Towards a Determination of the SecA Signal Peptide-Binding Site

by

William H. Beluch
Class of 2013

A thesis submitted to the faculty of Wesleyan University in partial fulfillment of the requirements for the Degree of Bachelor of Arts with Departmental Honors in Molecular Biology and Biochemistry
Abstract

SecA is an essential motor protein found in bacteria that uses ATP hydrolysis to transport preproteins across or into the cytoplasmic membrane. Before this translocation occurs, SecA must recognize and bind an N-terminal signal sequence of the newly synthesized preprotein. The location of this binding site on SecA and the orientation of the bound signal peptide is currently a matter of debate. Attempts at obtaining a high resolution X-ray crystal structure of signal peptide bound to SecA in order to elucidate the binding site have been inhibited by signal sequence hydrophobicity and aggregation. In this study, we have created a mutant E.coli SecA protein with the signal sequence KRR-LamB attached to the C-terminus using DNA recombination technology with the aim to obtain an X-ray crystal structure of SecA self-binding the attached signal peptide. Fluorescence anisotropy assays with a dye-labeled signal peptide and SecA or SecA-KRR-LamB proteins were used to analyze the extent of self-binding in two SecA-KRR-LamB constructs. Both the full length SecA-KRR-LamB and the truncated version lacking the C-terminus exhibit significant self-binding of the attached signal peptide. The functionality of the SecA-attached KRR-LamB signal peptide was tested by engineering a SecA-KRR-LamB-PhoA chimera in which the alkaline phosphatase gene lacking its endogenous signal sequence was fused downstream of the SecA-attached KRR-LamB signal sequence, and the ability of the latter signal peptide to transport alkaline phosphatase across the membrane where it is enzymatically active in vivo was assayed. Both the full-length and truncated construct show self-binding and are able to facilitate translocation of the fused preprotein substrate to different extents.
Acknowledgements

First and foremost, I would like to thank my advisor, Professor Donald Oliver, for giving me the opportunity to work in his lab and produce the work that turned into this thesis. I greatly appreciate his academic insight, life advice, and overall support.

A special thank you goes to the readers of my thesis – Professors Manju Hingorani and Ishita Mukerji – for taking the time to look over my work and provide feedback. I highly value their input.

Thank you to the MB&B department as a whole – faculty, staff, students-- it has been an enriching few years.

To the members of the Oliver Lab, past and present -- thanks for everything. My lab experience would not have been nearly as positive without you in it. Qi, it has been an exceptional pleasure collaborating with you on this project, I hope you didn’t mind my constant questions too much. Zeliang, Christine, Tithi, Sudipta, Jenn, Lorry and Stephanie, thank you.

This process would not have been tolerable without the Cave – Zack, Ryu, Jaewon, and Corey—and of course all those who associate with the Cave. It has been a blast living with you guys.

Thanks to my family for backing me in whatever I do, especially my Aunt Ellen and Uncle Lou. I appreciate beyond words your constant support, enthusiasm, and belief.

Finally, to Kat, the person who has kept me reasonably sane throughout my two years of often frustrating lab work and my semester of writing. Thank you. I can’t imagine these past years without you.
List of Figures and Tables

I. Introduction

Figure 1. Gram-positive and Gram-negative envelope. 3
Figure 2. Translocation pathways in bacteria.. 5
Figure 3. SRP-mediated targeting and translocation. 8
Figure 4. Signal sequence structure. 9
Figure 5. Sec-B targeting. 11
Figure 6. SecY crystal structure. 13
Figure 7. Membrane topology of SecYEG. 15
Figure 8. SecA crystal structure. 18
Figure 9. SecA-SecY crystal structure. 23
Figure 10. SecA signal peptide binding site. 27
Figure 11. Jablonski diagram. 29
Figure 12. Anisotropy assay theory. 30

II. Materials and Methods

Table 1. Oligonucleotide primers. 35
Table 2. Strains and plasmids. 37

III. Results

Figure 13. Chimera protein constructs. 46
Figure 14. CTL inhibiton of binding site. 48
Figure 15. Coomassie stained protein gel. 49
Figure 16. Anisotropy schematic. 51
Figure 17. E. coli anisotropy curves. 52
Figure 18. B. subtilis anisotropy curves. 56
Figure 19. Alkaline phosphatase assay layout. 58

Figure 20. Alkaline phosphatase assay results.
# Table of Contents

I. Introduction .......................................................................................................................... 1  

A. Protein Secretion Pathways ................................................................................................. 2  

B. SRP mediated export ........................................................................................................... 7  

C. Signal sequences .................................................................................................................. 9  

D. SecB preprotein targeting .................................................................................................. 10  

E. The SecYEG protein conducting channel ......................................................................... 12  

F. SecDFyajC .......................................................................................................................... 16  

G. SecA .................................................................................................................................. 16  
   G.1 Structure .......................................................................................................................... 16  
   G.2. Oligomeric State .............................................................................................................. 20  
   G.3. SecA interactions with SecB and SecYEG ...................................................................... 21  
   G.4. Mechanisms of SecA-mediated translocation ................................................................. 23  
   G.5. Signal sequence binding .................................................................................................. 25  

H. Fluorescence anisotropy ..................................................................................................... 28  

I. Rationale for work ............................................................................................................... 31  

II. Materials and Methods ..................................................................................................... 33  

A. Media and Chemicals ......................................................................................................... 34  

B. E. coli Strains and Plasmids ............................................................................................... 34  

C. Preparation of Competent Cells ......................................................................................... 34  

D. Polymerase Chain Reaction (PCR) .................................................................................... 38  

E. Visualization of PCR Product ............................................................................................. 38  

F. Transformation into Competent Cells ................................................................................ 39  

G. Preparation of Plasmid DNA ............................................................................................. 39  

H. Sequencing ......................................................................................................................... 39  

I. IPTG induction ..................................................................................................................... 40  

J. Protein Purification ............................................................................................................. 40
K. Fluorescence Anisotropy assay ........................................... 41
L. Alkaline Phosphatase assay ............................................... 42
M. Western Blotting of AP assay samples .............................. 43
III. Results ........................................................................... 44
   A. Overview ....................................................................... 45
   B. Generation of chimera proteins ...................................... 45
   C. Demonstrating self-binding in the chimera proteins ......... 47
      D. Demonstrating functionality of chimera protein signal
         sequence ...................................................................... 56
IV. Discussion ......................................................................... 63
V. References .......................................................................... 75
I. Introduction
A. Protein Secretion Pathways

Bacterial cells contain no internal membrane-bound organelles or nucleus. The only membrane present is the exterior plasma membrane, found in all living cells, composed of a phospholipid bilayer that separates the internal cytoplasm from the external environment. Bacteria can be divided into two groups based on the structure of their cellular envelope, which is detected by the interaction of cell wall components with a crystal violet-iodine complex in a Gram stain. Gram-positive bacteria have a single cytoplasmic membrane surrounded by a rigid, thick, cell wall composed of many layers of peptidoglycan (a polymer of sugars and amino acids) interspersed with techoic acids and lipotechoic acids (Figure 1A). In contrast Gram-negative bacteria have both the cytoplasmic membrane and an additional outer membrane that contains porins to allow passive transport of small molecules inward, as well as outward facing phospholipids and lipopolysaccharides (Figure 1B). The region formed in between the two membranes is termed the periplasm and contains a thin peptidoglycan layer. Also present in the periplasm are numerous ions, amino acids, and a high concentration of proteins involved in a variety of cellular functions, including nutrient degradation, binding, and uptake.

Many proteins synthesized by ribosomes function outside of the cytoplasm, and thus need to be integrated into or exported across the plasma membrane. Proteins that function in the periplasm or beyond usually contain some type of N-terminal signal sequence that marks them for secretion, whereas proteins destined to lie in the
cytoplasmic membrane have specific recognizable elements in their hydrophobic transmembrane segments. In bacterial cells, different pathways exist to translocate the proteins into the correct cellular compartment, and are generally classified by their cellular machinery. There are at least five distinct pathways utilized by bacterial cells: the twin arginine translocation (TAT) system, the ATP binding cassette (ABC) exporters, the Type II pathway, the Type III pathway, and the general secretion (Sec) pathway. The majority of proteins are translocated via the genetically conserved Sec pathway (Oliver and Galan 2000).

The twin arginine secretion system is a non-essential pathway used for the secretion of redox enzymes across the cytoplasmic membrane (Figure 2). It is used in all three domains of life and is completely independent from the general secretory pathway. A specific motif found within the signal peptide of the preproteins lends itself to the system’s name, as the sequence S/T-R-R-X-F-L-K (X represents any amino acid) codes for a conserved arginine pair (Oliver and Galan 2000). The integral membrane proteins TatA, TatB, and TatC comprise the machinery involved, and are capable of exporting fully folded proteins (Berks et al. 2005).

ABC transporters secrete a variety of toxins, proteases, and specific peptides across one or two membranes (in the case of Gram-negative bacteria) in a single step (Figure 2). The machinery generally consists of an ATP binding cassette protein with two membrane embedded hydrophobic domains and two ATP binding domains, along with a few accessory proteins (Oliver and Galan 2000). Adenosine triphosphate (ATP) binding and hydrolysis supplies the force behind the translocation across membranes. The substrates exported have a C-terminal secretion sequence, in
contrast to most other secretion pathways, indicating that proteins are transported post-translationally.

In comparison to the ABC secretion pathway and the TAT system, the main terminal branch (MTB) of the general secretory pathway, known as the Type II pathway, is only found in a small number of plant and animal pathogens. Its role is to

![Translocation pathways in bacteria](Wickner and Schekman 2005)

Figure 2. Translocation pathways in bacteria (Wickner and Schekman 2005)

secrete a small number of exoproteins from the periplasm in Gram-negative bacteria across the outer membrane (Figure 2). The exporters generally consist of a large number of proteins (sometimes over a dozen) that are capable of recognizing a diverse set of exoproteins, suggesting the recognition of common motifs on the
substrates to be exported. In addition, the protein must be correctly folded in order to be exported (Oliver and Galan 2000).

The type III secretion pathway is used by pathogenic bacteria to translocate bacterial virulence effector proteins into host cells (Figure 2). The substrate proteins have no signal sequence, but instead seem to require some sort of activating factor likely derived from bacterial interaction with host cells. The secretion machinery is comprised of at least 17 different proteins, including several membrane proteins and an ATPase. The complex as a whole takes the shape of a needle and is capable of spanning both the inner and outer membranes of Gram-negative bacteria, resulting in a single translocation event. There is little homology between proteins using type III pathways between species of bacteria, implying that the system is adapted on the species level to suit the specific needs of the pathogen in interrupting the host’s response (Oliver and Galan 2000).

The majority of protein translocation in bacteria is carried out by the highly conserved secretory (Sec) system (Figure 2). Proteins destined to lie in the membrane are generally translocated in a co-translational manner, in which the signal recognition particle (SRP) binds the hydrophobic transmembrane segments (TMS) of a nascent polypeptide chain emerging from the ribosome and targets the complex to the heterotrimeric SecYEG channel in the cytoplasmic membrane. Translation at the ribosome provides the driving force to push the preprotein into the periplasm. Proteins that function in the periplasm or beyond are recognized by the ATPase SecA via an N-terminal signal peptide and are generally translocated post-translationally (du Plessis et al. 2011). SecA is capable of binding the preprotein and targeting it to
the SecYEG channel, at which point ATP binding and hydrolysis pushes the preprotein across the membrane. In addition, the molecular chaperone SecB is capable of binding the newly formed protein and targeting it to the already SecYEG bound SecA. Rounding out the Sec translocase are the nonessential SecD, SecF, and YajC proteins that improve the efficiency of protein translocation by an as yet poorly defined process.

