.5], 10 mM NaCl and concentrated to 1.1 mM for analysis by ^{1}H-^{15}N-HSQC NMR spectroscopy using the Bruker 900 mHz instrument. Each peak represents a ^{1}H attached to a ^{15}N.

tains 221 amino acid residues of which there are 24 with a NH_2 in the side chain and 14 prolines, which means that approximately 70% of the protein was visible. This result suggested that the structure of YecA could be investigated using NMR spectroscopy.

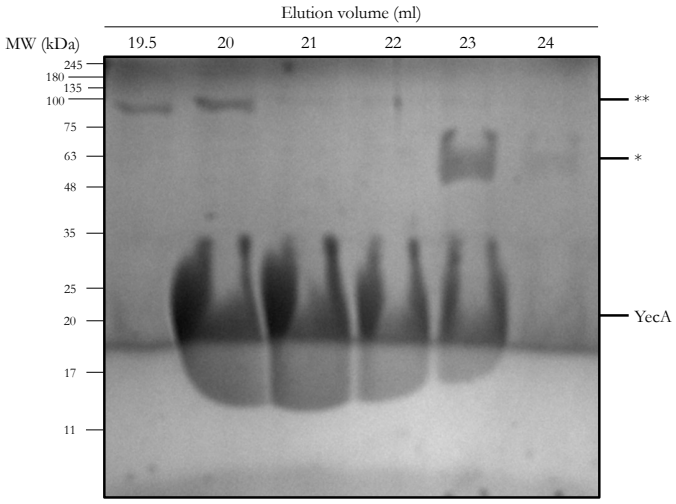
4.2.7 Preparation of ^{13}C-^{15}N-YecA for structural analysis by NMR spectroscopy

To analyse the structure of YecA, ^{13}C-^{15}N-YecA was prepared as described in Section 2.5.4 and 2.6.2 for analysis by NMR spectroscopy. Briefly, M9 media supplemented with $^{15}NH_4Cl$ and ^{13}C-D-glucose was used to grow *E. coli* strain BL21 DE3 that expressed recombinant YecA with a His_6-SUMO tag at its amino-terminus. Cells were grown to an OD_{600} of 0.8 and induced with IPTG overnight at 18°C. The cells were lysed using high pressure homogenisation. The lysate was centrifuged and filtered to remove cell debris and protein aggregates and applied to a nickel-affinity column that had been pre-equilibrated with Buffer 1. The column was washed and ^{1}H-^{15}N-His_6-SUMO-YecA was eluted from the column with imidazole. The washes and the imidazole elutions were analysed by SDS-PAGE and Coomassie staining (Figure 4.16). The gel showed that a major band resolved between 35 and 48 kDa, which corresponded to the molecular weight of recombinant YecA (37 kDa). The His_6-SUMO tag was cleaved by incubating overnight with His_6-Ulp1, a SUMO protease.

After the protein had been incubated with His_6-Ulp1, the sample was concentrated and anion exchange chromatography was used to separate the proteins according to their net surface change. The anion exchange column was pre-equilibrated with Buffer 1, which contains 100 mM KOAc, and the protein sample was injected on to the column manually. The concentration of KOAc in Buffer 1 was increased linearly to 1 M. YecA was eluted in 500 µl fractions from the column when the concentration of KOAc in buffer 1 was 680 mM (Figure 4.15a). The protein-containing fractions were analysed by SDS-PAGE and Coomassie staining (Figure 4.15b). Lanes 1 and 2 (elution volume



(a) Anion exchange chromatography elution trace of ^{13}C-^{15}N-YecA



(b) SDS-PAGE analysis of ^{13}C-^{15}N-YecA after anion exchange chromatography

Figure 4.15: Anion exchange chromatography of ^{13}C-^{15}N-His$_6$-YecA for structural analysis by NMR spectroscopy

To purify cleaved YecA after the incubation with SUMO protease, ^{13}C-^{15}N-YecA was concentrated and applied to an anion exchange column. The anion exchange column was pre-equilibrated with Buffer 1, which contains 100 mM KOAc, and the protein sample was injected on to the column manually. 4.15a shows the elution trace of YecA from the anion exchange column as the concentration of KOAc increased 100 mM KOAc to 1 M KOAc. The salt gradient that was used to elute the proteins according to their charge and is shown in green (right *y*-axis). The absorbance (left *y*-axis) at 280 nm is in red and the absorbance at 254 nm is in blue. ^{13}C-^{15}N-YecA was eluted in 500 µl fractions from the anion exchange column between 19.5 and 22 ml when the concentration of KOAc in buffer 1 was 680 mM. 4.15b: The protein-containing fractions (19.5 - 24 ml) were analysed by SDS-PAGE and Coomassie staining. YecA co-purified with a protein of approximately 100 kDa (Fraction 19.5 ml and 20 ml). YecA co-purified with a protein of approximately 63 kDa (Fraction 23 ml and 24 ml). Fractions that eluted between 20 ml and 22 ml were pooled and concentrated for further purification. * indicates an unknown protein of approximately 100 kDa that co-purified with YecA and ** indicates a protein an unknown protein of approximately 63 kDa that co-purified with YecA.

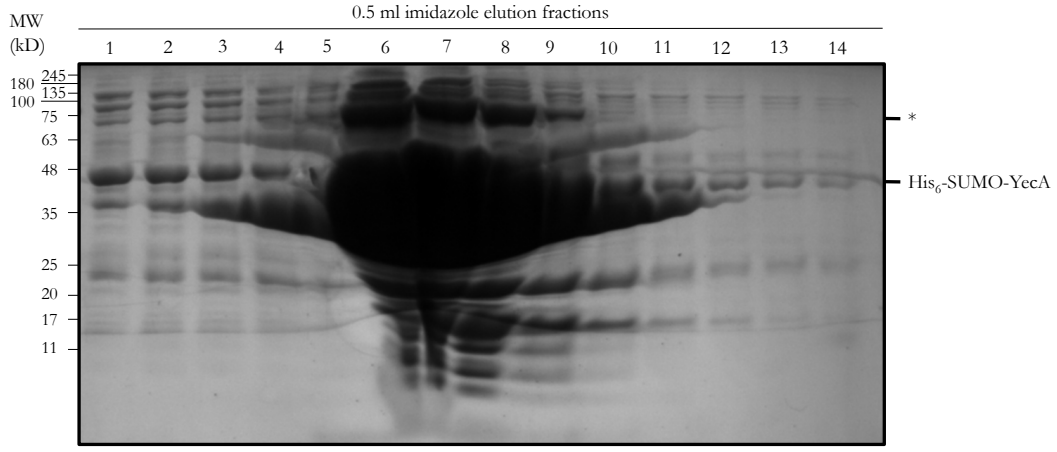


Figure 4.16: Nickel affinity chromatography of ^{13}C-^{15}N-His_6-SUMO-YecA for structural analysis by NMR spectroscopy

His_6-SUMO-YecA was expressed in *E. coli* strain BL21 DE3 in M9 media that was supplemented with $^{15}NH_4Cl$ and ^{13}C-D-glucose. Expression was induced overnight at 18°C. The IPTG-induced cells were harvested and lysed by high pressure homogenisation. The lysate was clarified by centrifugation and filtration to remove cell debris and aggregates before application to a nickel affinity column. A flow rate of 0.5 ml/min was used and the lysate passed over the column five times before the column was washed with five column volumes of Buffer 1 and five column volumes of High Salt Buffer. ^{13}C-^{15}N-His_6-SUMO-YecA was eluted with 0.5 mM imidazole in 0.5 ml fractions. The imidazole elutions were analysed by SDS-PAGE and Coomassie staining. ^{13}C-^{15}N-His_6-SUMO-YecA eluted in fractions 3 to 11, which was equivalent to an elution volume of between 1.5 and 8.5 ml. These fractions were pooled and incubated with SUMO protease to cleave the His_6-SUMO tag. Several co-purifying proteins also eluted with ^{13}C-^{15}N-His_6-SUMO-YecA. * indicates a co-purifying protein that resolved at approximately 75 kDa, which is possibly CdsA.

19.5 and 20 ml) of Figure 4.15b show that YecA had co-purified with a protein of approximately 100 kDa. Lanes 5 and 6 (Fraction 23 ml and 24 ml) of Figure 4.15b show that YecA had co-purified with a protein of approximately 63 kDa. The fractions that collected the elution volume of between 20 and 22 ml were pooled and concentrated and applied to a size exclusion column. After size exclusion chromatography, the protein-containing fractions were analysed by SDS-PAGE and Coomassie staining. Figure 4.17 shows that ^{13}C-^{15}N-YecA eluted between 55 and 80 ml, with the majority of the protein eluting in one fraction at 60 ml. Two co-purifying bands were resolved only after 48 h in Coomassie stain. One of the bands resolved at 37 kDa and it is possible that this protein was ^{13}C-^{15}N-His$_6$-SUMO-YecA. The other protein resolved at 48 kDa and is unidentified. The ^{13}C-^{15}N-YecA was approximately 95% pure and the protein was dialysed into 20 mM [MES]-NaOH [pH 6], 10 mM NaCl for analysis by NMR spectroscopy.

4.2.8 Assignment of the backbone of ^{13}C-^{15}N-YecA

To investigate the structure of the YecA-MeBD from NMR spectroscopy experiments, each of the resonances[1] were assigned to the backbone sequence of the protein. The first stage of the assignment process is clustering, which involves tracing each of the resonances that correspond to a particular residue. The HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, HNCO, and HN(CA)CO provide information about the chemical shifts of carbon atoms of the amino acid residue (CA, CB and CO) with respect to the hydrogen and the nitrogen of the backbone amide group. It is then necessary to identify the sequential arrangement of the amino acid residues. Each cluster provides information about the adjacent cluster, which facilitates the mapping of the residues. Using the program Sparky, it was possible to align the inter-residue carbon shifts of one

[1]The ^{1}H^{15}N-HSQC NMR spectrum of apo-YecA201 shown in Figure 5.8 was used to identify resonances in the ^{1}H^{15}N-HSQC NMR spectrum of the full-length YecA that were likely to correspond with the YecA-MeBD.

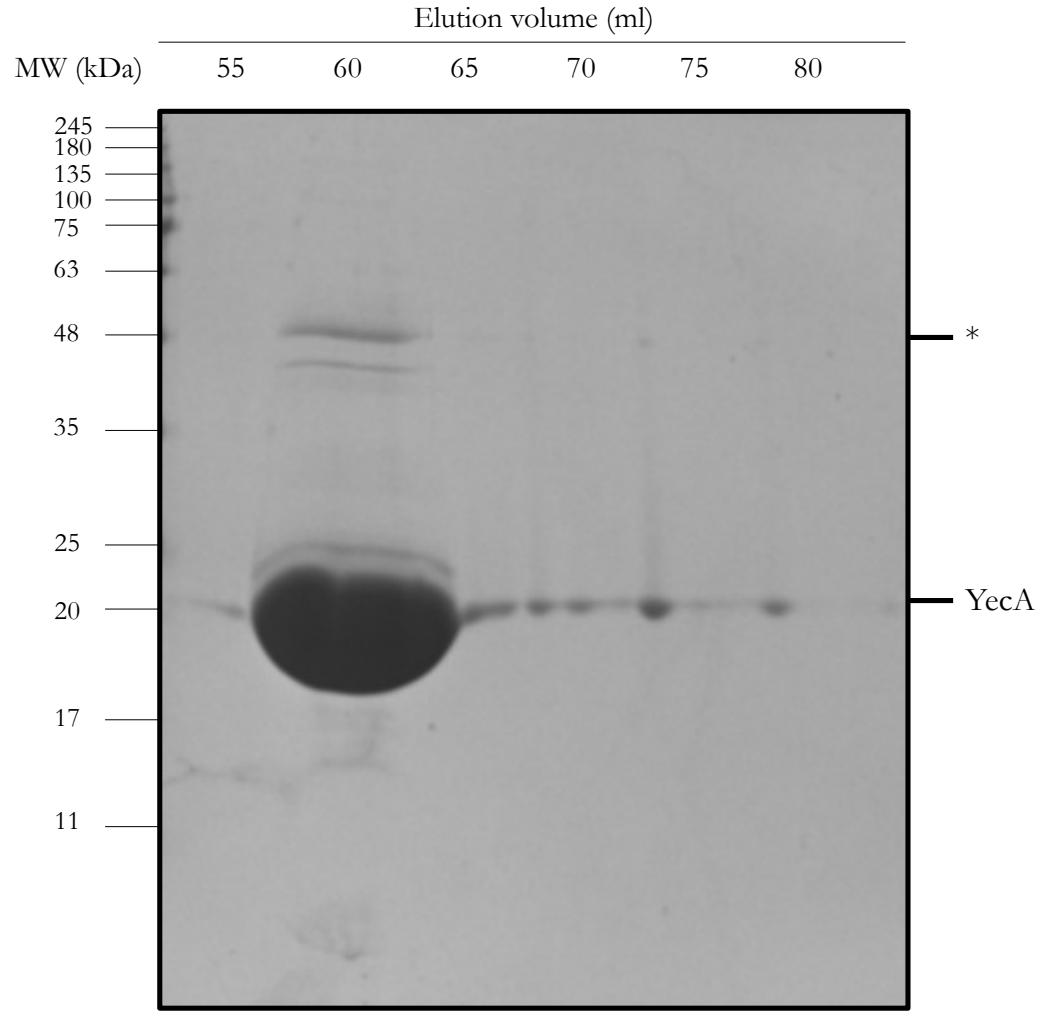


Figure 4.17: SDS-PAGE analysis of ^{13}C-^{15}N-YecA after gel filtration

Size exclusion chromatography was used as a final purification step to remove co-purifying proteins and remove aggregates prior to NMR spectroscopy experiments. ^{13}C-^{15}N-YecA applied manually to a S75 10/300 GL column that had been pre-equilibrated with Buffer 1 supplemented with 1 mM TCEP. The flow rate was 0.5 ml/min and the elutions were collected in 5 ml fractions. 10 µl from each protein-containing fraction was mixed with 10 µl Laemmli buffer and analysed by SDS-PAGE and Coomassie staining. 10 µl of each protein-containing fraction was analysed for purity by SDS-PAGE. A final concentration of 0.5 mM ^{13}C-^{15}N-YecA was achieved. *indicates two co-purifying proteins that resolved at 48 kDa and at 37 kDa. It is possible that the 37 kDa protein was His_6-SUMO-YecA, which is 37 kDa, while the other is unknown.

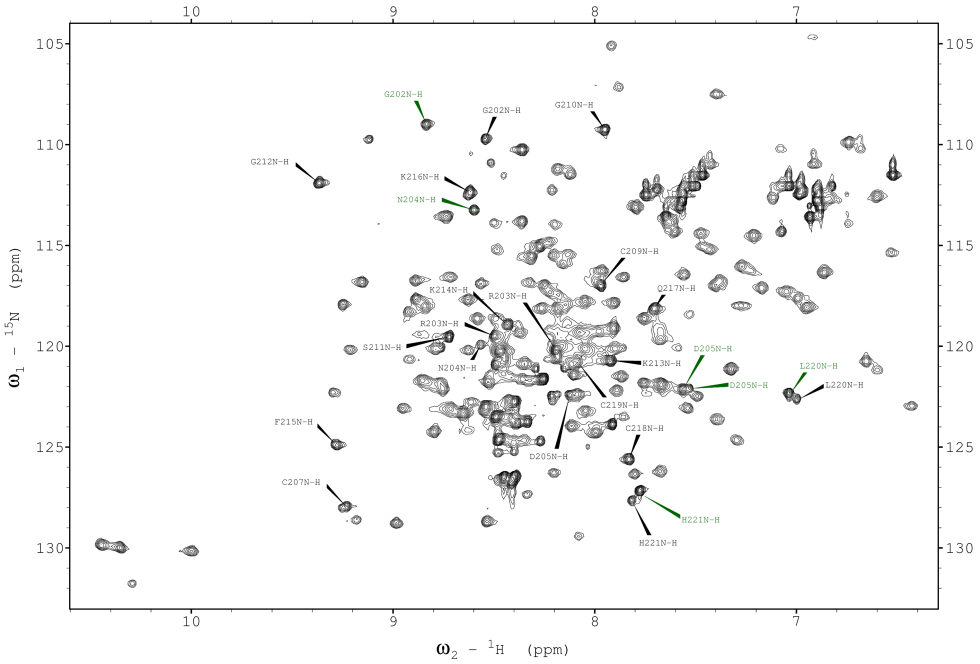


Figure 4.18: ^{1}H-^{15}N backbone assignments of ^{13}C-^{15}N-YecA

^{13}C-^{15}N-YecA, at a concentration of 0.5 mM in 20 mM [MES]-NaOH [pH 6.0], was analysed by ^{1}H-^{15}N-HSQC. The spectrum was recorded on a Bruker 900 mHz spectrometer. The chemical shift information from the HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, HNCO and HN(CA)CO experiments was correlated using clustering and then sequentially assigned using Sparky software. Black labels indicate the fully assigned conformation of the YecA-MeBD and the green labels indicate the second, partially assigned conformation of the YecA-MeBD.

cluster to the intra-residue carbon shifts of the adjacent cluster. Afterwards, the next cluster in the sequence was identified in the same way. In theory, this process can be repeated by assigning adjacent clusters to a residue until the full back-bone of the protein is assigned. However, proline is invisible because it does not have an amide proton in its backbone and some resonances are very poorly resolved. It is common to have fragments of backbone assignments (Knowles, 2005). After a sequence of resonances have been assigned, it is necessary to position the fragment to the correct location on the polypeptide sequence. The reposition function within Sparky was used identify the likely locations in the protein backbone for the fragment. The amino acid residues with characteristic chemical shift patterns were used to manually check the recommended position. Examples of amino acid residues with distinctive chemical shift patterns include serine (CA 63 ppm, CB, 67 ppm), threonine (CA 63 ppm, Cb 70ppm), alanine (CB 18 ppm) and glycine (CA 45 ppm).

The ^{1}H-^{15}N-HSQC spectrum of ^{1}H-^{15}N-YecA is shown in Figure 4.18. The aim of this process was to assign the backbone of the YecA-MeBD. 24 assignments were mapped to this region of YecA, which is only 20-amino acid residues long and contains two proline residues. These assignments were independently checked[2] and would not fit anywhere else in the protein backbone of YecA. The assignments suggested there were two YecA-MeBD species present in the sample. The entire YecA-MeBD was fully assigned in one of these conformations (G202-D205; C207; C209-H221), while the other conformation was only partially assigned (G202-D205; C219-H221). The two species showed considerable differences between their chemical shift patterns.

4.2.9 Small angle X-ray scattering

Small angle X-ray scattering (SAXS) was used to develop an understanding of the conformation of YecA in solution. The BM29 beamline at ESRF, Grenoble was used

[2]With many thanks to Dr. T. Knowles

to collect all SAXS data[3]. Data was collected every second from samples as they eluted from an in-line size exclusion chromatography column. YecA eluted from the column between 1080 and 1150 s (Figure 4.19). This was consistent with the elution of a protein of 25 kDa, which suggested that YecA was a monomer under these conditions (Dr. T. Knowles, personal communication). The SAXS profiles between 1087 and 1108 s were selected for further analysis to ensure the data was consistent. These profiles were buffer subtracted and the resulting SAXS profiles were averaged for further analysis (Figure 4.20).

Guinier analysis of the SAXS data was used to investigate whether the protein was aggregated or homogenous. The Guinier analysis is the scattering intensity curve in the low q region, which can be used to determine the radius of gyration (R_g). The linearity of the Guinier plot (Figure 4.21) indicated that the protein was not aggregated, which indicated that the data was of good quality and could be used for further analysis. The R_g was 2.16 ± 0.03 nm, which was determined automatically from the Guinier analysis using PRIMUS software (Konarev *et al.*, 2003).

SAXS MoW2 is an online, open-source platform that can be used to estimate the molecular weight of a protein from its SAXS data (Fischer *et al.*, 2010). SAXS MoW2 analysis of the averaged SAXS data of YecA indicated that the protein had a molecular weight of 33.217 kDa. The molecular weight of YecA is 25 kDa and this analysis was consistent with the elution profile that also suggested that YecA was monomeric.

Kratky analysis was used to determine whether YecA was folded. If a protein is folded, a bell-shaped peak in the low q region that converges to the q axis in the high q region (Mertens and Svergun, 2010). A protein that has high flexibility lacks this bell-shaped peak and instead has a plateau in the high q region. Multi-domain proteins sometimes

[3]Data was collected by Dr. Tim Knowles and Dr. Mohammed Jamshad

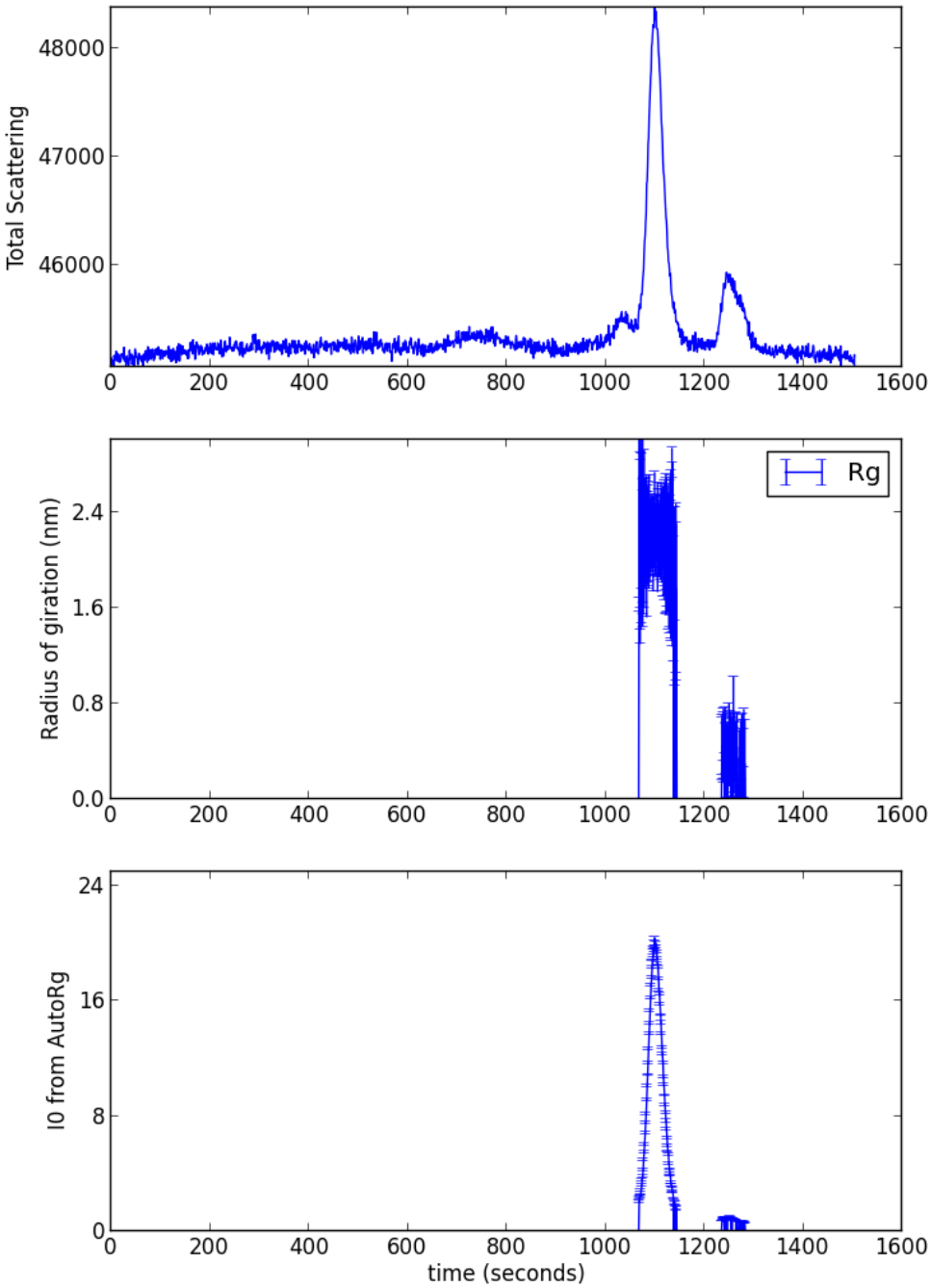


Figure 4.19: In-line gel filtration profile of YecA as a function of time

The total scattering (a.u.) of YecA as a function of time (s) was used to identify the elution of YecA. The scattering intensity between 1080 s and 1150 s was collated and used for further analysis. The radius of gyration (Rg) as a function of time was determined automatically at the beamline using the Guinier approximation.

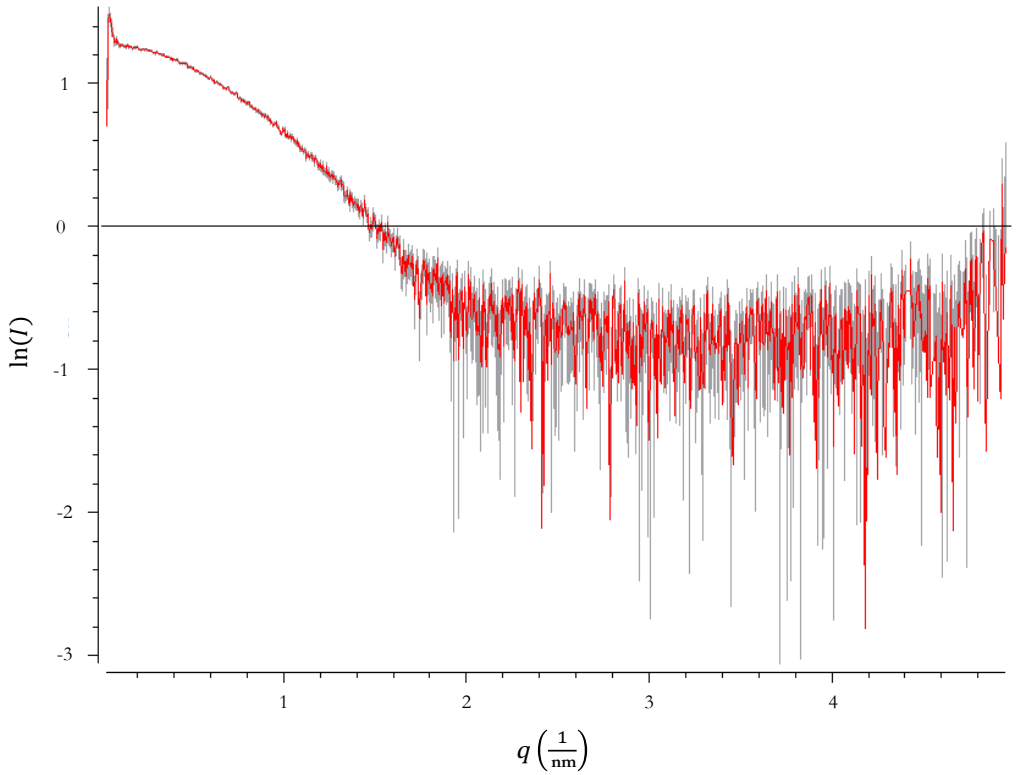


Figure 4.20: Average SAXS profile of YecA

The average log scattering intensity as a function of q of YecA. q is the vector of momentum transfer, or the distance the X-ray is scattered.

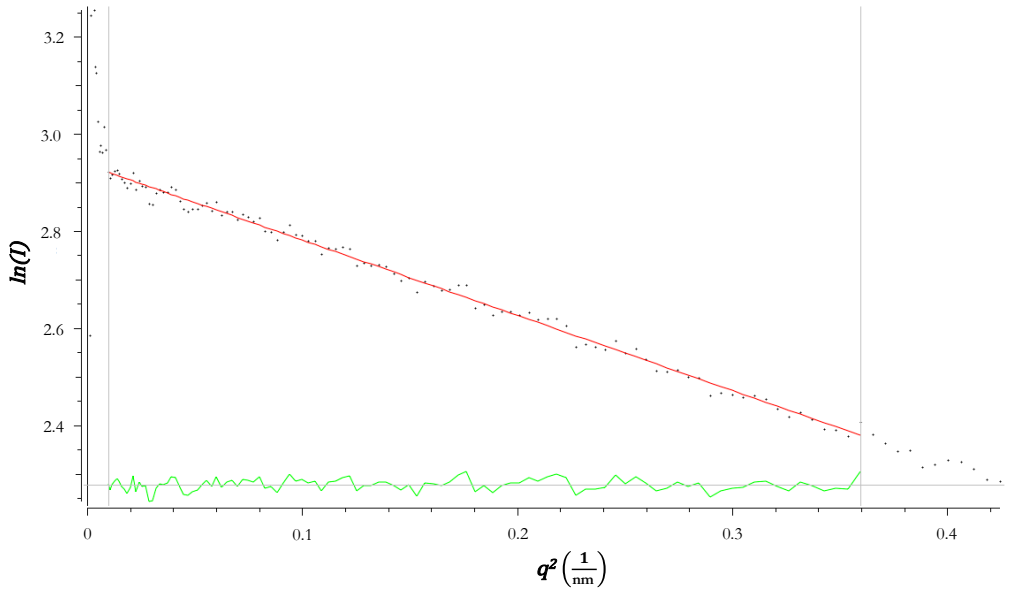


Figure 4.21: Guinier analysis of the low q region of the SAXS scattering curve of YecA

The Guinier plot was determined automatically using PRIMUS software (Konarev *et al.*, 2003). The linearity of the Guinier plot indicated that YecA was folded and not aggregating. The R_g was 2.16 ±0.03 nm.

