

Electronic Thesis and Dissertation Repository

---

4-29-2016 12:00 AM

## Interaction of SecA with Unfolded Polypeptides

Aliakbar Khalili Yazdi  
*The University of Western Ontario*

Supervisor  
Dr. Brian Shilton  
*The University of Western Ontario*

Graduate Program in Biochemistry  
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of  
Philosophy  
© Aliakbar Khalili Yazdi 2016

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Biochemistry Commons](#), and the [Structural Biology Commons](#)

---

### Recommended Citation

Khalili Yazdi, Aliakbar, "Interaction of SecA with Unfolded Polypeptides" (2016). *Electronic Thesis and Dissertation Repository*. 3739.  
<https://ir.lib.uwo.ca/etd/3739>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact [wlsadmin@uwo.ca](mailto:wlsadmin@uwo.ca).

Interaction of SecA with Unfolded Polypeptides

Thesis format: Monograph

by

Aliakbar Khalili Yazdi

Graduate Program in Biochemistry

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

The School of Graduate and Postdoctoral Studies  
The University of Western Ontario  
London, Ontario, Canada

© Aliakbar Khalili Yazdi 2016

## Abstract and Keywords

The evolutionarily well-conserved SecA is essential for bacterial post-translational translocation. SecA uses the energy of ATP to drive preproteins through the membrane pore. The functional oligomeric state of SecA and the molecular basis for recognition of unfolded polypeptides by SecA are major unresolved questions that must be addressed to understand preprotein targeting and the molecular mechanics of SecA-mediated translocation. This thesis will address three aspects of these questions. First, the role of unstructured termini in the oligomerization and function of various SecA constructs was elucidated. By re-examining the tetramerization of a truncated SecA construct (SecA-N68), it was shown that the unstructured polypeptides at its termini are mediating its oligomerization. In turn, by removal of the first 14 N-terminal residues of the functional SecA-N95 construct, dimerization was drastically weakened. Although the weakened dimerization did not significantly affect the solution ATPase activity of SecA-N95 *in vitro*, it was shown that SecA-N95 $\Delta$ N is not functional *in vivo*. Second, the interaction of unfolded polypeptides with SecA protein was investigated. It was revealed that preproteins contain sequences that can bind SecA *in vitro* through the binding site(s) located on the nucleotide binding domains and/or the stem region of the preprotein cross-linking domain. Both nucleotides and the N-terminal segment of SecA affected these interactions. Moreover, the sequences of several high affinity SecA binding peptides were analyzed and a possible “motif” on preproteins for interaction with SecA was identified. Finally, to acquire high-resolution insight into preprotein targeting mechanisms, we aimed at obtaining and improving the crystals of *E. coli* SecA-N95 and *M. tuberculosis* SecB (*mtSecB*). By removal of the unstructured termini of these proteins, larger crystals were obtained with higher frequency. Although the high-resolution structures of these molecules were not resolved, the advances made in the crystallization of these proteins paves the way for future efforts. In conclusion, this thesis has shown that SecA is able to interact with the unstructured polypeptides *in vitro*, and the presence of unstructured polypeptides at the termini of SecA have a profound affect on its function both *in vivo* and *in vitro*.

## Keywords

SecA, bacterial protein translocation, Sec system, oligomerization, unstructured polypeptides, preproteins, unfolded polypeptide binding, protein crystallization, X-ray crystallography, mycobacterial SecB-like, enzyme kinetics, and analytical ultracentrifugation.

## Acknowledgments

First and foremost, my sincere gratitude goes to my supervisor, Dr. Brian Shilton, for his constant guidance, encouragement, and support during my PhD studies. You have provided me with countless life lessons, and I am deeply indebted and immensely grateful for having had you as a supervisor, mentor, and friend. You will continue to be a source of inspiration in my future endeavors.

I would also like to thank the members of my advisory committee, Dr. Stan Dunn and Dr. Eric Ball, for all their knowledge and advice, and for taking an interest in my work. You always kept me focused, challenged my ideas, and guided me throughout my thesis project. I thank several research technicians in the Department of Biochemistry including, Guangxin Xing, Anne Brickenden, Lee-Ann Briere, and Yumin Bi, for their assistance and training. It would be impossible to learn many techniques without your help. I would also like to thank the current and past members of Shilton Lab, and all the members of biochemistry department.

I cherish all the friendships I have made during my time as a graduate student at Western University. Thank you especially to Dr. Kevin Leung, Ryan Killoran, Shahed Al'Massri, Andrew Maciejewski, and Joel Bierer for your friendship and guidance over the years. Thank you for your constant encouragement, support, coffee breaks, and being always there to listen whenever I had something to celebrate or complain about.

I would like to extend my gratitude to my whole family, especially to my beloved parents. You have encouraged me throughout my entire academic career, and your continual unconditional love and support always made me believe I can succeed.

Last but not least, Nasrin, I am immensely grateful for your continuous support and unwavering love. Thank you for being so patient during my thesis, and without your support, I would never have succeeded. You are my best friend, and it is a precious gift to have you in my life. Thank You.

# Table of Contents

Abstract and Keywords.....	ii
Acknowledgments.....	iv
Table of Contents.....	v
List of Tables.....	ix
List of Figures.....	x
List of Appendices.....	xiii
List of Abbreviations.....	xiv
Chapter 1.....	1
1 Introduction.....	1
1.1 General introduction.....	1
1.2 Bacterial protein translocation.....	2
1.3 Protein transport into and across the inner-membrane.....	3
1.4 Components of the bacterial Sec system.....	8
1.5 SecYEG; the bacterial membrane pore forming complex.....	10
1.6 Preproteins; substrates of the Sec system.....	12
1.7 SecA ATPase; the driving force for translocation.....	14
1.8 Interaction of SecA with preproteins.....	19
1.9 Oligomerization of SecA.....	22
1.10 SecB; a translocation dedicated chaperone.....	23
1.11 The mycobacterial Sec system.....	26
1.12 Accessory SecA2 translocation system.....	26
1.13 Mycobacterial SecB chaperone.....	28
1.14 Scope of thesis.....	29
Chapter 2.....	31

2	Oligomerization of SecA is Mediated by Unstructured Polypeptides .....	31
2.1	Introduction.....	31
2.2	Materials and methods .....	36
2.2.1	Molecular cloning.....	36
2.2.2	Expression of recombinant proteins.....	41
2.2.3	Purification of His-tagged SecA proteins with truncated N-terminus .....	41
2.2.4	Purification of His-tagged SecA proteins with the wild-type N-terminus .....	44
2.2.5	Purification of non-His-tagged SecA proteins.....	45
2.2.6	Analytical size exclusion chromatography .....	46
2.2.7	Analytical ultracentrifugation .....	47
2.2.8	ATPase measurements .....	49
2.2.9	<i>In vivo</i> SecA complementation assay .....	50
2.3	Results.....	51
2.3.1	SecA expression and purification .....	51
2.3.2	Oligomerization of SecA-N68 is mediated by its unstructured termini ...	54
2.3.3	H <sub>6</sub> -SecA-N68 forms monomers, dimers, and tetramers .....	58
2.3.4	Roles of the N- and C-terminal sequences in SecA-N68 oligomerization	60
2.3.5	The role of the extreme N-terminus in dimerization of SecA-N95 .....	65
2.3.6	Solution ATPase activity of dimeric and monomeric SecA-N95.....	69
2.3.7	Functionality of the N-terminally truncated SecA <i>in vivo</i> .....	73
2.4	Discussion.....	75
	Chapter 3.....	80
3	Interaction of of ATPase SecA with the Mature Regions of Preprotein Substrates ....	80
3.1	Introduction.....	80
3.2	Materials and methods .....	84
3.2.1	Purification of SecA proteins.....	84

3.2.2	Preparation of CNBr peptides from MBP.....	84
3.2.3	Surface plasmon resonance analysis.....	85
3.2.4	Sequence analysis .....	86
3.3	Results.....	88
3.3.1	Interaction of SecA with mature regions of MBP .....	88
3.3.2	Domain mapping of SecA-preprotein interactions .....	97
3.3.3	The extreme N-terminus of SecA is required for peptide binding.....	102
3.3.4	Analysis of the sequence specificity of SecA for peptide binding .....	106
3.4	Discussion.....	111
Chapter 4	.....	117
4	Towards High Resolution Analysis of the Sec System's Components .....	117
4.1	Introduction.....	117
4.2	Materials and methods .....	121
4.2.1	Molecular cloning.....	121
4.2.2	Purification of SecA-N95 $\Delta$ N(ER) .....	125
4.2.3	Purification of <i>mt</i> SecB proteins .....	125
4.2.4	Partial proteolysis and Edman sequencing .....	126
4.2.5	Analytical ultracentrifugation and size exclusion chromatography.....	126
4.2.6	Protein crystallization .....	127
4.3	Results.....	129
4.3.1	Purification and initial crystallization of SecA-N95 $\Delta$ N(ER).....	129
4.3.2	Optimization of SecA-N95 $\Delta$ N(ER) purification for greater stability.....	131
4.3.3	Crystallization of SecA-N95 $\Delta$ N(ER).....	134
4.3.4	Purification, characterization, and crystallization of <i>mt</i> SecB.....	136
4.3.5	Crystallization of an N-terminal truncation of <i>mt</i> SecB .....	140
4.4	Discussion.....	147



Chapter 5.....	149
5 Summery and conclusions .....	149
5.1 SecA oligomerization is mediated by unstructured termini.....	149
5.2 Interaction of SecA with unfolded preproteins.....	150
5.3 Crystallization of SecA and <i>mt</i> SecB.....	151
5.4 Future Directions .....	152
5.4.1 Structural models for preprotein binding site on SecA.....	152
5.4.2 Physiological roles of sequence specificity of SecA .....	153
5.4.3 Identification of binding site(s) for the N-terminal segment of SecA ....	154
5.4.4 Substrate specificities of mycobacterial SecA1 and SecA2 .....	154
5.4.5 The role of <i>mt</i> SecB in the Sec system .....	155
5.5 Conclusion .....	156
6 References.....	157
Appendices.....	169
Appendix A. The effect of the N-terminal segment on purification of SecA constructs	169
Appendix B. List of SecA binding peptides identified through oriented peptide library array.....	173
Curriculum Vitae .....	174

## List of Tables

Table 2-1. List of the selected research papers on dimerization of SecA.....	34
Table 2-2. List of the SecA constructs and expression vectors used in this chapter .....	39
Table 2-3. List of the primers used in this chapter. ....	40
Table 2-4. Kinetic parameters for ATP hydrolysis by SecA constructs.....	72
Table 4-1. List of the constructs used in these studies.....	123
Table 4-2. List of the primers used in these studies.....	124

## List of Figures

Figure 1-1. Schematic model of post-translational translocation by the Sec system.....	7
Figure 1-2. Structure of bacterial SecYEG pore forming complex. ....	11
Figure 1-3. Structure and domain arrangement of bacterial ATPase SecA.....	16
Figure 1-4. Crystal structure of SecA in complex with SecYEG.....	18
Figure 1-5. Conformational flexibility of PPXD domain of SecA.....	21
Figure 1-6. Crystal structure of SecB tetramer.....	25
Figure 2-1. SecA constructs.....	52
Figure 2-2. Purified SecA proteins used in this study.....	53
Figure 2-3. Size exclusion chromatography of H <sub>6</sub> -SecA-N68 and SecA-N68ΔNC.....	56
Figure 2-4. Sedimentation velocity analysis of SecA-N68ΔNC.....	57
Figure 2-5. H <sub>6</sub> -SecA-N68 participates in a monomer-dimer-tetramer equilibrium.....	59
Figure 2-6. Effect of the hexahistidine tag on oligomerization of SecA-N68.....	61
Figure 2-7. Size exclusion analysis of SecA-N68ΔC.....	62
Figure 2-8. Roles of unstructured termini in oligomerization of SecA-N68.....	64
Figure 2-9. Effect of the extreme N-terminus on oligomerization of SecA-N95.....	66
Figure 2-10. Sedimentation equilibrium analysis of SecA-N95 and SecA-N95ΔN.....	68
Figure 2-11. Kinetic analysis of SecA constructs.....	71
Figure 2-12. <i>In vivo</i> complementation analysis of SecA-N95 and SecA-N95ΔN.....	74
Figure 3-1. SecA constructs used in this study.....	89

Figure 3-2. The sequence of mature MBP and the unfolded polypeptides generated using CNBr digestion.....	90
Figure 3-3. Cyanogen bromide cleavage of MBP .....	91
Figure 3-4. Surface Plasmon Resonance (SPR) to study SecA-peptide interactions. ....	93
Figure 3-5. Binding of SecA-N68 and SecA-N95 to MBP fragments .....	95
Figure 3-6. Regeneration and reproducibility of the MBP biosensor chip.....	96
Figure 3-7. Domain mapping of SecA-preprotein interactions .....	100
Figure 3-8. The effects of different nucleotides on binding of SecA-N68 to the unfolded substrate	101
Figure 3-9. The extreme N-terminus of SecA is required for binding to unfolded substrate polypeptides <i>in vitro</i> .....	104
Figure 3-10. Role of the of SecA extreme N-terminal sequence in binding to an immobilized peptide. ....	105
Figure 3-11. Alignment of top SecA-binder peptides.....	108
Figure 3-12. Location of the identified motif in the top SecA binder peptides.....	109
Figure 3-13. The SecA-binding motif position along the preprotein substrates.....	110
Figure 3-14. Domain arrangement of <i>E. coli</i> SecA-N68 .....	114
Figure 4-1. SDS-PAGE analysis and initial crystallization of SecA-N95 $\Delta$ N(ER).....	130
Figure 4-2. Optimized purification of SecA-N95 $\Delta$ N(ER) for crystallization.....	132
Figure 4-3. Improved <i>in vitro</i> stability of SecA-N95 $\Delta$ N(ER) protein .....	133
Figure 4-4. Crystals of SecA-N95 $\Delta$ N(ER) .....	135
Figure 4-5. SDS-PAGE analysis of purified <i>mt</i> SecB protein and its initial crystallization .....	138

Figure 4-6 Sedimentation velocity analysis of wild-type <i>mtSecB</i> protein .....	139
Figure 4-7 Partial proteolysis of <i>mtSecB</i> protein.....	141
Figure 4-8 Removal of the N-terminal 15 residues leads to destabilization of <i>mtSecB</i> .....	143
Figure 4-9 Analysis of purified <i>mtSecB</i> $\Delta$ N8 protein.....	144
Figure 4-10 <i>mtSecB</i> $\Delta$ N8 crystals.....	146
Figure A-1. The effect of the N-terminal segment of SecA on purification of two SecA constructs. .....	171
Figure A-2. Analysis of SecA-N68 $\Delta$ C construct before removal of the affinity tag.....	172

## List of Appendices

Appendix A: The effect of the N-terminal segment on purification of SecA constructs .....	169
Appendix B: List of SecA binding peptides identified through oriented peptide library array .....	173

## List of Abbreviations

Å	Angstrom
ADP	Adenosine diphosphate
Amp	Ampicillin
AMP-PNP	Adenylyl-imidodiphosphate
ATP	Adenosine triphosphate
<i>B. subtilis</i>	<i>Bacillus subtilis</i>
BSA	Bovine serum albumin
CAPS	3-(cyclohexylamino)-1-propanesulfonic acid
CNBr	Cyanogen bromide
CTD	C-terminal domain
Cys	Cysteine
DegP	Periplasmic serine endoprotease
DM	DEAD motor
DNase I	Deoxyribonuclease I
DPCC	Diphenyl carbamyl chloride
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	Ethylenediaminetetraacetic acid
ER	Entropy reducing mutation
$f/f_{min}$	Frictional ratio
FRET	Fluorescence resonance energy transfer
GBP	D-galactose-binding periplasmic protein
Gn-HCl	Guanidine hydrochloride
<i>H. influenzae</i>	<i>Haemophilus influenzae</i>
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-tag	Hexa-histidine tag
HSD	Helix scaffold domain
HWD	Helix wing domain
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside

Kan	Kanamycin
K <sub>D</sub>	Dissociation constant
kDa	Kilodalton
LamB	Lambda receptor protein
LB	Luria-Bertani broth
Lys	Lysine
<i>M. tuberculosis</i>	<i>Mycobacterium tuberculosis</i>
m <sup>7</sup> Ino	7- methylinosine
MalE	Maltose-binding protein precursor
MBP	Maltose binding protein
<i>mtSecB</i>	<i>Mycobacterium tuberculosis</i> SecB
NBD	Nucleotide binding domain
NHS	N-hydroxysuccinimide
OD	Optical density
OmpA	Outer-membrane protein A
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDEA	2-(2-pyridinyldithio)ethaneamine hydrochloride
PMSF	Phenylmethanesulfonyl fluoride
PNPase	Purine nucleoside phosphorylase
PPXD	Preprotein cross-linking domain
<i>prl</i>	Protein localization mutation
psi	Pounds per square inch
PVDF	Polyvinylidene difluoride
rpm	Revolutions per minute
RSI	Relative Signal Intensity
RU	Response Units
S	Sedimentation coefficient
SAXS	Small-angle X-ray scattering
SDS	Sodium dodecyl sulfate
SecAts	Thermal sensitive SecA
SPR	Surface plasmon resonance
SRP	Signal recognition particle



<i>T. maritima</i>	<i>Thermotoga maritima</i>
<i>T. thermophilus</i>	<i>Thermus thermophilus</i>
TA	toxin-antitoxin
TAT	Twin arginine translocase
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
Tris	Tris(hydroxymethyl)aminomethane
UWO	University of Western Ontario
V-bar	Partial specific volume
ZBD	Zinc binding domain
β-Ala	Beta-alanine

## Chapter 1

### 1 Introduction

#### 1.1 General introduction

Biological membranes compartmentalize any living cell whether it belongs to the prokaryotic, archaeal, or eukaryotic domains of life. These membranes ensure that the composition and concentration of cytoplasmic molecules are maintained and they also provide a controlled exchange between cellular compartments and the outside environment. As lipid membranes are impermeable barriers to larger molecules, at some stage during evolution a set of mechanisms evolved to perform the task of translocation of larger molecules, such as proteins, through biological membranes. It is estimated that approximately 25-30% of the cellular proteins must be transported across or into the membranes (Kusters & Driessen, 2011; Pohlschröder *et al*, 2005), and as a result, translocation of proteins into and across the biological membranes is required for cell viability. Secreted proteins are essential for nutrient uptake, signaling, and membrane biogenesis (Yuan *et al*, 2010), and different organisms have devised a range of molecular machines for transport of these proteins. One of the most highly conserved translocation systems across all living organisms is the Sec (secretion) system, which is believed to be one of the first mechanisms of translocation that evolved (Pohlschröder *et al*, 2005). The secretory proteins are translocated in an unfolded state either co- or post-translationally through the Sec system. In the co-translational translocation, which is the predominant form of translocation in eukaryotes, the energy for movement of proteins through the membrane is provided by the ribosome (Park & Rapoport, 2012). However, the bulk transport of proteins across or into the inner membrane in prokaryotic systems happens post-translationally, and the ATPase SecA provides the driving force for the movement of secretory proteins through the membrane (Pohlschröder *et al*, 2005; Vrontou & Economou, 2004) (Figure 1-1).