B. SRP mediated export

Proteins destined to lie in the cytoplasmic membrane of bacterial cells are usually translocated via an interaction with a signal recognition particle (SRP) (Figure 3). Bacterial SRP consists of Ffh (fifty four homologue), a 48 kDa GTPase, and a 4.5S RNA molecule that increases the affinity of Ffh for preproteins (Driessen et al. 2001). The SRP complex recognizes and binds the hydrophobic transmembrane segments of a polypeptide emerging from the ribosome, with higher affinities for more hydrophobic segments (Luirink et al. 1997). The ribosome nascent chain complex is then targeted to FtsY, the membrane-bound SRP receptor. Binding of the SRP receptor to the complex increases both proteins affinity for GTP, and rearranges the position of an SRP domain, allowing the binding of the ribosome to the SecYEG translocon and the subsequent release of SRP after GTP hydrolysis (Beckmann et al. 2006). Translation at the ribosome then drives the protein through the SecYEG channel.

Both Ffh and FtsY have an NG domain strongly associated with the binding and hydrolysis of GTP. The third domain of FtsY is a central N domain, whereas in
Ffh is a methionine-rich M-domain with a deep hydrophobic groove, forming the signal sequence binding pocket. If the segments of the preprotein emerging from the ribosome are not sufficiently hydrophobic a ribosome associated trigger factor will inhibit Ffh from binding, and the protein will be translated to completion. At that stage, translocation would occur via SecB and/or SecA targeting (Driessen et al. 2001).

Figure 3. SRP-mediated targeting and translocation (Driessen et al. 2001)
1) SRP (purple) recognizes the signal sequence (black) of the preprotein being translated at the ribosome (yellow). 2) The ribosome nascent chain –SRP complex is targeted to the membrane bound receptor FtsY. 3) GTP binding to FtsY releases the ribosome complex to the SecYEG channel, where translation at the ribosome drives translocation. 4) GTP hydrolysis dissociates SRP from FtsY.
C. Signal sequences

Preproteins utilizing the general secretory system possess an N-terminal signal peptide of about 18-30 amino acids in length that serves as the targeting element for the Sec proteins. The sequence also delays the rate of preprotein folding, a crucial aspect considering protein translocation using the Sec system can only occur with unfolded proteins (Oliver and Galan 2000). Signal peptides share little sequence homology, yet all have a positively-charged N-terminus of about 1-5 residues, a hydrophobic core of 10-15 residues, and a short polar C-terminal region that contains the signal peptidase recognition site at residues -3 and -1 (Figure 4). Cleavage of the signal peptide occurs after translocation, and while not a necessary step for translocation, signal peptide retention will keep the protein anchored to the membrane if it is not removed by the signal peptidase (Rusch and Kendall 2007). Both the hydrophobic core and the charged amino terminus are important for successful translocation as alterations in either results in translocation defects (Inouye et al. 1982; Hofnung et al. 1980), however the impact of the hydrophobic core is greater as mutational increases in the hydrophobicity of the core can compensate for a lack of charge at the amino terminus (Hikita and Mizushima 1992).
D. SecB preprotein targeting

Many preproteins to be exported via the Sec system are SecB dependent. SecB functions as a molecular chaperone by binding the preprotein before it folds into its native state, usually post-translation, and directing it to the translocation apparatus present at the cytoplasmic membrane (Figure 5). SecB is a 17kD protein that functions as a tetramer (more specifically a dimer of dimers) with two 70Å long channels on each tetramer capable of binding a variety of preprotein substrates (Driessen et al. 2001). SecB does not recognize the N-terminal signal sequence of a preprotein, but rather multiple different binding sites throughout the protein sequence, about 20-30 residues apart, each of which is approximately 9 residues long and enriched in aromatic and basic residues (Knoblauch et al. 1999). It has been proposed that two different sub-sites exist on SecB for binding the preprotein substrates: one that binds very short, flexible regions, and another that binds hydrophobic regions. The latter sites are only bound after most of the flexible region binding sites are filled, thus helping to prevent aggregation of SecB. The nature of these binding sites help SecB bind secretory proteins, as the flexible regions and exposed interior hydrophobic regions are more characteristic of slowly folding proteins as opposed to rapidly folding proteins. In the same manner, the binding sites serve to prevent SecB from binding non-secretory proteins (Randall et al. 1998).

After binding the preprotein, the SecB complex is targeted to the SecYEG-bound motor protein SecA, at which point SecA binds the complex via interactions both with SecB and the preprotein (Hartl et al. 1990). A number of SecB
residues have been indicated as important for SecA binding, and have been shown to lie in an acidic, solvent-exposed surface present on both sides of the tetramer capable of electrostatically interacting with the C-terminus of SecA. The C-terminus of SecB has also been implicated in SecA binding. After binding, SecB is dependent on SecA’s ATP hydrolysis activity to transfer the preprotein and be released back into the cytoplasm (Driessen et al. 2001).
E. The SecYEG protein conducting channel

All preproteins translocated using the Sec system pass through the channel formed by the SecYEG protein complex, whether the translocation is co-translational via SRP-mediated targeting or post-translational via SecB and SecA. SecY is an essential 48kD protein comprised of 10 highly hydrophobic transmembrane alpha-helices that forms the largest component of the translocon channel. The transmembrane segments (TMS) are split into two halves, TMS 1-5 and TMS 6-10, and are hinged together at the cytoplasmic loop between TM5 and TM6 (Figure 6). The center of the hourglass-shaped channel features a narrow constriction a few angstroms wide made up of six hydrophobic isoleucine residues with their side chains pointing towards the center of the channel, which has been proposed to function as a seal around the translocating protein (Du Plessis et al. 2011). On the periplasmic side of the membrane, a small alpha helix TM2a folds back into the channel to block translocation, and is thus known as the “plug” domain. Signal sequence interaction with TM2 and TM7 helps to displace the plug from the channel and allow preprotein translocation to occur. Another important structural feature of SecY is the lateral gate, formed at the interface of TMS 2-3 and TMS 7-8. The opening of this interface is important in allowing signal sequences and hydrophobic TMS into the lipid bilayer, as well as playing a role in expanding the channel for preprotein translocation across the membrane (Kusters and Driessen 2011).
Figure 6. SecY crystal structure from *Methanococcus jannaschii* (van den Berg, 2004)  

A, View of the top of the Sec Y channel from the cytosol with SecE (Y) colored purple and SecG (B) colored magenta,  

B, Top view showing the two halves (red and blue) flanked by SecE and SecG (both gray),  

C, View from the back,  

D, View from the cytosol with the solid lines connecting the TMSs from N to C terminus; the dotted line is the axis of symmetry.
E. coli SecE is a 14kD protein also essential for protein translocation. The protein encompasses three TMSs (Figure 7), however only the third is necessary for cell viability. Without interacting with SecE and forming a complex, SecY is unstable and will be degraded. SecE-SecY interactions include the SecE cytosolic loop 2 (C2 loop) with the SecY C4 loop, the C-terminal TMS3 of Sec with TMS2, 7, and 10 of SecY, and the SecE periplasmic loop 1 (P2) with the SecY P1 loop (Veenendaal et al. 2004).

In contrast to SecY and SecE, E. coli SecG is not essential for protein translocation. The 12kD protein is comprised of two TMSs (Figure 7) and associates loosely with the SecYE complex. The role of SecG is to stimulate SecYE mediated preprotein translocation and SecA ATPase activity, the latter of which is hypothesized to be brought about by topological inversions of the TMSs of SecG (Veenendaal et al. 2004)

One point of debate concerning the translocon is the functional oligomeric state of the SecYEG channel, as there have been studies supporting both a monomer and a dimer. A study by the Rapoport group showed that a SecY channel deficient in translocation can be rescued by covalently linking another SecY molecule to it. It was thus hypothesized that one SecY molecule serves as a binding site for SecA, while the other SecY molecule forms the translocation pore (Osborne and Rapoport, 2007). Further evidence supporting a dimer is found by studies showing SecA only binds to SecYEG dimers, and that covalently linked SecYEG dimers are active in translocation (du Plessis et al 2011). However, support for the monomer can be found in a Thermotoga maritima SecA-SecY complex crystal structure, proposed to
be sufficient for translocation (Zimmer et al. 2008) and the crystal structure of the *Methanococcus jannaschii* SecYEG channel, also proposed to be sufficient for translocation (van den Berg et al. 2004). It is likely that different oligomeric states of SecYEG exist *in vivo*, however the functional state of the translocon, monomer or dimer, is not fully elucidated.

A question also exists as to the orientation of the SecYEG dimer. Two models have been proposed: the ‘front-to-front orientation, or the ‘back-to-back’ orientation. The former consists of the lateral gates facing each other in a manner that would enable the two channels to fuse and form one consolidated pore. In the ‘back-to-back’ orientation the two SecYEG complexes are connected via one of the TMSs of SecE (du Plessis et al. 2011).
F. SecDFyajC

The heterotrimeric membrane complex composed of SecD, SecF, and YajC associates with the SecYEG channel. SecD and SecF are both polytopic membrane proteins with large periplasmic domains. In contrast, the membrane protein YajC has a large cytosolic domain and consists of a single TMS (Driessen et al. 2008). SecD and SecF are not essential for protein translocation; however, their inactivation leads to cold sensitivity, severely inhibiting growth, as well as causing severe pleiotropic protein secretion defects (du Plessis et al. 2011). It has been proposed that the complex functions after the primary protein translocation reaction, possibly in clearing the pore of processed signal sequences, as reconstitution studies show the complex is not needed for the translocation itself. It has also been posited that the complex indirectly functions in the catalytic cycling of SecA during translocation, helping to drive the polypeptide chain through the SecYEG channel. Finally, another hypothesis is that the complex assists in forming an oligomeric SecYEG complex (Driessen et al. 2008). The function of YajC, which associates with the SecDF complex but is not needed for functionality, is not well understood.