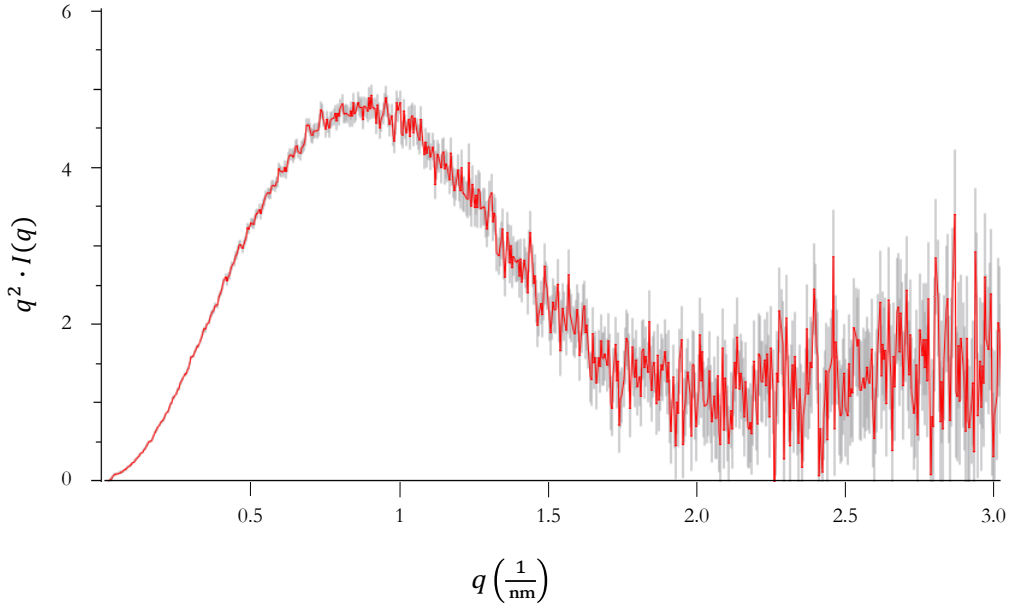
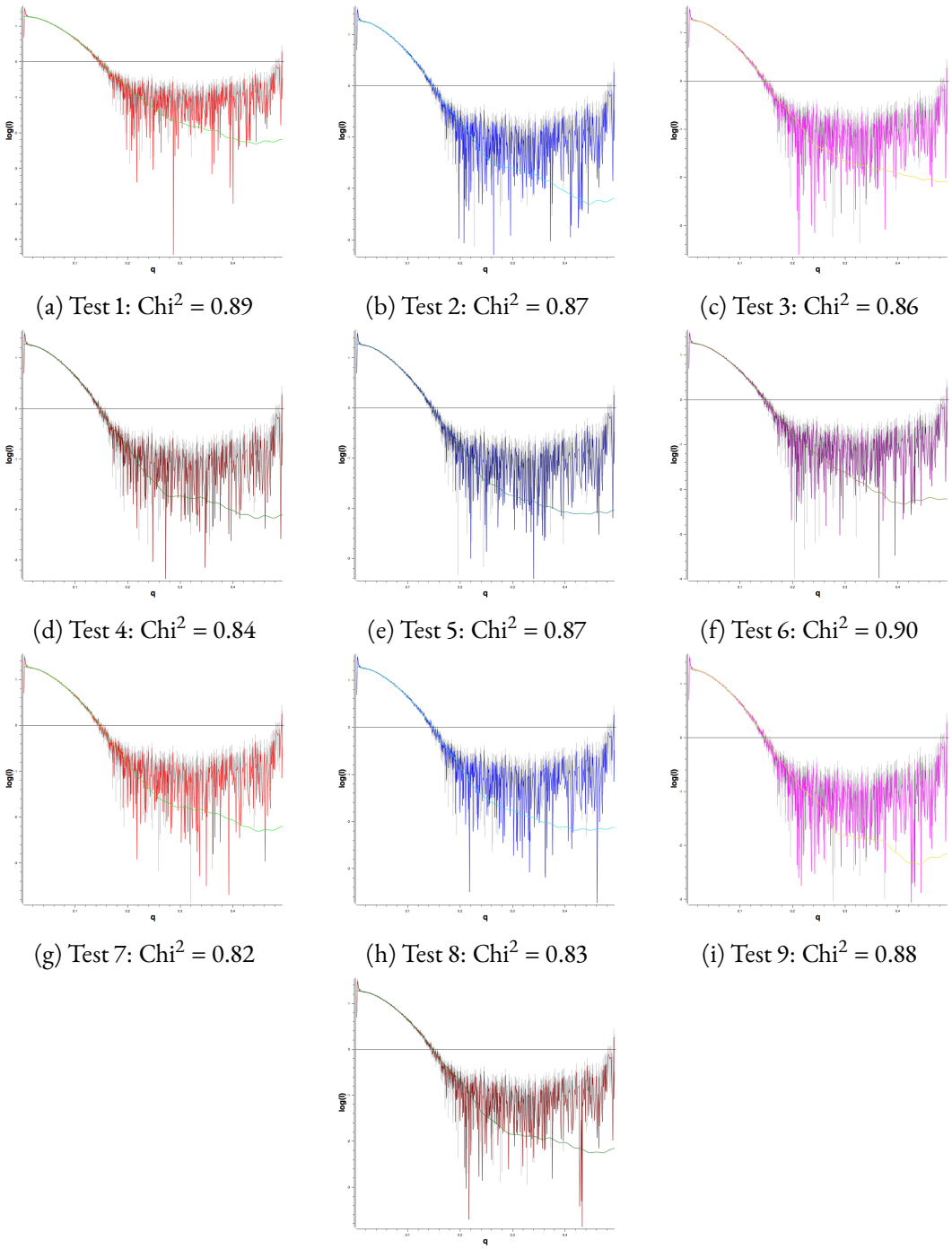


Figure 4.22: Kratky plot of the SAXS scattering curve of YecA

PRIMUS was used to generate a Kratky plot to ensure that YecA was folded. The $q^2 \times I(q)$ emphasised the shape at low intensities. A bell-shaped curve is indicative of a folded protein.



(a) Test 1: Chi^2 = 0.89 (b) Test 2: Chi^2 = 0.87 (c) Test 3: Chi^2 = 0.86

(d) Test 4: Chi^2 = 0.84 (e) Test 5: Chi^2 = 0.87 (f) Test 6: Chi^2 = 0.90

(g) Test 7: Chi^2 = 0.82 (h) Test 8: Chi^2 = 0.83 (i) Test 9: Chi^2 = 0.88

(j) Test 10: Chi^2 = 0.88

Figure 4.23: The fits of 10 independent CORAL models

YecA was analysed by SAXS using an in-line size exclusion chromatography column. The orientation of the two domains of YecA was modelled using the structures of two homologues, PDB ID: 1IZM (Galkin *et al.*, 2004) and PDB ID: 1TM6 (Matousek and Alexandrescu, 2004). Residues 11 to 192 were modelled using 1IZM and residues 201 to 221 were modelled using 1TM6. The remainder of the protein was modelled using dummy molecules. CORAL was to compare the scattering profile of YecA with the theoretical scattering profile of the model. The comparison of the experimental data with the theoretical fit is shown for ten models created with CORAL (a - j) where I is the intensity of scattering in arbitrary units and q is the scattering vector (Å^{-1}). The Chi^2 shows a good fit for all of the test fits (Chi^2 = 0.82-0.90).

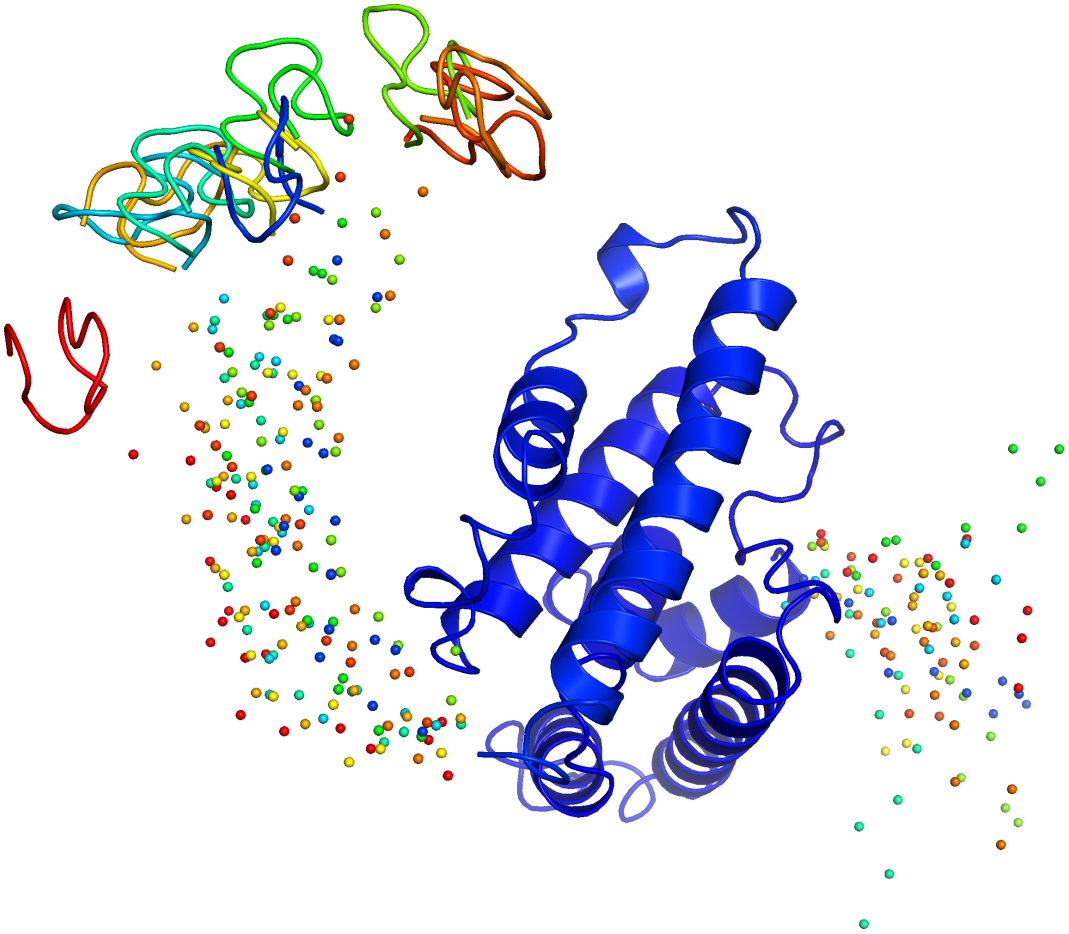


Figure 4.24: Ten structural models of YecA

The orientation of the YecA-MeBD in relation to the UPF0149 domain (blue) was reconstructed from the SAXS analysis using rigid body modelling in CORAL. The UPF0149 domain was modelled using the crystal structure of YgfB (PDB ID: 1IZM Galkin *et al.*, 2004) and the YecA-MeBD was modelled on the SecA-MeBD (PDB ID: 1TM6 Matousek and Alexandrescu, 2004). The two domains were rigid and flexible dummy molecules (dots) were used to model the linker region between the two domains and the amino terminus. The ten orientations of the YecA-MeBD, flexible linker and amino-terminus in relation to the UPF0149 domain are shown in red, green, blue, yellow, light blue, orange, tangerine, light green, dark orange and turquoise.

have a shoulder on the bell-shaped peak in the low q region but still converge to the q axis in the high q region, unless the domains are linked by a flexible region. The SAXS data was transformed automatically using PRIMUS software (Konarev *et al.*, 2003). The Kratky analysis of the averaged SAXS data of YecA showed the bell-shape peak in the low q that converged to the q axis in the high q region. This result indicated that YecA was in a folded, globular conformation (Figure 4.22).

CORAL is a program that is used to model complexes from SAXS data (Petoukhov *et al.*, 2012). Using CORAL, the averaged measured scattering profile of YecA was compared to theoretical profiles of a template molecule. The template molecule was composed of two domain homologues of YecA and dummy molecules (Petoukhov *et al.*, 2012). Dummy glycine residues were used to model the first 10-amino acid residues of the YecA molecule. Amino acid residues 11 to 189 of YecA were modelled using the crystal structure of YgfB (PDB ID: 1IZM Galkin *et al.*, 2004) Residues 201 to 221 of YecA were modelled using the NMR solution structure of the SecA-MeBD (PDB ID: 1TM6 Matousek and Alexandrescu, 2004). A flexible linker of 19 dummy glycine residues was used to attach the carboxy-terminus of 1IZM to the amino terminus of 1TM6. 1IZM was fixed in the template model, and the rest of the molecule was given freedom to move. 10 theoretical models were generated by CORAL (Chi^2 = 0.82-0.90). The simulated SAXS data for each model generated by CORAL is shown in Figure 4.23.

Figure 4.24 shows that the modelled YecA-MeBD was independent of the UPF0149 domain but had limited movement in solution. The model structures showed that the YecA-MeBD could rotate approximately 100° around the UPF0149 domain. The amino terminus of YecA, which was modelled using dummy molecules, showed large flexibility. The models were in agreement with the structural model of YecA that was generated using Phyre2. The models showed that the UPF0149 domain formed 7 α-helices, that were grouped in a 3-helix and a 4-helix bundle, like YgfB.

4.3 Discussion

The aim of this chapter was to investigate the structure of YecA. Although the structure of YecA was not resolved, progress towards this aim was made. The data presented in this chapter suggested that YecA was a monomer in solution under the experimental conditions used. Model structures of YecA that were generated suggested that YecA forms 7 α-helices. The CD spectra also suggested that YecA was mostly α-helical. The Phyre2 predicted model and the SAXS analysis suggested that the YecA-MeBD was independent of UPF0149. The two domains were predicted to be linked by residues with limited flexibility, which suggested that an interaction between the linker region and UPF0149 was possible. Assignment of residues from the NMR data suggested that the YecA-MeBD existed in two conformations.

The thermostability of YecA was investigated by CD spectroscopy to investigate whether YecA was stable at the temperature required for NMR spectroscopy. The CD analysis suggested that the melting temperature of YecA was 46.9°C. Above these temperatures, protein aggregation was observed by the increased absorbance at 200 nm. This data suggested that YecA was stable only to 39.6°C. The K2D analysis suggested that folded YecA was approximately 31% α-helical and approximately 10% β-sheet. The structural models of YecA both indicate that YecA is mostly α-helical. The SecA-MeBD has a ββα-fold when bound to zinc (Zhou and Xu, 2003; Dempsey *et al.*, 2004; Matousek and Alexandrescu, 2004) and it is possible that the YecA-MeBD forms the same fold. Additionally, the CD analysis provided the basis for repeating the citrate synthase assay that was described in Section 3.3 at 40°C in the future.

The investigation of the solution structure of YecA by ^{1}H-^{15}N-HSQC NMR spectroscopy revealed that the proton resonances were well-dispersed. The backbone assignments of YecA showed that the YecA-MeBD existed in two, distinct conformations

that were independent of the main body of the protein. The two species could be the result of disulphide bond formation because reducing agent was not included in the buffer that was used for the NMR experiments. However, it is also possible that the YecA-MeBD was structurally affected by iron. Fe^{2+} is a paramagnetic ion, which causes the intensity of the coordinating or neighbouring amide bonds to decrease. Therefore, it is possible that the absence of resonances that correspond to residues 206 to 216 (except proline) was caused by the proximity of an iron cofactor to these residues. The presence of iron could also explain the large changes in chemical shift of other residues. Studies of the SecA-MeBD indicated that the domain was unfolded in the absence of zinc and folded in the presence of zinc (Matousek and Alexandrescu, 2004). To determine the structure of YecA using NMR spectroscopy, the apo-protein will be prepared under reducing conditions to prevent the oxidisation of the cysteine residues. It is hoped that this method will allow the assignment of the YecA-MeBD in a single conformation.

The structural model of YecA that was determined using SAXS and CORAL indicated that the two domains, UPF0149 and YecA-MeBD, were independent of each other. This result suggested that the YecA-MeBD does not function as an auto-inhibitor. The carboxy-tail of SecA, which includes the SecA-MeBD and the flexible linker, has been shown to occlude the PPXD (Gelis *et al.*, 2007). In the future, the orientation of the YecA-MeBD with respect to the UPF0149 domain in response to ligands will be investigated. Similarly, determining whether YecA can dimerise, like the sequence homologues Lpg0076, YgfB and SecA, will provide an insight into the structural mechanisms that underlie the function of YecA.

YecA was submitted to crystal screen trials but no protein crystals were observed in any of the conditions. However, in some conditions brown or brown-red precipitates were formed. At the time, these precipitates were considered to be contaminants. However, subsequent purifications of YecA by nickel affinity chromatography resulted in the

nickel affinity column turning from turquoise-blue to brown, and the protein sample was yellow. Fe^{2+} is soluble but oxidation of this ion to Fe^{3+} frequently results in the formation a brown precipitates. It was hypothesised that the colours observed in the crystal screen trials and during the nickel affinity chromatography was caused by iron. The independent observations also suggested that the cognate cofactor of YecA could be iron and not zinc, as predicted by sequence similarity between the YecA-MeBD and the zinc-binding SecA-MeBD. If the red-brown precipitates are iron-bound protein crystals, and not contaminants, it is possible that the crystal screen trials described in this chapter have identified several conditions that may be appropriate for crystallising YecA. The data presented in this chapter suggested that YecA was a monomer in solution. UPF0149-containing proteins YgfB and Lpg0076 have both been crystallised as dimers. Therefore, crystallography trials could be repeated with a particular focus on the conditions that produced the brown or red precipitates.

The data presented in this chapter suggested that the cognate metal cofactor of YecA might not be zinc. Zinc compounds are generally colourless and zinc is diamagnetic. The presence of zinc causes the chemical shift of neighbouring protons to change, which would have been observed in the NMR spectra. Instead, resonances were absent from the NMR spectrum of the second conformation of the YecA-MeBD, which suggested the presence of a paramagnetic ion, such as iron. The investigation into the cognate metal cofactor of YecA is presented in Chapter 5.

Chapter 5

INTERACTION OF YECA WITH A METAL COFACTOR

5.1 Introduction

The coordination of zinc by the SecA-MeBD has been investigated using NMR and X-ray crystallography (Matousek and Alexandrescu, 2004; Dempsey *et al.*, 2004; Gelis *et al.*, 2007). The interaction between SecA and zinc is thought to be functionally relevant because the SecA-MeBD folds in the presence of zinc, which may contribute to the interaction between this domain and SecB (Fekkes *et al.*, 1999; Patel *et al.*, 2006; Matousek and Alexandrescu, 2004). The YecA-MeBD is predicted to be zinc-binding domain based on sequence homology to the SecA-MeBD but the qualitative data presented in Chapter 4 suggested that YecA may associate with iron.

Specificity of a metalloprotein for the cognate metal cofactor is affected by folding location, relative affinity, metal availability and chaperones that can deliver and insert metals (Waldron, 2009). However, mismetallation of a metalloprotein can occur during the purification process (Cotruvo Jr. and Stubbe, 2012). For example, preparations of azurin, a copper-binding protein, yields apo-azurin, copper-bound azurin and zinc-bound azurin (McLaughlin *et al.*, 2012). This occurs because the affinity of transition metals for metalloproteins often follows the Irving-Williams series such that, outside of metal-regulatory environment of the cell, metals that have a lower affinity for a binding site can be out competed by another metal ion (Waldron, 2009).

It is possible to analyse the trace metal content of a sample using inductively coupled plasma optical emission spectroscopy (ICP-OES). ICP-OES detects the electromagnetic radiation that is emitted by excited atoms and ions. Each metal emits electromagnetic radiation of a characteristic wavelength. By comparing the intensity of the emission to a set of standards the concentration of a particular metal can be determined. This technique does not provide information about the charge of the metal before analysis.

Ultra-violet visible (UV-vis) spectroscopy can be used to investigate the binding of a protein to a cofactor (Hamilton, 2009). UV-vis spectroscopy measures the absorbance of electromagnetic radiation in the UV-vis range (200 - 800 nm). The contribution of a metal cofactor to the UV-vis spectrum of a protein can be detected by comparison to the apoprotein. For example, UV-vis was used to investigate the interaction of TauD with its cognate metal cofactor, Fe^{2+}. This analysis revealed that Fe^{2+}-binding results in a small, reproducible increase in the absorbance of electromagnetic radiation compared with the apo-protein (Grzyska *et al.*, 2010).

NMR spectroscopy was used in Chapter 4 to investigate the structure of YecA. It is possible to use this technique to investigate the binding of YecA to a metal cofactor and potentially identify the coordinating amino acid residues. Zn^{2+} is an example of a diamagnetic ion. When a protein binds Zn^{2+}, the interacting and neighbouring residues would have a different chemical shift pattern compared with the apo-protein because of structural and electronic changes around those residues. Paramagnetic metal ions are characterised by unpaired electrons, which have large magnetic moments. There are strong interactions between the unpaired electrons of a paramagnetic metal ion and the nuclei of the surrounding protein (Jensen *et al.*, 2007). The coordination of a paramagnetic ion by a protein causes the proton NMR lines of coordinating and proximal nuclei to broaden. The nuclei closest to the paramagnetic ion are not detectable because of this phenomenon. Nuclei that are further away from the paramagnetic ion exhibit fewer proton NMR line broadening effects (Figure 5.1). The size of this effect depends on the nuclear relaxation of metal ion, which is affected by the number of unpaired electrons, the electron relaxation time, and the rotational correlation time of the molecule (Bertini *et al.*, 2005). An example of a paramagnetic ion is Fe^{2+}. The NMR ^{1}H-^{15}N-HSQC spectrum of apo-protein would have more peaks than the NMR ^{1}H-^{15}N-HSQC spectrum of iron-incubated protein.

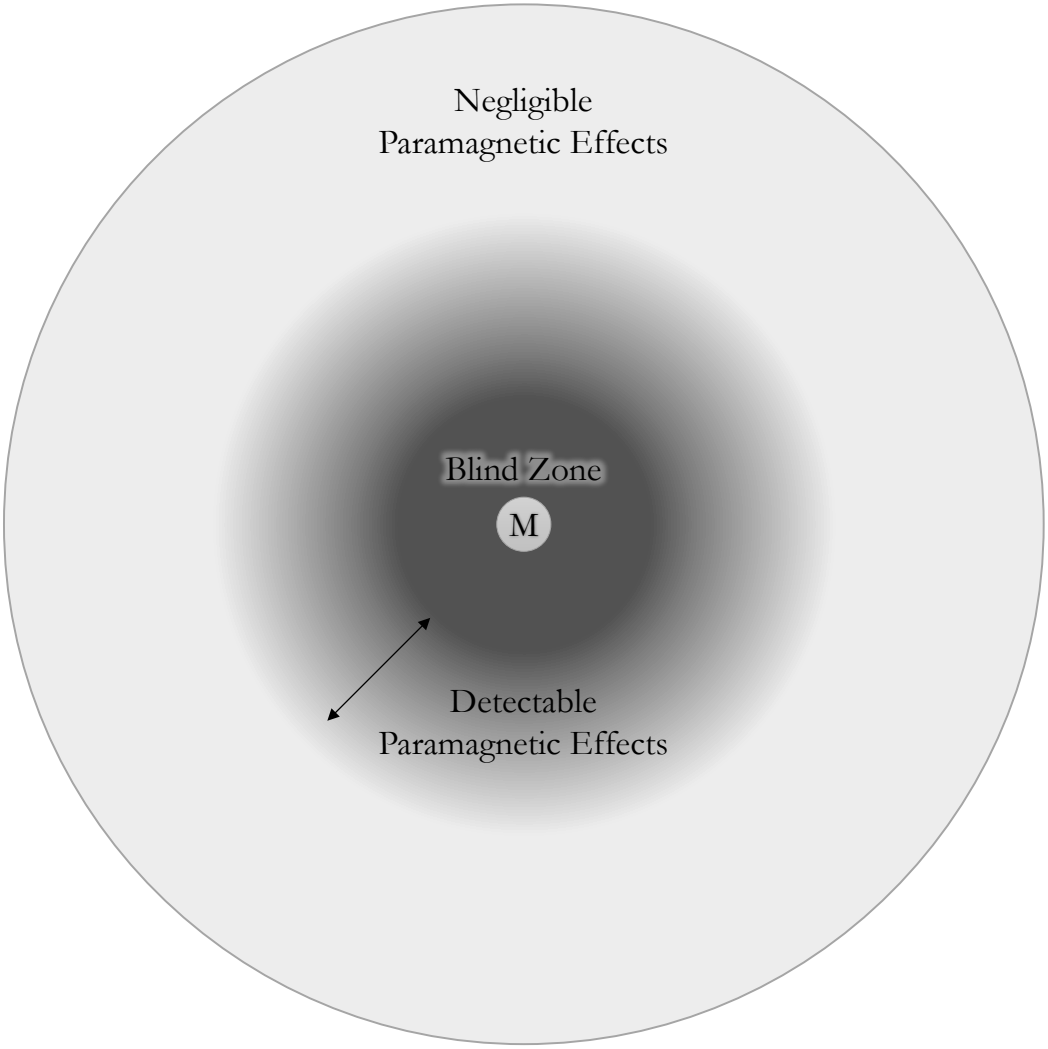


Figure 5.1: Diagram to illustrate the effect of a paramagnetic ion on the visibility of amino acid residues by NMR spectroscopy

The sphere around the paramagnetic metal ion is a blind zone because proton NMR signals are too broad to be detected. Outside the blind zine is another sphere, which represents the region where NMR signals are visible but still affected by paramagnetism. The region with detectable paramagnetic effects can be used to gather information on their position with respect to the metal ion. Figure adapted from Bertini *et al.* (2005)

EPR can be used to investigate metalloproteins. The spectroscopic signature of a metalloprotein is influenced by the number of ligands, the symmetry and the oxidation state of the bound metal. Every electron has a magnetic moment and a spin quantum number, $S = \frac{1}{2}$. When an external magnetic field, B_0, is applied to a sample, the spin quantum number (spin state) can either be $S = +\frac{1}{2}$ or $S = -\frac{1}{2}$ depending on whether the electron aligns against B_0 or with B_0. EPR depends on the energy difference (ΔE) between two spin states of an electron when B_0 is applied. An unpaired electron can flip between the two energy levels by either absorbing or emitting a photon of an energy that is equal to the difference between the two energy levels. The transition of an electron between the two energy levels can be induced by the application of an oscillating electromagnetic resonance that is equal to ΔE.

The spin state of iron has an effect on the EPR spectrum. Fe^{3+} is the only oxidation state of iron that is detectable by EPR because it is the only oxidation state that contains an unpaired electron. The *d* electrons are distributed across five orbitals. Electrostatic repulsion means that these electrons are likely to occupy different orbitals. This is known as the high spin state. The net spin of Fe^{3+} in the high spin state is $S = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{5}{2}$. The coordination of Fe^{3+} by a strong ligand causes the electrons to occupy the lower energy *d* orbitals such that four electrons are paired and one electron is not to give a net spin state of $S = \frac{1}{2}$. This is known as the low spin state. Fe^{2+} is not observable in either spin state because in the high spin state $S = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = 2$ and in the low spin state, the six electrons pair and cancel each other out to give a net spin of $S = 0$. The *g*-factor is the measure of the electronic interaction between an unpaired electron and B_0 and is characteristic of the spin state and the geometry of the ligand binding.

In this chapter, a combination of UV-vis spectroscopy, ICP-OES, ICP-MS, EPR spectroscopy and ^{1}H-^{15}N-HSQC NMR spectroscopy was used to investigate the metal cofactor of YecA. In Chapter 4, the observed colours during the purification and crystal screen

trials of YecA suggested that iron could be the cognate metal cofactor of YecA. The data presented in this chapter suggests that YecA binds iron.