The post-translational translocation system in *Escherichia coli* is composed of four main components: secretory proteins that usually carry an amino terminal signal sequence (Park & Rapoport, 2012), a membrane channel called SecYEG (Breyton *et al*, 2002; Van den Berg *et al*, 2004; Zimmer *et al*, 2008), the peripheral ATPase SecA (Kusters & Driessen, 2011), and a

chaperone called SecB (Bechtluft *et al*, 2010). Of these, the Sec channel and SecA are highly conserved. The functional role of these components and the steps involved in the protein translocation by the Sec system has been extensively studied. Briefly, it is known that after synthesis of the secretory proteins at the ribosome, SecB binds to these proteins, keeps them unfolded, and delivers them to the SecYEG-bound SecA located at the inner membrane (Vrontou & Economou, 2004). SecA, the motor protein of the Sec system, then drives the translocation of extended protein substrates into the SecYEG channel and across the cytoplasmic membrane (Kusters & Driessen, 2011). It is believed that SecA moves the unfolded polypeptide chain into the SecYEG channel in a stepwise manner by undergoing a series of conformational changes through cycles of ATP binding and hydrolysis (van der Wolk *et al*, 1997; Economou & Wickner, 1994a; Yuan *et al*, 2010). Cycles of ATP hydrolysis by SecA are repeated with a constant rate until the entire length of polypeptide has moved through the SecYEG channel (Uchida *et al*, 1995; Tomkiewicz *et al*, 2006).

In the current chapter, major concepts pertaining to the roles of components of bacterial Sec system in polypeptide substrate recognition and translocation will be presented. First, a brief overview of different bacterial secretion systems, especially regarding the systems involved in protein transport into or across the inner membrane, is provided. Next, the main components of the *E. coli* Sec system, the process involved in the transport of proteins by this system, and the key roles of SecA in these processes will be explained in greater detail. Finally, the mycobacterial Sec system will be introduced, which, as an alternative system, can provide invaluable insights into the modes of action of the Sec system.

## 1.2 Bacterial protein translocation

Proteins, which are usually synthesized in the cytoplasm, must be allocated to their intracellular or extracellular destinations, and it is estimated that around 25-30% of all proteins in a cell will at least cross through one biological membrane (Kusters & Driessen, 2011; Pohlschröder *et al*, 2005). Therefore, the presence of the systems allowing the transport of proteins through the membranes is essential for the origin of all living cells. Living organisms have devised many mechanisms of protein transport across and into their membranes, and with the increasing

accumulation of genomic data, knowledge of the complexity of these secretion systems is expanding rapidly (Tseng *et al*, 2009; Holland, 2004; Thanassi & Hultgren, 2000). For example, at least 16 transport systems have been discovered that prokaryotic organisms employ for the traffic of proteins to extracytoplasmic sites, each of which shows considerable diversity (Pohlschröder *et al*, 2005; Sardis & Economou, 2010). On one hand, most of these discovered systems are not found in all bacteria and many of them are specialized virulence factors in certain bacteria. For instance, some systems transport secreted proteins across the bacterial inner and outer membranes into the extracellular environment in a single step (such as Type 1 secretion system (T1SS) (Delepelaire, 2004; Thanassi & Hultgren, 2000), while other systems, such as T3SS, by forming an apparatus known as “injectisome” transfers proteins across the bacterial and host membranes directly into the cytoplasm of their hosts, where these effector proteins can modulate host functions (Cornelis, 2006; Grant *et al*, 2006). Other systems, such as T4SS, are even more specialized and can deliver nucleic acids in addition to proteins into the host cell (Christie & Cascales, 2005; Christie & Vogel, 2000). On the other hand, there are highly conserved secretion systems, such as the YidC and Sec pathways, that are present in all bacteria (Thanassi & Hultgren, 2000; Holland, 2004), and are mainly involved in the biogenesis of bacterial inner- and outer-membranes and for transport of periplasmic proteins. Interestingly, homologues of these systems are present in different domains of life; and hence it is believed that they are the first translocation systems that were evolved (Pohlschröder *et al*, 2005). These essential bacterial protein secretion machineries such as the Sec system, which is the topic of this text, will be described in the following sections.

### 1.3 Protein transport into and across the inner-membrane

In a typical bacterial cell, such as the Gram-negative bacterium *E. coli*, which contains four subcellular compartments (cytoplasm, inner membrane, periplasm, and outer membrane), many of the proteins that are synthesized in the cytoplasm must be actively transported to the extracytoplasmic compartments. All proteins that function outside of the cytoplasm either need to be inserted into the inner membrane or require crossing it in order to reach their final destination. For catalyzing these reactions, *E. coli* employs secretion systems such as the YidC and Sec systems as well as the TAT pathway (Keller *et al*, 2012; Fröderberg *et al*, 2004; Yuan *et*

*al*, 2010).

The YidC pathway is the simplest protein translocation pathway; it requires only a single conserved membrane protein, YidC (Pohlschröder *et al*, 2005), and is responsible for biogenesis of a subset of membrane proteins such as cytochrome *o* oxidase and F<sub>1</sub>F<sub>0</sub> ATP synthase (van der Laan *et al*, 2003; 2004). Although the molecular details of membrane insertion by YidC are not clear yet, it is believed that the YidC protein alone is sufficient for the insertion of its substrates by interaction with hydrophobic regions of proteins and stimulating their membrane insertion (Pohlschröder *et al*, 2005). There is also evidence that this protein can function cooperatively with the Sec system for insertion of certain membrane proteins (Fröderberg *et al*, 2004).

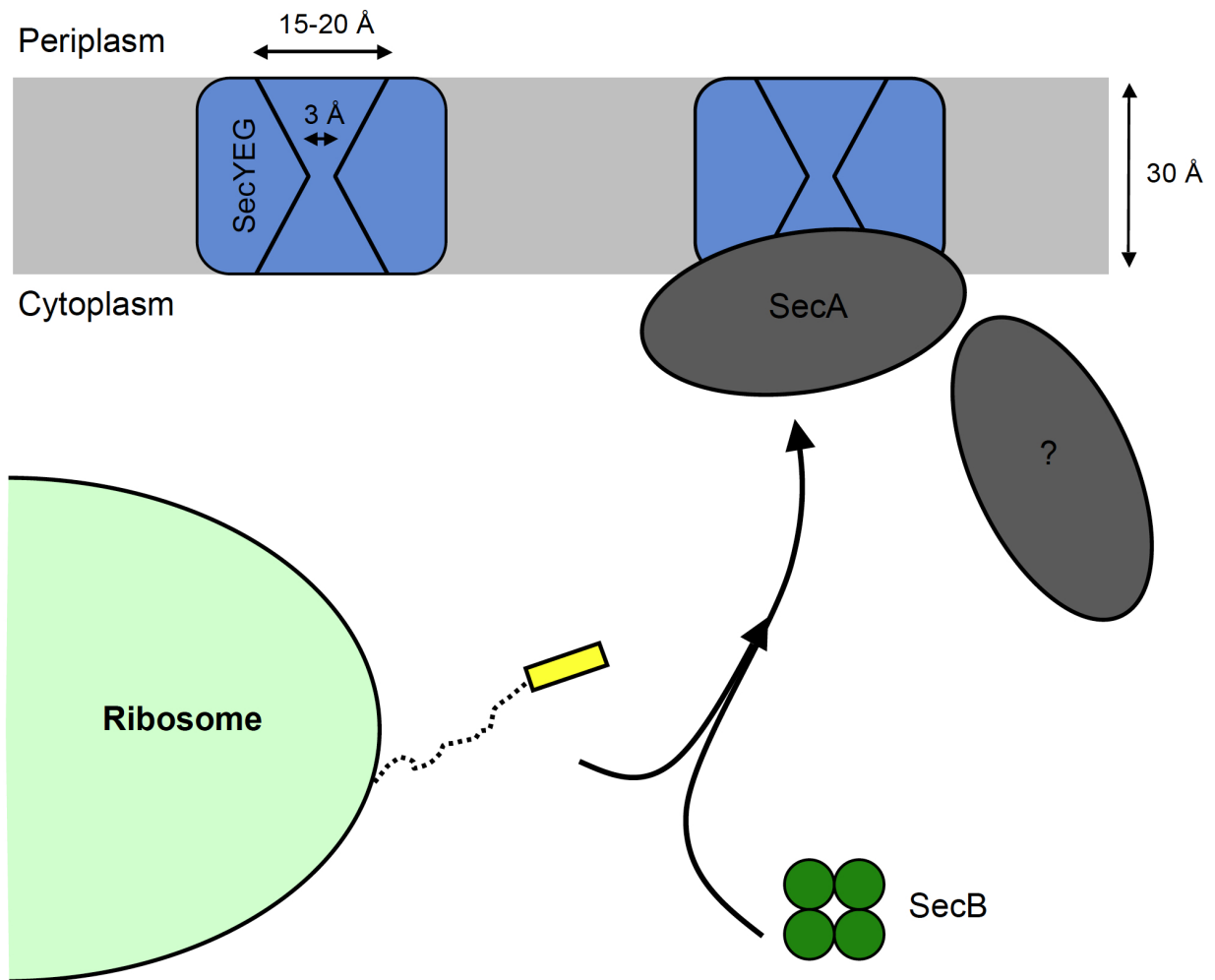
The second pathway, the Sec translocation system (also known as the General Secretion System (GSS) or “the Sec system”) is the only known universally conserved protein transport pathway (Pohlschröder *et al*, 2005); SecY and SecE components of the Sec pore complex are present in all living organisms (Pohlschröder *et al*, 2005; Hartmann *et al*, 1994). This system is responsible for the bulk transport of bacterial proteins (Pohlschröder *et al*, 2005; Vrontou & Economou, 2004). The central part of the Sec system is an inner-membrane pore-forming protein complex, SecYEG, where the secreted proteins are transported through or into the membrane. Proteins that are secreted by the Sec system usually carry an N-terminal and cleavable signal sequence, the most significant feature of which is a short segment of hydrophobic residues (7 to 12 amino acids). The most important feature of protein translocation through the Sec system is that this system translocates proteins in an “unfolded state” into or across the membrane (Kusters & Driessen, 2011; Park & Rapoport, 2012) This idea has been reinforced by extensive functional and structural studies as well as emergence of crystal structures of the SecYEG channel, which revealed that the Sec pore is barely large enough (~ 3Å) to accommodate an extended polypeptide chain (Breyton *et al*, 2002; Zimmer *et al*, 2008; Tanaka *et al*, 2015). Therefore, bacteria use a vast network of chaperones for preventing the premature folding of substrates of the Sec systems and keeping them in a translocation competent state (Economou, 1998). However, the SecYEG channel alone is a passive membrane pore and it has to interact with other partners that provide a driving force for protein transport (Park & Rapoport, 2012). Based on this, there are two different modes of transport of proteins through the Sec system; it can be either co-translational or post-translational. In co-translational translocation the main partner is

ribosome, while the driving force for the post-translational transport of the exported proteins is provided by cytosolic ATPase SecA protein (Park & Rapoport, 2012; Yuan *et al*, 2010).

Co-translational translocation is mainly used for the biogenesis of inner-membrane proteins, where the first step is recognition and targeting of proteins to the membrane by signal recognition particle (SRP) complex (Ulbrandt *et al*, 1997). In *E. coli*, SRP complex consists of Ffh and a 4.5S RNA, which binds to the ribosome-bound nascent chain by recognizing the signal sequence emerging from the ribosome before the nascent chain grows too long and reaches a folded conformation, therefore rendering it incompetent of being transported across the membrane (Luirink & Sinning, 2004). The primary sequence and length of the signal peptide are the key factor for recognition of nascent chain by SRP, as it binds specifically to long hydrophobic signal sequences (De Gier *et al*, 1997). Targeting of the ribosome-nascent chain-SRP complex to the membrane pore is achieved by recognition of this complex by the membrane-bound SRP receptor. In *E. coli*, this receptor is called FtsY, which interacts physically with SecYEG (Luirink & Sinning, 2004) and thus targets the nascent chain to the membrane channel. Thereafter, interaction of ribosome with the SecYEG stimulates the release of the SRP from the ribosome-bound nascent chain complex. Upon releasing from the SRP, the nascent chain is inserted into the membrane pore, and protein translocation by the ribosome through the SecYEG channel can then begin (Valent *et al*, 1998; Prinz *et al*, 2000). In *E. coli*, all members of this system including, Ffh, 4.5S RNA, and FtsY are essential (Luirink & Sinning, 2004).

On the other hand, however, the export of bacterial proteins through the Sec system, predominantly occurs with post-translational mechanism (Pohlschröder *et al*, 2005; Vrontou & Economou, 2004) (Figure 1-1). This pathway is mainly involved in the translocation of proteins that carry only moderately hydrophobic signal sequences or their signal sequences are too short to be effectively recognized by the SRP complex (Park & Rapoport, 2012). As a result, in this type of translocation, secretory proteins are entirely synthesized before entering into the translocation system, and, in *E. coli*, the chaperonic SecB is responsible for their targeting to the membrane. SecB binds to the secretory proteins in the cytoplasm and then targets them to SecA, localized at the Sec translocon at the membrane. Then, SecA by using the energy derived from the hydrolysis of ATP moves the proteins through the SecYEG channel. Although the SecYEG pore forming complex, SecA, and SecB are the main components of this system, some other

accessory proteins are also involved in this system. The details of the various components and the steps involved in post-translational translocation of proteins are described in the next sections.



**Figure 1-1. Schematic model of post-translational translocation by the Sec system.**

Bacterial protein translocation through Sec system occurs predominantly post-translationally. The main components of the bacterial post-translational translocation system are shown here. The pore forming SecYEG complex, the motor ATPase SecA, and SecB chaperon, are shown in blue, gray, and dark green, respectively. A secretory protein (preprotein) is shown with dashed line, and the yellow box represents its N-terminal signal sequence.



In addition to the above-mentioned translocation systems, bacteria have evolved an alternative secretory system, called Twin-arginine translocation pathway (briefly known as “TAT pathway”), that functions very differently from the YidC and Sec systems (Palmer & Berks, 2012). Unlike the Sec system, which translocates proteins in a mainly unfolded fashion, TAT pathway’s substrates must reach their tertiary structure before translocation. It is believed that bacteria have evolved secretion systems such as TAT pathway, simply because some secreted proteins may require the cytoplasmic factors (e.g. special chaperones, cytoplasmic co-factors, etc) for reaching their native folding, or they may fold almost instantly upon translation and exiting the ribosome tunnel (Pohlschröder *et al*, 2005; Palmer & Berks, 2012). Therefore, these proteins are incompetent for the translocation by the Sec system. In contrast to the Sec system, the TAT pathway is not crucial for viability of most bacteria (Pohlschröder *et al*, 2005; Yuan *et al*, 2010). Overall, since the bulk transport of bacterial proteins happens through the Sec system and the core members of this system are highly conserved, this transport system has been extensively studied, and it has remained an appealing target for development of novel antibacterial drugs.