G. SecA

G.1 Structure

SecA is a 102kD protein that is the final component of the general secretory system. The protein uses ATP hydrolysis in order to facilitate the transfer of preproteins across the membrane (Lill et al. 1989), and has been shown to be essential for protein translocation (Cabelli et al. 1988). SecA interacts with multiple different
substrates, including signal sequences, anionic phospholipids, SecB, and SecYEG. Numerous X-ray and NMR structures of SecA from various organisms have elucidated the structure and functional domains of SecA; these include two *Bacillus subtilis* crystal structures (Figure 8) (Hunt et al. 2002, Osborne et al. 2004), both a crystal structure and an NMR structure from *Escherichia coli* (Papanikolau et al. 2007, Gelis et al. 2007), a *Mycobacterium tuberculosis* crystal structure (Sharma et al. 2003), a *Thermus thermophilus* crystal structure (Vassylyev et al. 2006), and a *Thermotoga maritima* SecY-SecA crystal structure (Zimmer et al. 2008). The nucleotide-binding domain (NBD) consists of two nucleotide-binding folds, NBF1 and NBF2, that are essential for ATP binding and hydrolysis. The preprotein cross-linking domain (PPXD) is connected to NBF1 by two β-strands, and has been shown to be involved in signal peptide binding. The helical scaffold domain (HSD) consists of a long central helix that forms the backbone of SecA, and a two-helix finger region thought to be involved in driving protein translocation at the SecYEG channel. The helical wing domain (HWD) is an extension of the HSD with no known function. The carboxy-terminal linker domain (CTL) serves as the binding site for both acidic phospholipids and the molecular chaperone SecB (Figure 8).
Figure 8. **A**, crystal structure of *Bacillus subtilis* SecA (Hunt et al. 2002). The NBFs are in light blue and dark blue, with ADP (purple) bound in between. The PPXD is in orange and yellow, the HSD in green, and the HWD in cyan. The CTL is in red. This structure is of the ‘closed’ conformation. **B**, crystal structure of *Bacillus subtilis* SecA (Osborne et al. 2004). This structure is in the ‘open’ conformation. Notice the increased distance between the PPXD and HWD compared to the Hunt structure.
The motor region of SecA belongs to the DEAD-box family of proteins, as the folds involved in ATP binding and hydrolysis contain the conserved DNA/RNA helicase motifs present in this family. The Walker A and Walker B motifs characteristic of this superfamily of proteins, the latter of which incorporates the amino acid sequence D-E-A-D (thus lending its name to the family) are found within the binding folds in NBF1 and NBF2 (Kusters and Driessen 2011). Both regions are essential for preprotein translocation, however NBF1 binds ATP with a much higher affinity than NBF2 (Mitchell and Oliver 1993). Intramolecular-regulator-of-ATP-hydrolysis (IRA 1) is a helix-loop-helix structure (two-helix finger) that contacts both NBF2 and the PPXD, and has been shown to inhibit the ATPase activity of SecA, thus preventing ATP hydrolysis when SecA is not bound to and translocating a preprotein (Karamanou et al. 1999). A second domain equivalent to NBF2, IRA2, is a highly flexible region that via contacting NBF1 can activate ATP hydrolysis and promote nucleotide turnover (Sianidis et al. 2001). Preprotein binding at the PPXD has the ability to stimulate ATP hydrolysis via a conserved salt bridge known as Gate 1; this mechanism is only active after SecA is bound to SecYEG, thus leading to the coupling of ATP hydrolysis and translocation through the SecYEG channel (Du Plessis et al. 2011).

The various crystal structures have captured SecA in three different states: the ‘open’ conformation (Figure 8b), the ‘closed’ conformation (Figure 8a), and the SecYEG bound state (Figure 9). Compared to the closed form presented in the Hunt structure (Hunt et al. 2002), an ‘open’ conformation (Osborne et al. 2004) features a 60° rigid body rotation of the PPXD away from the tight packing with the HWD, thus
creating a large groove proposed to be a peptide binding site. In the SecYEG bound state, the PPXD has further rotated towards NBF2, thus closing the groove, also known as the clamp (Zimmer et al. 2008). Based on further work by the same group with SecA structures bound to ADP and bound to peptide (which had no similarity to functional signal sequences), it has been proposed that a translocation substrate enters the open clamp, and is captured by rotation of the PPXD (Zimmer and Rapoport 2009). A disulfide bridge cross-linking study demonstrated that this clamp does indeed come into contact with the mature region of the preprotein substrate (Bauer et al. 2009).

**G.2. Oligomeric State**

While SecA is mostly dimeric in the cytosol, its functional oligomeric state is a matter of debate. In addition, the arrangement of the dimer is also under discussion, as multiple structures with different dimer interfaces have been solved. SecA is positioned as an anti-parallel dimer in most of the structures, with the exception being one parallel dimer structure from *T. thermophilus*. The multiple interfaces solved could represent multiple dimeric states corresponding to the multiple functions of SecA, or could be the result of non-physiological dimer formation (Kusters and Driessen 2011).

The monomer-dimer equilibrium of SecA is affected by the binding of ligands, salt concentration, temperature, and the presence of detergents. Thus certain studies featuring high levels of detergent, for example, may compromise native conditions (Kusters and Driessen 2011). Support can be found for both functional
states. A study using a monomer-biased mutant showed that it is capable of translocation and sustaining 85% of wild type growth (Or et al. 2004). In addition, the structure of the SecYEG-bound SecA features a monomer bound potentially in a translocation state (Zimmer et al. 2008); however a high level of detergent used could have promoted artificial monomerization of SecA. Support for the dimer can be found from a crosslinking study linking two SecA protomers into a dimer, which was functional in translocation assays (de Keyzer et al. 2005), as well as from a study featuring a monomer-biased mutant that was inactive in protein translocation (Jilaveanu et al. 2005). The functional oligomeric state of the SecA protomer is still under debate.

G.3. SecA interactions with SecB and SecYEG

As mentioned before, the molecular chaperone SecB is capable of binding SecA with high affinity in order to deliver a preprotein to the translocation machinery (Hartl et al. 1990). SecA recognizes the chaperone SecB via a 22 amino acid region in its CTL domain, an interaction which is dependent on a zinc-finger motif found within the binding site (Fekkes et al. 1999). It is important to note that not all preproteins are SecB dependent; some are directly recognized by SecA and shuttled to the SecYEG channel.

An *in vivo* photo-crosslinking study demonstrated regions on the SecYEG channel important for binding SecA. It was found that loops C4-C6 of SecY all interact with SecA, while a C5 loop interaction is with the N-terminal region of SecA.
and likely activates SecA ATPase activity (Mori and Ito 2006). A study using the same methodology was carried out to locate the residues on SecA important for the binding interaction with SecYEG. It was found that the binding site on SecA is quite diverse, with two different binding sites corresponding to the binding of two different SecY molecules. One site incorporates the PPXD and HSD, while the other incorporates the two NBFs (Das et al. 2011). A crystal structure of the SecA-SecY complex with SecA bound to an ATP analogue found extensive interactions between the PPXD of SecA and loops 6-7 and 8-9 of SecYEG (Figure 9). In addition, the two-helix finger subdomain of the HSD in SecA was found to be inserted into the SecYEG channel, resting in a cleft between the 6-7 loop and C-terminal tail of SecY (Zimmer et al. 2008). The crosslinking study by Das et al. supports the presence of the two-helix finger interacting with the SecYEG channel. Yet another study used disulfide bridge-crosslinking to show that the translocating polypeptide passes from the two-helix finger into the SecYEG pore, utilizing mainly an important tyrosine residue in the loop connecting the two helices (Erlandson et al. 2008).
G.4. Mechanisms of SecA-mediated translocation

In addition to ATP hydrolysis, the proton motive force (PMF) is able to promote protein translocation. While not essential, the insertion and orientation of the signal sequence in the SecYEG channel and SecA de-insertion from the membrane are both amplified by the PMF. In addition, protein translocation can be driven by the PMF alone when the polypeptide is already inserted into the SecYEG
channel, and it prevents polypeptides from sliding back out of the channel (Kusters and Driessen 2011).

Multiple mechanisms have been proposed as to how SecA translocates preproteins. The Brownian ratchet hypothesis posits that SecA simply mediates channel opening, at which point the preprotein diffuses through the channel (Kusters and Driessen 2011). While there is support for this model, it cannot account for the observation that translocation occurs in steps. For example, a study by Economou and Wickner showed that SecA alternates between an inserted and deinserted state, with ATP hydrolysis driving a conformational change that inserts a 30kD fragment of SecA into the channel. Translocation of a 20-amino-acid portion of the preprotein was shown to be concomitant with the insertion cycles (Economou and Wickner 1994). This data, along with more recent data indicating that the two-helix finger interacts with the preprotein in the SecYEG channel (Zimmer et al. 2008, Erlandson et al. 2008), support the power-stroke motor hypothesis, in which ATP hydrolysis is coupled to the mechanical pushing of the polypeptide through the channel. The small conformational change of the two-helix finger, thought not to be large enough to account for the observed translocation of 2-2.5kD of preprotein, led to a refinement of the power-stroke model. In this modified situation, the translocation machinery would move the polypeptide via two different substrate binding sites (one on SecA and one on SecYEG, or dimeric SecA), in which when one binding site is bound, the other dissociates and grabs a segment further up the polypeptide chain, thus moving the chain in an “inchworm” fashion (Kusters and Driessen 2011).
G.5. Signal sequence binding

SecA has been shown through numerous studies to recognize preproteins directly, specifically via the preprotein’s N-terminal signal sequence. Importantly, SecA is capable of recognizing a multitude of different signal peptides that share little sequence homology. Early studies demonstrated this interaction by showing that translocation ATPase activity is dependent on interaction with the preprotein (Cunningham and Wickner 1989), and that synthetic signal peptides are capable of stimulating translocation ATPase activity in the absence of preprotein (Miller et al. 1998). A crosslinking study showed that the N-terminal positively-charged region of signal peptides is important for SecA binding, with the interaction being stronger the more positive charge present (Akita et al. 1990). Mutant preproteins lacking a positively-charged N-terminus but with sufficiently long hydrophobic cores were also shown to be capable of both in vitro and in vivo translocation, and were also shown to crosslink to SecA. As with positive charge at the N-terminus, the more hydrophobic the core is the stronger the interaction with SecA (Mori et al. 1997).

The next step in solving the SecA signal peptide interaction mystery was to elucidate the binding site on SecA. An initial study undertaken to solve this question cross-linked a presecretory protein with various overlapping N-terminal and C-terminal fragments of SecA (i.e. making use of SecA dimer structure), and resulted in the designation of a preprotein-binding site between residues 267-340 of SecA (Kimura et al. 1991). This region falls into the more recently determined PPXD, which is comprised of residues 221-377. A later study using deletion constructs of certain regions within the PPXD showed that the signal peptide binds between
residues 219-244 (Papanikou et al. 2005). A slightly different binding region was identified in a study by Kendall’s group, which implicated residues 269-322. This study utilized a signal peptide including the photo reactive label benzoylphenylalanine (Bpa) for UV crosslinking, and a C-terminus cysteine site for biotin linkage to serve as a reporter. A series of mutant SecAs with a Factor Xa cleavage site were engineered, cut after photo crosslinking, and the segment with the peptide cross-linked was detected via biotin (Musial-Siwek et al. 2007). As mentioned before, the two-helix finger was shown to play an important role in preprotein translocation, and the disulfide bridge cross-linking study demonstrated the importance of Tyr-794 in preprotein recognition (Erlandson et al. 2008). A nuclear magnetic resonance (NMR) structure of the artificially N-terminally-extended signal peptide KRR-LamB bound to E. coli SecA determined a binding groove along the interface of the PPXD and the HWD, with the signal peptide oriented perpendicular to the two-helix finger of the HSD (Gelis et al. 2007) (Figure 10A). A recent study utilizing Förster resonance energy transfer (FRET) identified the binding domain of alkaline phosphatase signal peptide on SecA to incorporate a region including portions of the PPXD, HSD, and NBF1 (Auclair et al 2010) (Figure 10A). This FRET study implicated the two-helix finger region as an important binding site much more so than the NMR structure data, although they do agree in approximate location. The FRET study also proposed a parallel orientation of the bound signal peptide in order to facilitate its entry into the SecY channel. The binding site proposed in the FRET study is also supported by the work by the Rapoport group demonstrating that the two-helix finger region of SecA interacts with the preprotein
during translocation (Zimmer et al. 2008) (Figure 10B). Finally, biochemical characterization of various SecA mutant proteins revealed residues 226, 310, 789, 806, and 808 as important for signal peptide binding. The first two residues listed are found in the PPXD, the third is in the two-helix finger subdomain, and the latter two
are in the second α-helix of the HSD. The study also noted that a similar set of residues are used to recognize two separate signal peptides (Grady et al. 2012).