5.2 Results

5.2.1 Inductively coupled plasma optical emission spectrometry to analyse the metal content of YecA

The identity of the metal that co-purified with YecA after nickel-affinity chromatography, as described in Chapter 4, was investigated using ICP-OES. YecA was purified as described in Section 3.2.6 and prepared at a concentration of 50 mg/l in 2% v/v nitric acid for ICP-OES. It was reasoned that the contaminant was most likely to be a divalent ion that had exchanged with nickel during the purification. The column was washed with 10 ml 0.1 M EDTA to elute the brown substance for analysis by ICP-OES. After this step, the column was white, which indicated that the substance had been eluted. The column eluate was prepared for ICP-OES by the addition of nitric acid to a final concentration of 2% (v/v) .

A calibration curve for each metal was determined using metal standards of 18 nM, 180 nM and 1.8 μM that were prepared in 2% (v/v) nitric acid. The three elements analysed were iron, manganese and zinc, which are the primary divalent transition metals in the *E. coli* cytoplasm (Imlay, 2014). The standards are shown in Figure 5.2. The molar concentration of each element in the protein and the column eluate samples were determined using the standard calibration curves (Figure 5.2). The concentration of zinc in the YecA sample was 6.8 μM and the concentration of zinc in the column eluate was 10.9 μM. The concentration of manganese in the YecA sample was 0.3 μM, while the manganese concentration was 0.8 μM in the column eluate. The concentration of iron in the YecA sample was 0.8 μM. The concentration of iron in the column eluate was 255.7 μM, which

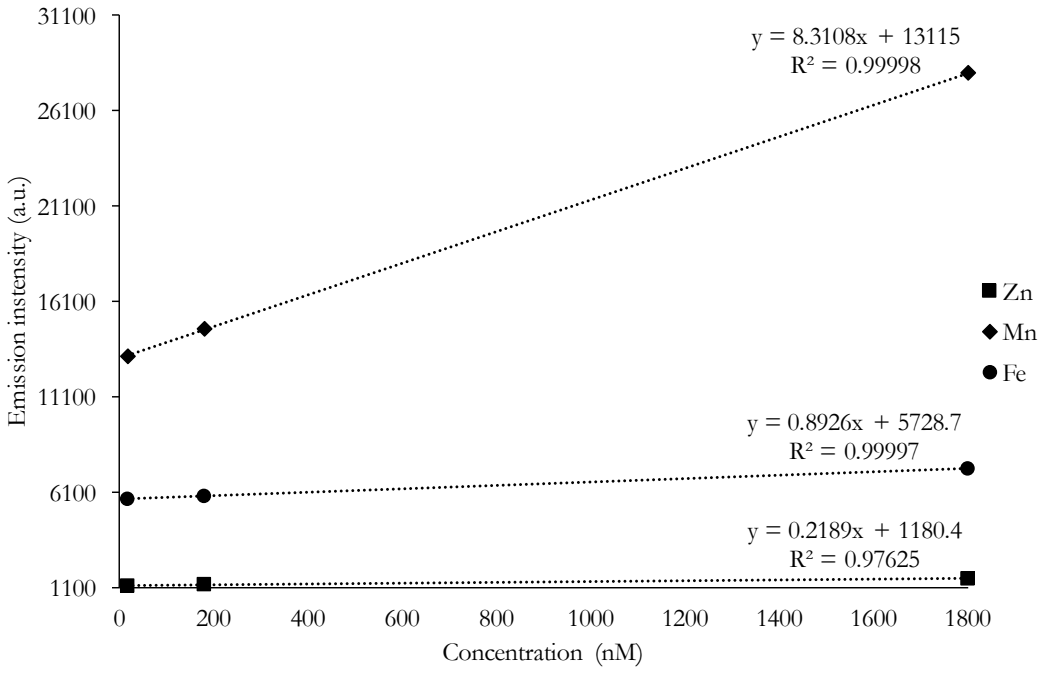
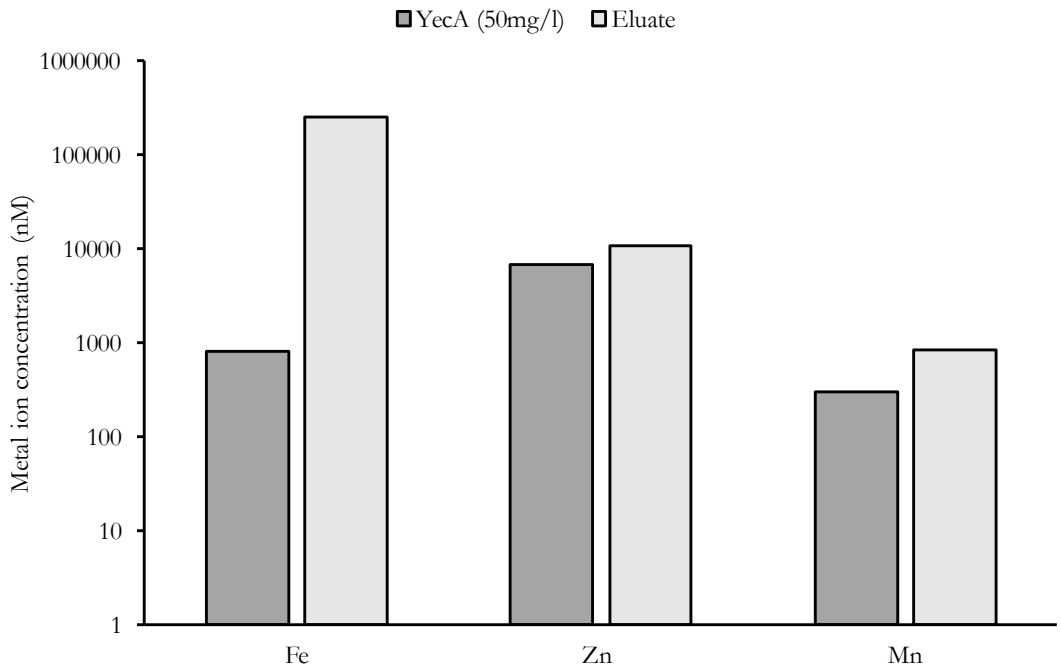
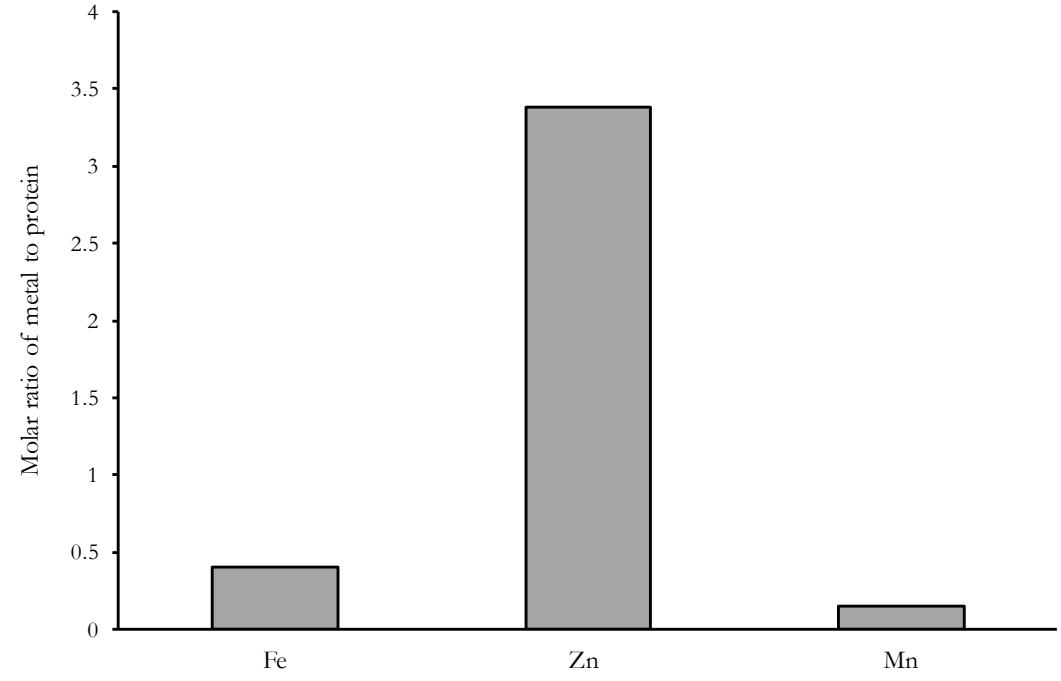


Figure 5.2: Three point calibration curves including linear regression coefficients for zinc, manganese and iron

For every wavelength (200 - 260 nm) the emission intensity was recorded (a.u.). The interval between each wavelength was 0.0038 nm. The wavelength that corresponded to the maximal emission intensity was identified manually. The wavelengths used were λ= 206.2011 nm for zinc, λ= 257.61 nm for manganese and λ= 238.204 for iron. The emission intensity of each of the metal standards at this wavelength was plotted as a function of concentration. The linear regression analysis was used to solve for the concentration of each element in the samples (equation 2.1).



(a) ICP-OES to identify the metal cofactor of YecA



(b) Molar ratio of metal to YecA

Figure 5.3: ICP-OES to analyse the metal content of YecA

YecA and the column eluate were analysed by ICP-OES. The concentration of iron, zinc and manganese in each sample was determined using the calibration curves. 5.3a shows the metal content of YecA and the column eluate. The *y*-axis log scale shows the concentration of each metal. The three metals analysed were iron, zinc and manganese. $n = 1$. 5.3b shows the molar ratio of metal cofactor to YecA. This was calculated by dividing the molar concentration of metal by the molar concentration of YecA.

was 96% of the total analysed metal content of this sample. The high concentration of iron in the eluate indicated that iron had caused the nickel affinity column to turn from blue to brown during the purification of YecA.

The molar ratio of each metal to YecA was determined (Figure 5.3b). This analysis indicated that YecA had a zinc:YecA molar ratio of 3.3:1, or a three-fold excess of zinc per molecule of YecA. The iron:YecA molar ratio was 0.4:1 and the manganese:YecA molar ratio was 0.1:1. The ICP-OES analysis of the column eluate suggested that the YecA-bound iron was lost during the nickel affinity purification. This result suggested that the identity of the metal that co-purified with YecA was iron.

5.2.2 UV-visible spectroscopy to characterisation of the interaction of YecA with iron

UV-vis spectroscopy was used to investigate the interaction between YecA and iron further. YecA was purified to homogeneity and then treated with 0.2 M EDTA to chelate any divalent cations. The EDTA was removed from the protein by dialysis into Buffer 1. The apo-YecA solution turned yellow when one equivalent molar concentration of $FeSO_4$ was added. Comparative analysis of the iron-incubated YecA with the apo-YecA by UV-vis spectroscopy showed that after the incubation of apo-YecA with $FeSO_4$ a broad shoulder with a maximum of 328 nm developed (Figure 5.4). Furthermore, this peak was intensified by the addition of 10 mM ascorbate, a reducing agent. The peak at 328 nm was attributed to the formation of an Fe^{2+}-YecA complex.

5.2.3 ^{1}H-^{15}N-HSQC NMR spectra to investigate the metal-binding site of YecA

^{1}H-^{15}N-HSQC NMR spectroscopy was used to investigate the iron-binding site of YecA. The sequence identity between the SecA-MeBD, a known metal binding domain, and

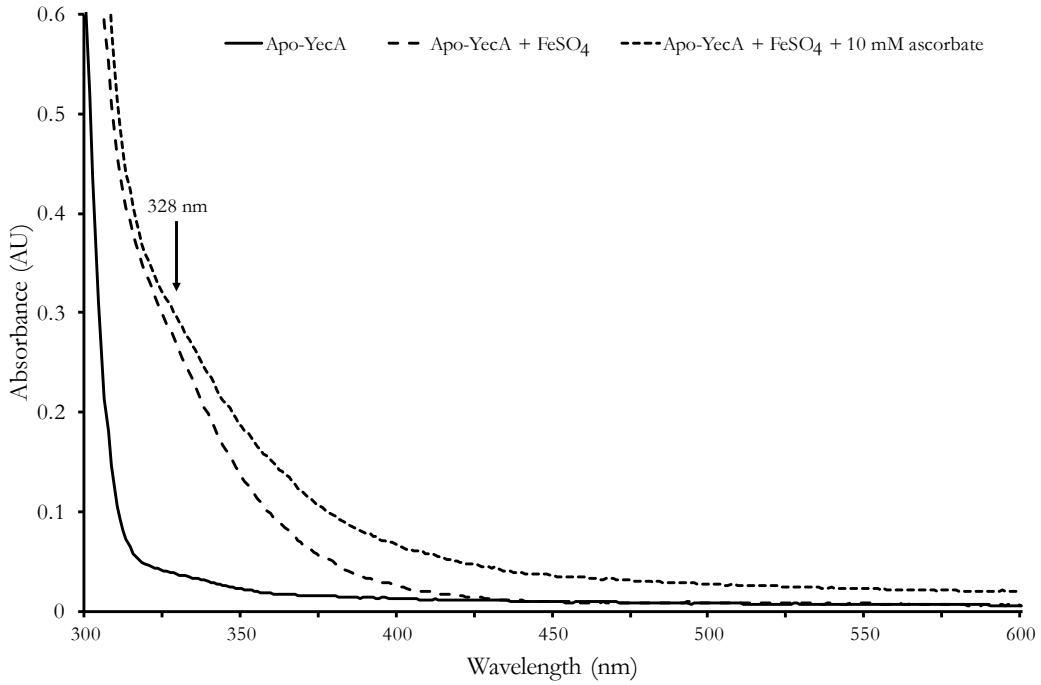


Figure 5.4: UV-visible spectra to show Fe^{2+}-binding of YecA

Apo-YecA was prepared by incubating YecA with 0.2 mM EDTA. The EDTA was removed by dialysis into Buffer 1 prior to analysis. The samples were analysed at room temperature under aerobic conditions. Solid line: the mean UV-vis spectrum of 0.8 mM apo-YecA. Long-dashed line: The mean UV-vis spectrum of apo-YecA after it was incubated with an equimolar concentration of $FeSO_4$. Short-dashed line: The mean UV-vis spectrum of the iron-bound protein plus 10 mM ascorbate. Data are buffer-corrected mean values. $n = 3$.

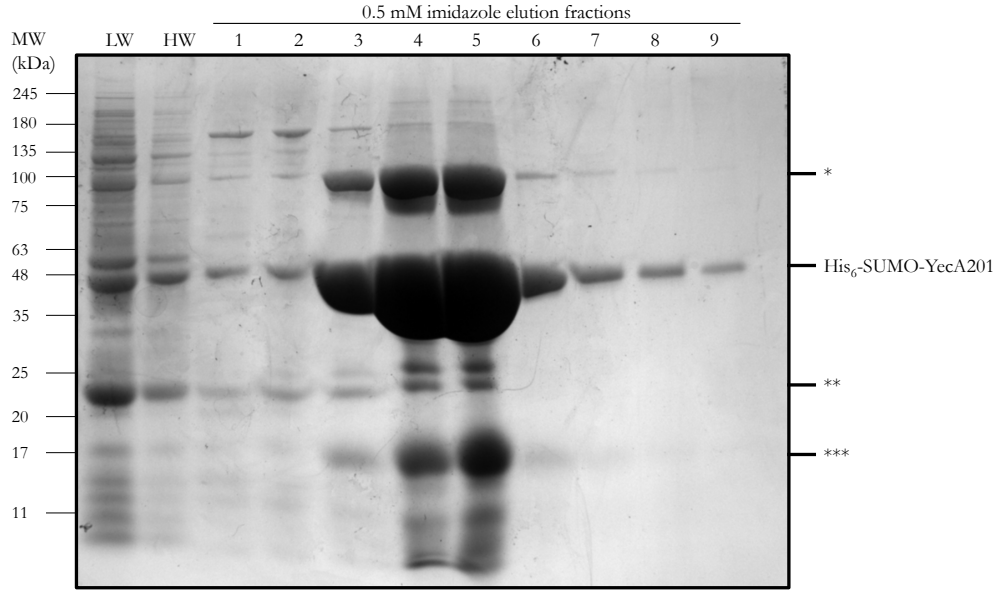


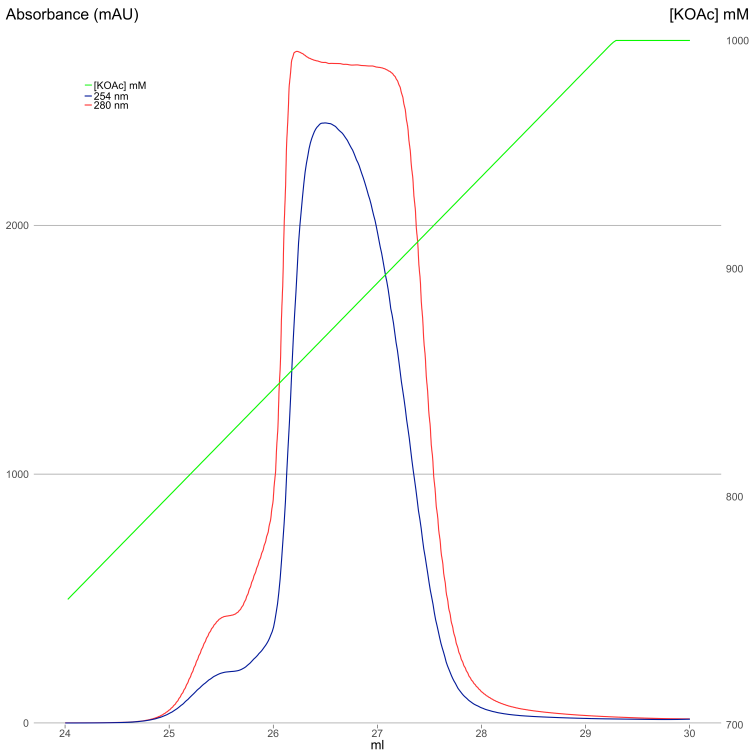
Figure 5.5: SDS-PAGE analysis of the nickel-affinity chromatography of ^{1}H-^{15}-N-His_6-SUMO-YecA201

His_6-SUMO-YecA201 was expressed in *E. coli* strain BL21 DE3 in M9 media that was supplemented with $^{15}NH_4Cl$. The IPTG-induced cells were pelleted and the pellet was resuspended in Buffer 1 that was supplemented with PMSF (to inhibit protein degradation) and Dnase I (to degrade the DNA) and lysed by high pressure homogenisation. The lysate was clarified by centrifugation and filtration before application to a nickel affinity column. The lysate passed through the column at least five times before the column was washed with five column volumes of Buffer 1 (LW). Subsequently, the column was washed with five column volumes of High Salt Buffer (HW). The protein was eluted with 0.5 mM imidazole in 0.5 ml fractions. The washes and imidazole elutions were analysed by SDS-PAGE and Coomassie staining. ^{1}H-^{15}-N-His_6-SUMO-YecA201 eluted in fractions 3 to 6. * indicates a protein of 75 kDa that co-purified with ^{1}H-^{15}-N-His_6-SUMO-YecA201, which was possibly CdsA. ** indicates a protein of between 20 and 25 kDa that was possibly the cleaved YecA201, the band the is indicated by *** is possibly the His_6-SUMO tag.

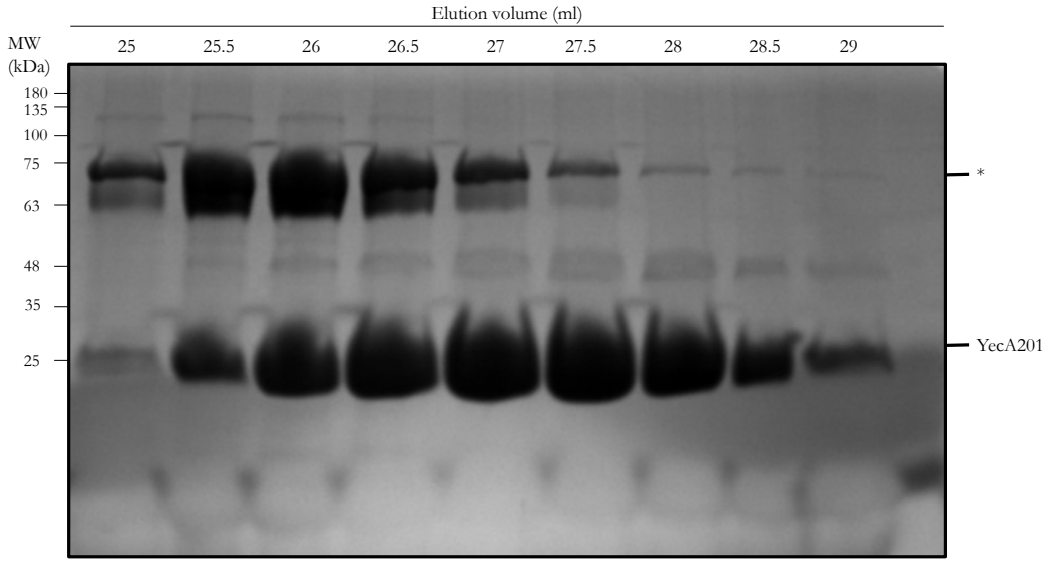
the YecA-MeBD suggested that the metal-binding site of YecA could be located in this domain. Both full length YecA and YecA201, a variant of YecA that lacked the YecA-MeBD, were analysed by ^{1}H-^{15}N-HSQC NMR spectroscopy to investigate whether the metal-binding-site was in the YecA-MeBD .

^{1}H-^{15}N-YecA201 was prepared by growing *E. coli* strain BL21 DE3 pCA528-*yecA201* in M9 media supplemented with $^{15}NH_4$ as the sole nitrogen source. After overnight induction, the cells were lysed and ^{1}H-^{15}N-His_6-SUMO-YecA201 was purified as described in Section 3.2.6. The washes and the imidazole elutions were analysed by SDS-PAGE and Coomassie staining, (Figure 5.5). The His_6-SUMO tag was cleaved by incubating overnight with His_6-Ulp1, a SUMO protease. After this step, the protein was concentrated and applied to an anion exchange column and eluted with a KOAc gradient of 100 mM to 1 M. The protein-containing fractions were analysed by SDS-PAGE. Figure 5.6 showed that after anion exchange chromatography cleaved textsuperscript1H-^{15}N-YecA201 had co-purified with a protein of 75 kDa, which was potentially CdsA, and the 48 kDa ^{1}H-^{15}N-His_6-SUMO-YecA201. The fractions from the anion exchange chromatography step that contained ^{1}H-^{15}N-YecA201 were pooled, concentrated and purified by size exclusion chromatography to remove these contaminants.

The protein-containing fractions were analysed by SDS-PAGE to determine the purity of ^{1}H-^{15}N-YecA201 after size exclusion chromatography. Analysis of the SDS-PAGE gel using Coomassie staining indicated that this step had resulted in 95% pure ^{1}H-^{15}N-YecA201 (Figure 5.7). A protein of 75 kDa had co-purified with ^{1}H-^{15}N-YecA201, and although the size exclusion chromatography had mostly separated the two proteins, a small amount of the co-purifying protein remained. It is possible that this contaminant was CdsA. The protein was treated with 0.1 M EDTA to chelate any divalent cations. The EDTA was removed by dialysis of the protein into 20 mM [MES]-NaOH [pH 6.5], 10 mM NaCl. ^{1}H-^{15}N-YecA201 was subsequently concentrated to 0.84 mM.



(a) Anion exchange chromatography elution trace of YecA201 of ^{1}H-^{15}N-YecA201



(b) SDS-PAGE analysis of the anion exchange chromatography of ^{1}H-^{15}N-YecA201

Figure 5.6: Anion exchange chromatography of ^{1}H-^{15}N-YecA201 for ^{1}H-^{15}N-HSQC NMR spectroscopy

Anion exchange chromatography was used to purify the cleaved ^{1}H-^{15}N-YecA201. A flow rate of 0.5 ml/min was used. 5.6a shows the elution profile of ^{1}H-^{15}N-YecA201 from an anion exchange column as the KOAc concentration increased from 100 mM to 1 M. The elution of the protein (left *y*-axis) was detected by measuring the absorbance at 280 nm (red) and the absorbance at 254 nm (blue). The concentration of KOAc is indicated in green (right *y*-axis). The protein was eluted from the anion exchange chromatography column between 800 mM to 950 mM KOAc. 5.6b shows the SDS-PAGE analysis and Coomassie staining of the eluted fractions. ^{1}H-^{15}N-YecA201, which was resolved at 25 kDa, eluted between 25.5 ml and 29 ml. These fractions were pooled and concentrated for further purification by size exclusion chromatography. * indicated a protein of 75 kDa that co-purified with ^{1}H-^{15}N-YecA201, which was possibly CdsA.

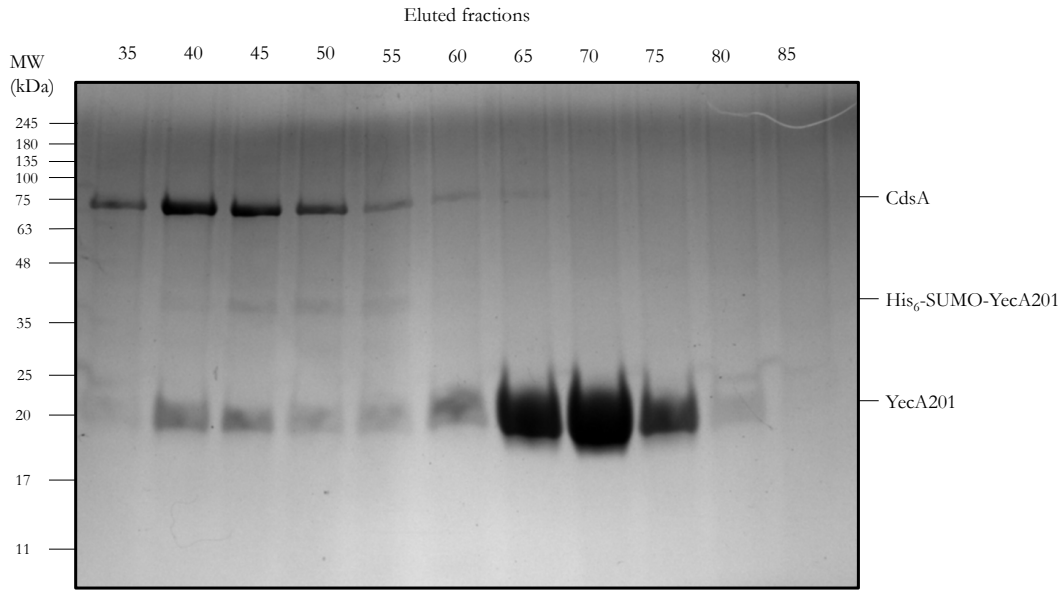


Figure 5.7: SDS-PAGE analysis of the size exclusion chromatography of ^{1}H-^{15}N-YecA201

Size exclusion chromatography was used to separate proteins according to their size. ^{1}H-^{15}N-YecA201 was applied to an S75 10/300 GL size exclusion chromatography column that had been pre-equilibrated in Buffer 1 plus 1 mM TCEP. A flow rate of 1 ml/min was used. The elution volume was collected in 5 ml fractions, which were analysed by SDS-PAGE and Coomassie staining. * indicates a protein of 75 kDa that co-purified with YecA201, which is possibly CdsA. Fractions 65 to 75 were pooled and dialysed into 20 mM [MES]-NaOH [pH 6.5], 10 mM NaCl and concentrated using a spin concentrator to 0.84 mM for NMR spectroscopy experiments

In initial experiments ^{1}H-^{15}N-YecA201 was analysed by ^{1}H-^{15}N-HSQC NMR spectroscopy to investigate whether the truncated variant of YecA could bind iron. The overlaid ^{1}H-^{15}N-HSQC spectra of ^{1}H-^{15}N-YecA201 in the presence and absence of $FeSO_4$ are shown in Figure 5.8. The incubation of apo-^{1}H-^{15}N-YecA201 with $FeSO_4$ did not result in the disappearance of any amide bond peaks. This result suggested that the MeBD was required for the binding of iron under these experimental conditions.