## 1.4 Components of the bacterial Sec system

In *E. coli*, the Sec translocase machinery is composed of several proteins, including the SecYEG pore forming complex, the SecB chaperone, the SecA ATPase, and the accessory components heterotrimeric SecDFYajC complex and YidC. The majority of genes encoding these proteins were discovered in a set of elegant genetic experiments in early 80s. For example, in the first experiment, Oliver and Beckwith used a MBP-LacZ fusion protein for screening the mutations that impaired the protein secretion (Oliver & Beckwith, 1981). Mutations that hampered the protein secretion, would therefore lead to the cytoplasmic accumulation of MBP-LacZ fusion protein and increase the  $\beta$ -galactosidase activity. Subsequent analyses of the mutants that had higher  $\beta$ -galactosidase resulted in identification of several Sec genes (*secA*, *secB*, etc.) (Kumamoto & Beckwith, 1983; Oliver & Beckwith, 1981). The second approach looked for suppressor mutations that restore secretion of a protein with an impaired signal sequence (Emr *et al*, 1981). By this approach, which seemingly broadens the specificity of the Sec machinery to permit recognition and transport of precursor proteins with impaired signal sequences, several

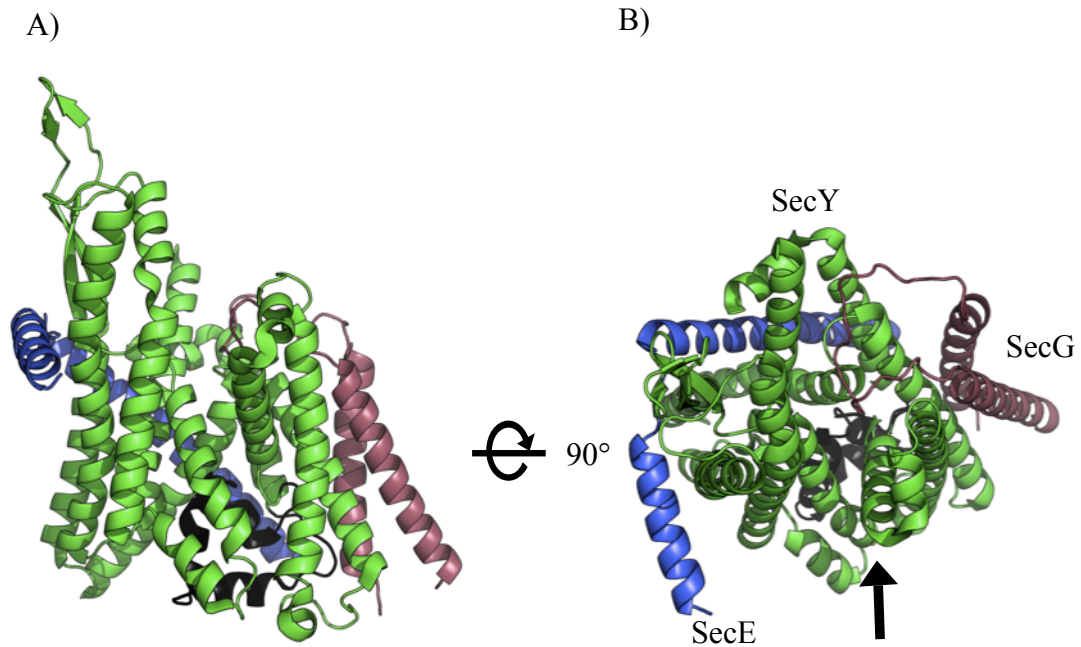
*prl* (protein localization) genes were identified that restored the translocation of the signal-less proteins including, *prlA*, *prlD*, *prlG*, and *prlH* (Huie & Silhavy, 1995). Subsequent complementation studies showed these genes are allelic with, *secY*, *secA*, *secE*, and *secG* genes, respectively (Huie & Silhavy, 1995; Stader *et al*, 1989; Riggs *et al*, 1988; Ito *et al*, 1983). Of these genes, *secA*, *secY* and *secE* are essential for survival of *E. coli*. Further biochemical analysis confirmed the critical role(s) of these three proteins in the translocation reaction; indeed, *in vitro* protein translocation experiments confirmed that SecA, SecE, SecY, and ATP are necessary and sufficient for the successful reconstitution of translocation of a model precursor protein into membrane vesicles (Brundage *et al*, 1990; Akimaru *et al*, 1991). Unlike these proteins, SecB is not essential for cellular growth of *E. coli*, and mutations in *secB* gene only abolished the transport of a subset of periplasmic proteins (Kumamoto & Beckwith, 1983). Deletion of the *secB* gene slows down or blocks the translocation of precursor of maltose binding protein (preMBP) and lead to the cytoplasmic accumulation of folded MBP, which is incompetent for translocation by the Sec system (Gannon & Kumamoto, 1993). Therefore, it has been suggested that SecB is responsible for stabilizing the secretory proteins in a translocation competent state. Other components, SecDFYajC complex, were identified as part of the Sec system in *E. coli* through the co-purification studies with SecYEG (Duong & Wickner, 1997; Schulze *et al*, 2014; Pohlschröder *et al*, 2005); the components of this complex can improve the translocation efficiency but they are not absolutely essential (Yuan *et al*, 2010; Pogliano & Beckwith, 1994). Although the exact role of these proteins in translocation is not yet clear, it is suggested that they may stabilize the SecYEG complex or facilitate the release of substrates from the SecYEG channel at the end of translocation process (Kato *et al*, 2003; Pohlschröder *et al*, 2005). Through other co-purifications studies with SecYEG, YidC protein was also identified as a partner of the Sec system (Luirink *et al*, 2005). This protein can function cooperatively with the SecYEG channel to promote the membrane insertion of certain Sec-dependent substrates (Scotti *et al*, 2000), but it typically does not play a significant role in the post-translational translocation of proteins in *E. coli* (Yuan *et al*, 2010). The main components of bacterial post-translational secretion and their roles in the translocation process are described in the following sections.

## 1.5 SecYEG; the bacterial membrane pore forming complex

One of the most important components of the Sec system is the pore complex, which is highly conserved across domains of life (Pohlschröder *et al*, 2005). In bacteria, the main two core components include SecY and SecE proteins, which together with a third protein (SecG) form the heterotrimeric complex of the Sec pore (commonly known as SecYEG). Although *E. coli* SecYEG is one of the most investigated Sec translocons, its high-resolution structure has not yet been determined. In addition, there is an ongoing debate regarding the functional *in vivo* oligomeric state of the SecYEG (Dalal *et al*, 2012; Mao *et al*, 2013; Duong, 2014). However, several crystal structures of SecYEG from various organisms in recent years have provided substantial information about the structure and function of the Sec system (Tanaka *et al*, 2015; Breyton *et al*, 2002; Van den Berg *et al*, 2004; Zimmer *et al*, 2008). These structures have revealed that the SecYEG channel is a 1:1:1 heterotrimer. SecY spans the membrane 10 times and forms the protein-conducting channel that is stabilized by SecE protein, placed at the back of SecY, and also the transmembrane (TM) segments of SecG tightly bound to SecY through hydrophobic interactions (Figure 1-2). This arrangement results in a relatively narrow pore across which an extended polypeptide can move. Furthermore, TM2 and TM7 segments of SecY form a lateral gate, where during the translocation and membrane insertion of membrane proteins the TM segments leave the channel and insert into the lipid phase.

The available structures have showed that only the SecY protein forms the pore structure through which the polypeptide chain has been speculated to be translocated (Park & Rapoport, 2012). In fact, the cross-linking studies demonstrated that the polypeptide substrate interacts with only a single copy of SecY molecule and moves through the center of the pore formed by SecY (Cannon *et al*, 2005), consistent with the observed crystal structure of SecYEG complex (Breyton *et al*, 2002; Van den Berg *et al*, 2004; Zimmer *et al*, 2008). Therefore, during the translocation process, the preprotein substrates likely move from the cytoplasmic channel, in the pore ring, and into the periplasmic channel (Breyton *et al*, 2002; Van den Berg *et al*, 2004; Zimmer *et al*, 2008).

Similar to any membrane channel, SecYEG has to be tightly regulated to maintain the membrane barrier for small molecules, such as metabolites or ions, and can open only to allow the passage of the polypeptide substrates. SecYEG is responsible for this task itself.



**Figure 1-2. Structure of bacterial SecYEG pore forming complex.**

The crystal structure of *Thermus thermophilus* SecYEG (PDB: 5AWW). **Panel A** gives a view from the side of SecYEG channel, and **panel B** is a view from the cytoplasmic side of channel. SecY is colored green, SecE is colored in blue, and SecG is red. The plug domain of SecY is in black. The lateral gate of SecY for integral membrane protein integration is indicated with an arrow.

The interior structure of the SecYEG complex has an hourglass shape with a constriction of hydrophobic residues about halfway across the pore. It has been proposed that this constriction, which forms a hydrophobic “pore ring” with a diameter of  $\sim 3\text{\AA}$ , makes a seal around the translocating polypeptide chain to prevent the nonspecific passage of molecules through the pore. While in the resting state, the SecYEG channel is sealed by the pore ring residues as well as the SecY plug domain, which is located at the periplasmic side of the channel and can block the membrane pore quite effectively (Figure 1-2). Cross-linking studies have shown that the plug domain can move away, therefore opening the SecYEG pore in which would permit the extended polypeptide to pass through (Harris & Silhavy, 1999).

It has been suggested that the opening of the pore happens in two stages (Park & Rapoport, 2012). In the first step, the binding of SecYEG partner (ribosome or SecA) would lead to displacement and partially opening of the plug domain and the lateral gate of the channel. The evidence for this comes from the crystal structure of SecYEG in complex with SecA in which the lateral gate is partially opened and the plug is displaced (Park & Rapoport, 2012; Zimmer *et al*, 2008) (Figure 1-2). The second step involves the insertion of the hydrophobic part of a signal sequence into the partially opened lateral gate, as it had been shown that the signal peptide bound to the translocating polypeptides can be photocross-linked to the lateral gate region as well as phospholipids (Plath *et al*, 1998). This, in turn, would further destabilize the plug interactions and move the plug domain out of the way, allowing the rest of the preprotein to move through the pore. Consistent with the role of signal sequences in the opening of SecYEG channel, many mutations that allow the secretion of proteins with defective or lacking signal sequences (called *prl* mutants) lead to relaxation and destabilization of the resting state of the translocation apparatus (Trueman *et al*, 2011; Park & Rapoport, 2012; Derman *et al*, 1993). The plug domain can seal the SecYEG again only when the polypeptide chain has left the channel (Park & Rapoport, 2012).

## 1.6 Preproteins; substrates of the Sec system

Secretory proteins that are targeted to the Sec transport system are called “preproteins”, for example the precursor for the maltose-binding protein (MBP) is called preMBP. They usually

carry a cleavable N-terminal signal sequence, and the main body of preproteins, which is translocated through the membrane, is known as the “mature region”.

Sec signal sequences are variable and lack primary structural homology. However, they usually share three common features; they (1) include a charged N-terminal region (usually are polar and positively charged residues), (2) They have a hydrophobic core (~10 amino acids long), and (3) have less hydrophobic but usually uncharged (around 6–9 amino acid) region that carries the signal peptide cleavage site (Gierasch, 1989). The cleavage site is usually located in position -3 to -1 of the mature preprotein, and an alanine residue is most abundant at the -1 and -3 positions (therefore, it is commonly called Ala-X-Ala site) (Choi & Lee, 2004). Although the exact role(s) of signal sequences is not clear yet, it is believed that signal sequences contain the information that dictates which type of translocation will be used for preprotein export (Grady *et al*, 2012); preproteins with a more hydrophobic signal sequence will be transported co-translationally, as SRP binds specifically to long hydrophobic signal sequences (Clérico *et al*, 2008; Park & Rapoport, 2012). There is also evidence that the signal sequence may play a role in the opening of SecYEG pore during the translocation reaction (see previous section). Furthermore, in terms of a post-translational translocation system, it has been suggested that the signal sequence may affect the folding rate of secreted proteins and actually slows down the folding of preproteins upon emerging from the ribosome (Clérico *et al*, 2008; Hardy & Randall, 1991). This would allow the SecB chaperone to bind to these proteins and direct them to the translocon.

However, the mere presence of the signal sequence at the N-terminal of a protein does not guaranty its successful transport through the Sec pathway (Lee & Bernstein, 2002; Moreno *et al*, 1980). It has been suggested that the mature region of different preproteins, which presumably lack any sequence homology, contains information that is used by the members of the Sec system for the transport of preproteins (Danese & Silhavy, 1998). In support of this idea, when the coding sequence of the signal sequence of some preproteins (such as periplasmic PhoA enzyme and the outer membrane protein LamB) are completely removed, there is still a low level of transport (Danese & Silhavy, 1998). Moreover, the translocation of these signal-less proteins is highly increased in *prl* mutant strains. The *prl* mutations, which are mapped to various components of the Sec machinery, especially SecY, restore the transport of preproteins with a deficiency in their signal sequence (Danese & Silhavy, 1998; Emr *et al*, 1981; Oliver &

Beckwith, 1981). In addition, it has been shown that the net charge distribution in a region of the mature proteins adjoining to the signal sequence, especially the first 18 residues of the mature protein, directly affects the translocation of protein across the cytoplasmic membrane by the Sec system (Kajava *et al*, 2000). Overall, the mature region of preproteins may contain information that assists their recognition and transport by the Sec system through the bacterial inner-membrane.

## 1.7 SecA ATPase; the driving force for translocation

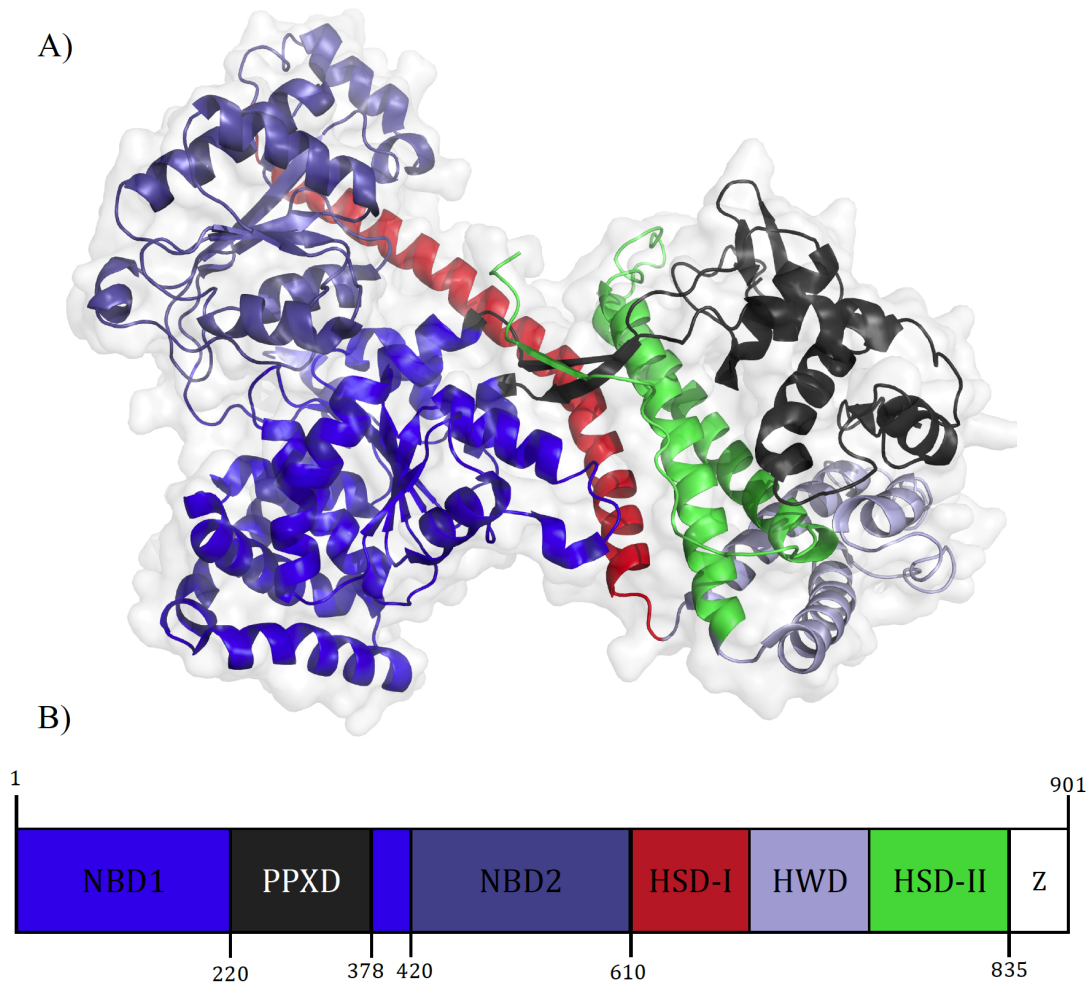
SecA provides the driving force for the export of proteins in the bacterial Sec system by coupling ATP hydrolysis to the preprotein translocation (Lill *et al*, 1989; Oliver & Beckwith, 1982). This protein is believed to be involved in the co-translational translocation of large periplasmic hydrophilic domains of inner membrane proteins; however, its main role is in stimulating post-translational protein transport. SecA is highly conserved and is essential for the survival of the all known bacteria (Kusters & Driessen, 2011; Vrontou & Economou, 2004), and was identified for the first time through genetic analysis (*prlD*) as a main component of the Sec system. Conditional-lethal mutants of SecA in *E. coli* accumulated precursors of several periplasmic proteins (including MBP, OmpA, PhoA, OmpF and LamB) in the cytoplasm (Oliver & Beckwith, 1981). *In vitro* translocation reactions also demonstrated that SecA is essential for the transport of preproteins of alkaline phosphatase and OmpA into membrane vesicles (Cabelli *et al*, 1988). Wild-type SecA alone in solution shows a very low level of ATPase activity, which likely does not have any physiological relevance (Price *et al*, 1996). However, while bound to the SecYEG, and in the presence of preproteins, SecA undergo cycles of ATP hydrolysis (Economou *et al*, 1995), where each ATP hydrolysis cycle results in the transport of approximately 5 kDa of preprotein polypeptide through the membrane (Schiebel *et al*, 1991; van der Wolk *et al*, 1997).

*E. coli* SecA is a 204 kDa homodimer with each protomer consisting of 901 residues. SecA contains several distinct domains, including two nucleotide-binding domains (NBDs; NBD1 and NBD2), a preprotein cross-linking domain (PPXD), C-terminal domains (CTDs), and a zinc-binding domain (ZBD) (Figure 1-3).

The SecA NBDs are homologous to Superfamily 2 (SF2) RNA helicases (Linder, 2006; Hunt *et al*, 2002). While SecA binds polypeptides, SF2 helicases interact with a different biological polymer, RNA, and their main biochemical function is the ATP-dependent directional translocation on single or double stranded nucleic acids (Singleton *et al*, 2007). SecA and the SF2 helicases use “Walker box” motifs, which are a set of conserved residues, for nucleotide binding. The SecA NBD1 carries the Walker A motif, which binds a single molecule of ATP. NBD2 contains the Walker B motif, which binds the cofactor for ATP binding,  $Mg^{2+}$ . (Papanikolau *et al*, 2007a; Hunt *et al*, 2002).

The PPXD domain is comprised of 150 residues and is internally fused to NBD1 (Figure 1-3). This domain was originally called the “preprotein cross-linking domain” (PPXD) because it was identified through cross-linking studies as the region of SecA where preprotein binds (Kimura *et al*, 1991). The PPXD is composed of a “bulb” region (residues 233-362), which is the main part of PPXD, and two highly conserved anti-parallel  $\beta$ -strands called “stem” region (residues 221-233 and 362-377), which connect the PPXD to NBD1 (Papanikou *et al*, 2005). Crystal structure of SecA in complex with SecYEG shows that the PPXD is responsible for the most critical interactions between SecA and the 6–7 and 8–9 loops of SecY (Zimmer *et al*, 2008). Indeed, SecA constructs lacking the PPXD are not able to form stable complexes with SecY (Zimmer *et al*, 2008), and mutations in these loops alter translocation by SecA (Mori & Ito, 2001). In addition, crystal structures of SecA from different bacteria revealed that PPXD is highly flexible; it is either packed against the HWD (helical-wing domain) domain or rotates away from it (Hunt *et al*, 2002; Osborne *et al*, 2004). The most significant change was observed in the crystal structure of SecA-bound SecYEG complex (Zimmer *et al*, 2008), where the PPXD is rotated even more, with a large rigid-body rotation by  $\sim 80^\circ$ , enabling contact with NBD2 (Figure 1-3). Based on this observation, it is suggested that SecA could capture the polypeptide substrate in a ‘clamp’ formed between NBD1 and PPXD. The back of this clamp is formed by the stem regions of the PPXD.