**H. Fluorescence anisotropy**

Certain molecules are capable of absorbing light energy, causing a transition from one vibrational energy level to a higher vibrational energy level (ground to excited state). The molecule must then lose the extra energy it has acquired in order to return to the ground state. Some molecules capable of absorbing light have a limited number of vibrational energy levels, and are unable to lose all of the absorbed energy in the form of heat. Instead, a portion of the absorbed energy is emitted as light in a process termed fluorescence. The energy not emitted is lost in smaller vibrational transitions, or is due to fluorescence quenching, a diminishing of the amount of light energy emitted (Figure 11). Internal quenching usually involves some sort of structural rearrangement of the molecule, whereas external quenching involves interaction of the fluorophore with another excited molecule. The length of the excited state of a fluorophore is usually very short, on the timescale of 0.5-8 ns (Sheehan 2009).
Figure 11. Jablonski diagram of absorption and emission of light. Fluorescence occurs as a result of the return to the ground state ($S_0$) from the excited state ($S_1$) after small vibrational translations drop the energy level from $S_2$ to $S_1$.

Anisotropy ($r$) refers to the polarization of the emission of a fluorophore. When a population of randomly oriented fluorophores is excited by polarized light, a certain subset of them that are oriented along the electric vector of the light are excited. However, the emission of the absorbed light is depolarized due to the rotational diffusion of the fluorophore in the excited state. This rotational rate is dependent on the viscosity of the solution and the shape and size of the fluorophore molecule. Anisotropy is defined according to the equation:

$$r = \frac{I_\parallel - I_\perp}{I_\parallel + 2I_\perp}$$

$I_\parallel$ is the detected polarized parallel emission, whereas $I_\perp$ is the detected polarized perpendicular emission when the sample is excited with vertical polarized light. In
such a case the electric vector is parallel to the vertical axis. The anisotropy value is independent of intensity, as the difference in emission (numerator) is divided by the total intensity (denominator) (Lakowicz 2006).

The use of anisotropy in binding assays is based on the different rates of rotational diffusion among molecules of different size. A small fluorophore molecule by itself will rotate very fast during its excited lifetime, resulting in a high depolarization. If the fluorophore molecule is bound to a larger molecule, for example a protein, the complex will rotate much slower and the depolarization will be smaller, leading to different anisotropy values (Figure 12).

![Figure 12. Schematic of fluorescence anisotropy binding assay (Guthrie et al. 2006).](image-url)
I. Rationale for work

Despite all of the work performed over the previous two decades on elucidating the signal peptide-binding site of SecA, the site and orientation of the bound signal peptide have not been conclusively determined due to conflicting studies and shortfalls of the various techniques used. The usefulness of many of the biochemical and genetic techniques is limited as they identify only a small number of residues important for binding, and it is difficult to tell if the residues directly or indirectly contribute to signal peptide binding. In addition, they generate an essentially linear map of a 3D binding domain by including and excluding certain regions. The NMR structure (Gelis et al. 2007) generated a high resolution structure, however it used a relatively high concentration of signal peptide not found under physiological conditions. The FRET study was able to use a lower concentration of signal peptide, but generated a broader, lower resolution binding site (Auclair et al. 2010). In addition, the site proposed by the Auclair study differs from the Gelis study in the proposed orientation of the signal peptide and the incorporation of more of the two-helix finger of SecA, a proposal consistent with the SecA-SecYEG co-crystal structure (Zimmer et al. 2008).

The generation of an X-ray crystal structure of SecA bound to a signal peptide would greatly help in elucidating the signal peptide binding site of SecA. However, such a structure has been difficult to obtain due to the abnormally high concentration of signal peptide required; the extremely hydrophobic signal peptide aggregates and precipitates.
In order to circumvent this problem, which prevents X-ray crystallization of a SecA-signal peptide complex, this study used PCR-directed mutagenesis/gene fusion techniques to attach the signal peptide KRR-LamB onto the C-terminus of *E. coli* SecA, with the ultimate goal of obtaining an X-ray crystal structure of the chimera protein with the attached signal peptide self-bound to SecA. This method does not need any exogenous signal peptide added. Such an approach has been previously utilized to obtain a crystal structure of a signal peptide-bound SRP chimera protein (Janda et al. 2010), as well as a crystal structure of an MHC Class II molecule-signal peptide chimera protein (Fremont et al. 1996).

Fluorescence anisotropy assays were performed to analyze the extent of self-binding of the chimeras, and an alkaline phosphatase assay was performed to analyze the *in vivo* functionality of the covalently attached signal sequence. The KRR-LamB signal peptide was used in this study as it was shown to bind SecA with an increased affinity compared to other signal peptides (Gelis et al. 2007). Corresponding work is also being performed with the PhoA signal peptide by another member of the lab, so comparisons of the binding site of different signal peptides on SecA can be made.
II. Materials and Methods
A. Media and Chemicals

LB broth supplemented with ampicillin (100ug/ml) and/or chloramphenicol (50ug/ml) was used for bacterial growth. LB broth and agar were purchased from EMD chemicals and Difco, respectively. IPTG, PNPP, and other chemicals used were obtained from Sigma or a comparable supplier.

B. E. coli Strains and Plasmids

Plasmid mutations were made using the QuikChange (Stratagene) mutagenesis kit. Oligonucleotide primers (Table 1) were designed on the Agilent website (www.genomics.agilent.com) and ordered from Integrated DNA Technologies. Strains and plasmids used in this study are listed in Table 2. All plasmids containing the SecA-KRR-LamB chimeras utilized the pT7 SecA-His plasmid as the starting vector.

C. Preparation of Competent Cells

An overnight culture of cells was subcultured at a 1:50 dilution into LB supplemented with the appropriate antibiotics. The cells were grown to an OD$_{600}$ of 0.25-0.3 and put on ice for 15 minutes. Cells were then spun down at 10,000x g for 5 minutes at 4°C. Cell pellets were re-suspended in 0.5 volumes of cold 0.05 M CaCl$_2$ and put on ice for 30 minutes. The cells were spun again as before, and suspended in 0.1 volumes of 0.05 M CaCl$_2$ + 15% glycerol. The suspension was aliquoted and the cells were quick-frozen on dry ice and stored at -80°C.
### Table 1. Oligonucleotide primers

A list of the sequences of the primers used for construction of the plasmids used in the study. Only the forward primer is displayed. The underlined portion represents the insertion or change in DNA sequence. A slash represents a deletion.

<table>
<thead>
<tr>
<th>Sequence (5’-3’)</th>
<th>Size (nt)</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>CATGGGCCGCTGCAAATGATGATTACTCTGCAC</td>
<td>60</td>
<td>Insertion of KRR-LamB residues 1-10 onto <em>E. coli</em> SecA</td>
</tr>
<tr>
<td>CGCAACGTCTCAAACCTCCTGCGGCTTGTGCCTCGAGCGGGCTCGAGCACACACC</td>
<td>60</td>
<td>Insertion of KRR-LamB residues 11-20 onto <em>E. coli</em> SecA</td>
</tr>
<tr>
<td>GTCGCAGCGGCGCTATTCTCTGCTAGAGCACACCTGCTGCGACGCACACC</td>
<td>48</td>
<td>Insertion of KRR-LamB residues 21-28 onto <em>E. coli</em> SecA</td>
</tr>
<tr>
<td>GTTCAGGTACGTATGCT/ATGATGATTACTCTGCGC</td>
<td>36</td>
<td><em>E. coli</em> SecA CTL deletion</td>
</tr>
<tr>
<td>CCATGGGCCGCTGCAAATGATGATTACCTTGTGCGAAGCGGTGTGAGGCACACC</td>
<td>57</td>
<td>Insertion of KRR-LamB residues 1-10 into SecA-PhoA mature</td>
</tr>
<tr>
<td>CTCTTGCAGCGAAGGTTCTGCAGTGTGCGAGGCGACACC</td>
<td>60</td>
<td>Insertion of KRR-LamB residues 11-20 into SecA-PhoA mature</td>
</tr>
<tr>
<td>CGTCGCAGCGGCGCTAATGCTGCTACAAGCACCCAGAGAAGGGCGACACC</td>
<td>50</td>
<td>Insertion of KRR-LamB residues 21-28 into SecA-phoA mature</td>
</tr>
<tr>
<td>TCGTGGCGCCGTACTGAAATGATTCCTGCTGCTACGTGCAACGTCGATACCCAC</td>
<td>58</td>
<td>Insertion of KRR-LamB residues 1-10 onto <em>B. subtilis</em> SecA</td>
</tr>
<tr>
<td>TCTGCGCGAACGTCTAAACCTCCTGCTGCGAAGCGGCGACACC</td>
<td>60</td>
<td>Insertion of KRR-LamB residues 11-20 onto <em>B. subtilis</em> SecA</td>
</tr>
<tr>
<td>GTTCAGGTACGTATGCTGCTGCTAGGCAATGGCTGAGCACACC</td>
<td>48</td>
<td>Insertion of KRR-LamB residues 21-28 onto <em>B. subtilis</em> SecA</td>
</tr>
<tr>
<td>Sequence</td>
<td>Position</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>----------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>ATGAAAGCTGAGATTGAAAAACAAAT/AT</td>
<td>47</td>
<td>B. subtilis CTL deletion</td>
</tr>
<tr>
<td>GATGATTACTCTCTCGCGCAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTGGCGGTGCCCAGTCAGCGGGCGTA</td>
<td>28</td>
<td>Point mutation in hydrophobic core of KRR-LamB</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CATCAGGATGACGACTC</td>
<td>17</td>
<td>Sequencing <em>E. coli</em> SecA</td>
</tr>
<tr>
<td>CTGGCAGCGATGGACTA</td>
<td>17</td>
<td>Sequencing <em>E. coli</em> SecA</td>
</tr>
<tr>
<td>CGACTAACATGCGCGGGGC</td>
<td>18</td>
<td>Sequencing <em>B. subtilis</em> SecA</td>
</tr>
<tr>
<td>TCTAGTGGATCTTTATCAA</td>
<td>18</td>
<td>Sequencing <em>B. subtilis</em> SecA</td>
</tr>
</tbody>
</table>
Table 2. List of strains and plasmids (*E. coli* unless otherwise mentioned)