Subsequently, ^{1}H-^{15}N-YecA was analysed by ^{1}H-^{15}N-HSQC NMR spectroscopy. The overlaid ^{1}H-^{15}N-HSQC spectra of the apo-^{1}H-^{15}N-YecA and iron-bound^{1}H-^{15}N-YecA are shown in Figure 5.9. The incubation of apo-YecA with $FeSO_4$ resulted in the disappearance of several amide proton peaks, which suggested that YecA bound iron. In Section 4.2.8, the YecA-MeBD was assigned in two, distinct conformations. Only one of these conformations was fully assigned. Although the resonances that coordinated iron were not annotated as being part of the YecA-MeBD, the ^{1}H-^{15}N-HSQC NMR spectrum of YecA201 does not contain several of amide proton peaks that disappeared or decreased in intensity in the ^{15}N-HSQC NMR spectrum of iron-incubated YecA (Figure 5.10). This result suggested that it is likely that the iron-binding site is located in the YecA-MeBD.

E. coli SecA coordinates zinc via three cysteine residues and a histidine residue, which suggests that YecA may bind to zinc. The NMR spectroscopy experiments were repeated with apo-^{1}H-^{15}N-YecA and ^{1}H-^{15}N-YecA in the presence of a two-fold molar excess of zinc sulphate to investigate whether YecA bound zinc (Figure 5.11). Zinc is diamagnetic, which means that structural changes to the environment around the amide bonds are detected by chemical shift perturbations. If YecA bound to zinc, the overlaid spectra of YecA in the presence and absence of zinc sulphate would be expected to show different positions of some amide bond peaks. However, no chemical shifts perturbations were

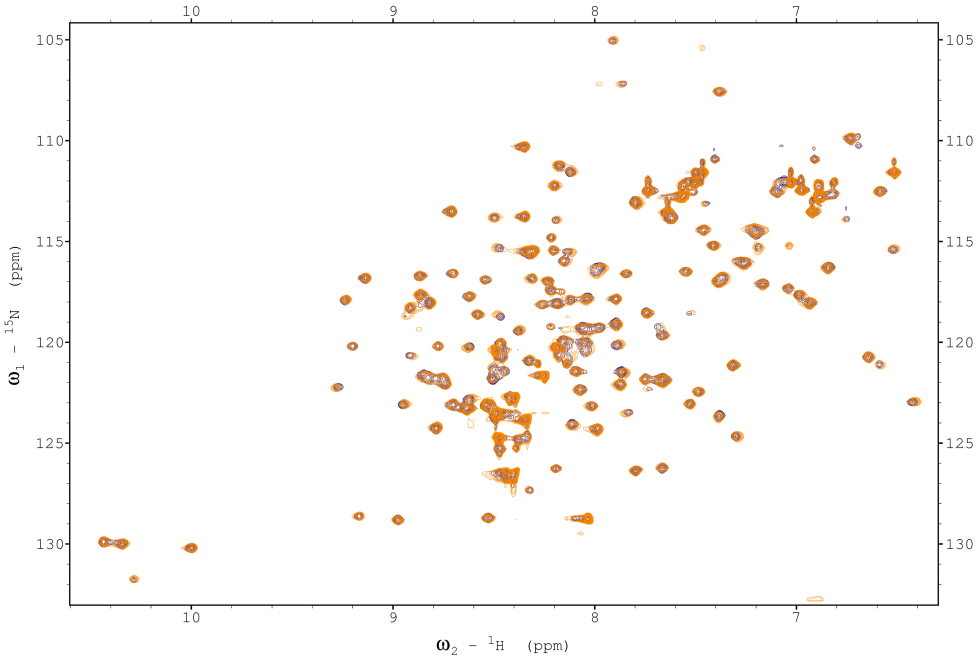


Figure 5.8: The effect of $FeSO_4$ on the ^{1}H-^{15}N-HSQC spectrum of apo-^{1}H-^{15}N-YecA201

The protein was dialysed into 20 mM [MES]-NaOH [pH 6.5], 10 mM NaCl and concentrated to 0.84 mM for analysis by ^{1}H-^{15}N-HSQC NMR spectroscopy using the Bruker 900 mHz instrument. Each peak represents a ^{1}H attached to a ^{15}N. The ^{1}H-^{15}N-HSQC spectrum of apo-^{1}H-^{15}N-YecA201 is shown in blue. The apo-^{1}H-^{15}N-YecA201 was incubated with a two-fold molar excess of $FeSO_4$. The ^{1}H-^{15}N-HSQC spectrum of ^{1}H-^{15}N-YecA201 after it had been incubated with $FeSO_4$ is shown in orange.

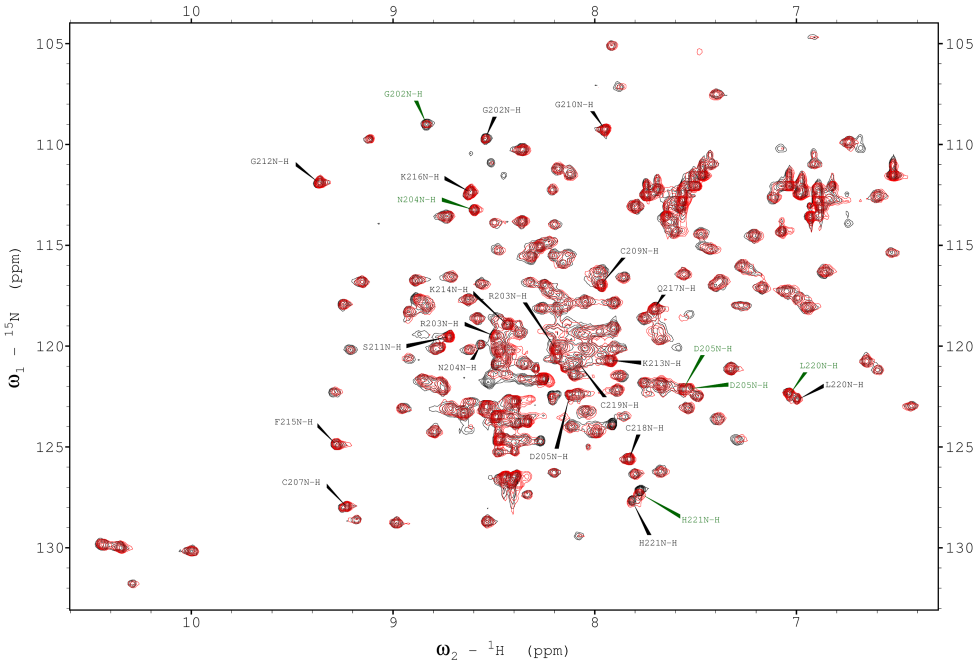


Figure 5.9: The effect of $FeSO_4$ on the ^{1}H-^{15}N-HSQC spectrum of apo-^{1}H-^{15}N-YecA

The protein was dialysed into 20 mM [MES]-NaOH [pH 6.5], 10 mM NaCl and concentrated to 1.1 mM for analysis by ^{1}H-^{15}N-HSQC NMR spectroscopy using the Bruker 900 mHz instrument. Each peak represents a ^{1}H attached to a ^{15}N. The ^{1}H-^{15}N-HSQC spectrum of apo-^{1}H-^{15}N-YecA is shown in black. The apo-^{1}H-^{15}N-YecA was incubated with a two-fold excess of $FeSO_4$. The ^{1}H-^{15}N-HSQC spectrum of ^{1}H-^{15}N-YecA after it had been incubated with $FeSO_4$ is shown in red. The two conformations of the YecA-MeBD are shown in green and black.

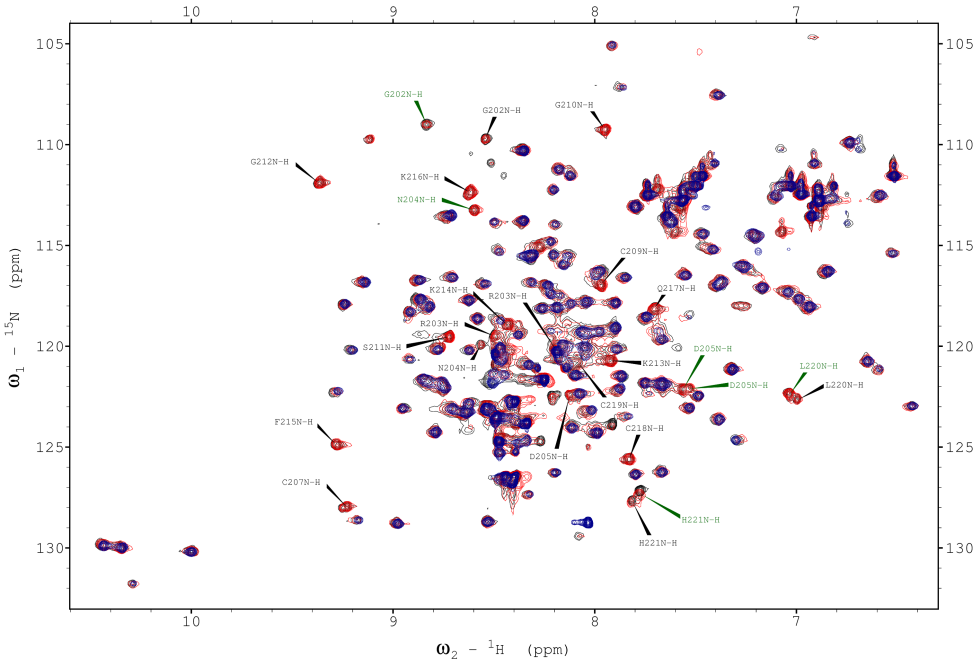


Figure 5.10: The effect of iron-binding on the YecA-MeBD

The protein was dialysed into 20 mM [MES]-NaOH [pH 6.5], 10 mM NaCl and concentrated to 1.1 mM for analysis by ^{1}H-^{15}N-HSQC NMR spectroscopy using the Bruker 900 mHz instrument. Each peak represents a ^{1}H attached to a ^{15}N. The ^{1}H-^{15}N-HSQC spectrum of apo-^{1}H-^{15}N-YecA is shown in black. The apo-^{1}H-^{15}N-YecA was incubated with a two-fold excess of $FeSO_4$. The ^{1}H-^{15}N-HSQC spectrum of ^{1}H-^{15}N-YecA after the addition of $FeSO_4$ is shown in red. The apo-^{1}H-^{15}N-YecA201 is shown in blue. The two conformations of the YecA-MeBD are shown in green and black.

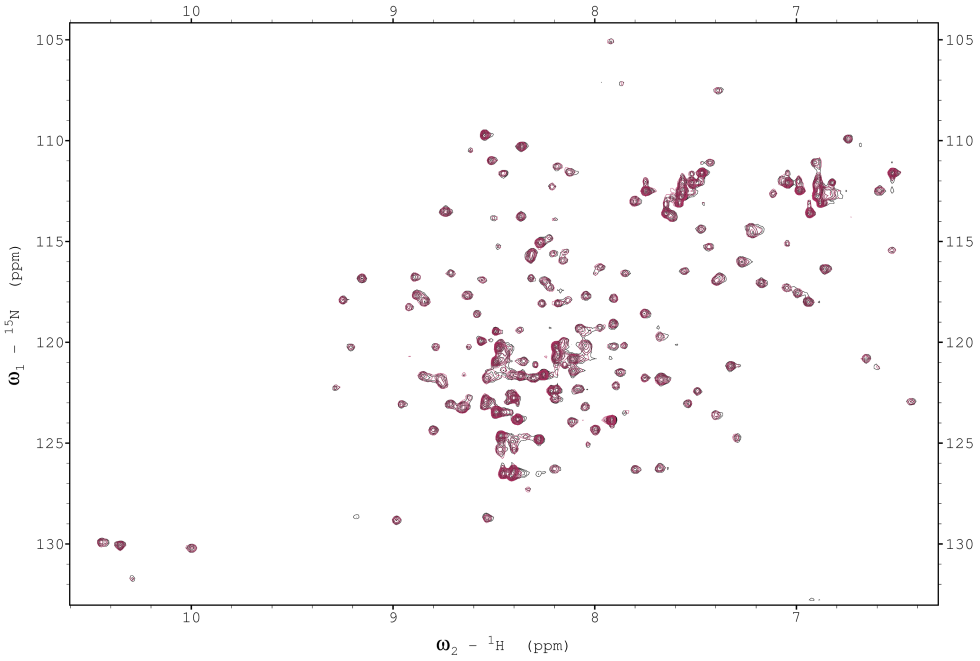


Figure 5.11: The effect of $ZnSO_4$ on the ^{1}H-^{15}N-HSQC spectrum of apo-^{1}H-^{15}N-YecA

After the iron-binding experiments, the protein was dialysed into 20 mM [MES]-NaOH [pH 6.5], 10 mM NaCl and concentrated to 1.1 mM for analysis by ^{1}H-^{15}N-HSQC NMR spectroscopy using the Bruker 900 mHz instrument. Each peak represents a ^{1}H attached to a ^{15}N. The overlay of the zinc-bound ^{1}H-^{15}N-YecA ^{1}H-^{15}N-HSQC spectrum is shown in purple and the apo-^{1}H-^{15}N-YecA ^{1}H-^{15}N-HSQC spectrum is shown in black. These spectra were taken on the same day.

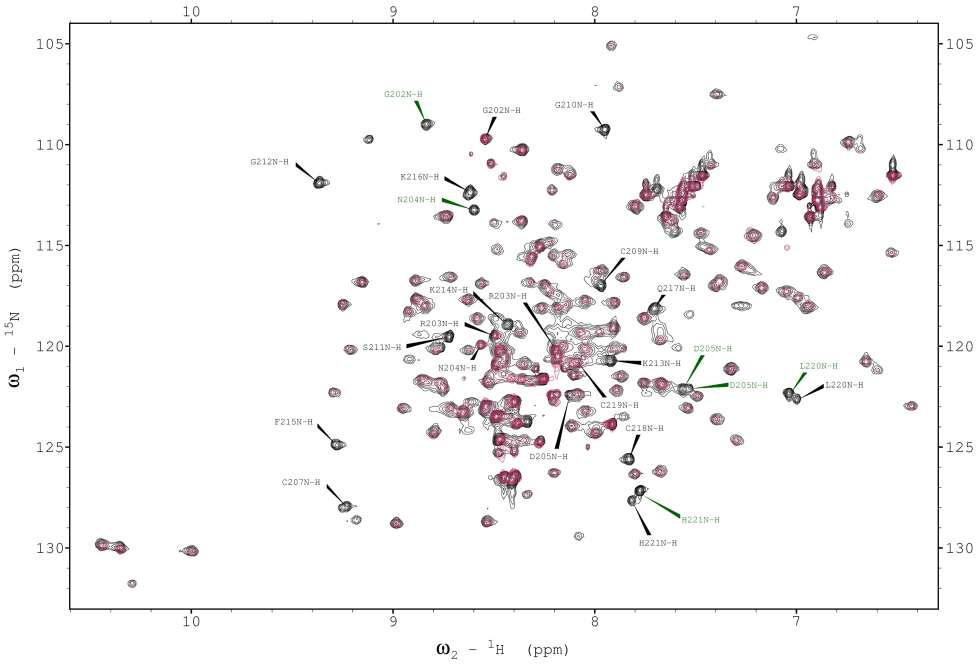


Figure 5.12: Overlaid ^{1}H-^{15}N-HSQC spectra to show the degradation of the YecA-MeBD

After the iron-binding experiments, the protein was dialysed into 20 mM [MES]-NaOH [pH 6.5], 10 mM NaCl and concentrated to 1.1 mM for analysis by ^{1}H-^{15}N-HSQC NMR spectroscopy using the Bruker 900 mHz instrument. Each peak represents a ^{1}H attached to a ^{15}N. The overlay of the zinc-bound ^{1}H-^{15}N-YecA ^{1}H-^{15}N-HSQC spectrum is shown in purple and the apo-^{1}H-^{15}N-YecA ^{1}H-^{15}N-HSQC spectrum is shown in black. These spectra were taken on the same day. 5.11 shows 5.12 shows the ^{1}H-^{15}N-HSQC spectra of the original apo-YecA (black) and the secondarily analysed ^{1}H-^{15}N-YecA (purple). The purple spectrum lacks many of the resonances that have been mapped to the YecA-MeBD.

observed between the overlaid spectra of ^{1}H-^{15}N-YecA in the presence or absence of zinc. This result indicated that YecA did not bind zinc under these conditions.

However, further analysis of this spectrum suggested that the protein had partially degraded between the iron-binding experiments and the zinc-binding experiments. When the apo-^{1}H-^{15}N-HSQC spectrum of YecA was overlaid with the original ^{1}H-^{15}N-HSQC spectrum of apo-YecA it was clear that the YecA-MeBD was not resolved in these later experiments (Figure 5.12). Therefore it could not be concluded that YecA did not bind zinc and these experiments should be repeated.

5.2.4 Analysis of the electronic state of iron in the presence of YecA by EPR spectroscopy

EPR was used to investigate the electronic state of iron when it was coordinated by YecA *in vitro*. YecA was incubated with $FeCl_3$ at a 1:1 molar ratio in Buffer 1 + 30% v/v glycerol and analysed at the Q-band frequency (34 GHz). The Q-band EPR spectrum of Fe^{3+} in the absence or presence of YecA is shown in Figure 5.13. In the presence of YecA, the maximum is at 6145 G and the *g* factor was determined as being 3.95.[1] The sharpness and intensity of the peak at 6000 G and the calculated *g*-factor was indicative of a mononuclear, rhombic, high-spin ($S = \frac{5}{2}$) Fe^{3+} ion (Bou-Abdallah and Chasteen, 2008; Duin, 2008; Gaffney, 2009).

5.3 Discussion

The results presented in this chapter suggested that YecA was an iron-binding protein. The YecA-MeBD is a predicted zinc-binding domain, based on sequence homology to the SecA-MeBD. ICP-OES analysis indicated that YecA was co-purifying with iron. UV-vis analysis showed that YecA could bind Fe^{2+} and ^{1}H-^{15}N-HSQC NMR

[1] With many thanks to Dr. J. Lovett, University of St. Andrews

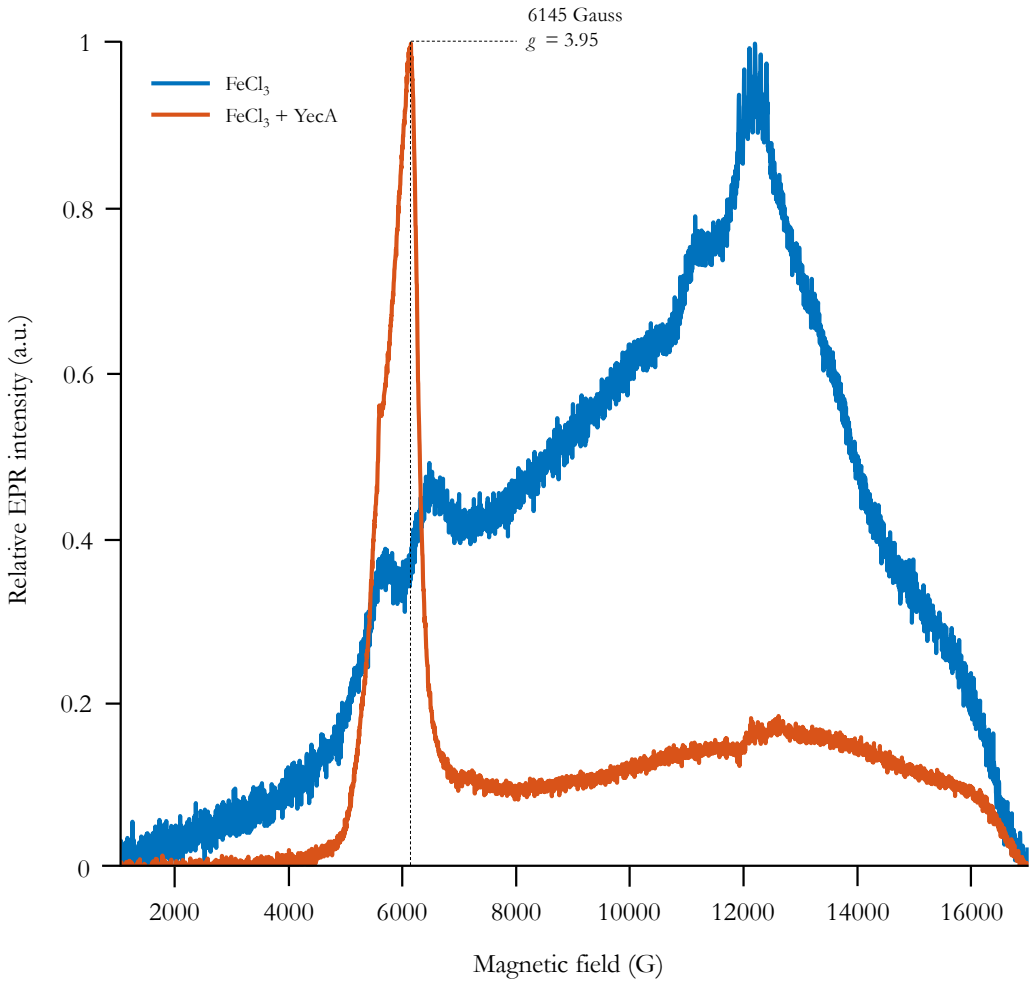


Figure 5.13: Q-band EPR spectra of $FeCl_3$ and Fe^{3+}-bound YecA

The EPR spectra are expressed as relative to their maximal EPR intensity, as determined by Dr. J. Lovett, University of St. Andrews. Blue: $FeCl_3$ in EPR buffer. Orange: equimolar concentration of YecA and $FeCl_3$ in EPR buffer. The "crown" at 12000 G in both spectra is typical of manganese, a contaminant.

spectroscopy suggested that the iron-binding site was located in the YecA-MeBD. EPR spectroscopy indicated that YecA bound a mononuclear Fe^{3+} ion. Together, these results suggested that YecA could bind labile iron cofactor.

Initially, ICP-OES was used to analyse the metal content of YecA. The analysis of the iron content of the protein revealed that iron was present in a substoichiometric concentration to YecA. This result was not unexpected because the observed colour change of the column suggested that the cognate metal cofactor was exchanging with nickel during the purification. Manganese was also observed at substoichiometric concentrations. However, the colours of the YecA sample and the ICP-OES analysis of the content of the nickel-affinity column both suggested that the cofactor was iron (Cadby *et al.*, 2016). Using UV-vis spectroscopy, it was possible to show that Fe^{2+}-reconstituted YecA had a broad shoulder with a maximum of 328 nm, which is indicative of iron binding by the protein (Ding and Clark, 2004; Lu *et al.*, 2010). The analysis of the interaction between YecA and iron by EPR spectroscopy suggested that YecA coordinated a mononuclear Fe^{3+} ion. Taken together, these data indicated that YecA coordinated a labile iron cofactor.

The analysis of the zinc content of YecA by ICP-OES indicated that zinc was present at a 3-fold excess of YecA. SecA coordinates a mononuclear divalent zinc ion via three cysteine residues and a histidine residue that are located in the SecA-MeBD (Breukink *et al.*, 1995; Fekkes *et al.*, 1999; Randall *et al.*, 2004; Dempsey *et al.*, 2004). It was predicted that the coordination of a metal ion by YecA would occur via equivalent residues to those that coordinate zinc in the SecA-MeBD. These residues are four cysteine residues in YecA. Without a second metal binding site, it is unlikely that YecA was bound to three zinc molecules, as the ICP-OES suggested. However, it is possible that a single zinc molecule was coordinated by the YecA-MeBD and that contaminating zinc that was present in the buffer was responsible for the excess. Azurin is a copper-binding protein

that purifies as apo-azurin, Cu^{2+}-bound azurin and Zn^{2+}-bound azurin. During the purification, the copper cofactor is outcompeted by zinc or lost, which could be the case with YecA and iron. A study of the interaction between the SecA-MeBD and Fe^{2+} suggested that the addition of Zn^{2+} could not outcompete the bound iron (Jamshad *et al.*, 2017). However, the oxidation of the Fe^{2+} to Fe^{3+} during the purification could affect the affinity of YecA for the metal cofactor.

Using NMR ^{1}H-^{15}N-HSQC experiments, it was possible to show that YecA bound to Fe^{2+} and that this interaction only occurred in the full-length protein. It had not been possible to assign all of the residues that interacted with Fe^{2+} in Chapter 4 because the peak intensity of the amide bonds in question were too weak. It is possible that YecA used both domains to coordinate the metal cofactor, however SAXS analysis of the full-length protein indicated that the YecA-MeBD and UPF0149 domain were independent of each other (described in Chapter 4) and comparison of the NMR ^{1}H-^{15}N-HSQC spectra of YecA and YecA201 showed that the amide bond peaks of residues 1 to 201 of YecA have the same chemical bond shifts in the presence or absence of the YecA-MeBD. These data indicated that the YecA-MeBD did not interact with the rest of the protein. These results suggested that YecA bound Fe^{2+} via the YecA-MeBD, which was consistent with the interaction between the SecA-MeBD and Fe^{2+} (Jamshad *et al.*, 2017) and Zn^{2+} (Dempsey *et al.*, 2004; Matousek and Alexandrescu, 2004).

The data presented in this chapter suggested that YecA could coordinate both Fe^{2+} and Fe^{3+}. The UV-vis spectroscopy suggested that YecA bound an Fe^{2+} ion and the EPR spectroscopy data suggested that mononucleic Fe^{3+} was bound by YecA in a tetrahedral coordination. Transferrin, ferrichrome or rubredoxin are examples of non-haem iron proteins with tetrahedral, mononuclear iron centres (Peisach *et al.*, 1971). Rubredoxin motifs have been found in nearly 2000 proteins, some of which have an unknown metal cofactor, while others have been characterised as a zinc-binding motif (Liu *et al.*, 2015).

Mononuclear Fe^{2+} cofactors are particularly sensitive to oxidation, the outcome of which is a poor understanding of the full extent of iron-binding proteins and their functions (Anjem and Imlay, 2012). Literature searches did not reveal any examples of proteins that detect fluctuations in the redox status of the cell via an iron cofactor. The coordination of high spin, mononuclear Fe^{3+} is indicative of a siderophore (Raymond and Carrano, 1979), however YecA is a cytoplasmic protein. The cytoplasm is a reducing environment, therefore the physiologically relevant state of the iron cofactor is probably Fe^{2+}.

The interaction between YecA and Zn^{2+} was analysed using NMR $^{1}H^{15}N$-HSQC experiments. Initially, the analysis of the interaction between Zn^{2+} and YecA by NMR spectroscopy did not indicate a binding event. However, the overlaid zinc-incubated ^{1}H-^{15}N-YecA HSQC spectrum and the original apo-^{1}H-^{15}N-YecA HSQC spectrum revealed that the YecA-MeBD had been degraded. Figure 5.12 shows that the degradation of the YecA-MeBD had not affected the chemical shift pattern of the rest of the protein. This result suggests that the YecA-MeBD and the main body of the protein, UPF0149, did not interact and that zinc was not bound to YecA in the absence of the YecA-MeBD. In the future, these experiments should be repeated to determine whether full-length YecA can bind zinc. This result may help to identify the YecA residues that coordinate the iron ion.

It has been suggested that role of zinc-binding by the SecA-MeBD is structural (Dempsey *et al.*, 2004; Matousek and Alexandrescu, 2004; Gelis *et al.*, 2007) and it is possible that the binding of iron by the YecA-MeBD has a similar function. However, the coordinated regulation of redox-dependent and oxidative pathways often requires a redox-sensitive functional group, like a mononuclear iron cofactor or a thiol group. Hsp33 is a redox-regulated chaperone that is activated by the oxidative unfolding of its redox-sensitive domain (Graf *et al.*, 2004; Reichmann *et al.*, 2012). In the case of Hsp33, the oxidation of the thiol groups that coordinate Fe^{2+} results in the release of its zinc cofactor and

the activation of the protein. The activity of the secreted PnkG of *Mycobacterium tuberculosis* is regulated by the redox-sensitive folding of the rubredoxin motif (Wittwer *et al.*, 2016). The oxidation of the iron cofactor bound in the rubredoxin motif of PnkG results in the unfolding of this domain and the increased activity of the catalytic site. It is possible that the activity of YecA could be regulated by either of these mechanisms. In the future, the effect of the iron cofactor on the activity of YecA should be investigated.