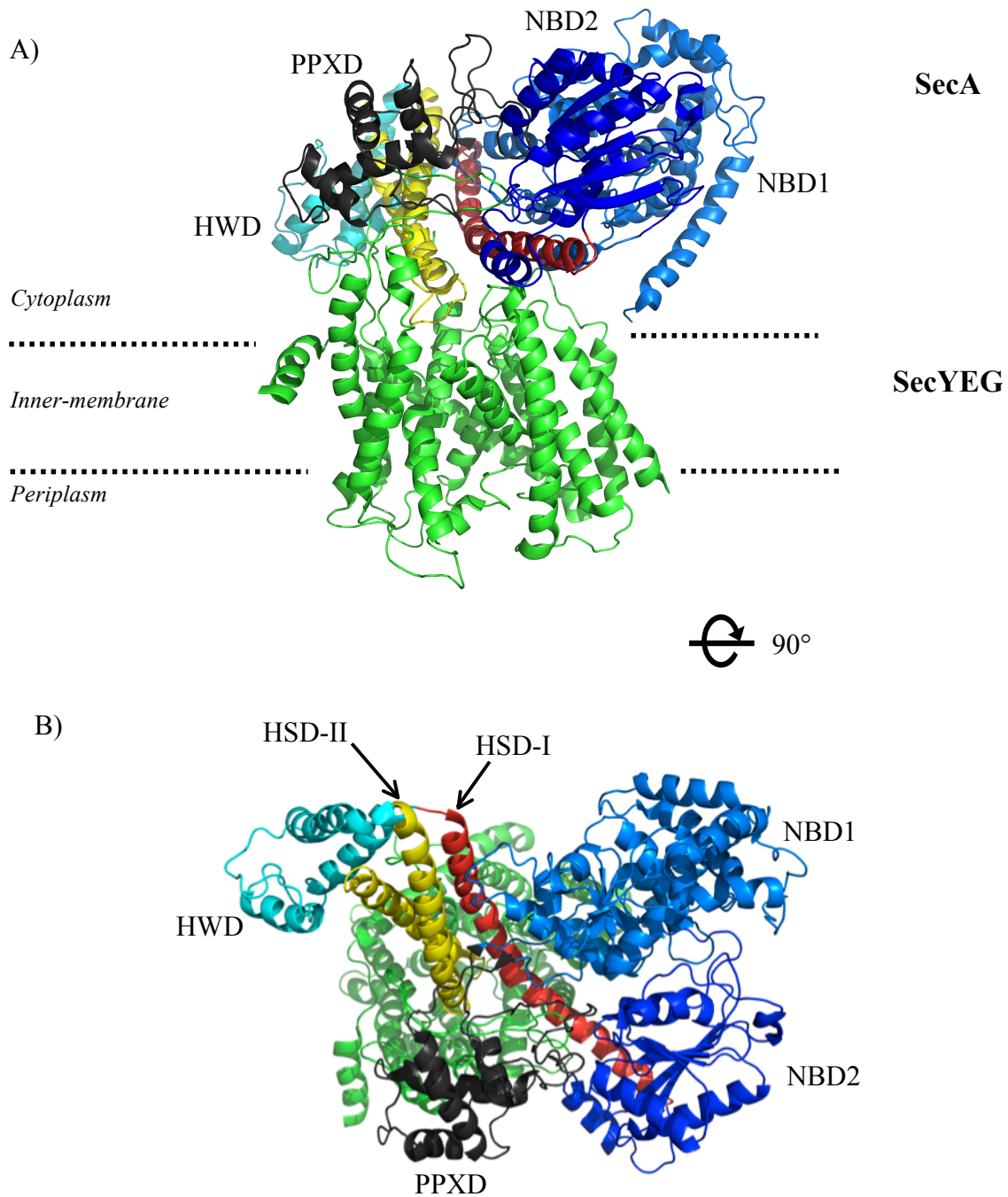


**Figure 1-3. Structure and domain arrangement of bacterial ATPase SecA**

**Panel A** shows the first atomic structure of SecA, which belongs to *B. subtilis* (PDB: 1M6N). In this crystal structure two protomers could be seen, but for better observation of domain arrangement in SecA, only one protomer is shown. The figure shows the nucleotide binding domains (NBD1 and NBD2) in blue, the preprotein cross-linking domain (PPXD) in black, helix scaffold domain 1 (HSD-I) in red, helical-wing domain (HWD) in light blue, and the helix scaffold domain 2 (HSD-II) in green. **Panel B** represents the corresponding domains on *E. coli* SecA, which is consisted of 901 amino acids. Both SecAs from *E. coli* and *B. subtilis* has a zinc-binding domain (ZBD; shown with Z) at their C-terminus that was not resolved in the structure of *B. subtilis* (panel A).

The CTDs of SecA consist of three distinct domains, the “helix scaffold domain I” (HSD-I), “helical wing domain” (HWD), and HSD-II (Figure 1-3). These domains have been implicated in the inter- and intra-molecular interactions between SecA protomers and other components of the Sec system (Karamanou *et al*, 1999; Zimmer *et al*, 2008; Hunt *et al*, 2002). The HSD-I, which is a long helix that runs the length of SecA, provides a scaffold upon which all the other domains of SecA are mounted (Figure 1-3), and the crystal structure of SecA bound to SecYEG showed that this domain is placed at the interface of SecA and SecYEG (Figure 1-4) (Park & Rapoport, 2012). This structure also showed that another domain, HSD-II, which is also known as the “two-helix finger” domain is deeply inserted into the SecYEG funnel (Zimmer *et al*, 2008) (Figure 1-4). Subsequent cross-linking experiments revealed that the translocating polypeptide comes close to the tip of HSD-II exactly above the entrance of SecYEG channel (Erlandson *et al*, 2008).

The zinc-binding domain (ZBD) is placed at the extreme C-terminus of SecA. In *E. coli*, it is very short (only 22 amino acids), highly conserved, and is attached to the main body of SecA by a long unstructured polypeptide linker composed of 59 residues (Dempsey *et al*, 2004; Bechtluft *et al*, 2010; Vrontou & Economou, 2004). The ZBD is the only domain that is not essential for the function of SecA. The structure of the ZBD is stabilized through binding a  $Zn^{2+}$  ion (Fekkes *et al*, 1998). The ZBD structure is not resolved in any of the available crystal structures of SecA; indeed, unpublished experiments suggested that the presence of this domain inhibits the growth of SecA crystals (Dr. B. Shilton, personal communication). This domain is mainly responsible for targeting of preproteins to the Sec translocon through interaction with SecB, which will be described later.



**Figure 1-4. Crystal structure of SecA in complex with SecYEG**

Crystal structure of *T. maritima* SecA -SecYEG complex is shown here (PDB: 3DIN). **Panel A** shows the side view of the complex. **Panel B** shows the complex as in panel A, but viewed from the cytoplasm.

## 1.8 Interaction of SecA with preproteins

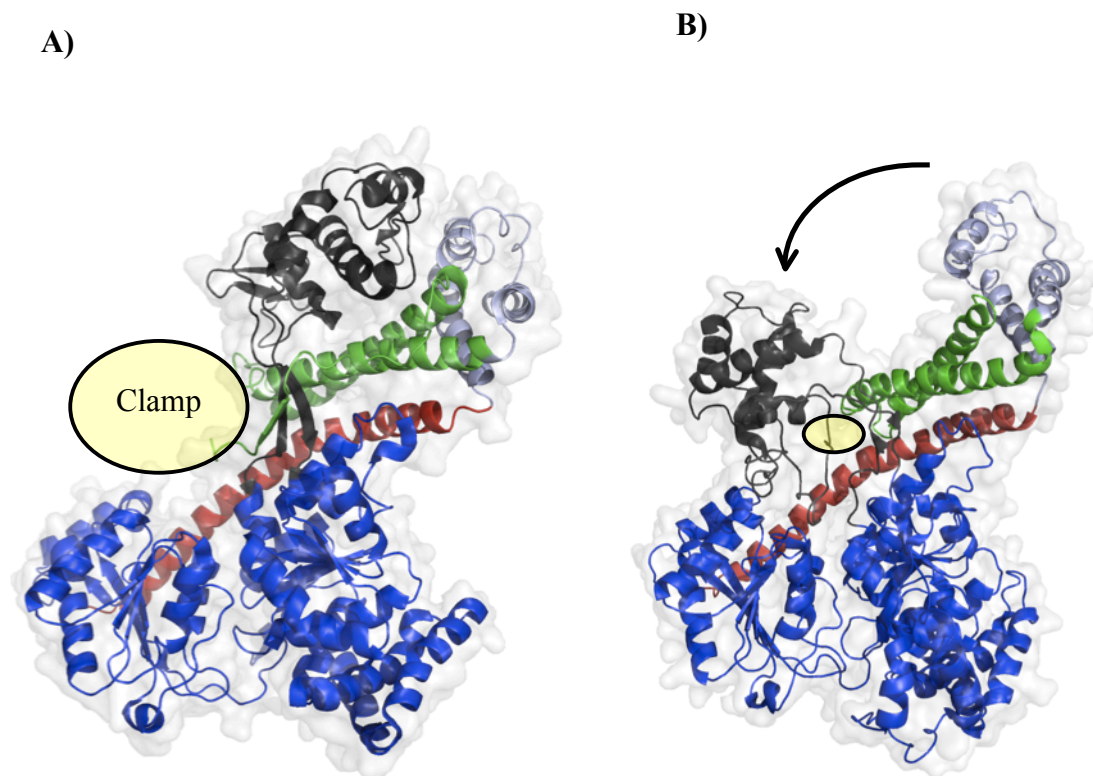
The minimal known biochemical function of SecA is perhaps the ATP dependent movement of unfolded preproteins through the SecYEG pore. The translocation process must inevitably involve binding and release of the preprotein substrates, and on this basis SecA is expected to possess sites that are capable of interacting with unfolded preproteins. However, despite many biophysical and biochemical investigations, the sites of interaction of SecA with the signal sequence as well as the mature regions of preprotein substrates are not clear yet.

It is believed that the cytoplasmic form of SecA can bind and recognize the signal sequence of preproteins to form an early complex with the preprotein substrate, and this initial complex must interact with SecYEG and deliver the signal sequence to the translocase (Clérico *et al*, 2008). Nevertheless, there is not a consensus on the exact location of the “signal sequence binding site” on SecA (Clérico *et al*, 2008). Based on the first crystal structure of SecA, it was proposed that a site located between CTD and the NBDs is the potential signal sequence binding site (Hunt *et al*, 2002). In contrast, based on another crystal structure of SecA, which had gone under major conformational changes and was crystallized as a monomer, it was proposed that signal sequences will bind SecA at the cleft between CTD and PPXD domain of SecA (Osborne *et al*, 2004). In addition, by using various truncated constructs of SecA and chemical cross-linking of signal peptides, a short beta strand region near the NBD1 at the base of the PPXD (called “stem region”) was proposed to be the signal peptide binding site of SecA (Papanikou *et al*, 2005; Baud *et al*, 2002). Later, by employing a synthetic signal peptide containing a photoreactive cross-linker, it was proposed that signal binding site lies between residues 269 to 322 of *E. coli* SecA, which is part of the PPXD domain (Musial-Siwiek *et al*, 2007). Overall, taking all these results together, it was suggested that the PPXD and/or central regions of SecA surrounding this domain are responsible for interaction with signal sequences (Clérico *et al*, 2008).

In addition to the signal sequence, SecA is required to interact with the mature region of preproteins. The evidence in support of this idea are as follows; first, substrates lacking a signal sequence can still be transported through the Sec system in the presence of the supportive *prl* mutations (Danese & Silhavy, 1998; Emr *et al*, 1981; Oliver & Beckwith, 1981). Second, simply adding a signal sequence to a non-secretory protein does not guarantee its translocation. Third,

SecA transports preprotein substrates through the membrane pore in a stepwise manner, and it is estimated that approximately 5 kDa of polypeptide is translocated for each round of ATP hydrolysis by SecA (Kusters & Driessen, 2011; van der Wolk *et al*, 1997). As a result, one can envision a mechanism in which SecA catalyzes multiple cycles of preprotein binding and release for translocation of substrates through SecYEG.

The nature of interaction of SecA with the mature region of preproteins is less defined, mainly because the unfolded preprotein substrates have a tendency to either refold or aggregate. Early cross-linking experiments that involved a preprotein substrate and a collection of deletion constructs of SecA, and suggested that a region located on NBD1 was the location of cross-links between SecA and a full-length preprotein substrate (Kimura *et al*, 1991). On the basis of these experiments, this domain of SecA was termed the “preprotein cross-linking domain” (PPXD). These results were later confirmed by showing that the PPXD is responsible for interaction of SecA with polypeptides that do not have any resemblance to the signal sequences (Papanikou *et al*, 2005). However, the crystal structure of SecA-SecYEG complex revealed that the PPXD makes major contacts with SecY (Figure 1-4), thus it seems that the PPXD is mainly responsible for interaction of SecA and SecYEG (Zimmer *et al*, 2008). Nonetheless, based on this structure, a groove between the stem region of PPXD, NBD2, and parts of the HSD, which is referred to as the “clamp”, was proposed to be involved in interaction with the unfolded preproteins (Figure 1-5). In complex with SecYEG, SecA had gone under major conformational changes; the PPXD of SecA made a large rigid-body movement and moved towards the NBDs, which was proposed to close the “clamp” for capturing the preprotein (Figures 1-4 and 1-5). In addition, the SecA-SecYEG complex indicated that another region of SecA may also be involved in the interaction with preproteins, as it was shown that the two-helix finger domain of SecA is deeply inserted into the SecYEG funnel (Figure 1-4). However, although further cross-linking experiments showed that the translocating polypeptide comes close to the tip of this domain above the entrance of the SecYEG channel (Erlandson *et al*, 2008), there is no direct evidence that this domain is responsible for recognizing and interacting with the mature region of preproteins. In summary, the exact site(s) on SecA for recognition and interaction with unfolded preprotein substrates has remained an open question. In Chapter 3, an investigation into the mechanism of interaction of SecA with the mature regions of preprotein substrates is presented.



**Figure 1-5. Conformational flexibility of PPXD domain of SecA**

Different position of the PPXD relative to the NBDs is shown here. **Panel A** *B. subtilis* SecA structure (PDB: 1M6N). **Panel B:** *T. maritima* SecA from the crystal structure of SecA-SecYEG complex (PDB: 3DIN). When bound to SecY, the PPXD domain of SecA makes a large rigid-body movement by  $\sim 80^\circ$  (Panel B), which is proposed to close the “clamp” (yellow circles) for capturing the preprotein. This view is from the cytoplasm, as in Figure 1-4.

## 1.9 Oligomerization of SecA

SecA exists as a dimer in solution (Gold *et al*, 2007), and at the cellular concentration, the majority of SecA proteins are expected to be dimeric (Woodbury *et al*, 2002). However, the functional oligomeric state of SecA for translocation has been a matter of considerable debate recently (Sardis & Economou, 2010). In regard to the mechanism of SecA-mediated translocation, a critical question is whether the translocating system incorporates a single SecA protomer, or whether multiple SecA protomers are cooperating during the translocation process. The evidence supporting involvement of a SecA monomer in the translocation process include experiments in which a dimerization defective SecA is still able to catalyze translocation reaction (Or *et al*, 2002; 2005), as well as experiments showing that cross-linked SecA dimers are not functional in translocation reaction (Or & Rapoport, 2007). Moreover, the available crystal structure of a SecA in complex with SecYEG (Figure 1-4) clearly indicates that at any given time only a single SecA is able to interact directly with SecYEG (Zimmer *et al*, 2008; Park & Rapoport, 2012). On the other hand, there is evidence supporting the involvement of SecA dimer in the translocation reaction, which include the experiments in which by alteration of the observed dimeric interfaces it was shown that dimerization of SecA is critical for translocation (Jilaveanu *et al*, 2005; Karamanou *et al*, 2005; Das *et al*, 2008). Furthermore, in most of the available crystal structures from different bacteria the SecA proteins are packed as dimers. However, despite the high degree of conservation of SecA between bacteria, these structures of SecA dimers from different or even the same organism do not contain the same dimer symmetry (Sardis & Economou, 2010). Remarkably, it was shown that the isolated N-terminal domains of SecA (called SecA-N68) participates in a monomer-tetramer equilibrium, but the physiological significance of the SecA-N68 tetramer to the translocation process has never been clear (Dempsey *et al*, 2002). Taken together, these observations have created controversy as to which structure represents the active oligomeric state of SecA. In turn, this has led to the belief that interaction between protomers of SecA may be very dynamic (Kusters & Driessen, 2011; Sardis & Economou, 2010), which allows SecA to utilize alternative dimeric interfaces, particularly during the translocation reaction. In summary, the available evidence does not provide a conclusive answer as to the functional oligomeric state of SecA or mechanisms of self-

association of this molecule. We have further investigated the molecular basis of self-association of SecA, which will be presented in Chapter 2.