<table>
<thead>
<tr>
<th>Strain</th>
<th>Description</th>
<th>Source or Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH5α</td>
<td><em>endA1 hsdR17 supE44 thi-1 recA1 gyrA relA1 Δ(argF-lac)U169 deoR φ80dlac(Δ(lazZ)M15</em></td>
<td>Laboratory Stock</td>
</tr>
<tr>
<td>BL21.19</td>
<td><em>secA13(Am) supF(Ts) trp(Am) zch::Tn10 recA::CAT clpA::kan</em></td>
<td>Laboratory Stock</td>
</tr>
<tr>
<td>BL21.20</td>
<td><em>secA13(Am) supF(Ts) trp(Am) zch::Tn10 recA635::kan</em></td>
<td>Laboratory Stock</td>
</tr>
<tr>
<td>MC1000 phoR</td>
<td>MC1000 <em>phoR</em></td>
<td>Laboratory Stock</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
<th>Source or Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pT7SecA-His</td>
<td>Ampicillin resistant plasmid containing <em>E. coli</em> SecA gene</td>
<td>Oliver Lab</td>
</tr>
<tr>
<td>SecA901-KRR-LamB</td>
<td>Full length <em>E. coli</em> SecA-KRR-LamB chimera protein</td>
<td>This study</td>
</tr>
<tr>
<td>SecA834-KRR-LamB</td>
<td>Truncated <em>E. coli</em> SecA-KRR-LamB chimera</td>
<td>This study</td>
</tr>
<tr>
<td>Div841-KRR-LamB</td>
<td>Full length <em>B. subtilis</em> SecA-KRR-LamB chimera</td>
<td>This study</td>
</tr>
<tr>
<td>Div783-KRR-LamB</td>
<td>Truncated <em>B. subtilis</em> SecA-KRR-LamB chimera</td>
<td>This study</td>
</tr>
<tr>
<td>SecA901-KRR-LamB-phoA mature</td>
<td><em>E. coli</em> SecA –KRR-LamB-PhoA mature fusion protein</td>
<td>This study</td>
</tr>
<tr>
<td>SecA834-KRR-LamB-phoA mature</td>
<td>Truncated <em>E. coli</em> SecA – KRR-LamB-PhoA mature fusion protein</td>
<td>This study</td>
</tr>
<tr>
<td>SecA901-KRR-LamB (V15D) – phoA mature</td>
<td><em>E. coli</em> SecA –KRR-LamB-PhoA mature fusion protein mutant</td>
<td>This study</td>
</tr>
<tr>
<td>SecA834-KRR-LamB (V15D) – phoA mature</td>
<td>Truncated <em>E. coli</em> SecA –KRR-LamB-PhoA mature fusion protein mutant</td>
<td>This study</td>
</tr>
</tbody>
</table>
D. Polymerase Chain Reaction (PCR)

PCR reactions were comprised of 1.25 μl of each primer (1 mg/ml), 1 μl of the DNA template (80 ng/μl), 17 μl of Master Mix (Stratagene), and 30 μl of water. Samples were prepared on ice, mixed thoroughly, and the tubes were placed in a thermal cycler. Following a 30 second heat activation at 95°C, samples were run for 18 cycles consisting of 95°C for 30 seconds, 60°C (or five degrees below T_M of primers) for 1 minute, and 68°C for 12 minutes, followed by perpetual incubation at 4°C.

E. Visualization of PCR Product

0.5 μl of the enzyme Dpn1 was added to each PCR product. Samples were mixed and incubated at 37°C for 60 minutes to destroy methylated parental plasmid. Samples were then stored on ice for immediate usage, or at -20°C for storage. Agarose gels (0.7%) were prepared by dissolving 0.7g agarose into 100 ml 1X TAE buffer (40mM Tris-acetate, 1mM EDTA). Next 5 μl of ethidium bromide (10 mg/ml) was added to the dissolved agarose while mixing, and the solution was poured into a gel casting device. The gel was allowed to solidify for 45 minutes before samples (10 μl Dpn1 digested DNA, 6.7 μl water, 3.3 μl 6x DNA loading dye) were added. The samples were run at 100V for approximately 70 minutes before visualization using UV light.
F. Transformation into Competent Cells

Ten microliters of Dpn-1 treated plasmid were added to 100 μl competent cells, followed by gentle mixing. Samples were incubated on ice for 30 minutes, followed by a two minute heat shock at 37°C. The samples were then placed on ice for a further two minutes. Cells were then diluted into 900 μl of LB media and grown at 37°C for one hour. Following the incubation, the cells were spun down at 14,000 x g for five minutes before being re-suspended in 200 μl LB media. 100 μl of cells were then plated on LB plates supplemented with the appropriate antibiotic and incubated at 37°C overnight. Individual colonies were then selected and re-streaked onto a fresh plate. This strain purification was then repeated once more.

G. Preparation of Plasmid DNA

Ten milliliter overnight cultures grown at 37°C were spun down at 10,000 rpm for five minutes. The plasmid DNA was prepared using the Wizard Preps Kit (Promega), following the manufacturer’s protocol and using the solutions provided. Concentrations of plasmid DNA were obtained by measuring the OD<sub>260</sub> and OD<sub>280</sub> using a Nanodrop. The DNA was then stored at -20°C.

H. Sequencing

Successful plasmid mutagenesis was verified via DNA sequence analysis performed by the University of Pennsylvania Department of Genetics Sequencing Facility. Samples were prepared with 6 μl of plasmid DNA diluted to 80 ng/μl and 3
μl of the appropriate sequencing primer at a concentration of 1.1 μM. Sequencing primers used are listed in Table 1.

I. IPTG induction

Overnight cultures were subcultured at a 1:50 dilution into 1 L of LB broth supplemented with the appropriate antibiotics. The cells were then grown to mid-log phase (OD600 0.4-0.6) at either 30°C or 37°C. 1M IPTG was then added to the culture to obtain a final concentration of 0.5 mM, and the cells were grown for one more hour. Cells were harvested by spinning down the cultures at 6,000 rpm for 15 minutes. The pellets were re-suspended in 20 ml TKM buffer (10 mM Tris-Cl pH 7.5, 50 mM KCl, 10 mM MgOAc, 1 mM PMSF) and either placed on ice to be used immediately, or stored at -80°C. One milliliter samples of the un-induced and induced cultures were collected, spun down, and re-suspended in 100 μl LB media. From these samples 15 μl were taken, mixed with 5 μl 4x SDS loading dye, boiled for 5 minutes, and run on an SDS-page gel at 150 volts for 1.5 hours. Induction was confirmed by visualization of the amount of SecA or chimera protein in the sample via a 20 minute Coomassie Blue stain.

J. Protein Purification

Previously induced cells suspended in TKM buffer were broken utilizing a French press at a pressure of 15,000 PSI. Unbroken cells were removed by spinning the samples at 10,000 rpm for 20 minutes at 4°C. The supernatant was then collected and spun at 62,000 rpm for 30 minutes at 4°C to pellet the cell membranes. The
cleared lysate supernatant containing the soluble SecA protein of interest was then harvested and added to an already prepared nickel column designed to bind His-tagged proteins. The column and associated reagents were obtained from a His-bind kit from Novagen; the purification was carried out according to the manufacturer’s protocol provided in the kit. Following purification, the eluate containing the protein (5 mL) was dialyzed for one hour in 2L TKE buffer (50 mM Tris-Cl pH 7.5, 50 mM KCl, 1 mM EDTA, 1 mM PMSF), then switched to a fresh preparation of 2L TKE buffer and dialyzed overnight. The purity of the protein was verified by visualization of a Coomassie Blue stained SDS-Page gel, containing the appropriate samples, run at 150V for 1.5 hours. The dialyzed sample was then concentrated by spinning at 5,000 rpm for 10-20 minutes in a centrifugal filter unit. Protein concentration was measured using the Bio-Rad protein assay kit (Bio-Rad). The protein was aliquoted and stored on ice in a 4°C cold room before being used for fluorescence anisotropy assays.

K. Fluorescence Anisotropy assay

Fluorescence measurements were performed on a FluoroMax-4 spectrofluorometer (Horiba Jovin Yvon) with a programmable water bath. The samples used in the experiment consisted of 1 μM IANBD-labeled alkaline phosphatase signal peptide, labeled with an artificially induced cysteine at the 22nd residue, in TKE buffer. For a description of the signal peptide labeling process (previously performed by other lab members) see Auclair et al. 2010. Purified SecA was added to the labeled signal peptide at a concentration range of 0 to 40 μM. The
samples were placed in quartz cuvettes with a 3mm path length in the fluorometer, and incubated for 15 minutes before data collection. The spectral bandwidths of the excitation and emission slits were 4 and 6 nm, respectively. Samples were excited at 480 nm and measured at 550 nm. Data were fit using ORIGIN, version 8.0, with the following equation:

\[
y = A_0 + (A_i - A_0) \left( \frac{[SP] + K_d + [P] - \sqrt{([SP] + K_d + [P])^2 - 4[SP][P]}}{2[SP]} \right)
\]

[SP] is the total concentration of signal peptide, [P] is the total concentration of SecA, \( K_d \) is the equilibrium dissociation constant, \( A_0 \) is the anisotropy value of the signal peptide in the absence of SecA, and \( A_i \) is the anisotropy under saturating binding conditions. The final data used in the figures presented are the average results of at least three separate experiments, each experiment generating three separate readings of a particular data point.

L. Alkaline Phosphatase assay

Overnight cultures were subcultured at a 1:50 dilution into LB broth supplemented with the appropriate antibiotics. Cells (BL21.20) were grown at 40°C, in order to turn off expression of chromosomal SecA, until an \( \text{OD}_{600} \) of 0.4. 1M IPTG was then added to each culture (except MC1000 phoR) to obtain a final concentration of 0.5 mM, and the cells were grown for one more hour. One milliliter of cell culture from each sample was spun down, washed in 1M Tris-Cl pH 8.0, and re-suspended in 1M Tris-Cl pH 8.0. The \( \text{OD}_{600} \) of the cells was recorded, and the cells were diluted into 1M Tris-Cl pH 8.0. One milliliter samples of diluted cells
were preheated at 37°C for ten minutes, before the addition of 100 μL 0.4% PNPP and further incubation at 37°C. When samples were sufficiently yellow, reactions were halted by addition of 200 μL 1M KH₂PO₄ and the OD₄₂₀ and OD₅₅₀ of the samples were recorded for data analysis.

M. Western Blotting of AP assay samples

Samples from the alkaline phosphatase assay were immediately run on a 10% SDS-PAGE gel at 100V for approximately 3 hours. The protein samples were transferred to a nitrocellulose membrane via an electroblotter (BioRad) at 15V for 20 minutes. The membrane was blocked overnight with 10% nonfat dry milk in TBS. The membrane was washed with TBS before primary antibody incubation with either alkaline phosphatase or SecA antibody for one hour, followed by another rinse with TBS. Secondary antibody incubation with IgG was also for one hour, followed by another TBS rinse. Membranes were visualized using an ECL kit (Thermo scientific) and imaged using a Syngene Gelbox system.
III. Results
A. Overview

The location of the signal peptide binding site on SecA is still a matter of debate. Numerous studies have narrowed down the prospective binding site to a region including the PPXD and the HSD, however multiple orientations of the bound signal peptide have been proposed. More specifically, an NMR structure of SecA showed the bound signal peptide perpendicular to the two-helix finger subdomain (Gelis et al. 2007), whereas a FRET study proposed a binding site parallel to the two-helix finger subdomain (Auclair et al. 2010). The latter is supported by evidence indicating that the two-helix finger of SecA inserts into the SecY channel during translocation (Zimmer et al. 2008), as a parallel orientation would facilitate the insertion of the signal sequence into the channel. An X-ray crystal structure of SecA bound to a signal sequence would help to further characterize the binding site. However attempts to do so have been inhibited by the high concentrations of signal peptide required; being very hydrophobic molecules, they tend to aggregate instead of binding to SecA.

The approach used in this study was to create chimera proteins with the signal sequence covalently attached to the C-terminus of SecA in the hopes that the signal sequence would self-bind to the SecA, thus eliminating the need for added signal peptide.