The studies of the interaction between SecA and zinc *in vivo* relied on the over-expression of SecA (Jamshad *et al.*, 2017). The affinity of a divalent metal ion for a ligand is correlated with the second ionisation enthalpy of the metal, which is the energy required to remove an electron from an atom with a single positive charge (Dudev and Lim, 2013). This is ranked by the Irving-Williams series. The cell machinery can regulate the cellular concentrations of these ions to promote the binding of a protein to the correct divalent metal ion and the balance between ion concentration, binding affinity and protein concentration is tightly regulated to facilitate the insertion of the correct metal. Therefore, the over-expression of a protein could result in mismetallation of the protein or production of the apo-protein. For example, the protein IscU was characterised as a zinc-binding protein (Liu *et al.*, 2005; Zhang, 2000; Ramelot *et al.*, 2004) but more recent work has identified IscU is an iron-binding protein (Ding and Clark, 2004; Lu *et al.*, 2010) and a recent study of the physiological metal cofactor of SecA revealed that SecA is an iron-binding protein (Jamshad *et al.*, 2017), which suggests that SecA has been misidentified as a zinc-binding protein. This reasoning is further strengthened by a genetic screen for *Acinetobacter baumanii* mutants defective for growth under iron limitation. The genetic screen identified an insertion that truncated the *secA* gene at the 3′ end, which deleted the SecA-MeBD (Fiester *et al.*, 2015). This implies that SecA has a role in iron acquisition, if not the direct coordination of iron by SecA.

The results presented in this chapter suggest that the cognate metal cofactor of YecA is

iron, which is likely to be coordinated via the four cysteine residues in the YecA-MeBD. The impact of the data described in this chapter related directly to the Sec pathway since a recent study has suggested that the physiological metal of SecA is Fe^{2+} (Jamshad *et al.*, 2017). With the knowledge that Fe^{2+} is sensitive to oxidation (Waldron, 2009), it is possible that SecA and YecA both monitor the redox status of the cell and alter the activity of the Sec pathway accordingly.

Chapter 6

FINAL DISCUSSION

In this thesis, the investigation of the structure and function of YecA, which is not essential for the growth of *E. coli*, was presented. The sequence homology between the SecA-MeBD and the YecA-MeBD suggested that YecA could interact with components of the Sec pathway. SecA binds to the ribosome (Huber *et al.*, 2011) and an unpublished study revealed that YecA also interacted with the ribosome (Dr. D. Huber, personal communication). The interaction of YecA with the ribosome was dependent on the YecA-MeBD and it was hypothesised that YecA could interact with the ribosome to facilitate the translocation of secretory substrates. This study indicated that YecA was a chaperone that improved the efficiency of the Sec-dependent translocation in *E. coli*. The functional characterisation of YecA was further supplemented by structural investigations that revealed that YecA was a monomeric protein that bound to a single iron ion. The results presented in this study indicate that YecA is the first component of the Sec pathway to be identified since the characterisation of YidC nearly two decades ago (Samuelson *et al.*, 2000; Scotti *et al.*, 2000).

6.1 The relationship between YecA and the Sec pathway

The Sec pathway is a pathway by which secretory proteins are translocated from the cytoplasm. In this thesis, a protein with sequence homology to SecA was investigated using *E. coli* as a model organism. YecA was previously unstudied. The main aim of this study was to investigate whether YecA was part of the Sec machinery of *E. coli*. Since the identification of YidC, much of the research in the field has focussed on the characterisation of the known components of the Sec pathway. This work has provided an insight into the mechanisms of both the cotranslational and posttranslational pathway, and how the two routes to translocation interact. However, by focussing on the core translocation machinery and the accessory proteins, the adaptability of the Sec pathway to either the diversity of secretory proteins or to changing environmental conditions, or indeed both might have been overlooked. The identification of a novel, non-essential

sec gene suggests that more proteins may be involved with protein translocation in *E. coli*.

One third of proteins are translocated by the Sec pathway. These proteins are directed to the Sec pathway by an amino-terminal signal sequence. The diversity of the cohort of secretory proteins is likely to necessitate a diverse network of chaperones. This study has presented data to support the hypothesis that YecA is a chaperone in *E. coli*. However the conditions under which *yecA* in *E. coli* is expressed is unknown. A study that explored global gene expression patterns in *Salmonella enterica* serovar Typhimurium found that *yecA* expression increased five-fold under anaerobic shock conditions (Kröger *et al.*, 2013). *E. coli* is a facultative anaerobe and the adaptation of this bacterium to changes in the availability of oxygen requires differential gene expression (Outten and Theil, 2009). It follows that the subset of secretory proteins that would be exported under aerobic conditions would be different to the subset of secretory proteins that would be exported under anaerobic conditions. Therefore, it is possible that YecA is a specific chaperone of a specific subset of proteins that are expressed under anaerobic conditions. In the future, Transposon Directed Insertion Sequencing (TraDIS) could be used to identify other chaperones that facilitate the translocation of a specific subset of secretory proteins or facilitate the translocation of secretory proteins under only a subset of conditions, such as anaerobic growth or oxidative stress. An *E. coli* strain containing a *secB* deletion could be screened under different conditions to identify potential *sec* genes. The data presented in Chapter 3 could also be verified using TraDIS because a *secB* deletion strain grown at cold temperatures should not have inactivating mutations in the *yecA* gene.

Currently, it is unknown how YecA interacts with SecA. SecA forms dimers under physiological conditions but the role of dimerisation for Sec-dependent translocation remains elusive (Ding *et al.*, 2003; Vrontou and Economou, 2004). It has been suggested that SecA can support translocation as a dimer (de Keyzer *et al.*, 2005; Jilaveani and Oliver, 2006), however other studies have suggested that the monomer is the functional

form of SecA (Or *et al.*, 2002, 2005; Or and Rapoport, 2007). It has been suggested that SecA dissociates into monomers in the presence of phospholipids or when it binds SecYEG (Or *et al.*, 2002; Zimmer *et al.*, 2008). It is possible that the oligomeric state of SecA in response to ligands enables the protein to fulfil its role at the core of the translocation machinery. With this in mind, it is possible that YecA interacts with SecA to induce conformational changes or regulate its activity. It is important that future work focusses on structurally and functionally characterising the interaction of these two proteins.

In the future, the mechanism by which YecA interacts with the Sec machinery, the ribosome and the mRNA degradation machinery could be investigated. Global inhibition of protein synthesis and mRNA turnover is a common response to stress conditions, including oxidative stress (Shen *et al.*, 2012). It is not currently known whether YecA interacts with translating ribosome. The effect of translation arrest in response to chloramphenicol-treatment followed by a ribosome-pelleting assay could provide an insight into the interaction of YecA with the translating ribosome. A similar experiment using rifampicin could help to determine whether transcription-arrest affected the interaction of YecA with the ribosome. Investigating the effect of YecA on the translocation of secretory proteins, either by pulse chase or using a YFP-fusion protein-containing a signal sequence, after translation arrest will help to determine whether YecA chaperones secretory proteins under translation arrest. The possibility that YecA interacts with the mRNA degradation machinery could be investigated by measuring the accumulation of RNA fragments in response to H_2O_2-treatment (Liu, 2012).

6.2 The relationship between the Sec pathway and iron

This study revealed that the metal cofactor of YecA was iron. The observation that YecA co-purified with iron prompted an investigation into the metal cofactor of SecA

that revealed that SecA was also an iron-binding protein (Jamshad *et al.*, 2017), which demonstrates the impact of the discovery that YecA is an iron-binding protein. The identification of two Sec proteins with iron cofactors implies that the Sec pathway is sensitive to the iron availability or responds to the presence of iron. *Lactobacillus plantarum* is one of the few known organisms that does not require iron for growth (Archibald, 1983). The homologue of SecA in *L. plantarum* lacks the 95 amino acid residues at the carboxy-tail that comprises the SecA-MeBD and the flexible linker region, which supports finding that the SecA-MeBD is an iron-coordinating domain. However, *L. plantarum* is not the only organism that lacks the carboxy-tail of SecA. A study found that a mutation that resulted in the truncation of the carboxy-terminus of SecA in *Acinetobacter baumanii* caused a poor growth phenotype in iron-depleted media, but not iron-rich media (Fiester *et al.*, 2015). The absence of key iron-receptors from the outer membrane was thought to be the cause of this phenotype. However, when this study was published, the SecA-MeBD was thought to be a zinc-binding domain and the possibility that SecA could be an iron-sensor was not investigated. It is possible that either the SecA-MeBD is required for the export of a subset of proteins under iron-limited conditions, or the SecA-MeBD has a iron-sensing role (Jamshad *et al.*, 2017).

The similarity between the YecA-MeBD and the SecA-MeBD amino acid sequences, combined with the role of both proteins in the Sec pathway, suggest a similar function for iron-binding by SecA and YecA. It is possible that structural changes that are caused by the presence or absence of iron, or changes induced by the redox state of the iron cofactor, could influence the activity of both YecA and SecA. Studies have suggested that the SecA-MeBD is folded in the presence of zinc and unfolded in the absence of zinc (Dempsey *et al.*, 2004; Matousek and Alexandrescu, 2004; Gelis *et al.*, 2007) and the presence of zinc stabilises the interaction between SecA and SecB (Fekkes *et al.*, 1999). Recently, it has been suggested that the interaction between SecA and SecB is more stable when SecA is bound to Fe^{2+} compared with Zn^{2+} (Jamshad *et al.*, 2017). It is

possible that iron could induce similar structural changes in both the YecA-MeBD and the SecA-MeBD. SAXS could be used to investigate the structure of apo-YecA compared with iron-bound YecA under reducing or oxidised conditions, which would help to determine whether YecA is a redox-sensitive protein.

The redox state of the iron could affect the function of the protein. An example of a redox-activated protein is Hsp33, which is a chaperone that responds to oxidative stress (Graumann *et al.*, 2001; Graf *et al.*, 2004). Hsp33, under the reducing conditions of the cytoplasm is a monomer that coordinates a single molecule of zinc via four conserved cysteine residues. Oxidative stress causes the four cysteine residues to form two disulphide bonds and release zinc, which induces the dimerisation of oxidised Hsp33 (Graumann *et al.*, 2001; Graf *et al.*, 2004). Dimeric Hsp33 is an active chaperone with two substrate-binding sites (Graumann *et al.*, 2001). The structural model described in this study suggests that YecA is a monomer in solution. However, this model was derived from data that was collected under aerobic and non-reducing conditions. It is possible that the chaperone activity of YecA is induced by the switch between aerobic and anaerobic conditions. Therefore to investigate whether YecA is induced by such a switch, crystal screen trials could be repeated to compare the X-ray crystal structure of aerobically and anaerobically prepared YecA, or to compare Fe^{2+}- and Fe^{3+}-reconstituted YecA.

This study has described a novel *sec* gene that had not been previously investigated. The results demonstrate that *yecA* is not essential for the growth of *E. coli* but provides the basis for further investigation into the link between the Sec pathway and stress response pathways.

Chapter 7

APPENDIX

Full list of mass spectrometry results

Table 7.1: Strep(II)-SUMO-YecA

Experiment	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# amino acids	MW (kDa)	calc. pI
1	DeaD/CdsA	416.84	66.61	1	36	36	168	629	70.5	8.72
1	YecA	340.04	49.77	2	10	12	150	221	25.0	4.59
2	SecA	276.93	57.94	2	48	49	109	901	102.0	5.60
2	FusA	225.56	59.80	1	31	31	81	704	77.5	5.38
2	YecA	200.95	64.25	2	10	11	94	221	25.0	4.59
2	DeaD/CdsA	200.43	58.66	1	30	30	81	629	70.5	8.72
2	InfB	162.07	40.00	1	35	35	73	890	97.3	6.07
1	RpsA	160.62	47.22	1	29	29	66	557	61.1	4.98
2	AceE	157.53	53.78	1	41	41	70	887	99.6	5.68
2	Pnp	145.13	56.54	1	31	31	58	711	77.1	5.21
2	AlaS	117.11	53.08	1	36	36	50	876	96.0	5.81
2	RpsA	103.77	50.81	1	26	26	39	557	61.1	4.98
2	AcnB	103.53	47.05	1	31	31	44	865	93.4	5.40
3	YecA	97.63	43.44	1	7	7	28	221	25.0	4.59
2	SucA	79.81	37.41	1	26	26	35	933	105.0	6.49
2	GyrB	75.17	36.94	1	23	23	31	804	89.9	6.06
2	DnaK	72.86	39.66	1	23	23	28	638	69.1	4.97
2	NuoG	70.34	33.81	1	24	24	32	908	100.2	6.25
2	Rnr	66.20	38.25	1	27	27	33	813	92.1	8.62
2	AceF	64.05	47.30	1	22	22	27	630	66.1	5.17
1	Pnp	60.98	37.97	1	20	20	24	711	77.1	5.21
2	ValS	52.85	27.02	1	21	21	23	951	108.1	5.34
2	Gcd	52.53	34.67	1	21	21	26	796	86.7	5.62
2	PflB	51.73	33.95	2	21	21	22	760	85.3	6.01
2	PheT	50.49	36.35	1	20	20	22	795	87.3	5.30
1	DnaK	49.89	33.70	1	17	17	20	638	69.1	4.97
2	ClpA	49.89	30.87	1	21	21	21	758	84.2	6.32
3	DnaK	49.86	19.28	1	10	10	14	638	69.1	4.97
2	Rne	49.42	27.90	1	20	20	23	1061	118.1	5.62
2	MalP	49.22	26.73	1	18	18	22	797	90.5	7.39
1	FusA	48.26	35.51	1	17	17	19	704	77.5	5.38
1	OmpF	46.87	37.85	4	11	11	18	362	39.3	4.96
2	AdhE	43.44	27.50	1	17	17	20	891	96.1	6.79
1	AtpA	42.72	37.62	1	15	16	18	513	55.2	6.13
2	GyrA	42.60	25.03	1	20	20	26	875	96.9	5.20
1	GroL	42.05	34.31	1	17	17	19	548	57.3	4.94
1	Tig	40.19	49.07	1	19	19	19	432	48.2	4.88
1	LpdA	40.07	35.23	1	12	12	16	474	50.7	6.15
2	Ppc	37.57	22.08	1	21	21	21	883	99.0	5.68
2	Pta	37.12	27.87	1	15	15	15	714	77.1	5.41
3	PpiD	36.71	27.29	1	12	12	12	623	68.1	5.07
2	IleS	36.36	27.40	1	22	22	22	938	104.2	6.01
2	TopA	36.02	26.59	1	22	22	24	865	97.3	8.46
3	RpsA	35.21	23.88	1	10	10	12	557	61.1	4.98
2	DeoD	32.91	51.88	1	11	11	18	239	25.9	5.66
2	LeuS	32.89	23.95	1	18	18	22	860	97.2	5.30
2	TolC	32.16	27.59	1	11	11	15	493	53.7	5.66
1	AceF	31.56	24.92	1	15	15	17	630	66.1	5.17
1	SucA	31.49	29.88	1	11	11	17	405	44.0	5.81
2	RpoD	30.47	21.70	1	13	13	16	613	70.2	4.79
2	TufA	30.30	41.62	2	12	12	12	394	43.3	5.45
3	FusA	29.70	16.90	1	8	8	8	704	77.5	5.38

Continued on next page

Table 7.1 – continued from previous page

Experiment	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# amino acids	MW (kDa)	calc. pI
2	RplB	29.24	37.00	1	9	9	13	273	29.8	10.93
2	TypA	26.41	18.45	1	10	10	11	607	67.3	5.33
2	PpsA	25.94	20.08	1	13	13	13	792	87.4	5.06
2	KatG	25.30	18.04	1	13	13	16	726	80.0	5.31
3	AceF	24.69	18.41	1	8	8	8	630	66.1	5.17
3	GroL	24.35	19.71	1	7	7	7	548	57.3	4.94
2	RpoC	23.10	9.38	1	11	11	11	1407	155.1	7.08
2	RpsC	21.91	48.50	1	9	9	9	233	26.0	10.27
2	MutS	20.94	13.48	1	10	10	10	853	95.2	5.55
2	GcvP	20.91	9.30	1	7	7	8	957	104.3	6.00
1	RpsC	20.52	30.04	1	6	6	7	233	26.0	10.27
2	RapA	20.21	12.71	1	11	11	13	968	109.7	5.17
3	SecA	19.31	7.77	1	5	5	6	901	102.0	5.60
3	RplF	18.75	42.37	1	7	7	7	177	18.9	9.70
1	AtpD	18.72	18.48	1	6	6	7	460	50.3	5.01
1	TypA	18.46	19.77	1	10	10	10	607	67.3	5.33
2	Prc	17.29	16.72	1	10	10	10	682	76.6	6.55
2	AcrB	17.25	9.72	1	8	8	8	1049	113.5	5.57
2	PepN	16.91	10.00	1	9	9	9	870	98.9	5.31
2	YhgF	16.42	15.14	1	11	11	12	773	85.1	6.30
1	LepA	16.19	14.36	1	8	8	8	599	66.5	5.59
1	NusA	15.34	17.78	1	8	8	8	495	54.8	4.64
2	FtsY	15.28	14.69	1	6	6	6	497	54.5	4.50
3	RpsC	14.92	33.91	1	5	5	6	233	26.0	10.27
3	Pnp	13.54	7.74	1	4	4	4	711	77.1	5.21
2	NrdA	13.42	18.27	1	12	12	12	761	85.7	6.18
2	LepA	12.72	11.02	1	7	7	7	599	66.5	5.59
2	OmpF	12.44	16.30	1	6	6	6	362	39.3	4.96
2	CarB	11.71	8.57	1	9	9	9	1073	117.8	5.34
3	LpdA	11.03	10.13	1	3	3	3	474	50.7	6.15
2	SucB	10.93	16.30	1	7	7	7	405	44.0	5.81
3	RplP	10.80	32.35	1	3	3	3	136	15.3	11.22
2	RecA	10.80	14.45	1	4	4	4	353	37.9	5.19
3	RpsD	10.55	20.87	1	3	3	3	206	23.5	10.05
1	SecD	10.54	9.11	1	4	5	5	615	66.6	8.60
1	RplB	10.24	30.04	1	6	6	6	273	29.8	10.93
2	AccB	9.95	23.72	1	3	3	6	156	16.7	4.65
3	Tig	9.92	12.50	1	4	4	4	432	48.2	4.88
2	RplO	9.61	34.03	1	4	4	4	144	15.0	11.18
3	DeaD/CdsA	9.58	10.81	1	5	5	5	629	70.5	8.72
2	MaeB	9.47	7.38	1	6	6	6	759	82.4	5.50
2	RpsM	9.43	38.14	1	3	3	3	118	13.1	10.78
3	RplD	9.34	24.38	1	3	3	3	201	22.1	9.73
3	TnaA	9.31	7.01	1	3	3	3	471	52.7	6.23
1	TufA	8.97	16.75	2	5	5	5	394	43.3	5.45
1	ClpA	8.93	6.60	1	5	5	5	758	84.2	6.32
2	RpsD	8.74	19.42	1	5	5	7	206	23.5	10.05
3	OmpC	8.63	7.90	1	3	3	3	367	40.3	4.82
1	RpsM	8.59	38.14	1	3	3	3	118	13.1	10.78
3	Enolase	8.33	11.81	1	3	3	3	432	45.6	5.48
3	TufA	8.30	10.15	2	3	3	3	394	43.3	5.45
1	SdhA	8.08	5.61	1	3	3	4	588	64.4	6.27
1	NuoG	8.07	14.93	1	8	8	8	596	68.2	6.42
2	GlyS	7.99	6.39	1	4	4	4	689	76.8	5.44
3	RplX	7.95	26.92	1	2	2	2	104	11.3	10.21

Continued on next page

Table 7.1 – continued from previous page

Experiment	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# amino acids	MW (kDa)	calc. pI
3	RplI	7.82	17.45	1	2	2	2	149	15.8	6.58
3	RplA	7.80	16.24	1	3	3	3	234	24.7	9.64
2	RpoD	7.64	4.69	1	7	7	7	1342	150.5	5.26
1	FtsH	7.51	6.52	1	3	3	3	644	70.7	6.24
2	SodB	7.48	31.61	1	5	5	5	193	21.3	5.95
3	RplB	7.48	12.09	1	2	2	2	273	29.8	10.93
2	RsxC	7.44	11.22	1	5	5	5	740	80.1	8.63
3	RpsB	7.22	12.45	1	2	2	2	241	26.7	7.14
3	RplO	7.02	18.06	1	2	2	2	144	15.0	11.18
2	RplA	7.01	16.24	1	3	3	3	234	24.7	9.64
3	RpsE	6.99	18.56	1	2	2	2	167	17.6	10.11
3	Lpp	6.90	33.33	1	2	2	2	78	8.3	9.25
3	RplN	6.77	21.14	1	2	2	2	123	13.5	10.42
2	TktA	6.61	6.64	1	4	4	4	663	72.2	5.67
3	RplE	6.46	17.88	1	2	2	2	179	20.3	9.48
1	PyrG	6.42	6.42	1	3	3	3	545	60.3	5.94
1	HslU	6.24	8.58	1	4	4	4	443	49.6	5.35
1	Der	6.17	10.82	1	5	5	5	490	55.0	5.85
2	MetE	6.14	4.52	1	3	3	3	753	84.6	5.92
3	RplJ	5.91	16.97	1	2	2	2	165	17.7	8.98
3	RplV	5.86	20.91	1	2	2	2	110	12.2	10.23
3	RplL	5.36	19.01	1	2	2	2	121	12.3	4.65
3	Hfq	5.28	24.51	1	2	2	2	102	11.2	7.65
2	BglX	5.08	5.23	1	3	3	3	765	83.4	6.21
3	RpsM	5.03	18.64	1	2	2	2	118	13.1	10.78
2	Ppa	4.93	14.20	1	2	2	3	176	19.7	5.17
3	NusA	4.85	5.45	1	2	2	2	495	54.8	4.64
3	RplQ	4.84	19.69	1	2	2	2	127	14.4	11.05
2	RplC	4.76	13.40	1	2	2	2	209	22.2	9.91
3	RplK	4.71	16.20	1	2	2	2	142	14.9	9.63
3	RplS	4.63	13.04	1	1	1	1	115	13.1	10.62
3	LacI	4.50	6.11	1	2	2	2	360	38.6	6.89
2	ClpP	4.31	14.01	1	2	2	2	207	23.2	5.80
2	GroL	4.21	5.11	1	3	3	3	548	57.3	4.94
2	LacI	4.11	5.28	1	2	2	2	360	38.6	6.89
2	MlaD	4.04	9.29	1	2	2	2	183	19.6	4.93
2	AtpD	3.97	7.39	1	3	3	3	460	50.3	5.01
2	HslU	3.83	8.13	1	4	4	4	443	49.6	5.35
2	SdhA	3.74	3.40	1	2	2	2	588	64.4	6.27
2	YgiQ	3.52	6.22	1	5	5	5	739	83.4	9.11
3	FtnA	3.45	7.88	1	1	1	1	165	19.4	4.83
3	OmpF	3.38	3.59	1	1	1	1	362	39.3	4.96
3	AtpD	3.27	4.13	1	1	1	1	460	50.3	5.01
3	RpsJ	3.26	13.59	1	1	1	1	103	11.7	9.69
3	RpmB	3.21	12.82	1	1	1	1	78	9.0	11.41
3	RplT	3.17	8.47	1	1	1	1	118	13.5	11.47
2	UvrA	3.15	1.81	1	1	1	1	940	103.8	6.64
2	IntQ	3.01	8.05	1	1	1	2	385	43.7	9.42
3	RpsK	2.82	12.40	1	1	1	1	129	13.8	11.33
3	Udp	2.81	4.35	1	1	1	1	253	27.1	6.21
3	YkgA	2.80	9.62	1	1	1	1	239	28.5	8.43
3	ClpB	2.78	1.52	1	1	1	1	857	95.5	5.52
3	RpsT	2.74	12.64	1	1	1	1	87	9.7	11.18
3	AceA	2.72	3.46	1	1	1	1	434	47.5	5.31
2	MgtA	2.70	2.00	1	1	1	2	898	99.4	5.96

Continued on next page

Table 7.1 – continued from previous page

Experiment	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# amino acids	MW (kDa)	calc. pI
3	GapA	2.70	4.23	1	1	1	1	331	35.5	7.11
3	RpsG	2.68	5.59	1	1	1	1	179	20.0	10.36
3	RplM	2.67	7.04	1	1	1	1	142	16.0	9.91
3	RplC	2.62	4.78	1	1	1	1	209	22.2	9.91
3	GrxD	2.59	11.30	1	1	1	1	115	12.9	4.75
3	Mdh	2.57	3.53	1	1	1	1	312	32.3	5.77
3	RpmD	2.54	15.25	1	1	1	1	59	6.5	10.96
3	YghJ	2.45	1.18	1	1	1	1	1520	167.1	5.03
2	FdoG	2.43	4.63	1	4	4	4	1016	112.5	7.28
2	RpsL	2.41	10.48	1	1	1	1	124	13.7	10.87
3	RpsP	2.35	13.41	1	1	1	1	82	9.2	10.55
2	PdxY	2.35	4.53	1	1	1	1	287	31.3	6.52
3	OmpA	2.34	4.05	1	1	1	1	346	37.2	6.42
3	YffS	2.32	4.09	1	1	1	1	269	29.7	5.52
3	RplW	2.31	12.00	1	1	1	1	100	11.2	9.94
3	RpsO	2.30	7.87	1	1	1	1	89	10.3	10.40
3	RpsQ	2.24	9.52	1	1	1	1	84	9.7	9.60
2	RplJ	2.22	7.88	1	1	1	1	165	17.7	8.98
2	RpsE	2.18	15.57	1	2	2	2	167	17.6	10.11
3	FabZ	2.13	5.30	1	1	1	1	151	17.0	7.39
2	RpmB	2.07	12.82	1	1	1	1	78	9.0	11.41
3	PckA	2.05	2.78	1	1	1	1	540	59.6	5.71
3	SucB	2.04	4.44	1	1	1	1	405	44.0	5.81
3	FucU	2.02	8.57	1	1	1	1	140	15.5	5.86
2	ParC	2.00	1.46	1	1	1	1	752	83.8	6.68
2	SpeB	1.98	7.19	1	2	2	2	306	33.5	5.39
3	SdhA	1.94	1.53	1	1	1	1	588	64.4	6.27
3	DamX	1.94	2.10	1	1	1	1	428	46.1	5.81
2	RplQ	1.93	6.30	1	1	1	1	127	14.4	11.05
3	RpsH	1.93	6.15	1	1	1	1	130	14.1	9.42
2	RplF	1.91	6.78	1	1	1	1	177	18.9	9.70
2	RpsI	1.89	15.38	1	2	2	2	130	14.8	10.95
2	RibF	1.85	5.75	1	1	2	2	313	34.7	9.23
2	PolA	1.85	3.02	1	3	3	3	928	103.1	5.58
2	RpsK	1.83	6.20	1	1	1	1	129	13.8	11.33
2	CysJ	1.81	2.50	1	2	2	2	599	66.2	5.02
2	CyaA	1.79	2.71	1	2	2	2	848	97.5	6.21
2	HisG	1.73	2.01	1	1	1	1	299	33.3	5.63
2	RplM	1.73	4.93	1	1	1	1	142	16.0	9.91
2	RplT	1.72	7.63	1	1	1	1	118	13.5	11.47
2	AtpA	1.71	3.31	1	2	2	2	513	55.2	6.13
2	AcnA	1.68	1.91	1	2	2	2	891	97.6	5.88
2	YdiJ	1.66	0.79	1	1	1	1	1018	113.2	7.11
2	FtsH	1.66	1.24	1	1	1	1	644	70.7	6.24
2	DadX	0.00	8.15	1	1	1	1	356	38.8	7.05
2	AmiB	0.00	5.84	1	1	1	1	445	48.0	9.29
2	CyoB	0.00	1.66	1	1	1	1	663	74.3	7.20
2	EptC	0.00	1.39	1	1	1	1	577	66.6	7.39
2	FecB	0.00	5.33	1	1	1	1	300	33.1	8.88
2	FliS	0.00	9.56	1	1	1	1	136	14.9	4.75
2	GlgX	0.00	1.83	1	1	1	1	657	73.5	6.06
2	IscS	0.00	1.98	1	1	1	1	404	45.1	6.37
2	MalE	0.00	2.53	1	1	1	1	396	43.4	5.71
2	MatH	0.00	2.36	1	1	1	1	1227	135.9	5.07
2	MnmE	0.00	1.76	1	1	1	1	454	49.2	5.02