## 1.10 SecB; a translocation dedicated chaperone

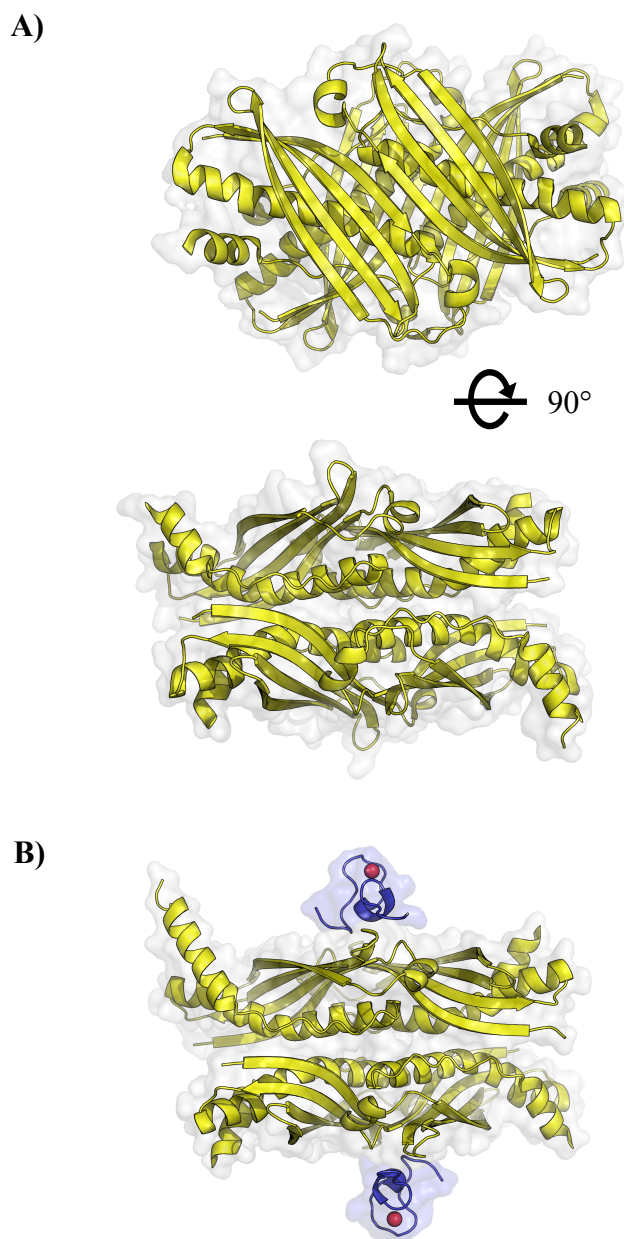
As mentioned before, the Sec system translocates solely unfolded proteins devoid of stable tertiary structures, therefore it ought to have a mechanism for either unfolding the secretory proteins or keeping them in an unfolded state after their synthesis at the ribosomes. However, there is no evidence for an unfolding activity (i.e. holdase activity) during the post-translational translocation of proteins by the Sec system (Randall & Hardy, 2002). Indeed, it has been shown that the Sec system employs a set of chaperones for capturing the preproteins before they reach to their final tertiary structure (Randall & Hardy, 1986). The best-known example of these translocation-dedicated chaperones is SecB, which is suggested to be present only in Proteobacteria (Scott & Barnett, 2006; Driessen, 2001). However, it has been shown that many other prokaryotes have evolved proteins with a similar function. For example, it has been proposed that in Gram-positive bacteria another chaperone called CsaA is involved in the targeting of preproteins to the Sec translocon (Kawaguchi *et al*, 2001; Müller *et al*, 2000). In addition, all mycobacteria contain a protein, which is somehow similar to SecB, and it is called *mtSecB*, which will be described later in more detail.

The main chaperone used for targeting the preproteins to the Sec translocon in *E. coli* is SecB, which is a stable tetramer in solution. Although this protein is not vital for *E. coli* cell survival, deletion of the *secB* gene leads to the cytoplasmic accumulation of some preproteins (Kumamoto & Beckwith, 1983). SecB can slow down the folding rate of preMBP *in vitro* (Collier *et al*, 1988), and SecB-bound polypeptides are devoid of any stable tertiary structure (Hardy & Randall, 1991). The crystal structure of *E. coli* SecB has been solved (Dekker *et al*, 2003), and it revealed that there is a long surface exposed groove, which has been proposed to be the unfolded polypeptide binding site on this chaperone protein (Dekker *et al*, 2003; Xu *et al*, 2000). SecB appears to interact with a variety of proteins provided that they are unfolded; therefore its selectivity for preprotein substrate binding has been unclear (Park & Rapoport, 2012). This selectivity of SecB towards its preprotein substrates can be explained by a “kinetic partitioning



model” (Bechtluft *et al*, 2010; Hardy & Randall, 1991). This model suggests that the presence of signal sequence lowers the folding rate of preproteins, and hence provides a larger time window for SecB to interact with preproteins. However, the exact molecular mechanisms of preprotein binding by SecB are not yet clear.

In addition to keeping the preproteins unfolded and competent for the transport by the Sec system, SecB also has the ability to direct preprotein substrates to SecA at the membrane (Hartl *et al*, 1990). Although SecB has a low affinity to soluble SecA (with a dissociation constant ( $K_D$ ) in micromolar range), the SecYEG-bound SecA complex can bind the SecB-preprotein complex with high affinity ( $K_D$  around 10 nM) (Bechtluft *et al*, 2010). Therefore, it has been suggested that SecB plays a major role in directing the preproteins towards SecA. Moreover, through a SecA truncation experiment, the “SecB-binding site” was mapped to the C-terminal region of SecA (Breukink *et al*, 1995). In addition, an atomic structure of the C-terminal polypeptide of SecA bound to SecB is available (Zhou & Xu, 2003), in which one tetramer of SecB binds to two SecA peptides (Figure 1-6C). However, the extreme N-terminal of SecA (residues 2 to 11) has also been shown to be involved in the interaction with SecB (Randall & Henzl, 2010). After completion of the task of delivering the preproteins to the SecYEG-bound SecA, SecB will need to disassociate. Indeed, it has been shown that after binding of ATP to SecA and initiation of the translocation, SecB is released from SecA and returns to the cytoplasm to salvage other preproteins from premature folding and delivering them to the Sec translocon (Fekkes *et al*, 1997).



**Figure 1-6. Crystal structure of SecB tetramer**

The crystal structure of bacterial SecB alone and in complex with SecA ZBD domain is shown here. **Panel A** shows the top and side views of the *E. coli* SecB (PDB: 1QYN). The crystal structure of *H. influenzae* tetramer with two bound SecA-ZBD (PDB: 1OZB) (**Panel B**). In both panels, the SecB is shown in light green, while the ZBD of SecA in panel B is shown in blue. The  $\text{Zn}^{2+}$  is shown as red spheres in panel B.

## 1.11 The mycobacterial Sec system

The *E. coli* Sec system is the most widely studied secretion system, however, other bacterial Sec systems, such as *Mycobacterium tuberculosis*, can provide alternative avenues for investigating the underlying molecular mechanisms of protein secretion by this system. In addition, due to their significance to the bacterial virulence and viability, *M. tuberculosis* secreted proteins and their respective translocation systems are considered excellent targets for drug development (Braunstein *et al*, 2003a; Ligon *et al*, 2012a; Swanson *et al*, 2015). In this bacterium, the main components of the Sec system including SecA motor protein, SecYEG pore forming complex, and the signal sequences of the preproteins are very similar to that of *E. coli*, however, other accessory elements of SecA dependent translocation are quite different. *M. tuberculosis* has two homologous SecAs with nonoverlapping functions. This organism also lacks the translocation dedicated SecB chaperone, instead containing a distantly related protein, *mtSecB*, which assists the Sec translocation reaction as well as functioning in other important cellular processes. These systems are briefly described below.

## 1.12 Accessory SecA2 translocation system

Sec translocation system in *M. tuberculosis* functions overall similar to *E. coli*. However, the Sec system in all mycobacteria is unusual in having two SecA paralogs; SecA1 and SecA2 (Ligon *et al*, 2012b). SecA1 (the one that is more similar to *E. coli* SecA (Rigel & Braunstein, 2008b)) is essential, while although SecA2 is the accessory SecA and it is not essential for viability (Braunstein *et al*, 2003b), it is necessary for the full virulence of *M. tuberculosis* (Feltcher & Braunstein, 2012). SecA2 functions in protein transport, but its role seems to be restricted to a much smaller number of exported proteins than those translocated by the housekeeping SecA1. Analysis of the exported proteins from *secA2* deficient mutant and wild-type strains revealed that superoxide dismutase (SodA) and catalase-peroxidase (KatG), which are involved in protecting *M. tuberculosis* from reactive oxygen intermediates generated by host phagocytes (Dussurget *et al*, 2001), are among the very few proteins that require SecA2 for their translocation (Braunstein

*et al*, 2003b). A similar study identified Msmeg1704 and Msmeg1712 lipoproteins as substrates of SecA2 system (Gibbons *et al*, 2007).

SecA2 and SecA1 proteins share some similarities, such as both are present in equivalent amounts in *M. tuberculosis* and ATPase activity is required for their function (Rigel *et al*, 2009a). Analysis of their sequence has also revealed that most domains are conserved between SecA1 and SecA2 (Feltcher & Braunstein, 2012; Swanson *et al*, 2015). However, they have some differences. First, SecA1 and SecA2 also have independent functions and they cannot substitute each other. Even when SecA2 is overexpressed, SecA1 cannot be deleted; similarly, overexpression of SecA1 does not rescue the SecA2 deletion mutant (Rigel & Braunstein, 2008b). Furthermore, SecA2 is smaller and it lacks the C-terminal region, where, in *E. coli*, SecA binds to SecB (Ligon *et al*, 2012b). Comparison of the available crystal structures of SecA1 and SecA2 also revealed that although these proteins are highly similar, the most important difference is the absence of the helix wing domain (HWD) in SecA2 (Swanson *et al*, 2015; Sharma *et al*, 2003). These structural differences have been attributed to the unique function and selectivity of SecA2 towards its substrates (Swanson *et al*, 2015).

It has been suggested that the two SecAs might work together to export SecA2 dependent substrates (Ligon *et al*, 2012b). In support of this idea, interestingly, it was shown that the translocation of Msmeg1712 in *M. smegmatis* in the absence of SecA1 was significantly compromised (Rigel *et al*, 2009b). As another evidence for interaction between SecA1 and SecA2, it was shown that while both proteins separately form homodimers in solution, but when both SecAs are present together, they can form dissociable heterodimers (Prabudiansyah *et al*, 2015). Given the controversial question regarding the functional role of dimer formation in the translocation process (see previous sections), these may be considered as another indirect indication of cooperation between two protomers in other bacterial Sec systems, such as *E. coli*, which only has a sole SecA.

### 1.13 Mycobacterial SecB chaperone

The proper interaction of unfolded preproteins with the Sec system is determined by the folding rate of these molecules. Many bacterial species have devised a dedicated chaperon, SecB, to maintain preproteins in an unfolded state and prevent their premature folding (Bechtluft *et al*, 2010). This protein is also believed to deliver preproteins directly to the membrane bound SecA. Despite the importance of protein secretion and outer membrane synthesis of *M. tuberculosis*, very little is known about its Sec system (Feltcher & Braunstein, 2012). It was believed that *M. tuberculosis* lacks a chaperonic SecB protein, and it is not clear yet how the preproteins are kept in the translocation competent state in this bacterium. A relatively recent study showed that this organism's genome contains the coding sequence for a putative SecB protein (Bordes *et al*, 2011a). Although *M. tuberculosis* SecB (*mtSecB*; also known as SecB-like or Rv1957) shares only 13% sequence similarity with *E. coli* SecB, it possesses the key structural elements of this protein; the key SecB amino acids important for its oligomerization, and its specific interaction with SecA are preserved in this protein (Sala *et al*, 2013). Further *in vivo* and *in vitro* assays showed that this protein exhibits chaperonic functions similar to SecB protein. For example, this protein was able to replace *E. coli* SecB protein *in vitro* and *in vivo*, and can prevent the *in vivo* aggregation proOmpC, which is a substrate of the Sec system. Interestingly, this protein was shown to be also involved in the functional stress-responsive hlgB-hlgA toxin-antitoxin (TA) system (Bordes *et al*, 2011a; Gerdes & Maisonneuve, 2012a). In TA systems, toxins usually inhibit cell growth, and are neutralized with their associated antitoxins from a TA system in normally growing cells. Environmental stresses will lead to degradation of the antitoxin, and hence activate the toxin protein. The activated toxin inhibits the bacterial growth, and this slower growth is a strategy to survive under environmental stress (Gerdes & Maisonneuve, 2012a). In *M. tuberculosis*, *mtSecB* binds and stabilizes the antitoxin protein, and therefore, prevents the action of the toxin protein. It was proposed that under protein export stress conditions, *mtSecB* would be mainly recruited by the Sec system to assist the effective targeting of preproteins to the Sec translocase. This would, in turn, likely compete with the other function of *mtSecB*, binding and stabilizing antitoxin protein, and as a result the unbound antitoxin will undergo degradation. Consequently, the free toxin will restrain mycobacterial growth until the stress is removed (Bordes *et al*, 2011b). The details of these interactions remain to be identified. The fact that *mtSecB* possesses the main predicted structural elements that are essential for the interaction of

*E. coli* SecB with preproteins and SecA suggests that this protein plays a role delivering the preproteins to SecA(s) in the mycobacterial Sec system.

## 1.14 Scope of thesis

The evolutionarily well-conserved ATPase SecA is critical for post-translational secretion. This is important, as a myriad of proteins that play important roles for virulence and bacterial viability are dependent on this system for their transport to the extracytoplasmic environment. However, despite the wealth of knowledge available regarding the post-translational translocation by the Sec system, the molecular basis for the coupling of ATP hydrolysis to protein translocation remains largely unknown. The chief question is, how does the SecA couple ATP binding and hydrolysis to the recognition and movement of unfolded substrates. This thesis will address aspects of this question as follows; first, by elucidating the role of unstructured termini in the oligomerization and function of various SecA constructs, second, through unveiling the molecular basis of unfolded polypeptide binding by SecA, and finally, by making advances towards obtaining high-resolution structures of *E. coli* SecA and *M. tuberculosis* mtSecB.

A critical question in synthesizing any potential mechanism for coupling of ATP hydrolysis to protein translocation is whether the translocating system incorporates a single SecA protomer, or whether multiple SecA protomers are involved. The functional oligomeric state of SecA and the domains involved in the self-association of this protein has been a matter of considerable controversy for several years. By re-examining the unusual tetramerization of a truncated SecA construct (SecA-N68), the results presented in Chapter 2 showed that the two unstructured polypeptides at its termini, the naturally occurring N-terminal polypeptide and a byproduct of truncation at C-terminus, are mediating these self-associations. Moreover, by investigating the role of the first 14 N-terminal residues of functional SecA-N95 construct it was shown that the removal of this region drastically weakened dimerization; the dissociation constant was increased from nanomolar ranges in the case of SecA-N95 to  $\sim 25\mu\text{M}$  in SecA-N95 $\Delta\text{N}$ . Although weakened dimerization did not significantly affect the solution ATPase activity of SecA-N95 *in vitro*, it was shown that SecA-N95 $\Delta\text{N}$  is not functional *in vivo*. This pointed towards the importance of the unstructured N-terminal segment of SecA for its function.

Furthermore, the binding site for this extreme N-terminal region of SecA was mapped to the N-terminal domains of SecA (DEAD motor domain and/or PPXD).

Another equally important question regarding the function of SecA concerns the nature of the interactions between SecA and the mature regions of preproteins. These interactions have been difficult to characterize mainly due to the fact that the unfolded preproteins have a tendency to either refold or aggregate *in vitro*. To investigate the direct interaction of SecA with the mature regions of preproteins, a new strategy was adopted that involved the immobilization of large peptides from maltose binding protein (MBP) to form a heterogeneous population of potential binding sites. Binding to these immobilized peptides was monitored using surface plasmon resonance (SPR). These studies were extended to a more defined system using an immobilized 15-residue peptide from the mature region of the periplasmic protein, FhuD. Our overall results investigating these interactions, which are presented in Chapter 3, revealed that preproteins contain sequences that can bind in a nucleotide dependent manner to SecA *in vitro*. The binding site(s) for these polypeptides were mapped to the NBD domains and/or stem region of PPXD. In addition, remarkably, the interaction between SecA and the immobilized peptides was almost completely dependent on the presence of the unstructured N-terminal region of SecA.

A complete understanding of the translocation mechanism requires high-resolution structures of the individual components and the molecular complexes that are formed during translocation. For example, structures of *E. coli* SecA-N95, both alone and in complex with a peptide representing a translocation substrate, have been longstanding goals in the field. The work presented in Chapter 4 showed that the diffraction quality of the SecA-N95 crystals could be improved by removal of the unstructured N-terminal segment. A similar strategy was used to improve the crystallization of *M. tuberculosis* Rv1957 or “*mtSecB*”. Rv1957 was cloned, expressed, and shown to form a stable tetramer in solution, in common with *E. coli* SecB. However, crystallization was very slow and difficult to reproduce, problems that were attributed to the N-terminal region of *mtSecB*. A *mtSecB* construct in which the N-terminus was truncated by eight residues yielded large crystals relatively rapidly. Overall, in the Chapter 4, it was shown that presence of unstructured termini have a profound effect on crystallization of these two members of the Sec system.

## Chapter 2

### 2 Oligomerization of SecA is Mediated by Unstructured Polypeptides

As a part of the Sec translocation system, the SecA ATPase is responsible for moving the unfolded preproteins through the SecYEG pore complex of the bacterial inner membrane. The molecular basis for interactions of SecA with unfolded polypeptides and the SecYEG channel as well as the coupling of ATP hydrolysis to the translocation by SecA are largely unknown. The functional oligomeric state of SecA and the domains involved in its self-association have been key questions for synthesizing a coherent model of these interactions. Here, by re-examining the unusual tetramerization of a truncated SecA construct (SecA-N68), it was shown that the two unstructured polypeptides at its termini, the naturally occurring N-terminal polypeptide and a by-product of truncation at C-terminus, are mediating these self-associations. Furthermore, it was shown that the tetramer assembly is most likely a dimer of dimers formed through interactions mediated by these unstructured termini. Next, by examining the role of the first 14 N-terminal residues of the functional SecA-N95 construct, it was shown that the removal of this region drastically weakened dimerization; the dissociation constant was increased from nanomolar range in SecA-N95 to  $\sim 25 \mu\text{M}$  in SecA-N95 $\Delta\text{N}$ . Although weakened dimerization did not significantly affect the solution ATPase activity of SecA-N95 *in vitro*, it was shown that SecA-N95 $\Delta\text{N}$  is not functional *in vivo*. Overall, these findings are important for understanding the molecular mechanism underlying the function of SecA in the translocation process.