B. Generation of chimera proteins

Chimera proteins were generated using site-directed mutagenesis. The plasmid modified contains the SecA gene under the pT7 promoter system, fused to a
C-terminal His tag in pET29b. The signal sequence KRR-LamB was added in three rounds of mutagenesis to the C-terminus of SecA, before the His tag (Figure 13A). Sequencing of the plasmids was performed in order to confirm the successful generation of the chimera proteins.

A

![Diagram of chimera protein constructs](image)

B

**KRR-LamB:** MMITLR\textbf{KRR}KLPLAVAAAGVMSAQAMA

**Figure 13.** A, Layout of chimera protein constructs. B, KRR-LamB signal peptide sequence. Positively charged residues are shown in red, while the underlined residues are the artificially added lysine and two arginines. The hydrophobic core residues are colored green.

KRR-LamB is a 28-amino-acid long artificially-extended signal peptide in which three positively charged residues, a lysine and two arginines, were added to the N-terminal region of the LamB signal peptide (Figure 13B). The idea behind this modification was to increase the affinity of the signal sequence for SecA, as positive charge in this region was shown to be important for translocation (Akita et al. 1990). The LamB protein is a maltoporin that lies in the outer membrane of gram-negative bacteria. Compared to the wild type-LamB sequence, KRR-LamB binds full length SecA with a 30-fold greater affinity (Gelis et al 2007). KRR-LamB was chosen for this study for two reasons, the first being its increased affinity for SecA. Theoretically this property would increase self-binding in the chimera leading to a
successful crystal structure. The second reason was so that ultimately comparisons could be made between different signal sequence binding sites on SecA, as similar work is being performed using PhoA signal sequence-containing chimera proteins. SecA is capable of recognizing hundreds of different signal sequences, which share no sequence homology, so whether the binding site remains the same between signal sequences is of interest.

In addition to the full length chimeras, truncated versions lacking the C-terminus of SecA were produced using site-directed mutagenesis. The resulting strains are labeled SecA834, as opposed to the full-length SecA901 (referring to the length of the protein in amino acids). The C-terminus of SecA has been shown to inhibit signal peptide binding by a factor of 4-30 fold depending on the signal peptide, indicating that the flexible CTL may swing back in towards the proposed signal peptide binding site (Gelis et al. 2007) (Figure 14). Thus, the hypothesis was that the truncated versions would result in stronger self-binding. In addition, previous crystallization attempts have often required the use of truncated proteins. SecA lacking the CTL is still functional (Karamanou et al. 2005).

C. **Demonstrating self-binding in the chimera proteins**

Once the strains were made, the first step was to confirm that the attached signal sequence was self-binding to the SecA. The plasmids encoding the chimera proteins were transformed into a BL21.19 background, which expresses SecA protein at levels 8 -fold greater than the normal chromosomal SecA level. The proteins were
**Figure 14. CTL can inhibit signal peptide binding site.** One of the crystal structures of *B. subtilis* SecA was able to resolve the CTL (red), and part of it lies among the two helix finger of the HSD (Hunt et al. 2002), a region shown to be important in signal peptide binding. The PPXD has been removed for better visualization.

overexpressed using IPTG induction, and purified on a nickel column via the attached His-tag. A high concentration of imidazole was used to elute the SecA protein bound to the column. Protein purity was confirmed by running the samples on an SDS-PAGE gel and visualized with Coomassie blue staining (Figure 15).

The purified proteins were then used in a fluorescence anisotropy assay to measure the extent of self-binding. In this case, the samples used were comprised of the chimera SecA protein and an exogenous fluorescently-labeled signal peptide. Both the attached signal peptide, KRR-LamB, and the exogenous signal peptide
Figure 15. Coomassie stained SDS-PAGE gel showing SecA chimera purification. Pellet 1 refers to the re-suspended unbroken cell pellet. Pellet 2 refers to the re-suspended membrane pellet. Lysate 1 is the supernatant after spinning to remove the membranes, before it is run on the column. Lysate 2 is the lysate after it has been run through the nickel column. Washes 1 and 2 are washes of the column with a low concentration of imidazole. Elution refers to the eluted SecA protein via a high concentration of imidazole.

compete for the signal peptide-binding site on SecA. Previous work has shown that only one signal peptide-binding site exists on SecA (Auclair et al. 2010), so only one of the two signal peptides can bind to a given SecA molecule. The exogenous signal peptide used, SP22, is the alkaline phosphatase signal sequence labeled with an IANBD fluorescent label via an additional cysteine at the 22nd residue (Figure 16A). This addition has negligible effects on the signal peptide binding, as it falls outside of both the positively charged N-terminus and the hydrophobic core, the two important
regions of the signal peptide required for binding. The samples are exposed to light, which excites the fluorophore on SP22. A very short time later, the light is re-emitted. Due to the small size of SP22 the molecule rotates very quickly, resulting in a large displacement even during the short excitation period. Thus the emitted light is depolarized compared to the initial polarization of the excited light. The extent of depolarization, or lack thereof, is what generates the anisotropy values (see introductory section on fluorescence anisotropy for a more detailed description).

When the SP22 is bound to SecA, indicating that the covalently attached KRR-LamB is not, the complex will rotate much slower due to the increased size. As a result the depolarization will be less, leading to a higher anisotropy value (Figure 16B). By reverse titrating in SecA from concentrations of 0-40 µM (at which concentrations SecA is a dimer, as it has a nanomolar monomer-dimer dissociation constant), a curve can be generated using the collected data, and a dissociation constant (binding constant, $K_d$) can be ascertained. The dissociation constant refers to the concentration of SecA required in order to bind half the available exogenous signal peptide, the concentration of which is kept constant at 1 µM. By comparing differences in the binding constants of the wild type proteins lacking any fused signal sequence with the chimera proteins, the extent of self-binding can be ascertained.
Figure 16. A, Alkaline phosphatase signal peptide sequence. Positively charged residues are shown in red, the hydrophobic core residues are colored green. The blue cysteine indicates the location of the conjugated IANBD label. B, Schematic of anisotropy assay. The SecA chimera (red) with covalently attached KRR-LamB (blue) is initially in solution with alkaline phosphatase (green, yellow IANBD label). Two options can occur: option 1, the attached KRR-LamB self-binds, preventing binding of the exogenous AP signal peptide, option 2, the AP signal peptide is able to bind preventing the attached KRR-LamB from binding.
SecA901 binds the SP22 signal peptide with a $K_d$ of 1.5 µM, whereas the SecA901-KRR-LamB chimera binds the SP22 with a $K_d$ of 7.3 µM. This corresponds to a 4.8 fold decrease in SP22 binding, a difference presumably the result of the covalently attached KRR-LamB signal peptide occupying the binding site. The difference in the truncated constructs is even more pronounced; the truncated SecA834 binds SP22 with a $K_d$ of 1.1 µM, whereas the truncated SecA834-KRR-LamB chimera binds SP22 with a $K_d$ of 11.4 (Figure 17). This corresponds to a 10.4 fold decrease in binding. The $K_d$ values are as follows:

- SecA901: $K_d = 1.5 \pm 0.3$ µM
- SecA901-KRR-LamB: $K_d = 7.26 \pm 2.1$ µM
- Truncated SecA834: $K_d = 1.1 \pm 0.2$ µM
- Truncated SecA834-KRR-LamB: $K_d = 11.4 \pm 2.9$ µM

Figure 17. Binding curves of SP22 signal peptide to full length and truncated *E. coli* SecA. A, SecA901 B, SecA901-KRR-LamB chimera. C, Truncated SecA834 D, Truncated SecA834-KRR-LamB chimera.
decrease in SP22 binding upon covalent attachment of the KRR-LamB signal peptide. Based on the data, the SecA901-KRR-LamB chimera appears to exhibit moderate self-binding, whereas the truncated SecA834-KRR-LamB chimera appears to exhibit rather strong self-binding. The shapes of the curves of the non-chimera and chimera proteins are markedly different, as both of the non-chimera proteins, SecA901 and SecA834, exhibit strong early self-binding (rapid rise in anisotropy values) and then a defined saturation of binding, resulting in clear hyperbolic curves. In contrast, the data points of the chimera proteins do not lend to a clear hyperbolic shape. It appears that at low concentrations of SecA, there is still rather strong binding of the SP22 signal peptide as the anisotropy values rise quickly, similar to the non-chimera protein curves. However, at concentrations of 2 or 3 µM the anisotropy values start to decrease compared to the corresponding non-chimera protein, implying increased inhibition of the SP22 signal peptide binding by KRR-LamB self-binding. This results in a biphasic data set especially pronounced in the truncated SecA834-KRR-LamB chimera, in which both SP22 signal peptide and KRR-LamB signal peptide self-binding is seen. Noticeably, the chimera protein curves do not saturate nearly as well as the non-chimera protein counterparts, also implying inhibition of SP22 signal peptide binding by the attached KRR-LamB peptide. It would be interesting to repeat the experiments with the chimera proteins using even higher concentrations of SecA (beyond 40 µM) in order to achieve saturation of the signal and generate a more accurate binding constant. However, the percentage of SecA that aggregates greatly rises with increasing SecA concentration, and such aggregation would interfere with the signal from the assay as it is based on molecule sizes.
Both the full-length and truncated *E. coli* chimera proteins exhibited a degree of self-binding. However, previous attempts at obtaining X-ray crystal structures have been more successful using *B. subtilis* SecA. In addition, the PPXD of the only *E. coli* SecA crystal structure so far published was largely unable to be determined from the data collected due to being weakly structured (Papanikolau et al. 2007). Considering the objective of this study, a crystal structure of the chimera proteins used without the PPXD being well-structured would be much less useful. Therefore the corresponding constructs were engineered using the *B. subtilis* SecA gene, Div, in the same plasmid and cell background as the *E. coli* constructs. The full-length version is 841 amino acids, while the corresponding truncated construct lacking the CTL is 783 residues.

SecA is an extremely well conserved protein among bacteria, and while *E. coli* and *B. subtilis* split phylogenetically an extremely long time ago, there is still about a 50% sequence homology at the amino acid level (for a sequence alignment see Zimmer et al. 2008 supplementary figures). There are no large structural differences between the SecAs of the two species, and the missing amino acids in the shorter *B. subtilis* sequence occur in small numbers at random intervals throughout the sequence.

The anisotropy assays were performed using the same conditions as before to determine the extent of self-binding, with the ultimate goal of attempting crystallization using *B. subtilis* SecA before trying the *E. coli* version. That being said, it would still be interesting to attempt crystallization of the *E. coli* chimeras given sufficient biochemical evidence of self-binding, as the binding of signal
peptide, which occurs at regions incorporating the PPXD, could help stabilize the PPXD and result in a structure with this region well-defined.

The *B. subtilis* SecA841 binds SP22 with a $K_d$ of 2.3 μM, and the corresponding SecA841-KRR-LamB chimera binds SP22 with a $K_d$ of 13.6 μM, a 5.9 fold difference. The truncated *B. subtilis* SecA783 binds SP22 with a $K_d$ of 2.0 μM, while the truncated SecA783-KRR-LamB chimera binds SP22 with a $K_d$ of 13.4 μM, a 6.7 fold difference (Figure 18). As with the *E. coli* chimeras, both *B. subtilis* chimeras exhibit self-binding, however in this case the difference in self-binding between the full-length construct and the truncated construct is not as pronounced. While this could be a result of the difference in binding affinity between *B. subtilis* and *E. coli* SecA, it could also be a result of the larger error seen in the non-chimera protein curves. With the smaller $K_d$ values seen in the non-chimera proteins, a larger error leads to the possibility of a much larger range of differences in binding between the chimera and corresponding non-chimera. Despite this error, on a more qualitative level, self-binding still seems present to a significant degree. Also similar to the *E. coli* curves is the biphasic nature and lack of saturation of signal present in the chimera protein curves compared to the corresponding non-chimera proteins.
Figure 18. Binding curves of SP22 signal peptide to full length and truncated $B.\ subtilis$ SecA.  a, SecA841  b, SecA841-KRR-LamB chimera.  c, Truncated SecA783  d, Truncated SecA783-KRR-LamB chimera.