Continued on next page

Table 7.1 – continued from previous page

Experiment	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# amino acids	MW (kDa)	calc. pI
2	MreB	0.00	3.46	1	1	1	1	347	36.9	5.26
2	MrcA	0.00	0.82	1	1	1	1	850	93.6	6.58
2	PtsI	0.00	3.83	1	1	1	1	575	63.5	4.87
2	RcsC	0.00	0.63	1	1	1	1	949	106.4	6.34
2	RplK	0.00	11.27	1	2	2	2	142	14.9	9.63
2	RplN	0.00	5.69	1	1	1	1	123	13.5	10.42
2	RplL	0.00	10.74	1	1	1	1	121	12.3	4.65
2	Rnb	0.00	1.71	1	1	1	1	644	72.4	5.62
2	RpsJ	0.00	9.71	1	1	1	1	103	11.7	9.69
2	RpsH	0.00	6.15	1	1	1	1	130	14.1	9.42
2	TamA	0.00	5.03	1	1	1	1	577	64.8	8.65
2	XapA	0.00	3.61	1	1	1	1	277	29.8	6.52
2	YceG	0.00	11.47	1	1	1	2	340	38.2	9.42
2	YegP	0.00	28.18	1	1	1	4	110	12.0	9.44
2	YffS	0.00	4.09	1	1	1	1	269	29.7	5.52
2	YghA	0.00	6.80	1	1	1	1	294	31.5	6.80
2	YhdP	0.00	2.13	1	1	1	1	1266	139.0	5.62
2	YnbC	0.00	2.05	1	1	1	1	585	65.4	8.85

Table 7.2: Strep(II)-SUMO-YecA201

Experiment	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# amino acids	MW (kDa)	calc. pI
1	DeaD/CdsA	275.59	59.94	1	32	32	112	629	70.5	8.72
2	YecA	229.30	51.58	1	7	7	71	221	25.0	4.59
2	DeaD/CdsA	197.69	52.46	1	24	24	59	629	70.5	8.72
1	RpsA	67.61	39.32	1	22	22	28	557	61.1	4.98
1	OmpF	52.51	30.66	4	12	12	22	362	39.3	4.96
3	SecA	51.73	27.64	1	25	25	27	901	102	5.6
1	FusA	49.37	36.36	1	18	18	19	704	77.5	5.38
3	DeaD/CdsA	48.94	32.43	1	17	17	26	629	70.5	8.72
1	YecA	46.28	33.94	1	6	6	20	221	25.0	4.59
1	Pnp	38.15	27.00	1	15	15	20	711	77.1	5.21
1	SecA	32.35	17.98	1	14	14	18	901	102.0	5.60
1	DnaK	30.33	23.20	1	14	14	15	638	69.1	4.97
2	FusA	27.61	17.90	1	8	8	8	704	77.5	5.38
1	TypA	27.30	22.73	1	11	11	12	607	67.3	5.33
1	AceF	24.53	22.70	1	12	12	13	630	66.1	5.17
2	RpsA	24.39	18.49	1	7	7	8	557	61.1	4.98
2	OmpF	22.85	15.75	1	4	4	6	362	39.3	4.96
2	Pnp	21.99	15.61	1	6	6	7	711	77.1	5.21
1	GroL	21.00	21.53	1	10	10	10	548	57.3	4.94
2	RpsM	20.63	53.39	1	5	5	6	118	13.1	10.78
2	RpsD	19.15	20.87	1	4	4	5	206	23.5	10.05
1	ClpA	18.33	12.53	1	10	10	10	758	84.2	6.32
1	NusA	18.27	15.15	1	7	7	9	495	54.8	4.64
1	LepA	17.89	16.86	1	10	10	10	599	66.5	5.59
1	Tig	16.91	25.46	1	10	10	10	432	48.2	4.88
2	GroL	16.15	12.96	1	4	4	4	548	57.3	4.94
3	Pnp	15.63	10.97	1	8	8	8	711	77.1	5.21
2	TufA	15.37	18.78	2	5	5	6	394	43.3	5.45
2	RplF	14.88	35.03	1	5	5	5	177	18.9	9.70
2	SecA	14.47	7.55	1	4	4	4	901	102.0	5.60
1	RpsC	12.21	33.91	1	6	6	6	233	26.0	10.27
2	LamB	11.58	9.42	1	3	3	3	446	49.9	4.98

Continued on next page

Table 7.2 – continued from previous page

Experiment	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# amino acids	MW (kDa)	calc. pI
1	SdhA	11.05	11.05	1	6	6	6	588	64.4	6.27
1	NuoC	10.86	10.57	1	6	6	6	596	68.2	6.42
2	RpsC	10.53	20.17	1	4	4	4	233	26.0	10.27
1	RpoD	10.47	7.99	1	5	5	5	613	70.2	4.79
2	DnaK	10.06	6.74	1	3	3	3	638	69.1	4.97
1	TufA	9.76	18.78	2	7	7	7	394	43.3	5.45
1	YecA	9.08	17.19	1	3	3	7	221	25.0	4.59
3	YecA	9.08	17.19	1	3	3	7	221	25	4.59
1	RplB	9.04	14.65	1	4	4	4	273	29.8	10.93
2	RplJ	8.92	24.24	1	3	3	3	165	17.7	8.98
1	PpiD	8.90	9.31	1	6	6	6	623	68.1	5.07
1	AtpA	8.81	9.55	2	5	5	5	513	55.2	6.13
1	RpsM	8.22	38.14	1	3	3	3	118	13.1	10.78
2	RpsE	7.58	20.96	1	2	2	2	167	17.6	10.11
2	RplD	7.46	24.38	1	3	3	3	201	22.1	9.73
2	NusA	7.32	6.26	1	2	2	2	495	54.8	4.64
2	AceF	7.27	4.60	1	2	2	2	630	66.1	5.17
2	LpdA	7.23	10.76	1	3	3	3	474	50.7	6.15
2	RplC	6.90	14.35	1	2	2	2	209	22.2	9.91
2	PpiD	6.66	5.78	1	2	2	3	623	68.1	5.07
2	AtpD	6.47	7.17	1	2	2	2	460	50.3	5.01
2	SdhA	5.85	3.91	1	2	2	2	588	64.4	6.27
1	KatG	5.77	6.61	1	5	5	5	726	80.0	5.31
2	AtpA	5.75	7.41	1	2	2	2	513	55.2	6.13
3	InfB	5.55	3.71	1	3	3	3	890	97.3	6.07
1	LamB	5.41	7.40	1	3	3	3	446	49.9	4.98
2	RpsK	5.38	33.33	1	3	3	3	129	13.8	11.33
2	RplT	5.33	15.25	1	2	2	2	118	13.5	11.47
2	RpsJ	5.28	20.39	1	2	2	2	103	11.7	9.69
2	RpsR	5.27	29.33	1	2	2	2	75	9.0	10.59
2	RpsG	5.10	15.64	1	3	3	3	179	20.0	10.36
2	RplI	4.56	18.12	1	2	2	2	149	15.8	6.58
1	RpsD	4.54	18.45	1	4	4	4	206	23.5	10.05
2	RplA	4.50	11.11	1	2	2	2	234	24.7	9.64
1	HscA	4.47	5.03	1	3	3	3	616	65.6	5.16
1	InfB	4.30	2.36	1	2	2	2	890	97.3	6.07
2	RplK	4.30	16.20	1	2	2	2	142	14.9	9.63
1	FtsH	4.25	6.21	1	4	4	4	644	70.7	6.24
3	FusA	4.24	2.7	1	2	2	2	704	77.5	5.38
1	RplA	4.22	11.11	1	2	2	2	234	24.7	9.64
1	RplX	4.09	17.31	1	1	1	1	104	11.3	10.21
2	SucC	3.99	3.61	1	1	1	1	388	41.4	5.52
2	RplP	3.95	11.76	1	1	1	1	136	15.3	11.22
1	SucB	3.94	5.19	1	3	3	3	405	44.0	5.81
1	HtpG	3.71	3.21	1	2	2	2	624	71.4	5.21
2	YghJ	3.70	1.25	1	1	1	1	1520	167.1	5.03
1	DamX	3.42	4.91	1	2	2	2	428	46.1	5.81
2	RplE	3.40	11.73	1	1	1	1	179	20.3	9.48
1	RecA	3.37	6.52	1	2	2	2	353	37.9	5.19
2	TnaA	3.26	2.34	1	1	1	1	471	52.7	6.23
2	FabZ	3.19	10.60	1	1	1	1	151	17.0	7.39
2	MreB	3.17	4.32	1	1	1	1	347	36.9	5.26
2	Eno	3.12	5.09	1	1	1	1	432	45.6	5.48
2	SucB	3.09	4.44	1	1	1	1	405	44.0	5.81
2	RpoA	2.87	3.65	1	1	1	1	329	36.5	5.06

Continued on next page

Table 7.2 – continued from previous page

Experiment	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# amino acids	MW (kDa)	calc. pI
1	AtpD	2.79	8.04	1	3	3	3	460	50.3	5.01
2	FtnA	2.79	7.88	1	1	1	1	165	19.4	4.83
2	RpmB	2.77	12.82	1	1	1	1	78	9.0	11.41
2	RplB	2.70	4.03	1	1	1	1	273	29.8	10.93
1	RplO	2.67	9.72	1	1	1	1	144	15.0	11.18
2	RplX	2.64	9.62	1	1	1	1	104	11.3	10.21
2	OmpC	2.58	2.45	1	1	1	1	367	40.3	4.82
2	Tig	2.53	4.40	1	1	1	1	432	48.2	4.88
2	GapA	2.52	4.23	1	1	1	1	331	35.5	7.11
2	RpsH	2.50	6.15	1	1	1	1	130	14.1	9.42
2	RpsB	2.49	6.64	1	1	1	1	241	26.7	7.14
1	TufA	2.43	2.54	2	1	1	1	394	43.3	5.45
3	TufA	2.43	2.54	2	1	1	1	394	43.3	5.45
2	DnaJ	2.38	5.59	1	1	1	1	376	41.1	7.84
2	RpsT	2.33	12.64	1	1	1	1	87	9.7	11.18
1	RpsL	2.29	10.48	1	1	1	1	124	13.7	10.87
2	ClpX	2.27	2.36	1	1	1	1	424	46.3	5.35
2	AccB	2.25	5.77	1	1	1	1	156	16.7	4.65
1	RpsN	2.18	18.81	1	1	1	1	101	11.6	11.17
2	Udp	2.07	9.49	1	1	1	1	253	27.1	6.21
2	RpmI	2.03	20.00	1	1	1	1	65	7.3	11.78
1	Rne	2.02	0.85	1	1	1	1	1061	118.1	5.62
2	MppA	2.01	3.54	1	1	1	1	537	59.9	8.41
1	RpmB	1.99	12.82	1	1	1	1	78	9.0	11.41
1	FabZ	1.86	5.30	1	1	1	1	151	17.0	7.39
1	EptC	1.78	1.39	1	1	1	1	577	66.6	7.39
1	RplF	1.76	10.73	1	2	2	2	177	18.9	9.70
1	OmpF	1.73	1.11	1	1	1	1	810	90.5	5.12
1	GlpD	1.72	1.80	1	1	1	1	501	56.7	7.44
2	Rne	1.71	1.13	1	1	1	1	1061	118.1	5.62
2	MacA	1.70	3.50	1	1	1	1	371	40.6	8.98
1	RplM	1.67	4.93	1	1	1	1	142	16.0	9.91
1	RpsA	1.62	1.97	1	1	1	1	557	61.1	4.98
1	LpdA	1.62	2.32	1	1	1	1	474	50.7	6.15
3	RpsA	1.62	1.97	1	1	1	1	557	61.1	4.98
1	PtsI	0.00	1.57	1	1	1	1	575	63.5	4.87
1	PflB	0.00	0.92	1	1	1	1	760	85.3	6.01
1	HslU	0.00	1.81	1	1	1	1	443	49.6	5.35
1	RpsE	0.00	4.79	1	1	1	1	167	17.6	10.11
1	GapA	0.00	2.42	1	1	1	1	331	35.5	7.11
1	AccB	0.00	5.13	1	1	1	1	156	16.7	4.65
1	FolX	0.00	6.67	1	1	1	1	120	14.1	7.06
1	MalE	0.00	2.53	1	1	1	1	396	43.4	5.71
1	AceE	0.00	2.82	1	3	3	3	887	99.6	5.68
1	YejH	0.00	2.22	1	1	1	1	586	66.4	7.72
1	YhjJ	0.00	1.41	1	1	1	1	498	55.5	6.02
1	RplE	0.00	4.47	1	1	1	1	179	20.3	9.48
1	YffS	0.00	4.09	1	1	1	1	269	29.7	5.52
1	YfhM	0.00	0.60	1	1	1	2	1653	181.5	5.43
2	RplO	0.00	9.72	1	1	1	1	144	15.0	11.18
2	RplU	0.00	11.65	1	1	1	1	103	11.6	9.85
2	TraG	0.00	2.24	1	1	1	1	938	102.4	6.34
3	FolX	0	6.67	1	1	1	1	120	14.1	7.06
3	AceE	0	2.82	1	3	3	3	887	99.6	5.68
3	YejH	0	2.22	1	1	1	1	586	66.4	7.72

Continued on next page

Table 7.2 – continued from previous page

Experiment	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# amino acids	MW (kDa)	calc. pI
3	YfhM	0	0.6	1	1	1	2	1653	181.5	5.43

Table 7.3: Strep(II)-SUMO

Experiment	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# amino acids	MW (kDa)	calc. pI
3	SecA	112.18	29.86	1	26	26	50	901	102	5.6
1	DnaK	63.74	27.74	1	14	14	21	638	69.1	4.97
1	SecA	38.44	17.76	1	13	13	13	901	102.0	5.60
1	GroL	32.14	24.64	1	10	10	10	548	57.3	4.94
2	RpsA	25.57	22.44	1	11	11	13	557	61.1	4.98
1	FusA	25.50	12.22	1	6	6	7	704	77.5	5.38
1	LpdA	23.59	20.04	1	7	7	7	474	50.7	6.15
1	AceF	19.57	13.97	1	7	7	7	630	66.1	5.17
1	YecA	18.85	27.60	1	4	4	7	221	25.0	4.59
2	OmpF	16.13	16.85	1	7	7	8	362	39.3	4.96
1	TufA	15.52	19.80	2	6	6	7	394	43.3	5.45
1	OmpF	14.22	10.77	1	3	3	4	362	39.3	4.96
1	RplB	13.23	20.88	1	4	4	4	273	29.8	10.93
1	RpsA	12.57	7.90	1	3	3	4	557	61.1	4.98
1	NusA	12.45	9.90	1	4	4	4	495	54.8	4.64
1	RplI	12.05	26.85	1	3	3	3	149	15.8	6.58
1	OmpC	11.82	14.44	1	4	4	4	367	40.3	4.82
1	LamB	10.98	9.19	1	3	3	3	446	49.9	4.98
2	DeaD/CdsA	10.86	12.4	1	7	7	7	629	70.5	8.72
1	RpsE	10.70	26.95	1	3	3	4	167	17.6	10.11
1	LacI	10.03	10.28	1	3	3	3	360	38.6	6.89
2	DnaK	9.99	10.03	1	7	7	7	638	69.1	4.97
1	AccC	9.62	9.58	1	4	4	4	449	49.3	7.11
1	RpsD	9.57	16.50	1	3	3	3	206	23.5	10.05
1	TnaA	9.45	6.37	1	3	3	3	471	52.7	6.23
2	GroL	9.41	12.23	1	7	7	8	548	57.3	4.94
1	Tig	8.98	9.72	1	3	3	3	432	48.2	4.88
1	RplD	8.91	18.91	1	3	3	3	201	22.1	9.73
1	RpsK	8.80	42.64	1	4	4	5	129	13.8	11.33
1	HemL	8.75	10.09	1	3	3	3	426	45.3	4.84
1	RpsG	8.69	21.23	1	3	3	3	179	20.0	10.36
1	RpsC	8.54	12.88	1	3	3	3	233	26.0	10.27
1	LacZ	8.49	3.71	1	3	3	3	1024	116.4	5.50
1	SucA	8.28	3.32	1	3	3	3	933	105.0	6.49
1	Mdh	7.98	10.58	1	3	3	3	312	32.3	5.77
1	RplF	7.84	18.08	1	3	3	3	177	18.9	9.70
1	RplJ	7.47	27.88	1	3	3	3	165	17.7	8.98
1	PpiD	7.06	4.17	1	2	2	2	623	68.1	5.07
1	Lpp	7.02	33.33	1	2	2	2	78	8.3	9.25
1	MreB	6.90	7.78	1	2	2	2	347	36.9	5.26
1	RpsB	6.85	12.86	1	2	2	2	241	26.7	7.14
1	RpsJ	6.54	23.30	1	2	2	2	103	11.7	9.69
3	FusA	6.48	8.66	1	5	5	5	704	77.5	5.38
1	Pgi	6.43	5.28	1	2	2	2	549	61.5	6.29
1	RplA	6.24	11.11	1	2	2	2	234	24.7	9.64
1	RpsM	6.20	17.80	1	2	2	2	118	13.1	10.78
2	SecA	5.97	4.11	1	3	3	3	901	102	5.6
1	Crp	5.89	10.48	1	2	2	2	210	23.6	8.25
3	RpsC	5.87	8.58	1	2	2	3	233	26	10.27

Continued on next page

Table 7.3 – continued from previous page

Experiment	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# amino acids	MW (kDa)	calc. pI
1	Hfq	5.72	24.51	1	2	2	2	102	11.2	7.65
1	ClpA	5.69	3.03	1	2	2	2	758	84.2	6.32
1	Prs	5.69	7.94	1	2	2	2	315	34.2	5.44
1	RplL	5.63	19.01	1	2	2	2	121	12.3	4.65
1	GapA	5.62	8.76	1	2	2	2	331	35.5	7.11
1	DeaD/CdsA	5.51	4.93	1	2	2	2	629	70.5	8.72
1	RplT	5.36	15.25	1	2	2	2	118	13.5	11.47
1	HslU	5.26	5.42	1	2	2	2	443	49.6	5.35
1	AceE	5.20	2.25	1	2	2	2	887	99.6	5.68
2	AtpA	5.14	7.21	2	4	4	4	513	55.2	6.13
2	TufA	4.89	8.12	2	3	3	3	394	43.3	5.45
2	RplA	4.71	11.54	1	2	2	2	234	24.7	9.64
1	RplN	4.64	13.82	1	1	1	1	123	13.5	10.42
2	SucB	4.48	7.65	1	3	3	3	405	44	5.81
1	FabZ	4.45	10.60	1	2	2	2	151	17.0	7.39
3	RpsA	4.18	8.8	1	5	5	5	557	61.1	4.98
2	RpsC	4.16	9.87	1	2	2	2	233	26	10.27
2	AceF	4.1	7.78	1	5	5	5	630	66.1	5.17
2	RpsD	4.09	8.74	1	2	2	2	206	23.5	10.05
1	SucB	3.84	4.44	1	1	1	1	405	44.0	5.81
1	RplV	3.84	11.82	1	1	1	1	110	12.2	10.23
1	Pta	3.82	2.24	1	1	1	1	714	77.1	5.41
1	FtsY	3.70	2.41	1	1	1	1	497	54.5	4.50
1	FtnA	3.55	7.88	1	1	1	1	165	19.4	4.83
1	Eno	3.55	3.24	1	1	1	1	432	45.6	5.48
1	RplO	3.54	9.72	1	1	1	1	144	15.0	11.18
1	HtpG	3.45	2.24	1	1	1	1	624	71.4	5.21
2	HslU	3.36	5.87	1	3	3	3	443	49.6	5.35
1	RpsI	3.34	9.23	1	1	1	1	130	14.8	10.95
1	FtsZ	3.34	3.92	1	1	1	1	383	40.3	4.78
1	RpmB	3.30	12.82	1	1	1	1	78	9.0	11.41
1	RpoA	3.29	3.65	1	1	1	1	329	36.5	5.06
1	Pgk	3.29	3.62	1	1	1	1	387	41.1	5.22
1	PflB	3.26	1.45	1	1	1	1	760	85.3	6.01
1	FucU	3.23	10.00	1	1	1	1	140	15.5	5.86
1	RplQ	3.16	11.81	1	1	1	1	127	14.4	11.05
1	MalE	3.13	3.03	1	1	1	1	396	43.4	5.71
1	RplP	2.98	11.76	1	1	1	1	136	15.3	11.22
1	GuaB	2.97	3.28	1	1	1	1	488	52.0	6.42
1	InfC	2.82	8.33	1	1	1	1	180	20.6	9.52
1	AccA	2.75	4.08	1	1	1	1	319	35.2	6.04
1	RplX	2.74	9.62	1	1	1	1	104	11.3	10.21
1	RplM	2.74	7.04	1	1	1	1	142	16.0	9.91
1	GatZ	2.74	2.38	2	1	1	1	420	47.1	5.77
1	KatG	2.72	1.93	1	1	1	1	726	80.0	5.31
1	SdhA	2.66	1.53	1	1	1	1	588	64.4	6.27
1	Rho	2.63	2.39	1	1	1	1	419	47.0	7.25
1	Pnp	2.60	2.11	1	1	1	1	711	77.1	5.21
3	TufA	2.55	2.54	2	1	1	1	394	43.3	5.45
1	MscM	2.54	2.26	1	1	1	4	1107	123.9	7.01
1	AcnB	2.52	1.27	1	1	1	1	865	93.4	5.40
1	AceA	2.50	3.46	1	1	1	1	434	47.5	5.31
2	RplB	2.5	4.03	1	1	1	1	273	29.8	10.93
1	RecA	2.48	3.40	1	1	1	1	353	37.9	5.19
3	AceE	2.38	3.16	1	3	3	3	887	99.6	5.68

Continued on next page

Table 7.3 – continued from previous page

Experiment	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# amino acids	MW (kDa)	calc. pI
1	RplE	2.37	4.47	1	1	1	1	179	20.3	9.48
3	YecA	2.36	6.33	1	1	1	1	221	25	4.59
1	LysS	2.36	2.38	2	1	1	1	505	57.6	5.24
1	DeoD	2.33	4.60	1	1	1	1	239	25.9	5.66
2	NusA	2.32	4.04	1	2	2	2	495	54.8	4.64
1	RpsR	2.27	13.33	1	1	1	1	75	9.0	10.59
1	YbeZ	2.20	4.05	1	1	1	1	346	39.0	5.97
1	GltA	2.14	2.58	1	1	1	1	427	48.0	6.68
3	InfB	2.13	2.02	1	2	2	2	890	97.3	6.07
1	EvgS	2.12	0.84	1	1	1	1	1197	134.7	6.21
2	NuoC	2.04	1.51	1	1	1	1	596	68.2	6.42
1	TreC	2.03	1.63	1	1	1	1	551	63.8	5.85
1	AccB	1.92	5.77	1	1	1	1	156	16.7	4.65
2	AtpD	1.71	4.78	1	2	2	2	460	50.3	5.01
2	KatG	1.67	3.31	1	3	3	3	726	80	5.31
3	RpsD	1.63	4.37	1	1	1	1	206	23.5	10.05
2	PpiD	1.63	1.44	1	1	1	1	623	68.1	5.07
2	TypA	1.62	1.81	1	1	1	1	607	67.3	5.33
3	MalE	1.61	2.53	1	1	1	1	396	43.4	5.71
2	FabZ	1.61	5.3	1	1	1	1	151	17	7.39
1	AtpD	0.00	2.39	1	1	1	1	460	50.3	5.01
1	SecD	0.00	5.53	1	1	1	1	615	66.6	8.60
1	YeaH	0.00	4.97	1	1	1	1	644	74.4	5.85
3	AceF	0	1.43	1	1	1	1	630	66.1	5.17
3	RplC	0	8.61	1	1	1	1	209	22.2	9.91
2	RmlA1	0	2.39	1	1	1	1	293	32.7	5.55
2	YafC	0	3.62	1	1	1	1	304	33.8	7.4
2	YecA	0	6.79	1	1	1	1	221	25	4.59

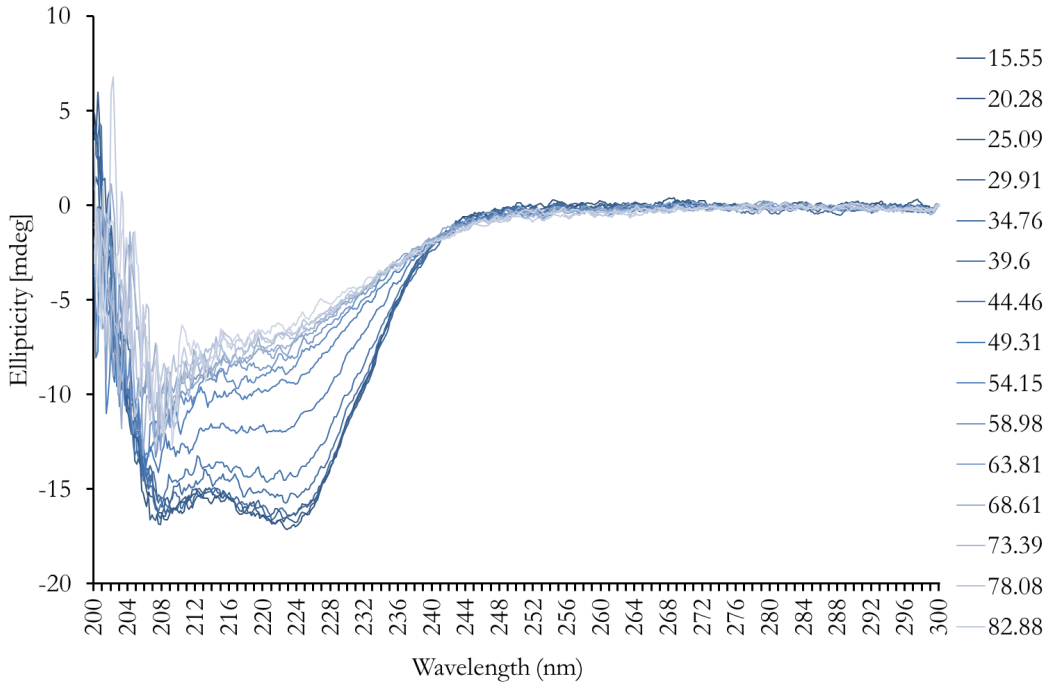
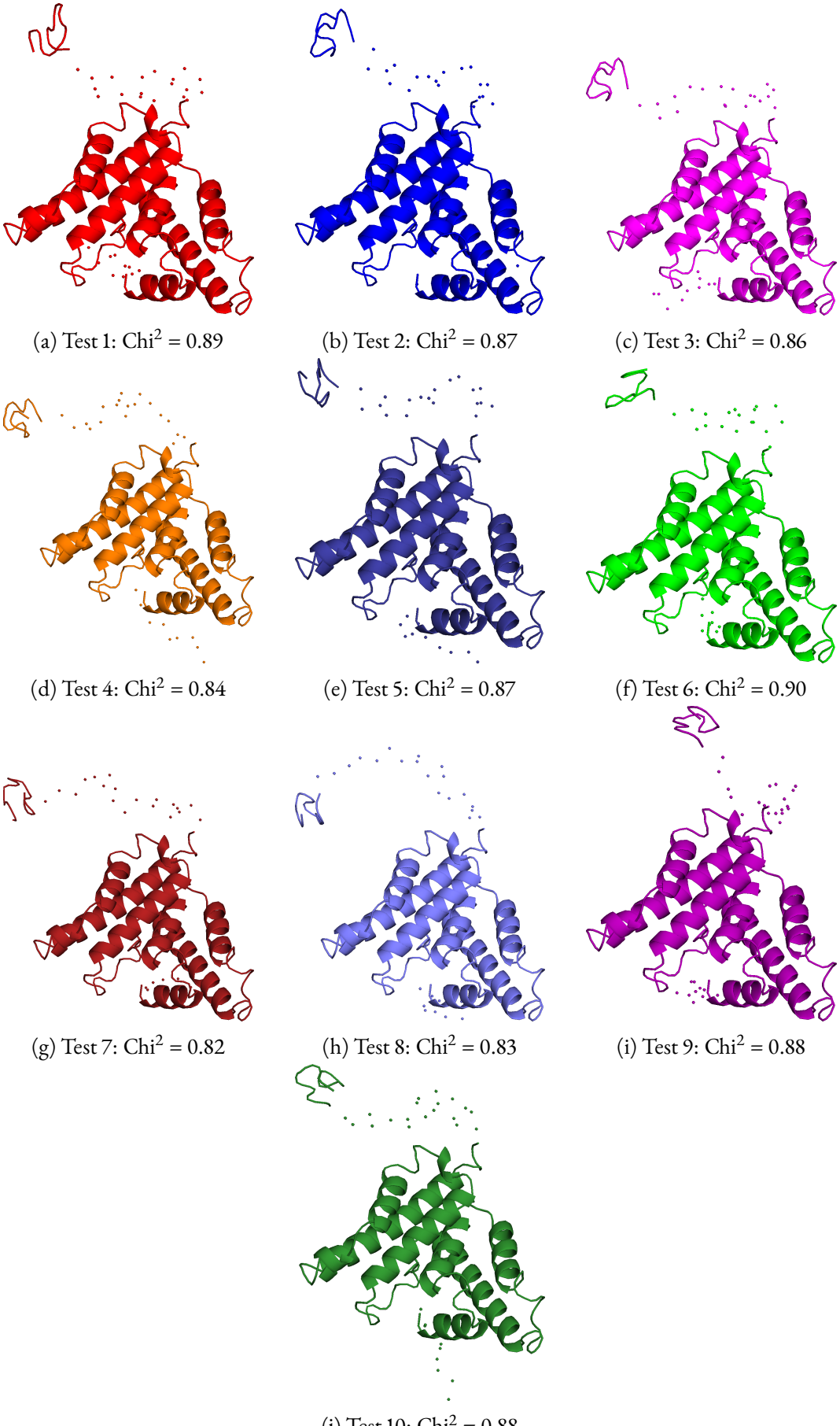