#### 2.1 Introduction

Bacteria use the evolutionarily well-conserved secretion (Sec) system for the bulk transport of extracytoplasmic proteins into and through their inner-membrane (Kusters & Driessen, 2011). The central components of the Sec system are the pore forming trimeric complex (SecYEG) and the peripheral ATPase SecA, and it has been shown that these two components are sufficient for translocation of preproteins *in vitro* (Brundage *et al*, 1990; Akimaru *et al*, 1991). In this system, SecA is responsible for the translocation of unfolded preprotein substrates through the SecYEG



channel by coupling the energy derived from the ATP hydrolysis to the mechanical movement of preproteins (Park & Rapoport, 2012). SecA is a large protein (102 kDa as a monomer), which forms a stable dimer in solution with a sub-micromolar dimer dissociation constant (Wowor *et al*, 2011; Kusters *et al*, 2011). SecA is composed of several distinctive domains (Figure 1-3) which include: two NBD1 and NBD2 domains that are responsible for nucleotide binding, the preprotein cross-linking domain (PPXD) that is internally fused to NBD1 and was identified through *in vitro* cross-linking studies as the domain responsible for the preprotein binding (Kimura *et al*, 1991), C-terminal domains (CTDs) including several domains that have been shown to be implicated in the inter- and intra-molecular interactions between SecA protomers and other components of the Sec system (Karamanou *et al*, 1999; Zimmer *et al*, 2008; Hunt *et al*, 2002) and a zinc binding domain that is involved in the interaction with SecB and is the only “non essential” domain of SecA (Xu *et al*, 2000).

In post-translational translocation, SecA in concert with SecYEG moves the unfolded preproteins in a stepwise fashion through the membrane (Erlandson *et al*, 2008; Papanikou *et al*, 2007; Park & Rapoport, 2012; Uchida *et al*, 1995). It has been shown that upon interaction with SecYEG and preproteins at the membrane, the ATPase activity of SecA is significantly increased. Moreover, it appears that the conformational changes in SecA, which are derived from ATP binding and hydrolysis, are responsible for the movement of preproteins through the membrane channel (Economou & Wickner, 1994b). However, the molecular details of the translocation steps as well as the chief question regarding the coupling of ATP hydrolysis to preprotein translocation by SecA are still largely unknown.

The interaction of SecA with unfolded preproteins and SecYEG during translocation is further complicated by the controversial observation that the dimerization of SecA is important for its function. Specifically, although SecA exists as a dimer in solution, the functional oligomeric state of SecA during the translocation process has been a matter of considerable debate recently (Sardis & Economou, 2010). A critical question, therefore, in any potential mechanism is whether the translocating system incorporates a single SecA protomer, or whether multiple SecA protomers are involved. On the one hand, the evidence supporting involvement of a SecA *monomer* in the translocation process includes experiments in which dimerization defective SecA is still able to catalyze translocation reaction (Or *et al*, 2002; 2005), as well as the crystal

structure of a SecA-SecYEG complex, which clearly indicates that at any given time only a single SecA is able to interact directly with SecYEG (Zimmer *et al*, 2008; Park & Rapoport, 2012). On the other hand, there is evidence supporting the involvement of SecA *dimer* in the translocation reaction, which include the observation of dimeric forms in several crystal structures of SecA from various bacteria (Hunt *et al*, 2002; Zimmer *et al*, 2006; Sharma *et al*, 2003), and also experiments in which alteration of the observed dimeric interfaces showed that dimerization of SecA is critical for the translocation reaction (Jilaveanu *et al*, 2005; Karamanou *et al*, 2005; Das *et al*, 2008). In addition, different dimeric interfaces have been observed in the available crystal structures of SecA, and as a result, it has been suggested that SecA may acquire different oligomeric states during various steps of the translocation process (Kusters & Driessen, 2011; Sardis & Economou, 2010). Perhaps the strongest evidence for the involvement of two SecA protomers in the translocation reaction comes from investigation of *Mycobacterium tuberculosis* Sec system, which contains two SecA genes, *secA1* and *secA2*. SecA1 is essential and resembles the single, critical SecA present in *E. coli*, whereas SecA2 is non-essential for survival of this organism, has a slightly altered domain structure, and is crucial for the secretion of only a few specific preproteins (Swanson *et al*, 2015; Rigel & Braunstein, 2008a; Braunstein *et al*, 2003b). Interestingly, it has been shown that SecA2 dependent preprotein translocation requires SecA1, and therefore at least two SecA molecules are involved for the translocation of these SecA2 dependent preproteins (Rigel *et al*, 2009b). However, in the case of the *E. coli* Sec system, the available evidence does not provide a conclusive answer as to the number of SecA protomers required for the translocation process (Sardis & Economou, 2010). A comprehensive list of selected research papers regarding the nature of dimerization and the role of self-association of SecA in the translocation process is provided in Table 2-1.

Experiments with SecA deletion constructs have provided insight into the mechanisms of its oligomerization. Full-length SecA was recalcitrant to the crystallization attempts, and to acquire a shorter, better-behaved construct, a version of SecA truncated at residue 610 was generated and termed “SecA-N68” because of its 68 kDa molecular weight (Dempsey *et al*, 2002). This truncated SecA formed tetramers in solution. The SecA-N68 tetramer was characterized

**Table 2-1. List of the selected research papers on dimerization of SecA**

<b>Title</b>	<b>Ref.</b>	<b>Year</b>	<b>Dimer/monomer</b>
SecA, an essential component of the secretory machinery of <i>Escherichia coli</i> , exists as homodimer.	(Akita <i>et al</i> , 1991)	1991	dimer
SecA, the peripheral subunit of the <i>Escherichia coli</i> precursor protein translocase, is functional as a dimer	(Driessen, 1993)	1993	dimer
<i>Escherichia coli</i> SecA shape and dimensions	(Shilton <i>et al</i> , 1998)	1998	dimer
Dissociation of the dimeric SecA ATPase during protein translocation across the bacterial membrane	(Or <i>et al</i> , 2002)	2002	monomer
Complex behavior in solution of homodimeric SecA	(Woodbury <i>et al</i> , 2002)	2002	dimer
Nucleotide control of interdomain interactions in the conformational reaction cycle of SecA	(Hunt <i>et al</i> , 2002)	2002	dimer
<i>Bacillus subtilis</i> SecA ATPase exists as an antiparallel dimer in solution	(Ding <i>et al</i> , 2003)	2003	dimer
Binding, activation and dissociation of the dimeric SecA ATPase at the dimeric SecYEG translocase	(Duong, 2003)	2003	monomer
A large conformational change of the translocation ATPase SecA	(Osborne <i>et al</i> , 2004)	2004	monomer
Covalently dimerized SecA is functional in protein translocation	(de Keyzer <i>et al</i> , 2005)	2005	dimer
Dimeric SecA is essential for protein translocation	(Jilaveanu <i>et al</i> , 2005)	2005	dimer
The bacterial ATPase SecA functions as a monomer in protein translocation	(Or <i>et al</i> , 2005)	2005	monomer
<i>Escherichia coli</i> SecA truncated at its termini is functional and dimeric	(Karamanou <i>et al</i> , 2005)	2005	dimer
SecA dimer cross-linked at its subunit interface is functional for protein translocation	(Jilaveanu & Oliver, 2006)	2006	dimer
Cross-linked SecA dimers are not functional in protein translocation	(Or & Rapoport, 2007)	2007	monomer
Reexamination of the role of the amino terminus of SecA in promoting its dimerization and functional state	(Das <i>et al</i> , 2008)	2008	dimer
Structure of a complex of the ATPase SecA and the protein-translocation channel	(Zimmer <i>et al</i> , 2008)	2008	monomer
Energetics of SecA dimerization	(Wowor <i>et al</i> , 2011)	2011	dimer
Defining the <i>Escherichia coli</i> SecA dimer interface residues through <i>in vivo</i> site-specific photo-cross-linking	(Yu <i>et al</i> , 2013)	2013	dimer
Defining the solution state dimer structure of <i>Escherichia coli</i> SecA using Förster resonance energy transfer	(Auclair <i>et al</i> , 2013)	2013	dimer
Mapping of the SecA signal peptide binding site and dimeric interface by using the substituted cysteine accessibility	(Bhanu <i>et al</i> , 2013)	2013	dimer
SecAAA trimer is fully functional as SecAA dimer in the membrane	(Wang <i>et al</i> , 2014)	2014	dimer
Structural similarities and differences between two functionally distinct SecA proteins: the <i>M. tuberculosis</i> SecA1 and SecA2	(Swanson <i>et al</i> , 2015)	2015	dimer

in terms of its stability and shape using ultracentrifugation and small-angle X-ray scattering (SAXS), which indicated that SecA-N68 is in equilibrium between monomer and tetramer with a  $K_D$  of 63  $\mu\text{M}$  for the formation of the tetramer. Given the important question regarding the number of SecA protomers involved in translocation, the formation of this tetramer was an interesting observation (Dempsey *et al*, 2002), but the physiological significance of the SecA-N68 tetramer to the translocation process has never been clear.

Here, this work has been extended to show that, surprisingly, the self-association of SecA-N68 is mediated solely by its unstructured terminal polypeptides. By further investigating these interactions it was shown that the short N-terminal polypeptide has a critical role in the dimerization of a functional SecA construct. In order to study the significance of this N-terminal polypeptide in the functionality of SecA, it was shown that the removal of the N-terminal sequence does not affect the solution ATPase activity of SecA, but this region is critical for the function of this protein *in vivo*.

## 2.2 Materials and methods

The QIAquick Gel Extraction Kit and QIAprep Miniprep Kit were obtained from Qiagen. Restriction enzymes were ordered from Thermo Scientific and NEB (New England Biolabs). T4 DNA Ligase was from Promega. PfuTurbo DNA polymerase was ordered from Agilent. The Antarctic phosphatase was from NEB. Lysozyme and DNase I were obtained from Sigma. Primers were ordered from UWO Oligo Factory. Chromatographic media were from GE Healthcare.

The Cibacron Blue resin used for the affinity purification of SecA protein was prepared by coupling Cibacron Blue 3GA dye (Sigma) to the Sepharose CL-6B resin. In short, Sepharose CL-6B was rinsed with water and then the excess water was removed by vacuum filtration. Approximately 100 g of the wet Sepharose was resuspended in 150 mL of 0.5 M NaCl and then 5 g of Cibacron Blue 3GA dye and 6 mL of 10 M NaOH were added. This suspension was mixed on an orbital shaker for an hour at room temperature and then was incubated for another 45 min at 37 °C. The resulting gel was washed consecutively with 1 M NaOH, water, 60% ethanol, and water. The final washed blue resin was stored in 20% ethanol.

### 2.2.1 Molecular cloning

Constructs used for this study are listed in Table 2-2, and the list of primers used for the DNA amplification and sequencing reactions are provided in Table 2-3. Molecular cloning was carried out according to the standard protocols (Sambrook & Russell, 2001).

#### 2.2.1.1 Cloning of pAK-SecA-N68 $\Delta$ C

The coding sequence for *SecA-N68 $\Delta$ C* (residues 1-590 of wild-type SecA) was amplified from pZ52-SecA (originally pT7SecA2 (Rajapandi *et al*, 1991)) using primers F-AK-N68 $\Delta$ C and R-AK-N68 $\Delta$ C (Table 2-3). PCR was carried out using PfuTurbo DNA polymerase with the following procedure: an initial denaturation step at 96 °C for 5 min was followed by the first cycle (96 °C for 3 min, 54 °C for 4 min, and 72 °C for 6 min); the reaction was continued for 30 cycles (96 °C for 45s, 51 °C for 45s, and 72 °C for 4 min 25s); the reaction was completed with

an extension step for 15 min at 72 °C. The amplified DNA band was gel-purified and ligated between the *Bam*HI and *Ssp*DI cut sites of the expression vector pProEX-HTa (Invitrogen) with T4 DNA ligase. The resulting plasmids were transformed into DH5 $\alpha$  cells. DNA sequencing using primers listed in Table 2.3 confirmed that no unwanted mutations were introduced.

### 2.2.1.2 Cloning of pAK-SecA-N68 and pAK-SecA-N68 $\Delta$ NC

Both constructs were made in a pProEX HTa background vector. For creating pAK-SecA-N68 $\Delta$ NC, the *Eco*RI fragment from pAK-SecA-N68 $\Delta$ C was inserted in *Eco*RI sites of pJH-N68-ER1-ER2. For constructing pAK-SecA-N68, the *Eco*RI fragment of pZ52-70H (His<sub>6</sub>-SecA-N68) vector was inserted in *Eco*RI sites of pAK-SecA-N68 $\Delta$ C. Vectors were treated with Antarctic Phosphatase to reduce the self-ligated vector background. The correctly oriented vectors were selected by analyzing the restriction digestion pattern of vectors (by *Xho*I and *Bgl*II double digestion pattern). Both vectors were sequenced to verify the accuracy of the coding region of the SecA construct.

### 2.2.1.3 Construction of pAK-SecA-N68 $\Delta$ N

For cloning of pAK-SecA-N68 $\Delta$ N, the *Nco*I fragment of pZ52-70H (His<sub>6</sub>-SecA-N68) vector was inserted into *Nco*I site of pJH-N68-ER1-ER2 (construct with truncated 14 N-terminal residues). The correctly oriented SecA-N68 $\Delta$ N vectors were selected by analyzing the restriction digestion pattern by *Eco*RI and further confirmed by DNA sequencing.

### 2.2.1.4 Cloning of pAK-SecA-N95 $\Delta$ N

The *Eco*RI fragment from pBSWT plasmid (His<sub>6</sub>-SecA-N95 (Shilton *et al*, 1998)) was ligated between the *Eco*RI sites of pAK-SecA-N68 $\Delta$ N. pAK-SecA-N68 $\Delta$ N vector's backbone was treated with Antarctic Phosphatase to reduce the self-ligated vector background. After successful transformation, vectors with the right orientation of the insert were selected by analyzing the SDS-PAGE pattern of protein expression (as the correctly oriented segments were expected to express a protein around 95 kDa (SecA-N95 $\Delta$ N)). Consequently, the plasmid extraction was performed on strains that showed the desired protein pattern, and the plasmids were sequenced.

### 2.2.1.5 Cloning of pAK-SecA-N95 $\Delta$ N(T109N)

pAK-SecA-N95 $\Delta$ N(T109N) was produced as follows. The *Nco*I segment of pBST109N (His<sub>6</sub>-N95(T109N) (Shilton *et al.*, 1998)) was ligated into the *Nco*I site of pProEX-N95 $\Delta$ N vector. Vector's backbone segment was treated with Antarctic Phosphatase to reduce the self-ligated vector background. The resulted DNA mixture was transformed in *E. coli*. The right oriented constructs were selected by analyzing the protein expression pattern, and subsequent DNA sequencing.

### 2.2.1.6 Cloning of pAK-SecA-N95(T109N)

For generating the pAK-SecA-N95(T109N) construct, the *Nco*I fragment of pBST109N (His<sub>6</sub>-SecA-N95(T109N) (Shilton *et al.*, 1998)) was inserted into the *Nco*I cut site of pJZ7-N95 vector (which is a derivative of pZ52-SecA with a stop codon after 835th amino acid). After transformation, the correctly oriented constructs were selected by analyzing the protein expression pattern as described above. Further DNA sequencing confirmed the desired results.

### 2.2.1.7 Construction of pAK-SecA-N95 $\Delta$ N14

For generation of the pAK-SecA-N95 $\Delta$ N14 vector, which does not carry the cleavable N-terminal affinity tag, for *in vivo* complementation analyses, a PCR mutagenesis method was used. Primers F-AK-N95 $\Delta$ N14 and R-AK-N95 $\Delta$ N14 (See Table 2-3 for the sequences of primers) were used in a PCR based on Klock and Lesley's procedure (Klock & Lesley, 2009) on the pJZ7-SecA-N95 plasmid (pZ52-SecA that carries a stop codon after 835<sup>th</sup> residue). After optimization of the PCR reaction, the generated DNA was treated with *Dpn*I for 2 hours at 37 °C. Then the vectors were transformed into the DH5 $\alpha$  *E. coli* cells.