D. Demonstrating functionality of chimera protein signal sequence

The next step was to demonstrate that the signal sequence attached to the chimera proteins was capable of directing translocation of a protein substrate, proving that the signal peptide is binding in a physiological manner. A colorimetric alkaline phosphatase assay was used for this purpose featuring new constructs in which the
mature region of the alkaline phosphatase protein, with its native signal sequence removed, was fused to the KRR-LamB signal sequence in the chimera protein (Figure 19A). Alkaline phosphatase is only active in the periplasm of the cell due to the inability of the protein to form disulfide bonds and properly fold in the cytoplasm. The assay measures the presence of the enzyme via its breakdown of added para-nitrophenylphosphate (PNPP) to form a colored substrate (Figure 19B). The cells were grown in a BL21.20 background, so that when the cultures were shifted to 39°C only plasmid SecA was expressed. This is important as the only SecA present in the cell was then part of the chimera-PhoA mature fusion protein, and the assay could directly measure the functionality of the attached KRR-LamB signal peptide. The presence of functional alkaline phosphatase indicates the successful translocation of the protein attached to the chimera.
Figure 19. A, Layout of chimera + AP constructs. B, Schematic of AP assay

The MC1000phoR strain is a healthy cell line that produces an abundance of alkaline phosphatase, and was used as a positive control. It generated an alkaline phosphatase (AP) activity of 483 units. The SecA901-KRR-LamB-PhoA mature protein scored 327 AP units, while the truncated SecA834-KRR-LamB-PhoA mature protein averaged 100 AP units. Full length and truncated negative controls featuring a point mutation in the hydrophobic core of KRR-LamB that changed a valine to an aspartic acid, a mutation estimated to disrupt signal sequence functionality by greater
than 95% (Donald Oliver 1985), were used as negative controls. These constructs produced -13 and -3 AP units, respectively (Figure 20A).

While both constructs showed significantly more alkaline phosphatase activity than the negative controls, implying that the attached KRR-LamB signal sequence is capable of binding in a physiological manner, there was a significant difference in activity between the SecA901-KRR-LamB-PhoA mature construct and the truncated SecA834-KRR-LamB-PhoA mature construct. To further investigate whether this was truly a difference in translocation ability between the two constructs, or whether the full-length strain simply was producing more protein and thus had more to translocate, a continuation of the assay was performed. Immediately after the activities were calculated, samples of the cell cultures of SecA901-KRR-LamB-PhoA mature, SecA834-KRR-LamB-PhoA mature and the MC1000phoR, normalized to the same amount of AP units, were run on an SDS-PAGE gel. The samples were ultimately visualized by performing two separate western blots on the same samples, one using a SecA antibody (Figure 20B), and the other using an alkaline phosphatase antibody (Figure 20C).

All three of the strains display a band corresponding to the SecA monomer in Figure 20B. The SecA-KRR-LamB-PhoA chimera protein is also able to be visualized as a band higher in the gel in the two constructs. Since the only SecA protein being produced is
Figure 20. A, AP assay results. B, Western blot of AP assay samples probed using SecA antibody. C, Western blot of AP assay samples probed using AP antibody.
plasmid SecA, all of it begins as a fusion protein, and the SecA by itself results from the cleavage of the AP complex. Noticeably, a much higher proportion of the SecA in the SecA834-KRR-LamB-PhoA mature construct is in the form of the fusion protein, compared to the SecA901-KRR-LamB-PhoA mature construct in which most of the SecA is in the cleaved form. This seems to imply that there is some problem with cleavage of the preprotein in the truncated construct – one possible explanation is that that the SecA has a harder time releasing the bound signal peptide into the SecYEG channel, as needs to occur to begin translocation. This is supported by the relatively tight binding demonstrated in the truncated constructs with the anisotropy assay.

Looking at the alkaline phosphatase probe, it is apparent that the SecA901-KRR-LamB-PhoA mature construct required more alkaline phosphatase than the MC1000PhoR strain to achieve similar AP activity, and that the SecA834-KRR-LamB-PhoA mature construct required even more than the full-length construct. This is evident by the darker bands seen when comparing similar amounts of AP units (for example, lanes 3, 7, and 11). This result demonstrates that the SecA901-KRR-LamB-PhoA mature has a lower specific activity than wild type alkaline phosphatase produced by the MC1000-phoR strain, and the SecA834-KRR-LamB-PhoA mature has a lower specific activity than the full-length. This result proves that the difference in AP activity is not due to a smaller amount of alkaline phosphatase being produced. The AP antibody was also able to detect a significant amount of the SecA-PhoA fusion protein still intact in the SecA834-KRR-LamB-PhoA mature construct, whereas virtually none is detectable in the SecA901-KRR-LamB-PhoA mature
construct, further validating the results from the first probe. This also partially explains why this construct has a lower AP activity— it is stuck in the fusion protein state, unable to translocate the AP and thus contribute to the assay.

Another result revealed in the Western blot is the potential for multiple cleavage sites of the PhoA protein in the full-length construct. Looking at the SecA only band in the SecA901-KRR-LamB-PhoA mature lanes, there seems to be not one, but two bands overlapping each other. Furthermore, in the AP probe, looking at the same lanes, there are noticeable yet faint bands above the defined AP band. These results imply that the cleavage of the AP may be occurring in two different sites, resulting in two slightly different SecA molecules, and two slightly different AP molecules. If one of the AP proteins is not functional or poorly functional, then this could contribute to the reduced specific activity seen.
IV. Discussion
The binding of the N-terminal signal sequence of a preprotein to the DEAD motor protein SecA is an essential step in post-translational protein translocation in bacteria, the path used by most proteins destined to function in the periplasm and outer membrane. However, the precise signal peptide-binding site of SecA and the orientation of the bound signal peptide have not been determined conclusively by previous approaches. Biochemical and genetic studies often can only identify a small subset of residues involved in the binding process, and fail to achieve an overarching, three-dimensional representation of the binding site. The NMR structure features a very high concentration of signal peptide that could have facilitated non-physiological signal peptide binding (Gelis et al. 2007), whereas the FRET study proposed a less-defined, broader binding site, and an orientation opposite that seen in the NMR structure (Auclair et al. 2010). An X-ray crystal structure of SecA bound to a signal peptide would help to more conclusively characterize the signal peptide binding site. However, traditional methods of obtaining such a crystal involve extremely high concentrations of signal peptide, resulting in the molecule aggregating and precipitating. In order to overcome this hurdle and attempt to obtain an X-ray structure of SecA bound to signal peptide, we designed chimera proteins with the KRR-LamB signal sequence fused to the C-terminus of SecA. If the signal sequence self-binds it eliminates the need for added signal peptide, thus potentially avoiding the aggregation issue.

The anisotropy assay proved that self-binding is evident in both the *E. coli* and *B. subtilis* full-length SecA chimera proteins (Table 3). The extent of self-binding
### Table 3. Anisotropy assay results of SecA chimera proteins.

<table>
<thead>
<tr>
<th>Chimera Protein</th>
<th>Dissociation constant (µM)</th>
<th>X Fold Difference in Binding</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em> SecA901 (WT)</td>
<td>1.5 ± 0.3</td>
<td>4.9</td>
</tr>
<tr>
<td><em>E. coli</em> SecA901-KRRLamB</td>
<td>7.3 ± 2.1</td>
<td></td>
</tr>
<tr>
<td><em>E. coli</em> SecA834</td>
<td>1.1 ± 0.2</td>
<td>10.4</td>
</tr>
<tr>
<td><em>E. coli</em> SecA834-KRRLamB</td>
<td>11.4 ± 2.9</td>
<td></td>
</tr>
<tr>
<td><em>B. subtilis</em> SecA841</td>
<td>2.3 ± 1.3</td>
<td></td>
</tr>
<tr>
<td><em>B. subtilis</em> SecA841-KRRLamB</td>
<td>13.6 ± 1.8</td>
<td>5.9</td>
</tr>
<tr>
<td><em>B. subtilis</em> SecA783</td>
<td>2.0 ± 1.1</td>
<td></td>
</tr>
<tr>
<td><em>B. subtilis</em> SecA783-KRRLamB</td>
<td>13.4 ± 3.8</td>
<td>6.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chimera Protein</th>
<th>Dissociation constant (µM)</th>
<th>X Fold Difference in Binding</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em> SecA901 (WT)</td>
<td>1.8 ± 0.4</td>
<td>1.5</td>
</tr>
<tr>
<td><em>E. coli</em> SecA901-PhoA</td>
<td>2.73</td>
<td></td>
</tr>
<tr>
<td><em>E. coli</em> SecA834</td>
<td>1.1 ± 0.2</td>
<td>6.4</td>
</tr>
<tr>
<td><em>E. coli</em> SecA834-PhoA</td>
<td>7.0 ± 1.0</td>
<td></td>
</tr>
<tr>
<td><em>B. subtilis</em> SecA841</td>
<td>4.6 ± 0.9</td>
<td>0.4</td>
</tr>
<tr>
<td><em>B. subtilis</em> SecA841-PhoA</td>
<td>1.9 ± 0.4</td>
<td></td>
</tr>
<tr>
<td><em>B. subtilis</em> SecA783</td>
<td>0.5 ± 0.25</td>
<td>9</td>
</tr>
<tr>
<td><em>B. subtilis</em> SecA783-PhoA</td>
<td>4.5 ± 0.6</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Anisotropy assay results of SecA chimera proteins. a, KRR-LamB chimera proteins. b, PhoA chimera proteins (data provided by Qi Zhang)
was much more pronounced in the *E. coli* truncated SecA834-KRR-LamB chimera than the SecA901-KRR-LamB-chimera, and slightly more pronounced in the corresponding *B. subtilis* chimera protein (Table 3). However as mentioned before, the anisotropy curves of the chimera proteins are generally biphasic, and the analysis of self-binding would benefit greatly from an alternative curve fit that accounts for the exogenous SP22 signal peptide binding. With such an analysis the three different states of SecA present in the assay (free SecA, SP22 signal sequence bound, KRR-LamB bound) could be more definitively ascertained.

Similar work on chimera proteins with the PhoA signal sequence attached, as opposed to the KRR-LamB sequence used in this study, have been performed by Qi Zhang as one of the overarching goals of this study was to compare the binding site on SecA of different signal peptides. With the present results such a comparison is not yet feasible, however comparisons can be made as to the extent of self-binding seen from the anisotropy assay.