Figure 7.1: Raw CD data to show the secondary structure of YecA as a function of temperature (°C)



(a) Test 1: Chi^2 = 0.89

(b) Test 2: Chi^2 = 0.87

(c) Test 3: Chi^2 = 0.86

(d) Test 4: Chi^2 = 0.84

(e) Test 5: Chi^2 = 0.87

(f) Test 6: Chi^2 = 0.90

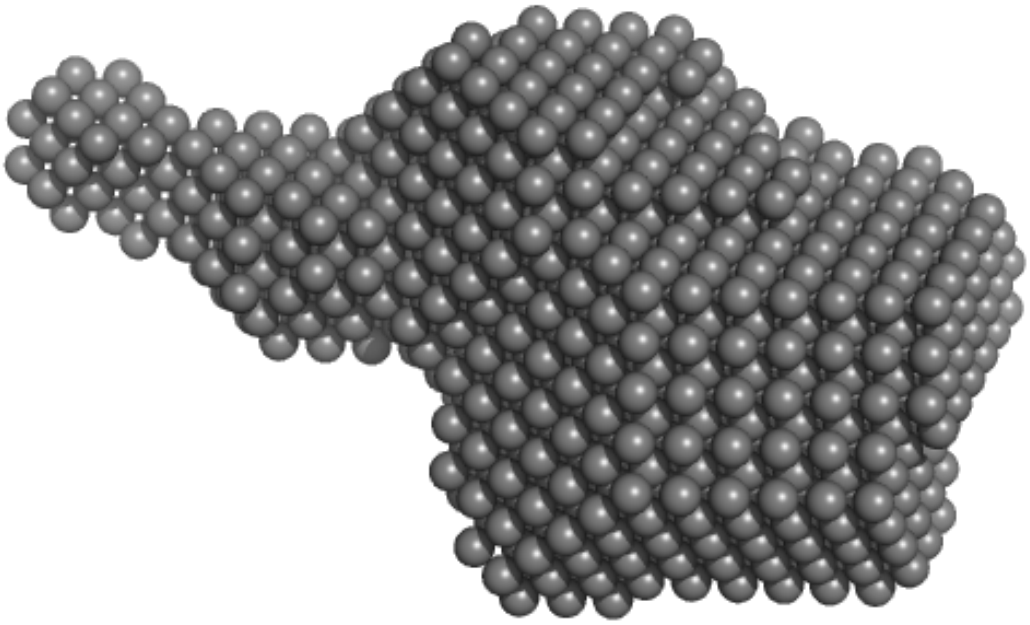
(g) Test 7: Chi^2 = 0.82

(h) Test 8: Chi^2 = 0.83

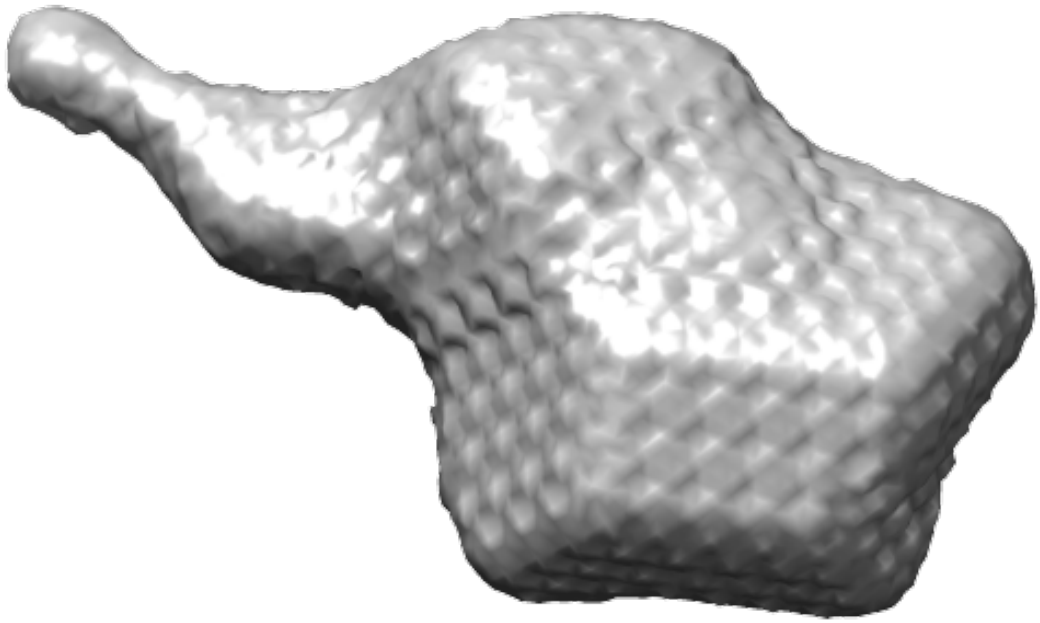
(i) Test 9: Chi^2 = 0.88

(j) Test 10: Chi^2 = 0.88

Figure 7.2: 10 independent structural models of YecA compiled using CORAL



(a) Dummy atom representation of YecA (DAMAFILT)



(b) *Ab initio* surface model of YecA (DAMAFILT)

Figure 7.3: *Ab initio* DAMMIN model of YecA generated from small angle X-ray scattering data

The program DAMMIN implements a method to restore an *ab initio*, low resolution shape of randomly oriented particles in solution (e.g., biological macromolecules) from its small angle X-ray scattering data using dummy molecules (Svergun, 1999). Chi^2 = 0.791, calculated by Dr. Tim Knowles. Images were rendered using UCSF Chimera (Pettersen *et al.*, 2004).

BIBLIOGRAPHY

P. Abrusci, M. A. McDowell, S. M. Lea, and S. Johnson. Building a secreting nanomachine: a structural overview of the T3SS. *Current Opinion in Structural Biology*, 5:111 – 117, 2014.

D. Akopian, K. Shen, X. Zhang, and S. Shan. Signal recognition particle: an essential protein-targeting machine. *Annual Review of Biochemistry*, 82:693–721, 2013.

W. J. Allen, R. A. Corey, P. Oatley, R. B. Sessions, S. A. Baldwin, S E Radford, R Tuma, and I Collinson. Two-way communication between SecY and SecA suggests a Brownian ratchet mechanism for protein translocation. *eLife*, 5:e15598, 2016.

S. F. Altschul, W. Gish, W. Miller, E. W. Miller, and D. J. Lipman. Basic local alignment tool. *Journal of Molecular Biology*, 215(3):403–410, 1990.

M. A. Andrade, P. Chacón, J. J. Merelo, and F. Morán. Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised learning neural network. *Protein Engineering, Design and Selection*, 6(4):383–389, 1993.

C. Andréasson, J. Fiaux, H. Rampelt, M. P. Mayer, and B. Bukau. Hsp110 is a nucleotide-activated exchange factor of Hsp70. *The Journal of Biological Chemistry*, 283:8877–8884, 2008.

A. Anjem and J. A. Imlay. Mononuclear iron enzymes are primary targets of hydrogen peroxide stress. *Journal of Biological Chemistry*, 287(19):15544–15556, 2012.

F. Archibald. *Lactobacillus plantarum*, an organism not requiring iron. *FEMS Microbiology Letters*, 19:29–32, 1983.

P. Artimo, M. Jonnalagedda, K. Arnold, D. Baratin, G. Csardi, E. de Castro, Duvaud S., V. Flegel, A. Fortier, E. Gasteiger, A. Grosdidier, C. Hernandez, V. Ioannidis, D. Kuznetsov, R. Liechti, S. Moretti, K. Mostaguir, N. Redaschi, G. Rossier, I. Xenarios, and H. Stockinger. Expasy: SIB bioinformatics resource portal. *Nucleic Acids Research*, 40(W1):W597–W603, 2012.

T. Baba, T.i Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, and H. Mori. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol.*, 2, 2006.

J. Bakelar, S. K. Buchanan, and N. Noinaj. The structure of the β-barrel assembly machinery complex. *Science*, 315:180–186, 2016.

T. B. Bartinkas and J. D Gitlin. How to make a metalloprotein. *Nature Structural Biology*, 8:733–734, 2001.

P. J. Bassford Jr., T. J. Silhavy, and J. R. Beckwith. Use of gene fusion to study secretion of maltose-binding protein into *Escherichia coli* periplasm. *Journal of Bacteriology*, 139(1):19–31, 1979.

J. Beckwith. The Sec-dependent pathway. *Research in Microbiology*, 164(6):497–504, 2013.

K. Beena, J. B. Udgaonkar, and R. Varardarajan. Effect of signal peptide on the stability and folding kinetics of maltose binding protein. *Biochemistry*, 43(12):43527–43535, 2004.

B. C. Berks, T. Palmer, and F. Sargent. Protein targeting by the bacterial twin-arginine translocation (Tat) pathway. *Current Opinion in Microbiology*, 8(2):174–181, 2005.

H. M. Bermand, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000.

I. Bertini, C. Luchinat, G. Parigi, and R. Pierattelli. NMR spectroscopy of paramagnetic metalloproteins. *ChemBioChem*, 6:1536–1546, 2005.

M. J. Betts and R. B. Russell. *Bioinformatics for Geneticists (Hierarchical Exotoxicology Mini Series)*, chapter Amino acid properties and consequences of substitutions. John Wiley & Sons Ltd, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, P019 8SQ, England, 1 edition, 2003.

J. M. Bollinger Jr. Getting the metal right. *Nature*, 465:40–41, 2010.

M. Botte, N. R. Zaccai, J. L. à. Nijeholt, R. Martin, K. Knoops, G. Papai, J. Zou, A. Deniaud, M. Karuppasamy, Q. Jiang, A. S. Roy, K. Schulten, P. Schultz, J. Rappsilber, G. Zaccai, I. Berger, I. Collinson, and C. Schafitzel. A central cavity within the holotranslocon suggests a mechanism for membrane protein insertion. *Scientific Reports*, 7(6):38399, 2016.

F. Bou-Abdallah and N. D. Chasteen. Spin concentration measurements of high-spin ($g' = 4.3$) rhombic iron(III) ions in biological samples: theory and application. *Journal of Biological and Inorganic Chemistry*, 13(1):15–24, 2008.

D. Boyd, D. S. Weiss, J C. Chen, and J. R. Beckwith. Towards single-copy gene expression systems making gene cloning physiologically relevant: Lambda InCh, a simple *Escherichia coli* plasmid-chromosome shuttle system. *Journal of Bacteriology*, 182(3):842–827, 2000.

E. Breukink, N. Nouwen, A. van Raalte, S. Mizushima, J. Tommassen, and B de Kruijff. The C terminus of SecA is involved in both lipid binding and SecB binding. *The Journal of Biological Chemistry*, 270:7902–7907, 1995.

F. Briani, T. Carzaniga, and G. Dehò. Regulation and functionsof bacterial PNPase. *Wiley Interdisciplinary Reviews: RNA*, 7(2):241–58, 2016.

L. Brundage, J. P. Hendrick, E. Schiebel, A. J. Driessen, and W. Wickner. The purified *E. coli* integral membrane protein SecY/E is sufficient for reconstitution of SecA-dependent precursor protein translocation. *Cell*, 62(4):649–657, 1990.

B. Bukau, E. Deuerling, C. Pfund, and E. A. Craig. Getting newly synthesized proteins into shape. *Cell*, 101:119–122, 2000.

B. Bukau, J. Weissman, and A. Horwich. Molecular chaperones and protein quality control. *Cell*, 125(3):443–451, 2006.

D. Büttner. Protein export according to schedule: architecture, assembly, and regulation of type III secretion systems from plant- and animal-pathogenic bacteria. *Microbiology and Molecular Biology Reviews*, 76(2):262–310, 2012.

I. T. Cadby, S. A. Ibrahim, M. Faulkner, D. J. Lee, D. Browning, S. J. Busby, A. L. Lovering, M. R. Stapleton, J. Green, and J. A. Cole. Regulation, sensory domains and roles of two *Desulfovibrio desulfuricans* ATCC27774 Crp family transcription factors, HcpR1 and HcpR2, in response to nitrosative stress. *Molecular Microbiology*, 102(6):1120–1137, 2016.

J. Charollais, M. Dreyfus, and I. Iost. CsdA, a cold-shock RNA helicase from *Escherichia coli*, is involved in the biogenesis of 50S ribosomal subunit. *Nucleic Acids Research*, 32(9):2751–2759, 2004.

K. E. Chatzi, M. Sardis, S. Karamanou, and A. Economou. Breaking on through to the other side: Protein export through the bacterial Sec system. *Biochemical Journal*, 449:25–37, 2013.

K. E. Chatzi, M. F. Sardis, A. Economou, and S. Karamanou. SecA-mediated targeting

and translocation of secretory proteins. *Biochimica et Biophysica Acta*, 1843:1466–1474, 2014.

R. T. Clubb, V. Thanabel, and G. Wagner. A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C′ chemical shifts in 15n 13c-labelled proteins. *Journal of Magnetic Resonance*, 92:213–217, 1992.

D. N. Collier, V. A Bankaitis, J. B Weiss, and P. J Bassford Jr. Activity of SecB promotes the export of the *E. coli* maltose-binding protein. *Cell*, 53:273–283, 1988.

R. A. Corey, W. J. Allen, J. Komar, S. Masiulis, S. Menzies, A. Robson, and I. Collinson. Unlocking the bacterial SecY translocon. *Structure*, 24(4):518–527, 2016.

J. A. Cotruvo Jr. and J. Stubbe. Metallation and mismetallation of iron and manganese proteins *in vitro* and *in vivo*: the class I ribonucleotide reductases as a case study. *Metallomics*, 4(10):1020–1036, 2012.

T. Cranford-Smith and D. Huber. The way is the goal: how SecA transports proteins across the cytoplasmic membrane in bacteria. In review, January 2018.

R. E. Dalbey and A. Kuhn. Protein traffic in Gram-negative bacteria – how exported and secreted proteins find their way. *FEMS Microbiology Review*, 36:1023–1045, 2012.

J. de Keyzer, E. O. van der Sluis, R. E. Spelbrink, N. Nijstad, B. de Kruijff, N. Nouwen, and A. J. M. Driessen. Covalently dimerized SecA is functional in protein translocation. *Journal of Biological Chemistry*, 280:35255–35260, 2005.

S. Deiterman, G. S. Sprie, and H. G. Koch. A dual function for SecA in the assembly of single spanning membrane proteins in *Escherichia coli*. *Journal of Biological Chemistry*, 280(47):39077–39085, 2005.

B. R. Dempsey, M. Wrona, J. M. Moulin, G. B. Gloor, F. Jalilehvand, G. Lajoie, G. S. Shaw, and B. H. Shilton. Solution NMR structure and X-ray absorption analysis of

the C-terminal zinc-binding domain of the SecA ATPase. *Biochemistry*, 43:9361–9371, 2004.

K. Denoncin and J. F. Collet. Disulfide bond formation in the bacterial periplasm: Major achievements and challenges ahead. *Antioxidants & Redox Signaling*, 19(1):63–71, 2013.

H. Ding and R. J. Clark. Characterization of iron binding in IscA, an ancient iron–sulphur cluster assembly protein. *Biochemical Journal*, 379:433–440, 2004.

H. Ding, J. F. Hunt, I. Mukerji, and D. Oliver. *Bacillus subtilis* SecA ATPase exists as an antiparallel dimer in solution. *Biochemistry*, 42(8729-8738), 2003.

A. J. M. Driessen and N. Nouwen. Protein translocation across the bacterial cytoplasmic membrane. *Ann. Rev. Biochem.*, 77:643–667, 2008.

D. du Plessis, N. Nouwen, and A. Driessen. The Sec translocase. *Biochimica et Biophysica Acta*, 1808:851–865., 2011.

T. Dudev and C. Lim. Competition among metal ions for protein binding sites: Determinants of metal ion selectivity in proteins. *Chemical Reviews*, 114:538–556, 2013.

E. Duin. Electron paramagnetic resonance theory [online]. *url: https://coraifeartaigh.files.wordpress.com/2008/04/epr.pdf*, Last checked:21.07.2017, 2008.

F. Duong and W. Wickner. Distinct catalytic roles of the SecYE, SecG and SecDFyajC subunits of preprotein translocase holoenzyme. *The EMBO Journal*, 16(10):2756–2768, 1997.

R.S. Dwyer, J.C. Malinverni, D. Boyd, J. Beckwith, and T.J. Silhavy. Folding LacZ in the periplasm of *Escherichia coli*. *J. Bacteriol.*, 196(18):3343–3350, 2014.

A. Economou., J. A. Pogliano., J. Beckwith., D. B. Oliver, and W. Wickner. SecA membrane cycling at SecYEG is driven by distinct ATP binding and hydrolysis events and is regulated by SecD and SecF. *Cell*, 83:1171–1181, 1995.

J. Eichler and W. Wickner. Both an N-terminal 65-kDa domain and a C-terminal 30-kDa domain of SecA cycle into the membrane at SecYEG during translocation. *Proceedings of the National Academy of Sciences*, 94(11):5574–5581, May 1997.

J. Eichler. Evolution of the prokaryotic protein translocation complex: a comparison of archaeal and bacterial versions of SecDF. *Molecular Phylogenetics and Evolution*, 27(3):504–509, 2003.

S J. Facey and A. Kuhn. Membrane integration of *E. coli* model membrane proteins. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1694(1-3):55–66, 2004.

P. Fekkes, C. van der Does, and A. J. M. Driessen. The molecular chaperone SecB is released from the carboxy-terminus of SecA during initiation of precursor protein translocation. *The EMBO Journal*, 16:6105–6113, 1997.

P. Fekkes, J. G. de Wit, J. P. van der Wolk, H. H. Kimsey, C. A. Kumamoto, and A. J. M Driessen. Preprotein transfer to the *Escherichia coli* translocase requires the co-operative binding of SecB and the signal sequence to SecA. *Molecular Microbiology*, 29(5):1179–1190, September 1998.

P. Fekkes, J. G. de Wit, A. Boorsma, R. H. E. Friesen, and A. J. M. Driessen. Zinc stabilizes the SecB binding site of SecA. *Biochemistry*, 38(16):5111–5116, 1999.

S. E. Fiester, C. C. Nwugo, W. F. Penwell, J. M. Neary, A. C. Beckett, B. A. Arivett, R. E. Schmidt, S. C. Geiger, S. M. Connerly, S. L. Menke, A. P. Tomaras, and L. A. Actis. Role of the carboxy terminus of SecA in iron acquisition, protein translocation, and virulence of the bacterial pathogen *Acinetobacter baumannii*. *Infection and Immunity*, 83(4):1354–1365, 2015.

H. Fischer, M. d. O. Neto, H. B. Napolitano, A. F. Craievich, and I. Polikarpov. The molecular weight of proteins in solution can be determined from a single SAXS measurement on a relative scale. *Journal of Applied Crystallography*, 43:101–109, 2010.

B. J. Gaffney. EPR of mononuclear non-heme iron proteins. *Biological Magnetic Resonance*, 19(28):233–268, 2009.

A. Galkin, E. Sarikaya, C. Lehmann, A. Howard, and O. Herzberg. X-ray structure of HI0817 from *Haemophilus influenzae*: protein of unknown function with a novel fold. *Proteins*, 57(4):874–877, 2004.

C. Gardel, S. Benson, J. Hunt, S. Michaelis, and J. R. Beckwith. *secD*, a new gene involved in protein export in *Escherichia coli*. *Journal of Bacteriology*, 169(3):1286–1290, 1987.

C. Gardel, K. Johnson, A. Jacq, and J. R. Beckwith. The *secD* locus of *E. coli* codes for two membrane proteins required for protein export. *The EMBO Journal*, 9(10):3209–3216, 1990.

I. Gelis, A. M. Bonvin, D. Keramisanou, M. Koukaki, G. Gouridis, S. Karamanou, A. Economou, and C. G. Kalodimos. Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell*, 131:756–769, 2007.

B. L. Geller and H. M. Green. Translocation of Pro-OmpA across inner membrane vesicles of *Escherichia coli* occurs in two consecutive energetically distinct steps. *The Journal of Biological Chemistry*, 264(28):16465–16469, 1989.

L.M. Gierasch. Signal sequences. *Biochemistry*, 28(3):923–930, 1989.

V.A. Gold, A. Robson, A. R. Clarke, and I. Collinson. Allosteric regulation of SecA: Magnesium-mediated control of conformation and activity. *The Journal of Biological Chemistry*, 282:17424–17432, 2007.

P. C. F. Graf, M. Martinez-Yamout, S. VanHaerents, H. Lilie, H. J. Dyson, and U. Jakob.

Activation of the redox-regulated chaperone Hsp33 by domain unfolding. *The Journal of Biological Chemistry*, 279(19):20529–20538, 2004.

J. Graumann, H. Lilie, X. Tang, K. A. Tucker, J. H. Hoffmann, J. Vijayalakashami, M. Saper, J. C. A. Bardwell, and U. Jakob. Activation of the redox-regulated molecular chaperone Hsp33—a two-step mechanism. *Structure*, 9(5):377–387, 2001.

E. R. Green and J. Mecsas. Bacterial secretion systems – an overview. *Microbiology Spectrum*, 4(1):10.1128, February 2016.

S. Grzesiek and A. Bax. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *Journal of Magnetic Resonance*, 96(2):432–440, 1992.

P. K. Grzyska, R. P. Hausinger, and D. A. Proshlyakov. Metal and substrate binding to Fe(II) dioxygenase resolved by UV spectroscopy with global regression analysis. *Analytical Biochemistry*, 399(1):64–71, 2010.

Y. Gu, H. Li, H. Dong, Y. Zeng, Z. Zhang, N. G. Paterson, P. J. Stansfeld, Z. Wang, Y. Zhang, and W. Wang. Structural basis of outer membrane protein insertion by the BAM complex. *Nature*, 531:64–69, 2016.

D. H. Hamilton. Using proteins in a bioinorganic laboratory experiment: Iron loading and removal from transferrin. *Journal of Chemical Education*, 86(8):969 – 972, 2009.

C.R. Harris and T.J. Silhavy. Mapping an interface of SecY (PrlA) and SecE (PrlG) by using synthetic phenotypes and *in vivo* cross-linking. *Journal of Bacteriology*, 181:3438–3444, 1999.

F. U. Hartl, S. Lecker, E. Schiebel, J. P. Hendrick, and W. Wickner. The binding cascade of SecB to SecA to SecY/E mediates preprotein targeting to the *E. coli* plasma membrane. *Cell*, 63(2):269–279, 1990.

D. Huber, D. Boyd, Y. Xia, M. H. Olma, M. Gerstein., and J. R. Beckwith. Use of thioredoxin as a reporter to identify a subset of *Escherichia coli* signal sequences that

promote signal recognition particle-dependent translocation. *Journal of Bacteriology*, 187(9):2983–91, 2005.

D. Huber, N. Rajagopalan, S Preissler, M.A. Rocco, F. Merz, G. Kramer, and B. Bukau. SecA interacts with ribosomes in order to facilitate posttranslational translocation in bacteria. *Molecular Cell*, 41(3):343–353, 2011.

D. Huber, M. Jamshad, R. Hanmer, D. Schibich, K. Döring, I. Marcomini, G. Kramer, and B. Bukau. SecA cotranslationally interacts with nascent substrate proteins *in vivo*. *Journal of Bacteriology*, 199(2):e00622–00616, 2016.

J. F. Hunt, S. Weinkauf, L. Henry, J. J. Fak, P. McNicholas, D. B. Oliver, and J. Deisenhofer. Nucleotide control of interdomain interactions in the conformational reaction cycle of SecA. *Science*, 297(2018-2026), 2002.

S. Hussain and H. D. Bernstein. The bam complex catalyzes efficient insertion of bacterial outer membrane proteins into membrane vesicles of variable lipid composition. [ePub ahead of print], January 2018.

M. G. Iadanza, A. J. Higgins, B. Schiffrin, A. N. Calabrese, D. J. Brockwell, A. E. Ashcroft, S. E. Radford, and N. A. Ranson. Lateral opening in the intact β-barrel assembly machinery captured by cryo-em. *Nature Communications*, 7:12865, 2016.

J. A. Imlay. The mismetallation of enzymes during oxidative stress. *Journal of Biological Chemistry*, 289(41):28121–28128, 2014.

H. M. N. H. Irving and R. J. P. Williams. The stability of transition-metal complexes. *Journal of the Chemical Society*, 0:3192–3210, 1953.

A. Ishihama, T. Shimada, and Y. Yamazaki. Transcription profile of *Escherichia coli*: genomic SELEX search for regulatory targets of transcription factors. *Nucleic Acids Research*, 44(5):2058–2074, 2016.

G. Isom. *A role for proteobacterial mammalian cell entry domains in phospholipid trafficking and infection*. PhD thesis, University of Birmingham, Birmingham, U.K., 2017.

M. Jamshad, R. Chandler, M. Jeeves, A. Robinson, F. Alam, T. Cranford-Smith, Anokhi Shah, O. Daubdney, K. A. Dunne, N. Nabi, A. Iqbal, S. Ahmed, A. Peacock, J. E. Lovett, T. Knowles, I. Henderson, and D. Huber. A genetic screen suggests an alternative mechanism for inhibition of SecA by azide. *Manuscript submitted for publication*, 2017.

M. R. Jensen, M. A. S. Hass, D. F. Hansen, and J. J. Led. Investigating metal-binding in proteins by nuclear magnetic resonance. *Cellular and Molecular Life Sciences*, 64:1085–1104, 2007.

L. B. Jilaveani and D. Oliver. SecA dimer cross-linked at its subunit interface is functional for protein translocation. *Journal of Bacteriology*, 188:335–338, 2006.

Joint Center for Structural Genomics. Crystal structure of hypothetical protein (YP_265345.1) from *Psychrobacter Arcticum* 273-4 at 1.75 åresolution. *To be published*, 2006.

P. G. Jones, M. Mitta, Y. Kim, W. Jiang, and M. Inuoye. Cold shock induces a major ribosomal-associated protein that unwinds double-stranded RNA in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 93:76–80, 1996.

C. M. Kaiser, H. Chang, V. R. Agashe, S. K. Lakshmipathy, S. A. Etchells, M. Hayer-Hartl, F. U. Hartl, and J.M. Barral. Real-time observation of trigger factor function on translating ribosomes. *Nature*, 444:455–460, 2006.

S. Karamanou, G. Sianidis, G. Gouridis, C. Pozidia, Y. Papanikolau, E. Papanikou, and A. Economou. *Escherichia coli* SecA truncated at its termini is functional and dimeric. *FEBS Letters*, 579(5):1267–1271, 2005.

L. E. Kay and T. Yamazaki G. Y. Xu. Enhanced-sensitivity triple-resonance spectroscopy with minimal H_2O saturation. *Journal of Magnetic Resonance*, 109(1):129–133, 1992.

L. E. Kay, P. Keiffer, and T. Saarinen. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *Journal of The American Chemical Society*, 114(26):10663–10665, 1992.

J. Keeler. Understanding NMR spectroscopy [online]. *http://www-keeler.ch.cam.ac.uk/lectures/Irvine/*, Last checked:24.09.2017, 2002.

L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg. The phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6):845–858, 2015.

S. M. Kelly, T. J. Jess, and N. C. Price. How to study proteins by circular dichroism. *Biochimica et Biophysica Acta*, 1751(10):119–139, 2005.

D. Keramisanou, N. Biris, I. Gelis, G. Sianidis, S. Karamanou, A. Economou, and C. G. Kalodimos. Disorder-order folding transitions underlie catalysis in the helicase motor of SecA. *Nature Structural and Molecular Biology*, 13(7):594–602, 2006.

A. Kihara, Y. Akiyama, and K. Ito. FtsH is required for proteolytic elimination of uncomplexed forms of SecY, an essential protein translocase subunit. *Proceedings of the National Academy of Sciences*, 92:4532–4536, 1995.