**Table 2-2. List of the SecA constructs and expression vectors used in this chapter**

<b>Construct</b>	<b>Plasmid</b>	<b>Features</b>
SecA-N68 $\Delta$ NC-ER	pJH-N68-ER1-ER2	Residues 15-590 in pProEX-HTa; carries entropy-reducing (ER) mutations: E55A, K56A, E58A, E196A, and E197A. Has a cleavable N-terminal His-tag
SecA	pZ52-SecA	Residues 1-901. Originally pT7SecA2
His <sub>6</sub> -SecA-N68	pZ52-N70H	Residues 6-609. Derived from pZ52-SecA. Has a non-cleavable N-terminal His-tag, which has modified the very N-terminal sequence from the wild-type sequence of MLIKLLTKVFG to MHHHHHHLTKVFG
SecA-N68	pAK-SecA-N68	Residues 1-609 in pProEX-HTa; cleavable N-terminal His-tag
SecA-N68 $\Delta$ N	pAK-SecA-N68 $\Delta$ N	Residues 15-609 in pProEX-HTa; cleavable N-terminal His-tag
SecA-N68 $\Delta$ NC	pAK-SecA-N68 $\Delta$ C	Residues 15-590 in pProEX-HTa; cleavable N-terminal His-tag
SecA-N68 $\Delta$ C	pAK-SecA-N68 $\Delta$ NC	Residues 1-590 in pProEX-HTa; cleavable N-terminal His-tag
SecA-N95	pJZ7-SecA-N95	Residues 1-835. Derived from pZ52-SecA.
His <sub>6</sub> -SecA-N95	pBSWT	Residues 6-835. Has a non-cleavable N-terminal His-tag, which has modified the N-terminus sequence from the wild-type sequence of MLIKLLTKVFG to MHHHHHHLTKVFG
SecA-N95 $\Delta$ N	pAK-SecA-N95 $\Delta$ N	Residues 15-835 in pProEX-HTa; cleavable N-terminal His-tag
SecA-N95(T109N)	pAK-SecA-N95(T109N)	Residues 1-835. Derived from pZ52-SecA. Mutation T109N
SecA-N95 $\Delta$ N(T109N)	pAK-SecA-N95 $\Delta$ N(T109N)	Residues 15-835 in pProEX-HTa; cleavable N-terminal His-tag. Mutation T109N
SecA(T109N)	pZ52-SecA(T109N)	Residues 1-901. Derived from pZ52-SecA. Mutation T109N



**Table 2-3. List of the primers used in this chapter.**

<b>Primer</b>	<b>Sequence</b>	<b>Length</b>	<b>Use</b>
F-AK-N68 $\Delta$ C	5'- ATCTGAGGCGCCATGCTAATCAAATTGTTAACTAAAGTTTTCGGTAG	47	Cloning
R-AK-N68 $\Delta$ C	5'- CCTGATGGATCCTTACATCGACAGGTAGAAACGGGAAG	38	Cloning
F-AK-N95 $\Delta$ N14	5'- TGAGATTTTATTATGGATCGCACCCCTGCGCCGGATG	36	Cloning
R-AK-N95 $\Delta$ N14	5'- CATAATAAAATCTCAAACGCCCCGCGTTGC	30	Cloning
F-AK-pProEX	5'- TAACAATTTACACAGGAAACAGACC	26	Sequencing
R-AK-pProEX	5'- TTCTCTCATCCGCCAAAACAGC	22	Sequencing
4362-1	5'- GAGATGGAAAACTCTCCGACG	22	Sequencing
4313-2	5'- GAATACGGCTTTGACTACCTG	21	Sequencing
4313-3	5'- GCTGCGCGCTCATGCGC	17	Sequencing
4313-4	5'- GCCAGCCGGTGCTGGTGG	18	Sequencing
4313-5	5'- GGGGATGCTGGTTCTTCC	18	Sequencing
4313-6	5'- TGGGATATTCCGGGGCTGC	19	Sequencing

## 2.2.2 Expression of recombinant proteins

For high-level expression, all the plasmids with wild-type SecA sequence were transformed into *E. coli* BL21(DE3), while constructs carrying the T109N mutation were transformed into *E. coli* BL21.19(DE3) strain, in which the endogenous expression of SecA is suppressed at elevated temperatures (Mitchell & Oliver, 1993). This was to avoid contamination of SecA(T109N) with endogenous wild-type SecA.

Induction of protein expression in *E. coli* BL21(DE3) was performed as follows. A single bacterial colony of an overnight *E. coli* culture was transferred to 2 mL of 2xYT/Amp broth (2xYT medium containing 100 mg/L ampicillin) and grown at 37 °C with shaking. When this starter culture looked cloudy, it was removed from incubator and kept on ice for a few hours. In the evening, the starter culture was diluted 1000 times with fresh 2xYT/Amp. This inoculum was afterwards used to inoculate of a few liters of fresh 2xYT/Amp broth with 1:1000 dilutions. The inoculated broth, was then grown at 28 °C over night while shaking at 220 rpm. Next morning, at an OD<sub>600</sub> of ~0.6, protein expression was induced by addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and the cells were allowed to grow at 28 °C while shaking at 220 rpm. The cells were harvested by centrifugation at 6,000 rpm for 10-15 min at 4 °C, and were flash frozen and stored at -80 °C.

For protein expression in *E. coli* BL21.19, overnight cultures were prepared and inoculated into fresh LB/Amp/Kan (LB media containing 100 mg/L ampicillin and 35 mg/L kanamycin) essentially as described above. At OD<sub>600</sub> of 0.6, temperature was shifted to 40 °C (to suppress wild-type SecA expression) while shaking, and 0.5 mM IPTG was added to induce the expression of SecA(T109N) constructs. After 10 hours, cells were harvested, flash frozen, and then stored at -80 °C. Cell yield for T109N mutants was much lower than other SecA because of the elevated ATPase level of this protein.

## 2.2.3 Purification of His-tagged SecA proteins with truncated N-terminus

All the His-tagged SecA proteins, which carried a truncated N-terminus (which lacked the first

14 amino acids of the wild-type SecA), were purified by the following methods with minor modifications for each construct.

### 2.2.3.1 *E. coli* cell lysis preparation

Cells were resuspended in a cold lysis buffer (20 mM sodium phosphate, 10 mM imidazole, 500 mM NaCl, 1 mM EDTA, pH 7.4). 300 µg/ml Lysozyme and 1.5 mM PMSF (Phenylmethylsulfonyl fluoride) was added to the cell suspension. The resulting mixture was incubated room temperature for 15 min, and was stirred for 20 min at 4 °C. Thereafter, MgCl<sub>2</sub> (from 2M stock solution (100x)) and DNase I (from a 3 mg/ml stock (100x)) were added to the viscous suspension at final concentrations of 10 mM and 30 µg/ml, respectively. PMSF (1 mM) was added, and the solution was stirred for another 10 min at 4 °C. This suspension was passed twice through a French pressure cell (Thermo Scientific) at 18,000 psi. The resulting solution was centrifuged at 38,000 rpm (45Ti rotor) at 4 °C for 30 min. The supernatant was collected and the pellets were discarded. The cell lysate was kept on ice until the next purification step.

### 2.2.3.2 Nickel column purification

For purification of N-terminal His-tagged SecA proteins from the cell lysate, immobilized metal affinity chromatography (IMAC) was used. Briefly, the clarified lysate was passed through the Chelating Sepharose Fast Flow resin (GE Healthcare) using a gravity flow column. The flow through solution was collected and kept on ice for further analysis. Then the column was washed by several column volume of wash buffer (20 mM sodium phosphate, 10 mM imidazole, 500 mM NaCl, pH 7.4) to remove the unbound proteins from the affinity column. When there was no significant amount of protein in the wash through solution, the trapped His-tagged SecA proteins were eluted in a single step using elution buffer (20 mM sodium phosphate, 500 mM imidazole, 500 mM NaCl, pH 7.4) and kept on ice for further purification steps. The purity of the IMAC purified proteins was assessed by SDS-PAGE analysis. The amount of the purified protein was estimated by the Bradford reagent (Biorad).

### 2.2.3.3 Removal of the affinity tag

A His-tagged TEV (Tobacco Etch Virus) protease (which was prepared as described before (Tropea *et al*, 2009)) was used for the removal of the cleavable hexa-histidine affinity tags from

SecA constructs. After estimating the amount of purified protein, DTT (1 M Stock solution) and EDTA (0.5 M pH 8.0 Stock solution) were added to the final concentrations of 10 mM and 2 mM, respectively. Then, TEV protease was added to the His-tagged SecA solution at a 1:50 ratio (10 mg/ml stock solution), and the solution was dialyzed overnight against two liters of 50 mM Tris-HCl pH 7.5, 200 mM NaCl, 1 mM EDTA, and 14 mM  $\beta$ -mercaptoethanol at 4 °C. Dialysis buffer was exchanged the next day, and after taking a sample of digestion reaction, the SecA solution was supplemented with fresh TEV enzyme. The process of digestion was monitored by SDS-PAGE, and was continued until almost all of the affinity tags were removed from the target SecA protein. At the end of the digestion process, the protein solution was dialyzed against the aforementioned IMAC wash buffer overnight at 4 °C, and then the proteins were passed through the IMAC column to remove both the His-tagged TEV protease and any traces of SecA proteins that still carried the N-terminal hexa-histidine tag.

#### 2.2.3.4 Ion exchange chromatography

To further remove the unwanted proteins from the affinity purified SecA proteins, they were subjected to ion exchange chromatography at 4 °C. The protein solution was dialyzed against buffer A (50 mM Tris-HCl, 2 mM EDTA, 14 mM  $\beta$ -mercaptoethanol, pH 8.2) overnight at 4 °C. Then, a 1.6 x 13 cm Q-Sepharose HP column (GE Healthcare) was equilibrated with buffer A at 2.5 mL/min (75 cm/hr). The protein solution was filtered through a 0.45  $\mu$ m syringe filter and injected into the Q-Sepharose column using a 50 mL Superloop (GE Healthcare) at the same flow rate. The column was washed with buffer A until UV absorbance (280 nm) returned to baseline. Finally, the bound proteins were eluted in a 240 mL linear gradient from 20% buffer B (buffer A containing 0.5 M NaCl) to 100% buffer B at 2 mL/min (60 cm/hr). The eluted proteins were collected in 6 mL fractions and were analyzed afterwards by SDS-PAGE. The fractions that contained the desired SecA protein, as judged by SDS-PAGE, were combined and kept on ice for the next step.

#### 2.2.3.5 Preparative size exclusion chromatography

As a final polishing step in preparation of SecA proteins for structural analysis, size exclusion chromatography was used. Accordingly, a 2.6 x 65 cm Superdex 200 preparative grade column (GE Healthcare) with an approximate volume of 350 mL was equilibrated with two column

volumes of running buffer (50 mM Tris-HCl, 300 mM NaCl, 2 mM EDTA, and 15 mM  $\beta$ -mercaptoethanol at pH 7.4) at flow rate of 2 mL/min (21 cm/hr). The protein solution was concentrated to 3-6 mL at 4 °C using a 50 mL stirred ultrafiltration unit (Amicon) or 15 mL Amicon ultra centrifugal filters (Millipore) with a MW cutoff of 10,000 kDa, and clarified using a 0.45  $\mu$ m syringe filter. Using a pre-cooled 10 mL Superloop (GE Healthcare), the concentrated protein was injected on column at the 2 mL/min. The elution of proteins from the column was monitor by recording the UV absorbance at 280 nm. 6 mL fractions of the main elution peak(s) were collected and analyzed by SDS-PAGE. The protein-containing fractions were combined, concentrated, flash frozen in liquid nitrogen, and stored in small aliquots at -80 °C.

#### 2.2.4 Purification of His-tagged SecA proteins with the wild-type N-terminus

Since the His-tagged SecA constructs with the wild-type N-terminus behaved differently during the purification process using the abovementioned common purification techniques (See Appendix A for more details), after several purification attempts a modified method was devised as follows.

*E. coli* cell disruption and preparation of soluble lysate was performed essentially similar to the previous section. Then, the histidine-tagged proteins were captured by the IMAC method, as described before. The affinity-purified proteins were supplemented with 3 mM EDTA and 5 mM DTT and then spin down at 35,000 rpm (45Ti rotor) for 45 min at 4 °C. The clear supernatant was concentrated to 4 mL at 4 °C using a 15 mL Amicon Ultra Centrifugal Filters device (Millipore) with a MW cutoff of 10,000 kDa. Thereafter, the Superdex 200 preparative grade column (2.6 x 65 cm) was pre-equilibrated with two column volumes of running buffer (50 mM Tris-HCl, 500 mM NaCl, 3 mM EDTA, and 20 mM  $\beta$ -mercaptoethanol at pH 7.5) at flow rate of 2.5 mL/min (~26 cm/hr). The concentrated His-tagged protein (~ 3 mL) was injected into the column at a flow rate of 1.5 mL/min and the fractions containing the soluble target protein were collected and analyzed by SDS-PAGE, as described before. Thereafter, the affinity tag was removed by TEV protease as mentioned before. Finally, after removing the histidine tag, SecA proteins were applied once again to the preparative size exclusion chromatography to remove

any remaining aggregates. The fractions that contained the clarified soluble SecA protein were combined and concentrated using centrifugal ultrafiltration, aliquoted, flash-frozen in liquid N<sub>2</sub>, and stored at -80 °C.

## 2.2.5 Purification of non-His-tagged SecA proteins

All the non-histidine tagged SecA proteins were purified from *E. coli* by the following methods.

### 2.2.5.1 Crude extraction and ammonium sulphate precipitation

Cell lysis was carried out as described in the previous section with a minor modification that the cell lysis was performed in a different buffer (100 mM Tris-HCl, pH 8.0). After centrifugation of the cell lysate at 38,000 rpm for 60 min using a 45Ti rotor, ammonium sulfate precipitation was performed. PMSF was added throughout this procedure at a final concentration of 1 mM. Briefly, supernatants were pooled and incubated in an ice bath for 20 min while rapidly stirred. Then, 29.5 g of ammonium sulfate was added per 100 mL of cell lysate (which will give a 50% saturation at 0 °C), which was in the ice water bath. Ammonium sulfate was added gradually to the solution, especially the second half, as this was when the proteins began to precipitate. Thereafter, the solution was stirred on ice for another 30 min to fully equilibrate. The resulting solution was transferred to pre-chilled centrifuge tubes and spin down at 12,000 rpm (45Ti rotor) for 45 min. The supernatant was removed and 25 mL of 100 mM Tris-HCl, pH 8.0 was added to each white pellet in the centrifuge tubes. Tubes were put on their sides on ice on an orbital shaker for almost an hour to resuspend the pellets. The resulting crude extracts that contained SecA protein were flash frozen in liquid nitrogen and stored in -80 °C.

### 2.2.5.2 Hydrophobic interaction chromatography

A 2.6 x 25 cm Phenyl Sepharose FF (Hi-Sub) column was used for separation of SecA molecules from the crude extract obtained in the previous step. First, the column was pre-equilibrated with running buffer (50 mM Tris-HCl, 1 M KCl, pH 8.0) at 5 mL/min. Meanwhile, the frozen crude extract was thawed in cold water, and was supplemented with 1 M KCl. To remove any solids, the crude extract was centrifuged at 20,000 rpm for 30 minutes. Then the clarified crude extract was applied to the column at the same flow rate. The flow through was saved, and the column

was washed with the running buffer until the absorbance at 280 nm returned to the baseline. The bound proteins were eluted afterward by using elution buffer (10 mM Tris-HCl, pH 8.0) in a single step. All the protein fractions were analyzed by SDS-PAGE. The elution fraction that contained the SecA protein was kept on ice for further purification.

### 2.2.5.3 Cibacron blue affinity chromatography

Cibacron blue resin (that was prepared as described before) was packed in a 2.6 x 25 cm column (120-140 mL), and was used for affinity purification of SecA proteins. The column was equilibrated with buffer A (50 mM Tris-HCl, pH 8.5) at 5 mL/min. Then the elution fraction from the last step (hydrophobic interaction chromatography) that was supplemented with 200 mM KCl was applied to the column at the same flow rate. The column was washed with 10% buffer B (buffer A with 1.8 M KCl) until the absorbance at 280 nm returned to baseline. Both the flow through and the wash through were saved for SDS-APGE analysis. The bound proteins were then eluted with a 10-100% buffer B gradient over 1.2 L. 10-20 mL fractions were collected. Subsequently, the elution fractions that contained the desired SecA protein, as judged by SDS-PAGE, were pooled and dialyzed against 50 mM Tris-HCl, 100 mM NaCl, pH 7.5 in preparation for the ion exchange chromatography.

### 2.2.5.4 Ion exchange chromatography

As a final polishing step for purification of these SecA proteins ion exchange chromatography was used, which was performed essentially as described previously. Purified SecA proteins were then aliquoted, flash-frozen in liquid N<sub>2</sub>, and stored at -80 °C.

## 2.2.6 Analytical size exclusion chromatography

Analytical size exclusion chromatography was carried out using a Superdex 200 HR 10/30 column. The running buffer was 50 mM Tris-HCl pH 7.5, 100 mM KCl, 1 mM EDTA, 5 mM MgCl<sub>2</sub>, and 5 mM β-mercaptoethanol, and the column was equilibrated at flow rate of 0.7 mL/minute at room temperature. Samples (50 μL) of the analyte or molecular weight standards were injected onto the column and absorbance was monitored at 280 nm. The molecular weight standards used were obtained from Sigma and included catalase (250 kDa), alcohol

dehydrogenase (150 kDa), BSA (66 kDa), and carbonic anhydrase (29 kDa). Acetone (10 mg/mL) and Blue Dextran 2000 (GE Healthcare; 1 mg/mL) were used to determine the included and void volumes, respectively.

## 2.2.7 Analytical ultracentrifugation

### 2.2.7.1 Sedimentation velocity analysis

For performing the sedimentation velocity experiments, all the protein samples were extensively dialyzed against a buffer containing 50 mM Tris pH 7.5, 100 mM KCl, 2 mM EDTA, and 5 mM MgCl<sub>2</sub>. This buffer also included 5 mM TCEP-HCl to ensure the full reduction of the free Cys residues. The final dialysis buffer was used as the “reference solution” in the centrifugation experiments.

The ultracentrifugation experiments were conducted at 20 °C in a Beckman Optima XL-A analytical ultracentrifuge using an An-60 Ti four-place analytical rotor. Standard two channel (double-sector) epon-charcoal centerpieces equipped with quartz windows were used to run the experiments. Typically, 380 and 400 μL of protein solution (at various concentrations) and the dialysis solution were injected into the sample and reference cells, respectively. After thermal equilibration of the rotor, samples were subjected to a high speed ranging from 25,000 rpm to 40,000 rpm depending on the expected size of the sedimenting protein species (higher speeds for smaller protein species such as SecA-N68ΔNC (approximately 65 kDa), while using slower speeds for larger proteins such as dimeric SecA-N95 (approximately 190 kDa)).

Absorbance measurements were made at 280 nm where possible. Alternatively, for more concentrated protein samples, longer wavelengths (295 nm or higher) were used to adjust the initial absorbance readings into the range from 0.15 to 0.6. Thereafter, the absorbance measurements were collected in a 0.002 cm radial step and averaged over three readings. Overall 30 scans were collected with intervals of 10 min.

Afterwards, the data were processed using software package SedFit. The partial specific volume ( $\bar{V}$ ) of each protein was calculated from their amino acid composition by SEDNTERP



software. Using the same software, the solvent viscosity and density were calculated to be 0.01 and 1.005 g/mL, respectively. The reported *s*-values throughout the thesis are observed sedimentation coefficient in experimental conditions (*s*<sub>obs</sub>).