Firstly, the dissociation constants seen in the KRR-LamB chimeras tend to be much higher than those seen in the PhoA chimeras, even if the difference in binding between the chimera and corresponding non-chimera is more similar (brought about by a lower WT $K_d$, as seen in the *B. subtilis* SecA783/SecA783-PhoA pair). Beyond the simple explanation of different experimenters performing the assay, the nature of the signal sequences themselves offers a potential answer; PhoA signal sequence has only one positively charged residue in its N-terminus, whereas KRR-LamB has five. The increased positive charge presumably increases the affinity of the signal peptide for SecA.
The other major difference seen between the two different signal sequence chimera sets is the effect of the CTL on signal sequence binding. For the KRR-LamB constructs, self-binding is evident in all of the chimeras, with the removal of the CTL only exerting a strong influence on the *E. coli* chimeras. In contrast, the full-length PhoA constructs exhibit virtually no self-binding. However upon removal of the CTL self-binding increases greatly. This varying effect of the CTL can be explained based off of the data presented in the Gelis et al. study that proposed the CTL inhibits the signal peptide binding site. The difference in binding calculated in that study between the SecA901 and SecA834 varied greatly depending on the signal sequence used. LamB binding increased 30 fold, KRR-LamB binding increased 10 fold, and PhoA binding increased 4 fold (Gelis et al. 2007). While the differences seen in this study don’t exactly corroborate those differences (namely it seems the PhoA is more affected by removal of the CTL), the takeaway point is that the CTL inhibition varies greatly depending on the signal peptide that is binding. Since the differences seen in our study are with different signal peptides, this presents a plausible explanation.

On a similar note, the lack of noticeable difference in exogenous binding of SP22 signal peptide seen in the KRR-LamB chimeras compared to the differences seen in the Gelis et al. study can be explained by the limited range of detection in the assay. With the amount of error present, there easily could be a four-fold difference in binding (the number presented by Gelis et. al) between SecA901 and SecA834, or SecA841 and SecA783. Another difference between the results of this study and the Gelis et al. study is the dissociation constants presented for the non-chimera binding of PhoA. In this study, $K_d$ of 1.5 µM and 1.1 µM were calculated for SecA901 and
SecA834, respectively. The corresponding numbers for the same \( K_d \) presented in the Gelis et al. study were 30 µM and 7.5 µM. The differences can be attributed to the assay conditions; their study used ITC calorimetry featuring a slightly higher concentration of SecA (50 µM), and a much higher concentration of signal peptide (up to 0.5 mM, compared to the 1 µM used in this study). These higher concentrations could have led to an abnormally high dissociation constant. In addition, the dissociation constant presented in this study for SecA901 affinity for PhoA signal peptide is corroborated by other studies (Auclair et al. 2010, Grady et al. 2011, Wang et al. 2000).

The alkaline phosphatase assay demonstrated that the KRR-LamB signal sequence is capable of physiologically binding to SecA and facilitating translocation of a fused preprotein, as both the SecA901-KRR-LamB-PhoA mature and the SecA834-KRR-LamB-PhoA mature constructs scored significantly higher than the corresponding constructs with a point mutation in the hydrophobic core of the signal sequence.

To investigate the large difference in AP activities between the full-length and truncated chimera proteins, Western blots probing for both SecA and alkaline phosphatase were performed. The results revealed that the specific activity in the SecA901-KRR-LamB-PhoA mature protein was lower than that of the wild type MC1000-phoR strain, and that the specific activity of the t SecA834-KRR-LamB-PhoA mature was even lower than that of the full-length protein. The blot also showed that a large proportion of the alkaline phosphatase in the SecA834-KRR-
LamB-PhoA mature construct is still fused to the chimera protein, indicating an issue with the successful translocation of the substrate.

A possible explanation for the lowered specific activity of the truncated strain compared to the full-length strain is that some of the alkaline phosphatase showing up in the Western and contributing to the AP activity of the SecA834-KRR-LamB-PhoA mature construct is not actually translocated across the periplasm, but instead has gained enzymatic activity in the cytoplasm. Although normally the disulfide bond formation required for enzymatic activity does not normally occur until the enzyme reaches the periplasm, it has been reported that it can occur slowly in the cytoplasm in non-growing cells, thus facilitating alkaline phosphatase folding (Derman and Beckwith 1995).

In order to further investigate this phenomenon, an experiment can be conducted in which the contents of the cytoplasm and the periplasm are separated, and both the alkaline phosphatase assay and Western blot are performed on each separate collection. By comparing the results of the two compartments, perhaps it will be evident that in the periplasm the full-length and truncated construct have similar specific activities, and that indeed some of the AP activity in the SecA834-KRR-LamB-PhoA construct is a result of cytoplasmic activity.

The other intriguing question presented by the results of the Western blot is why the SecA834-KRR-LamB-PhoA protein seems to stall in the fusion protein state and is unable to cleave the alkaline phosphatase successfully. The intact chimera-PhoA protein could be in the cytoplasm in a pre-translocation state, or it could be membrane associated and in the process of translocating. In the latter situation,
perhaps the AP is translocated through the SecYEG channel, however is unable to be cleaved and released into the periplasm. Thus it would be floating in the periplasm, yet still attached to the SecA on the cytoplasmic side. In order to ascertain the location of the stalled SecA834-KRR-LamB-PhoA protein, a protease experiment can be performed. By treating the cells with lysozyme and EDTA, thus opening holes in the outer membrane, a protease can be added to the periplasm. If the alkaline phosphatase is floating in the periplasm yet still attached to the SecA, the protease will cleave it and release it into the periplasm. Since alkaline phosphatase is a rather protease resistant protein, it will not be further degraded. If the SecA834-KRR-LamB-PhoA protein is completely in the cytoplasm, the addition of the protease will have no effect. Again, analyzing the respective compartments with a Western blot will determine whether any alkaline phosphatase has been released due to protease treatment.

Beyond ascertaining the location of the 834-KRR-PhoA chimera protein, there remains the question of why the attached alkaline phosphatase is not being translocated. In comparing the 834-KRR-PhoA chimera with the 901-KRR-PhoA chimera, the latter of which has a much higher specific activity and seems to cleave alkaline phosphatase much better, the only apparent difference is the presence of the CTL. A previous study showed that the CTL of SecA is important for the translocation of certain preproteins, and that truncation of it can lead to translocation defects (Rajapandi et al. 1994). The CTL was then later shown to be the SecB binding site (Fekkes et al. 1999). Finally, while alkaline phosphatase is normally not dependent on SecB for efficient translocation, a weak signal sequence can necessitate
the chaperone activity of SecB in order to rescue translocation (Kim et al. 2000). The hypothesis drawing on these studies is that the covalently bound KRR-LamB signal peptide is binding SecA very tightly, preventing it from being an efficient promoter of translocation. In the 901-KRR-PhoA chimera, SecB is able to rescue this defect. In the 834-KRR-LamB-PhoA chimera the lack of CTL prevents SecB binding, and thus translocation cannot occur as efficiently. This hypothesis could theoretically be tested by running the assay in a SecB knockout strain, and seeing if the specific activity of the 901-KRR-LamB-PhoA chimera is lowered to a similar level to that of the 834-KRR-LamB-PhoA chimera.

Finally, as mentioned in the results section, there is a light second band directly above the alkaline phosphatase band in the 901-KRR-LamB-PhoA chimera, potentially corresponding to a slightly larger alkaline phosphatase protein as a result of an alternate cleavage site. Such an alternate cleavage site, as opposed to the standard Ala-X-Ala motif found starting at the -3 position of the signal sequence, has been demonstrated to be viable in both maltose-binding protein and alkaline phosphatase if the normal site is inaccessible to the signal peptidase (Fikes et al. 1990, Laforet and Kendall 1991). If the alkaline phosphatase is indeed being cleaved in an alternate manner, in either the 834-KRR-LamB-PhoA chimera or the 901-KRR-LamB-PhoA chimera, the alkaline phosphatase activity could be reduced. This is supported by work showing that mutations in the processing site of PhoA signal peptide lead to reduced catalytic activity of alkaline phosphatase (Nesmeyanova et al. 1996).
Since both the full-length and truncated constructs displayed significant self-binding in the anisotropy assay, and signal peptide activity in the alkaline phosphatase assay (although how physiological the binding is remains to be seen), both the SecA841-KRR-LamB and the truncated SecA783-KRR-LamB chimeras from *B. subtilis* SecA were sent to the collaborator in Rapoport’s group as crystallization candidates. As of now, screening of various crystallization conditions is taking place. The *B. subtilis* strains are being tried first due to the success with using this organism in the past, however attempts at an *E. coli* structure have not been ruled out. In fact, one hypothesis is that a bound signal peptide will help stabilize the PPXD and allow for better definition in this area when obtaining a crystal structure, compared to the previous structure of *E. coli* SecA. Ultimately, comparison of a SecA–KRR-LamB structure to a SecA-PhoA signal sequence structure is a main goal, however even just one of the structures would be extremely helpful in solving the signal peptide binding site mystery.

In order to continue to investigate the orientation of the signal peptide, specifically if it binds parallel or orthogonal to the two-helix finger subdomain of SecA, a Förster resonance energy transfer study has been designed (Qi Zhang). It features the placement of chromophores at specific locations that will result in substantially different FRET signals for the proposed orientations of the signal peptide. The locations were picked by using the Auclair et al. results mapped out on the Hunt et al. structure to test the parallel binding orientation, and the Gelis et al. structure to test the orthogonal binding orientation. This experiment can be performed for both the PhoA and the KRR-LamB signal sequence. While it would
not generate as definitive a picture as an X-ray crystal structure would, it should still provide some insight as to the nature of signal peptide binding without being reliant on successful crystallization.

In conclusion, the chimera protein approach is a technique novel to the field of SecA crystallography that hopefully will enable the determination of the signal peptide binding domain of SecA at a very high resolution. In addition to the use of chimera proteins as a way to achieve signal peptide binding without extremely high concentrations of exogenous signal peptide, the chimera proteins could also be used to study the importance of certain regions of SecA in signal peptide binding in a different way. The CTL is inherently flexible, and provides a way for the covalently attached signal sequence to self-bind SecA. In fact, the addition of a linker sequence in the SecA-PhoA chimera proteins was shown to have a negligible effect on signal sequence binding (Qi Zhang). Considering self-binding was seen in the truncated KRR-LamB constructs, a linker sequence was not pursued. However, even though the sequence is binding, perhaps it is binding in a non-physiological site or promoting non-physiological folding due to the immediate attachment of the signal sequence to the HSD, since the CTL has been removed. Thus, the effects of various linker sequences on translocation activity of the truncated chimeras would be interesting to see. For example, if the addition of a ten amino acid linker improves the AP activity of the SecA834-KRR-LamB-PhoA mature construct. In addition, different truncations of SecA could lead to results as to how flexible certain regions of SecA are (i.e. which truncations exhibit translocation activity of fused signal sequence without GS linker).
The SecA-KRR-LamB-PhoA chimeras could also be used beyond this study. The isolation and crystallization of a translocation intermediate, a SecA-preprotein complex that is trapped with the preprotein somewhere in the SecYEG channel is of great interest. Such an intermediate may have been found naturally with the SecA834-KRR-LamB-PhoA chimera as seen in the results section, however even if not found naturally alteration of the chimera could perhaps promote this translocation intermediate state. If the chimeras without the PhoA mature region themselves are successful at crystallizing, perhaps the PhoA mature fusion proteins will provide easier crystal growth of the large complex required for a translocation intermediate.
V. References


Rusch, S. L., & Kendall, D. A. (2007). Current Topics Interactions That Drive Sec-Dependent Bacterial Protein Transport †, *46*(34).


*Biochimica et biophysica acta, 1808*(3), 851-65.