T. J. Knowles. *Solution NMR studies of the Escherichia coli methionine repressor, MetJ*. Thesis (ph.d.), Biochemistry and Microbiology, Leeds, 2005.

P. V. Konarev, V. V. Volkov, A. V. Sokolova, M. H. J. Koch, and D. I. Svergun. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography*, 36:1277±1282, 2003.

S. Kozak, L. Lercher, M. N. Karanth, R. Meijers, T. Carlomagno, and S. Boivin. Optimization of protein samples for NMR using thermal shift assays. *Journal of Biomolecular NMR*, 64:281–289, 2016.

G. Kramer, T. Rauch, W. Rist, S. Vorderwülbecke, H. Patzelt, A. Schulze-Specking, N. Ban, E. Deuerling, and B. Bukau. L23 protein functions as a chaperone docking site on the ribosome. *Nature*, 419(6903):171–174, 2002.

C. Kröger, A. Colgan, S. Srikumar, K. Händler, S. K. Sicasankaran, D. L. Hammarlöf, R. Canals, J. E. Grisson, T. Conway, K. Hokamp, and J. C. D. Hinton. An infection-relevant transcriptomic compendium for *Salmonella enterica* serovar Typhimurium. *Cell Host and Microbe*, 14(6):683–695, 2013.

R. Kudva, K. Denks, P. Kuhn, A. Vogt, M. Müller, and H-G. Koch. Protein translocation across the inner membrane of Gram-negative bacteria: the sec and tat dependent protein transport pathways. *Research in Microbiology*, 164:505–534, 2013.

C. A. Kumamoto and J. R. Beckwith. Mutations in a new gene, *secB*, cause defective protein localization in *Escherichia coli*. *Journal of Bacteriology*, 154(1):253–260, 1983.

I. Kusters, G. van den Bogaart, A. Kedrov, V. Krasnikov, F. Fulyani, B. Poolman, and A. J. M. Driessen. Quaternary structure of SecA in solution and bound to SecYEG probed at the single molecule level. *Cell*, 19(3):430–439, 2011.

S. H. Lecker, A. J. M. Driessen, and W. Wickner. ProOmpA contains secondary and tertiary structure prior to translocation and is shielded from aggregation by association with SecB protein. *The EMBO Journal*, 9(7):2309–2314, 1990.

H. C. Lee and H. D. Bernstein. The targeting pathway of *Escherichia coli* presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. *Proceedings of the National Academy of Sciences*, 98:3471–3476, 2001.

E. Lescop, P. Schanda, and B. Brutscher. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *Journal of Magnetic Resonance*, 187(1):163–169, 2007.

D. L. Leyton, A. E. Rossiter, and I. R. Henderson. From self sufficiency to dependence: mechanisms and factors important for autotransporter biogenesis. *Nature Reviews Microbiology*, 10(3):213–225, 2012.

W. Li, S. Schulman, D. Boyd, K. Erlandson, J. Beckwith, and T.A. Rapoport. The plug domain of the SecY protein stabilizes the closed state of the translocation channel and maintains a membrane seal. *Molecular Cell*, 26:511–521, 2007.

R. Lill, K. Cunningham, L. A. Brundage, K. Ito, D. Oliver, and W. Wickner. SecA protein hydrolyzes ATP and is an essential component of the protein translocation ATPase of *Escherichia coli*. *The EMBO Journal*, 8(3):961–966, 1989.

R. Lill, W. Dowhan, and W. Wickner. The ATPase activity of SecA is regulated by acidic phospholipids, SecY, and the leader and mature domains of precursor proteins. *Cell*, 60(2):271–280, 1990.

G. Liu, T. B. Topping, and L. L. Randall. Physiological role during export for the retardation of folding by the leader peptide of maltose-binding protein. *Proceedings of the National Academy of Sciences*, 86:9213 – 9217, 1989.

J. Liu, N. Oganesyan, D.H. Shin, J. Jancarik, H. Yokota, R Kim, and Kim S.H. Structural characterization of an iron-sulfur cluster assembly protein IscU in a zinc-bound form. *Proteins*, 59(4):875–881, 2005.

F. Liu, J. Geng, R. H. Gumpper, A. Barman, I. Davis, A. Ozarowski, D. Hamelberg, and A. Liu. An iron reservoir to the catalytic metal: The rubredoxin iron in an extradiol dioxygenase. *Journal of Biological Chemistry*, 290(25):15621 – 15634, 2015.

S. Liu, C. Skory, N. Qureshi, and S. Hughes. The *yajC* gene from *Lactobacillus buchneri* and *Escherichia coli* and its role in ethanol tolerance. *Journal of Industrial Microbiology and Biotechnology*, 43(4):441–450, 2016.

M. Liu. *RNA Oxidative Damage and Ribosomal RNA Surveillance under Oxidative Stress*. PhD thesis, Florida Atlantic University, Boca Raton, Florida, U.S.A., 2012.

J. Lu, J. P. Bitoun, G. Tan, W. Wang, W. Min, and H. Ding. Iron-binding activity of human iron–sulfur cluster assembly protein hIscA1. *Biochemical Journal*, 428:125–131, 2010.

J. Lubkowski and C. Barinka. Structural basis of interaction between urokinase-type plasminogen activator and its receptor. *Journal of Molecular Biology*, 363:428–495, 2006.

J. Luirink and I. Sinning. SRP-mediated protein targeting: structure and function revisited. *Biochimica et Biophysica Acta*, 1694(1-3):17–35, 2004.

J. A. Lycklama a Nijeholt and A. J. M. Driessen. The bacterial Sec-translocase: structure and mechanism. *Philosophical Transactions of The Royal Society B*, 376:1016–1028, 2012.

M. Malecki, C. Bárria, and C. M. Arraiano. Characterization of the RNase R association with ribosomes. *BMC Microbiology*, 14(34), 2014.

Y. Mao, M. P. Doyle, and J. Chen. Insertion mutagenesis of *wca* reduces acid and heat tolerance of enterohemorrhagic *Escherichia coli* O157:H7. *Journal of Bacteriology*, 183(12):3811–3815, 2001.

C. Mao, S. J. S. Hardy, and L. L. Randall. Maximal efficiency of coupling between ATP hydrolysis and translocation mediated by SecB requires two protomers of seca. *Journal of Bacteriology*, 191(3):978–984, 2009.

W. M. Matousek and A. T. Alexandrescu. NMR structure of the C-terminal domain of SecA in the free state. *Biochimica et Biophysica Acta*, 1702(2):163–171, 2004.

I. Matsumara and L. A. Rowe. Whole plasmid mutagenic PCR for directed protein evolution. *Biomolecular Engineering*, 22(1-3):73–39, 2005.

L. P. McIntosh and F. W. Dahlquist. Biosynthetic incorporation of ^{15}N and ^{13}C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Quarterly Review of Biophysics*, 23(1):1–38, 1990.

M. P. McLaughlin, M. Retegan, E. Bill, T. M. Payne, H. S. Shafaat, S. Peña, J. Sudhamsu, A. A. Engsign, B. R. Crane, F. Neese, and P. L. Holland. Azurin as a protein scaffold for a low-coordinate non-heme iron site with a small-molecule binding pocket. *Journal of The American Chemical Society*, 134(48):19746–19757, 2012.

A. McLennan, A. Bates, P. Turner, and M. White. *Molecular biology*. Garland Science, 711 Third Avenue, 8th Floor, New York, NY 10017, USA, and 3 Park Square, Milton Park, Abingdon, OX14 4RN, UK, fourth edition, 2013.

H. D. T. Mertens and D. I. Svergun. tructural characterization of proteins and complexes using small-angle X-ray solution scattering. *Journal of Structural Biology*, 172(1):128–141, 2010.

K. Michalska, X. Xu, S. Savchenko, and A. Joachimiak. Crystal structure of lpg0076 protein from *Legionella pneumophila*. *To be published*, 2012.

A. Micsonai, F. Wien, L. Kernya, Y. Lee, Y. Goto, M. Réfrégiers, and J. Kardos. Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proceedings of the National Academy of Sciences*, 112(24):e3095–3103, 2015.

A. Miller, L. Wang, and D. A. Kendall. SecB modulates the nucleotide-bound state of SecA and stimulates ATPase activity. *Biochemistry*, 41:5325–5332, 2002.

J. H. Miller. *Experiments in Molecular Genetics*. Cold Spring Habor Laboratory Press, New York, USA, 11 edition, 1972.

H. Mori and K. Ito. Different modes of SecY–SecA interactions revealed by site-directed in vivo photo-cross-linking. *Proceedings of the National Academy of Sciences*, 103(44):16159–16164, 2006.

H. Nakatogawa, A. Murakami, and K. Ito. Control of SecA and SecM translation by protein secretion. *Current Opinion in Microbiology*, 7(2):145–150, 2004.

P. Natale, T. Brüser, and A. J. M. Driessen. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—Distinct translocases and mechanisms. *Biochimica et Biophysica Acta*, 1778:1735–1756, 2008.

E. Oh, A. H. Becker, A. Sandikci, D. Huber, R. Chaba, F. Gloge, R. J. Nichols, A. Typas, C. A. Gross, G. Kramer, J. S. Weissman, and B. Bukau. Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor *in vivo*. *Cell*, 147(6):1295–1308, 2011.

D.B. Oliver and J. Beckwith. *E. coli* mutant pleiotropically defective in the export of secreted proteins. *Cell*, 25:765–772, 1981.

D. B. Oliver and J. Beckwith. Regulation of a membrane component required for protein secretion in *Escherichia coli*. *Cell*, 30(1):311–319, 1982.

D. B. Oliver, J. Norman, and S. Sarker. Regulation of *Escherichia coli secA* by cellular protein secretion proficiency requires an intact gene X signal sequence and an active translocon. *Journal of Bacteriology*, 180(19):5240–5242, 1998.

E. Or and T. A. Rapoport. Cross-linked SecA dimers are not functional in protein translocation. *FEBS Letters*, 581(4):2616–2620, 2007.

E. Or, A. Navon, and T. Rapoport. Dissociation of the dimeric SecA ATPase during

protein translocation across the bacterial membrane. *The EMBO Journal*, 21:4470–4479, 2002.

E. Or, D. Boyd, S. Gon, J. Beckwith, and T. A. Rapoport. The bacterial ATPase SecA functions as a monomer in protein translocation. *Journal of Biological Chemistry*, 280:9097–9105, 2005.

A. R. Osborne, W. M. Clemons Jr., and T. A. Rapoport. A large conformational change of the translocation ATPase SecA. *Proceedings of the National Academy of Sciences*, 101(30):10937–10942, 2004.

F. W. Outten and E. C. Theil. Iron-based redox switches in biology. *Antioxidants & Redox Signaling*, 11(5):1029–1046, 2009.

V. G. Pans, P. Vogel, W. E. Trommer, and R. Varadarajan. A thermodynamic coupling mechanism for the disaggregation of a model peptide substrate by chaperone secB. *Journal of Biological Chemistry*, 275(25):18698–18703, 2000.

E. Papanikou, S. Karamanou, C. Baud, M. Frank, G. Sianidis, D. Keramisanou, C. G. Kalodimos, A. Kuhn, and A. Economou. Identification of the preprotein binding domain of SecA. *Journal of Biological Chemistry*, 280:43209–43217, 2005.

E. Papanikou, S. Karamanou, and A. Economou. Bacterial protein secretion through the translocase nanomachine. *Nature Reviews Microbiology*, 5(11):839–851, 2007.

E. Park and T. A. Rapoport. Preserving the membrane barrier for small molecules during bacterial protein translocation. *Nature*, 473:239–242, 2011.

C. N. Patel, V. F. Smith, and L. L. Randall. Characterization of three areas of interactions stabilizing complexes between SecA and SecB, two proteins involved in protein export. *Protein Science*, 15(6):1379–1386, 2006.

S. H. Payne, S. Bonissone, S. Wu, R. N. Brown, D. N. Ivankov, D. Frishman, L. Paša-Tolić,

R. D. Smith, and P. A. Pevzner. Unexpected diversity of signal peptides in prokaryotes. *mBio*, 3(6):e00339–e003312, 2012.

J. Peisach, W. E. Blumberg, E. T. Lode, and M. J. Coon. An analysis of the electron paramagnetic resonance spectrum of pseudomonas oleovorans rubredoxin. a method for determination of the liganids of ferric iron in completely rhombic sites. *Journal of Biological Chemistry*, 246:5877–5881, 1971.

K. Pereira de Jesus, L. Serre, C. Zelwer, and B. Castaing. Structural insights into abasic site for Fpg specific binding and catalysis: comparative high-resolution crystallographic studies of Fpg bound to various models of abasic site analogues-containing DNA. *Nucleic Acids Research*, 33:5936–5944, 2005.

M. V. Petoukhov, D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. T. Mertens, P. V. Konarev, and D. I. Svergun. New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography*, 45:342–350, 2012.

E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D.M. Greenblatt, E. C. Meng, and T. E. Ferrin. UCSF Chimera–a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.

K. J. Pogliano and J. R. Beckwith. The Cs *sec* mutants of *Escherichia coli* reflect the cold sensitivity of protein export itself. *Genetics*, 133(4):763–73, 1993.

K. J. Pogliano and J. R. Beckwith. SecD and SecF facilitate protein export in *E. coli*. *The EMBO Journal*, 13(3):554–561, 1994.

M. E. Pohlschroder, N. J. Hartmann, K. Dilks Hand, and A. Haddad. Diversity and evolution of protein translocation. *Annual Review of Microbiology*, 59:91–111, 2005.

T. A. Ramelot, J. R. Cort, S. Goldsmith-Fischman, G. J. Kornhaber, R. Xiao, R. Shastry, T. B. Acton, B. Honig, G. T. Montelione, and M.A. Kennedy. Solution NMR struc-

ture of the iron-sulfur cluster assembly protein U (IscU) with zinc bound at the active site. *Journal of Molecular Biology*, 344(2):567–583, 2004.

L. L. Randall and S. J. S. Hardy. SecB, one small chaperone in the complex milieu of the cell. *Cellular and Molecular Life Sciences*, 59:1617–1623, 2002.

L. L. Randall, T. B. Topping, S. J. Hardy, M. Y. Pavlov, D. V. Freistroffer, and M. Ehrenberg. Binding of SecB to ribosome-bound polypeptides has the same characteristics as binding to full-length, denatured proteins. *Proceedings of the National Academy of Sciences*, 94(3):802–807, February 1997.

L. L. Randall, J. M. Crane, G. Liu, and S. J. S. Hardy. Sites of interaction between SecA and the chaperone SecB, two proteins involved in export. *Protein Science*, 13(4):1124–1133, 2004.

L. L. Randall, J. M. Crane, A. A. Lilly, G. Liu, C. Mao, C. N. Patel, and S. J. S. Hardy. Asymmetric binding between SecA and SecB two symmetric proteins: Implications for function in export. *Journal of Molecular Biology*, 348(2):479–489, 2005.

L. L. Randall. Translocation of domains of nascent periplasmic proteins across the cytoplasmic membrane. *Cell*, 33:231–240, 1983.

K. N. Raymond and C. J. Carrano. Coordination chemistry and microbial iron transport. *Accounts of Chemical Research*, 12(5):183–190, 1979.

M. E. Regonesi, M. Del Favero, F. Basilico, F. Briani, L. Benazzi, P. Tortora, P Mauri, and G. Dehò. Analysis of the *Escherichia coli* RNA degradosome composition by a proteomic approach. *Biochemie*, 88(2):151–161, 2006.

D. Reichmann, Y. Xu, C. M. Cremers, M. Ilbert, R. Mittelman, M. C. Fitzgerald, and U. Jakob. Order out of disorder: Working cycle of an intrinsically unfolded chaperone. *Cell*, 148(5):947–957, 2012.

J. Ren, L. Wen, X Gao, C. Jin, Y. Xue, and X. Yao. Dog 1.0: Illustrator of protein domain structures. *Cell Research*, 19:271–273, 2009.

A. Robson, B. Carr, R. B. Sessions, and I. Collinson. Synthetic peptides identify a second periplasmic site for the plug of the SecYEG protein translocation complex. *FEBS Letters*, 583:207–212, 2009.

A. Robson, V. A. M. Gold, S. Hodson, A. R. Clarke, and I. Collinson. Energy transduction in protein transport and the atp hydrolytic cycle of seca. *Proceedings of the National Academy of Sciences*, 106(13):5111–5116, 2009.

I. Sachelaru, N. A. Petriman, R. Kudva, P. Kuhn, T. Welte, B. Knapp, F. Drepper, Warscheid B., and H. G. Koch. YidC occupies the lateral gate of the SecYEG translocon and is sequentially displaced by a nascent membrane protein. *Journal of Biological Chemistry*, 288(3):16295–16307, 2013.

I. Sachelaru, L. Winter, D. G. Knyazec, M. Zimmerman, A. Vogt, R. Kuttner, N. Ollinger, C. Siligan, P. Phl, , and H-G. Koch. YidcC and SecYEG form a heterotetrameric protein translocation channel. *Scientific Reports*, 7(101):1–15, 2015.

M. Salzmann, K. Pervushin, G. Wider, H. Senn, and K. Wüthrich. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proceedings of the National Academy of Sciences*, 93(23):13585–13590, 1998.

M. Salzmann, G. Wider, K. Pervushin, H. Senn, and K. Wüthrich. TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins. *Journal of The American Chemical Society*, 121(4):844–848, 1999.

J. C. Samuelson, M. Chen, F. Jiang, I. Möller, M. Wiedmann, A. Kuhn, G. J. Phillips, and R. E. Dalbey. YidC mediates membrane protein insertion in bacteria. *Nature*, 406(6796):637–647, 2000.

I. Saraogi, D. Akopian, and S. O. Shan. Regulation of cargo recognition, commitment, and unloading drives cotranslational protein targeting. *Journal of Cell Biology*, 205(5):693–706, 2014.

P. Schanda, H Van Melckebeke, and B. Brutscher. Speeding up three-dimensional protein NMR experiments to a few minutes. *Journal of The American Chemical Society*, 128(28):9042–9043, 2006.

F. Schlünzen, D. N. Wilson, P. Tian, J. M. Harms, S. J. McInnes, H. A. Hansen, R. Albrecht, J. Buerger, S. M. Wilbanks, and P. Fucini. The binding mode of the trigger factor on the ribosome: implications for protein folding and SRP interaction. *Structure*, 13(11):1685–1694, 2005.

M. G. Schmidt and D. B. Oliver. SecA protein autogenously represses its own translation during normal protein secretion in *Escherichia coli*. *Journal of Bacteriology*, 171(2):643–649, 1989.

R. J. Schulze, J. Komar, M. Botte, W. J. Allen, S. Whitehouse, V. A. M. Golda, J. A. Lycklama a Nijeholt, K. Huard, I. Berger, C. Schaffitzel, and I. Collinson. Membrane protein insertion and proton-motive-force-dependent secretion through the bacterial holo-translocon SecYEG–SecDF–YajC–YidC. *Proceedings of the National Academy of Sciences*, 111(13):4844–4849, 2014.

P. A. Scotti, M. L. Urbanus, J. Brunner, J. L. de Gier, G. von Heijne, C. van der Does, A. J. M. Driessen, B. Oudega, and J. Luirink. YidC, the *Escherichia coli* homologue of mitochondrial Oxa1p, is a component of the Sec translocase. *The EMBO Journal*, 19(4):mediates membrane protein insertion in bacteria, 2000.

J. Serek, G. Bauer-Manz, G. Struhalla, L. van den Berg, D. Kiefer, R. Dalbey, and A. Kuhn. *Escherichia coli* YidC is a membrane insertase for sec-independent proteins. *The EMBO Journal*, 23(2):294–301, 2004.

R. Shah, A. T. Large, A. Ursinus, B. Lin, P. Gowrinathan, J. Martin, and P A. Lund. Replacement of GroEL in *Escherichia coli* by the group ii chaperonin from the archaeon *Methanococcus maripaludis*. *Journal of Bacteriology*, 198(19):2692–2700, 2016.

V. Sharma, A. Arockiasamy, D. R. Ronning, C. G. Savva, A. Holzenburg, M. Braunstein, W. R. Jacobs Jr., and J. C. Sacchettini. Crystal structure of *M. tuberculosis* SecA, a preprotein translocating ATPase. *Proceedings of the National Academy of Sciences*, 100:2243–2248, 2003.

K. Shen, S. Arslan, D. Akopian, D. Ha, and S. O. Shan. Activated GTPase movement on an rna scaffold drived co-translational protein targeting. *Nature*, 492(7428):271–275, 2012.

E. Slasi, E. Farah, Z. Netter, J. Dann, and D. N. Ermolenko. Movement of Elongation Factor G between compact and extended conformations. *Journal of Molecular Biology*, 427(2):454–467, 2015.

M. A. Smith, W. M. Clemons Jr., C. J. DeMars, and A. M. Flower. Modeling the effects of *prl* mutations on the SecY complex. *Journal of Bacteriology*, 187(18):6454–6465, 2005.

M. S. Smyth and J. H. J. Martin. X-ray crystallography. *Journal of Clinical Pathology*, 53:8–14, 2000.

T. Soderberg. *Organic Chemistry With a Biological Emphasis*, volume I. Chemistry Publications, UM-Morris Digital Commons, 1 [ebook] edition, 2016.

D. I. Svergun. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical Journal*, 76(6):2879–2886, 1999.

The UniProt Consortium. UniProt: the universal protein knowledge database. *Nucleic Acids Research*, 45(D1):D158–D169, 2016.

S. Thomas, I. B. Holland, and Schmitt L. The Type 1 secretion pathway - the hemolysin system and beyond. *Biochimica et Biophysica Acta*, 1843(8):1629 – 1642, 2014.

T. B. Topping, R. L. Woodbury, D. L. Diamond, S. J. S. Hardy, and L. L. Randall. Direct demonstration that homotetrameric chaperone SecB undergoes a dynamic dimer-tetramer equilibrium. *Journal of Biological Chemistry*, 276:7437–7441, 2001.

S. Tottey, K. J. Waldron, S. J. Firbank, B. Reale, C. Bessant, K. Sato, T. R. Cheek, J. Gray, M. J. Banfield, C. Dennison, and N. J. Robinson. Protein-folding location can regulate manganese- binding versus copper- or zinc-binding. *Nature*, 455(7216):1138–1142, 2008.

T. Tseng, K. S. Gratwick, J. Kollman, D. Park, D. H. Nies, A. Goffeau, and M. H. Saier Jr. The RND permease superfamily: An ancient, ubiquitous and diverse family that includes human disease and development proteins. *Journal of Molecular Microbiology and Biotechnology*, 1(1):107–125, 1999.

A. Tsirigotaki, J. De Geyter, N. Šoštarić, A. Economou, and S. Karamanou. Protein export through the bacterial Sec pathway. *Nature*, 15(1):21–36, 2017.

T. Tsukazaki, H. Mori, Y. Echizen, R. Ishitani, S. Fukai, T. Tanaka, A. Perederina, D. G. Vassylyev, T. Kohno, A. D. Maturana, K. Ito, and O. Nureki. Structure and function of a protein export-enhancing membrane component SecDF. *Nature*, 474(7350):235–238, 2013.

R. S. Ullers, J. Luirink, N. Harms, F. Schwager, C. Georgopoulos, and P. Genevaux. SecB is a bona fide generalized chaperone in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 101(20):7583–7588, 2004.

R. S. Ullers, D. Ang, F. Schwager, C. Georgopoulos, and P. Genevaux. Trigger factor can antagonize both SecB and DnaK/DnaJ chaperone functions in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 104(9):3101–3106, 2007.

M. L. Urbanus, L. Fröderberg, D. Drew, P. Björk, J. W. L. de Gier, J. Brunner, B. Oudega, and J. Luirink. Targeting, insertion, and localization of *Escherichia coli* YidC. *The Journal of Biological Chemistry*, 277(15):12718–12723, 2002.

B. van den Berg, W. M. Clemons Jr., I. I. Collinson, Y. Modis, S. C. Harrison E. Hartmann, and T.A. Rapoport. X-ray structure of a protein-conducting channel. *Nature*, 427:36–44, 2004.

K. H. M. van Wely, J. Swaving, M. Klein, R. Freudl, and A. J. M. Driessen. The carboxyl terminus of the *Bacillus subtilis* SecA is dispensable for protein secretion and viability. *Microbiology*, 146:2573–2581, 2000.

D. G. Vassylyev, H. Mori, M. N. Vassylyeva, T. Tsukazaki, Y. Kimura T. H. Tahirov, and K. Ito. Crystal structure of the translocation ATPase SecA from *Thermus thermophilus* reveals a parallel, head-to-head dimer. *Journal of Molecular Biology*, 364:248–258, 2006.

G. von Heijne. Signal sequences. The limits of variation. *Journal of Molecular Biology*, 184(1):99–105, 1985.

G. von Heijne. Membrane proteins: from sequence the structure. *Annu. Rev. Biophys. Biomol. Struct*, 23:167–192, 1994.

E. Vrontou and A. Economou. Structure and function of SecA, the preprotein translocase nanomotor. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1694(1-3):67–80, 2004.

K. J. Waldron. Metalloproteins and metal sensing. *Nature*, 460:823–830, 2009.

H. Wang, B. Na, H. Yang, and P. C. Tai. Additional *in vitro* and *in vivo* evidence for SecA functioning as dimers in the membrane: Dissociation into monomers is not essential for protein translocation in *Escherichia coli*. *Journal of Bacteriology*, 190(4):1413–14–18, 2008.

L. Whitmore and B. A. Wallace. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research*, 32(Web server issue):W668–W673, 2004.

L. Whitmore and B. A. Wallace. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers*, 89(5):392–400, 2008.

J. Wild, E. Altman, T. Yura, and C. A. Gross. DnaK and DnaJ heat shock proteins participate in protein export in *Escherichia coli*. *Genes and Development*, 6(7):1165–1172, 1992.

M. P. Williamson. Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 73:1–16, 2013.

M. Wittwer, Q. Luo, V. R. I. Kaila, and S. A. Dames. Oxidative unfolding of the rubredoxin domain and the natively disordered n-terminal region regulate the catalytic activity of *M. tuberculosis* Protein Kinase G. *Journal of Biological Chemistry*, 291(32):27062–27072, 2016.

R. L. Woodbury, S. J. Hardy, and L. L. Randall. Complex behavior in solution of homodimeric SecA. *Protein Science*, 11(4):875–882, 2012.

C.H. Wu, S. Chen, M.R. Shortreed, G.M. Kreitinger, Y. Yuan, B.L. Frey, Y. Zhang, S. Mirza, L.A. Cirillo, M. Olivier, and L.M. Smith. Sequence-specific capture of protein-DNA complexes for mass spectrometric protein identification, 2011.

J. Ye, A. R. Osborne, M. Groll, and T. A. Rapoport. RecA-like motor ATPases–lessons from structures. *Biochimica et Biophysica Acta - Bioenergetics*, 1659(1):1–18, 2004.

J. W. Zhang, G. Butland, J. F. Greenblatt, A. Emili, and D. B. Zamble. A role for SlyD in the *Escherichia coli* hydrogenase biosynthetic pathway. *The Journal of Biological Chemistry*, 280(6):4360–4366, 2005.

D. Zhang, K. Yan, Y. Zhang, G. Liu, X. Cao, G. Song, Q. Xie, N. Gao, and Y. Qin. New insights into the enzymatic role of EF-G in ribosome recycling. *Nucleic Acids Research*, 21:10525–10533, 43.

J. Zhang. Protein-length distributions for the three domains of life. *Trends in Genetics*, 16(3):107–109, 2000.

J. Zhou and Z. Xu. Structural determinants of SecB recognition by SecA in bacterial protein translocation. *Nature Structural Biology*, 10(11):942–947, 2003.

J. Zimmer and T. A. Rapoport. Conformational flexibility and peptide interaction of the translocation ATPase SecA. *Journal of Molecular Biology*, 394(4):606–612, 2009.

J. Zimmer, Y. Nam, and T.A. Rapoport. Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature*, 455:936 –943, 2008.

C. R. Zito and D. Oliver. Two-stage binding of SecA to the bacterial translocon regulates ribosome-translocon interaction. *Journal of Biological Chemistry*, 278(42):40640–40646, 2003.

D. Zopf, H. D. Bernstein, A. E. Johnson, and P. Walter. The methionine-rich domain of the 54 kD protein subunit of the signal recognition particle contains an RNA binding site and can be crosslinked to a signal sequence. *The EMBO Journal*, 9(13):4511–4517, 1990.