### 2.2.7.2 Sedimentation equilibrium experiments

Sedimentation equilibrium experiments were performed employing the same ultracentrifuge and rotor type, but this time six-channel cells with epon-charcoal centerpieces equipped with quartz windows (path length of 1.2 cm) were used. The buffer used for sedimentation velocity was also used for the equilibrium experiments. First, the protein solutions were dialyzed exhaustively against this buffer, and then, equilibrium was reached at rotor speeds of 7,000 rpm, 10,000 rpm, 12,000 rpm, and 16,000 rpm at 20 °C. After reaching equilibrium, the absorbance data were collected at 0.002 cm radial steps and averaging over ten readings. The absorbance data were analyzed using models built in Prizm 5 (Graphpad) with the following equations (Briere & Dunn, 2006).

A single ideal protein model was defined by equation 1:

#### Equation 1. Single ideal species model

$$C = C_o \cdot \exp\left[\frac{\omega^2}{2RT} \cdot M_{obs}(1 - \bar{v}\rho) \cdot (x^2 - x_o^2)\right] + I_o$$

In this expression, *C* is the concentration at radius *x*, *C*<sub>o</sub> refers to the concentration at reference radius *x*<sub>o</sub>, *ω* is the angular velocity of the rotor, *v̄* refers to the partial specific volume of the analyzed protein, *M*<sub>obs</sub> is the molecular weight of the protein, *ρ* is the solvent density, *T* refer to the temperature in degrees Kelvins, *R* is the ideal gas constant, and *I*<sub>o</sub> is the baseline offset.

Then, for the self-associating proteins, a series of protomer: *n*-mer models (i.e. monomer-dimer, monomer-tetramer, etc.) with different values of *n* were also built; the association constants, *K*<sub>A</sub>, for these models would be described by equation 2:

### Equation 2. Association constants ( $K_A$ )

$$K_A = \frac{C_{n-mer}}{(C_{monomer})^n}$$

Equilibrium data of SecA proteins were then fit to the protomer:  $n$ -mer models using the following expression (equation 3):

### Equation 3. Protomer: $n$ -mer model

$$C = \left\{ C_o \cdot \exp \left[ \frac{\omega^2}{2RT} \cdot M (1 - \bar{v}\rho) \cdot (x^2 - x_o^2) \right] \right\} + \left\{ C_o^n \cdot K_A \cdot \exp \left[ \frac{\omega^2}{2RT} \cdot nM (1 - \bar{v}\rho) \cdot (x^2 - x_o^2) \right] \right\} + I_o$$

In this equation,  $M$  is the molecular weight of the protomer and other terms are as described before. The obtained values by this equation which were in absorbance units, were then converted to the molar dissociation constant in molar by the equation 4:

### Equation 4. Conversion of $K_A$ to $K_D$

$$K_D = \frac{1}{K_A(C)} = \frac{2}{\epsilon d} \cdot \frac{1}{K_A(A)} = \frac{1}{0.6\epsilon} \cdot \frac{1}{K_A(A)}$$

In this expression, which was used for calculating the  $K_D$  of dimerization,  $K_A(A)$  and  $K_A(C)$  refer to the association constants in terms of absorbance and molar, respectively.  $\epsilon$  is the extinction coefficient of monomer. Pathlength,  $d$ , is 1.2 cm for the cells used in the sedimentation equilibrium experiments.

## 2.2.8 ATPase measurements

A coupled phosphate detection assay was used to determine the kinetics of SecA-mediated ATP turnover in real-time (Rieger *et al*, 1997). In this test, purine nucleoside phosphorylase (PNPase) is used for consuming the phosphates produced by ATPase activity of SecA to convert 7-methylinosine ( $m^7Ino$ ) into hypoxanthine, which will results in a decrease of 2.1 absorbance

units at 291 nm per millimolar change in concentration of phosphate. The buffer assay either consisted of 50 mM Tris pH 7.5, 100 mM KCl, and 10 mM MgCl<sub>2</sub> for SecA-N95(T109N) and SecA-N95ΔN(T109N), or 100 mM HEPES, 10 mM MgCl<sub>2</sub>, 50 mM KCl, pH 7.5 for SecA(T109N). The reactions also contained 1 unit of PNPase (Sigma), 500 μM m<sup>7</sup>Ino, and ADP at different concentrations. Assays were performed at 20 °C in a 2 mL volume, with stirring, using a Cary 100 spectrophotometer. Enough SecA protein was added so that there was a measureable change in absorbance, but less than 10% of the ATP was consumed during the reaction. This corresponded to 50 μg for SecA-N95(T109N) and SecA-N95ΔN(T109N) or 25 μg for SecA(T109N) per milliliter. Before starting the reactions by adding ATP, they were allowed to thermally equilibrate for 4 minutes, and then followed for 5 minutes. Initial rates, in terms of ΔAbs/s, were obtained by least-squares fits using the Cary WinUV software. The rates were converted to a change in phosphate concentration using the differential extinction coefficient of 2100 M<sup>-1</sup>cm<sup>-1</sup>. The initial rate data were fitted to kinetic models using *Prism* (Graphpad Software).

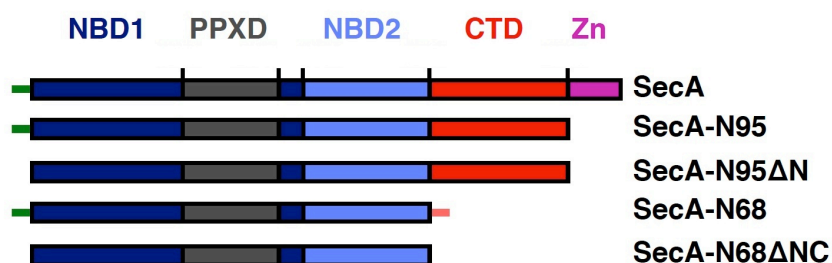
### 2.2.9 *In vivo* SecA complementation assay

pAK-SecA-N95ΔN14 and pJZ7-N95 were transformed into *E. coli* BL21.19(DE3) strain [*secA13(Am) supF(Ts) trp(Am) zch::Tn10 recA::CAT clpA::KAN*] (Mitchell & Oliver, 1993) for *in vivo* analysis as described before (Jilaveanu *et al*, 2005). Briefly, these constructs were grown in LB/Amp/Kan media for 12 hours at 27 °C while shaking. As controls, BL21.19(DE3) strain without any plasmid and BL21(DE3) carrying pJZ7-N95 were grown in LB/Kan and LB/Amp media, respectively, under the same condition. Then, the cell cultures were normalized to an OD<sub>600</sub> of 1 by adding LB broth. Afterwards, all the cell cultures were serially diluted 1:10 with LB broth to make 6 concentrations. 5 μL of each dilution was placed on LB agar plates containing 100 μg/mL and 35 μg/mL of ampicillin and kanamycin, respectively. The plates were incubated at 28 °C or 42 °C for 30 hrs.

## 2.3 Results

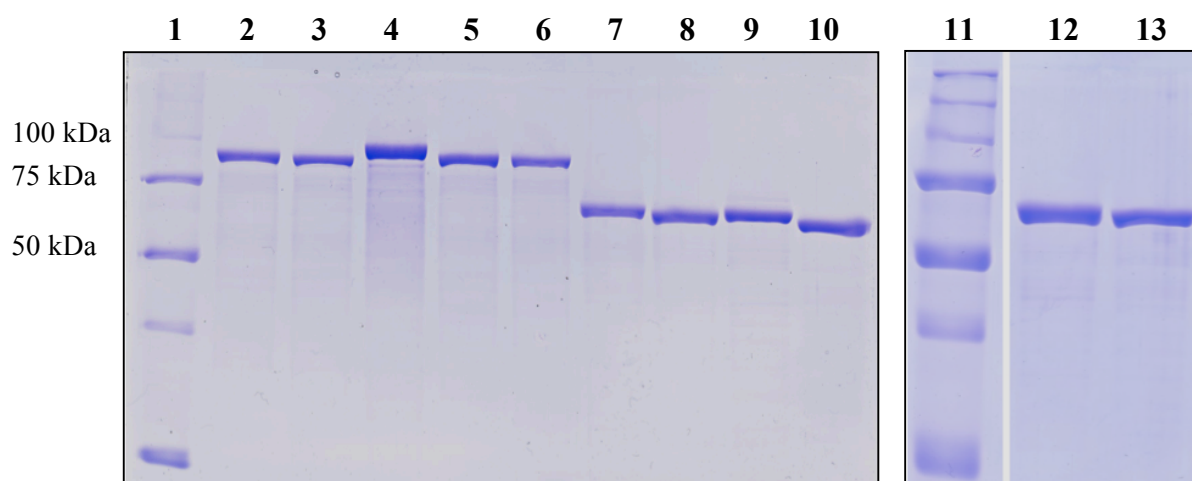
### 2.3.1 SecA expression and purification

For the structural and functional studies of SecA oligomerization, the complete list of constructs used is provided in Table 2-2. Domain arrangement of SecA constructs is also presented in Figure 2-1. All of the SecA proteins were purified to near homogeneity as described in the Methods section. The purity of the preparations was assessed by dilution to between 200 and 400  $\mu\text{g/mL}$  and analysis using 10% SDS-polyacrylamide gels. Figure 2-2 shows the Coomassie-stained gels of the purified SecA constructs.



**Figure 2-1. SecA constructs**

Full-length SecA consists of nucleotide binding domain 1 (NBD1), the preprotein cross-linking domain (PPXD), NBD2, the C-terminal domains (CTD), and the zinc binding domain (Zn), which consists of a 22 residue zinc-binding motif at the extreme C-terminus, connected to residue 835 by an unstructured linker. The unstructured N-terminus of SecA is illustrated by the short green line. The unstructured polypeptide at the C-terminal region of SecA-N68 and SecA-DM constructs is showed by an orange line. For the complete list of constructs, refer to table 2-2.



**Figure 2-2. Purified SecA proteins used in this study**

SecA protein constructs were purified as described in the Material and Methods section. Coomassie blue stained SDS-PAGE gels (10%) are shown here. *From left to right:* Lane 1 is the molecular weight marker (Precision Plus Protein Standards, Bio-Rad); Lane 2, SecA-N95; Lane 3, SecA-N95ΔN; Lane 4, SecA(T109N); Lane 5, SecA-N95(T109N); Lane 6, SecA-N95ΔN(T109N); Lane 7, SecA-N68; Lane 8, SecA-N68ΔC; Lane 9, His<sub>6</sub>-SecA-N68; Lane 10, SecA-N68ΔNC. Lanes 11, 12, and 13 are molecular weight markers, SecA-N68ΔN, and SecA-N68ΔNC(ER), respectively.

### 2.3.2 Oligomerization of SecA-N68 is mediated by its unstructured termini

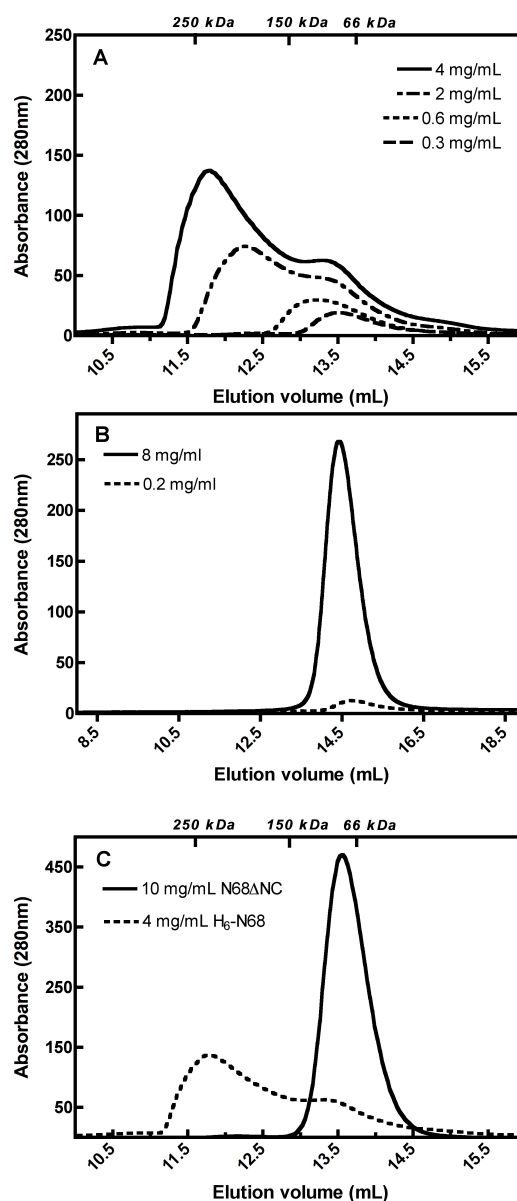
Previous efforts to obtain diffraction-quality crystals of dimeric SecA were not successful, and therefore to create a monomeric SecA for crystallization and structural studies, the C-terminal “dimerization” domains of SecA were removed by truncation at residue 610 to produce H<sub>6</sub>-SecA-N68 (Dempsey *et al*, 2002). The truncation clearly affected oligomerization, but not in the manner that was expected: H<sub>6</sub>-SecA-N68 participated in a monomer-tetramer equilibrium in solution with a  $K_D$  of 63  $\mu$ M for the formation of the tetramer (Dempsey *et al*, 2002). Unfortunately, H<sub>6</sub>-SecA-N68 did not yield diffraction-quality crystals; however, crystal structure analysis of a related construct in which the “preprotein cross-linking domain” (PPXD) was deleted, “H<sub>6</sub>-SecA-DM”, indicated that both the extreme N- and C-terminal regions were unstructured (Nithianantham & Shilton, 2008). To be specific, the H<sub>6</sub>-SecA-DM construct had an extreme N-terminal sequence MHHHHHHLTKVFGSRNDRTL (the wild-type sequence is MLIKLLTKVFGSRNDRTL) but electron density was evident only for residues from S12 onwards, indicating the first 11 residues were disordered in the crystal. Similarly, at the extreme C-terminal sequence, MEDALMRIFASDRVSGMMRK (residues 590 to 610) only residues up to M595 were apparent in the electron density maps. These unstructured residues are underlined.

Unstructured regions are generally not favorable for crystallization, and therefore to obtain diffraction-quality crystals of SecA-N68 a new construct was created, SecA-N68 $\Delta$ NC, in which the unstructured regions at both N- and C-termini were removed. The SecA-N68 $\Delta$ NC construct comprises residues 15 to 590 of SecA, with an additional GA sequence at the N-terminus remaining after removal of the affinity tag. Surprisingly, the purified SecA-N68 $\Delta$ NC appeared to be completely monomeric even at high concentrations. The oligomeric states of H<sub>6</sub>-SecA-N68 and SecA-N68 $\Delta$ NC were analyzed by size exclusion chromatography (Figure 2.3). Consistent with previous work, H<sub>6</sub>-SecA-N68 eluted with a molecular weight varying from 68 kDa to 260 kDa, depending on the concentration (Figure 2.3A), which is consistent with a monomer-tetramer equilibrium, while SecA-N68 $\Delta$ NC eluted at 68 kDa irrespective of the concentration. In fact, SecA-N68 $\Delta$ NC eluted as a monomer at 20 mg/mL, the highest concentration tested,

indicating that without the flexible termini, the SecA-N68 protein has no tendency to self-associate.

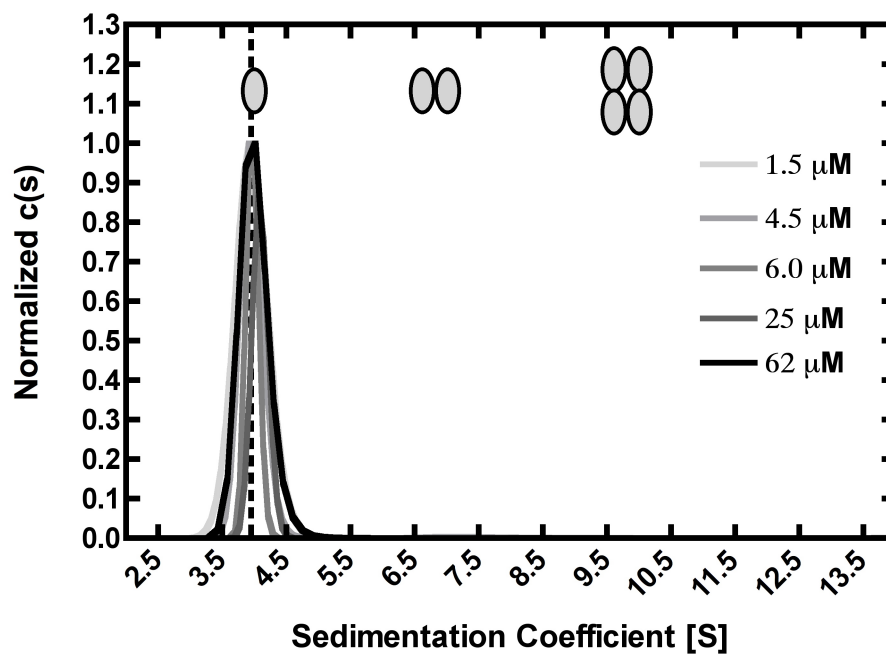
The monomeric nature of SecA-N68 $\Delta$ NC was further characterized by sedimentation velocity experiments performed at protein concentrations from 0.1 mg/mL (1.5  $\mu$ M) to 4.0 mg/mL (62  $\mu$ M; Figure 2-4). SecA-N68 $\Delta$ NC sedimented with a coefficient of 3.9 S, which did not change over the range of protein concentrations used. The SecA-N68 $\Delta$ NC monomer has a frictional ratio ( $f/f_{min}$ ) of 1.31 indicating a globular shape in solution.





**Figure 2-3. Size exclusion chromatography of H<sub>6</sub>-SecA-N68 and SecA-N68ΔNC**

The oligomerization of H<sub>6</sub>-SecA-N68 and SecA-N68ΔNC was examined by analytical size exclusion chromatography. The position of elution for several molecular weight standards is indicated on the top axis of the plots. (A) Elution of H<sub>6</sub>-SecA-N68 at concentrations of 0.3, 0.6, 2, and 4 mg/mL. (B) Elution of SecA-N68ΔNC at 0.2 and 8 mg/mL. (C) The elution of H<sub>6</sub>-SecA-N68 at 4 mg/mL is overlaid with that of SecA-N68ΔNC at 10 mg/mL to facilitate comparison of their chromatographic behavior. Note that the results presented in panel B are obtained from experiments on a different size exclusion chromatography column.



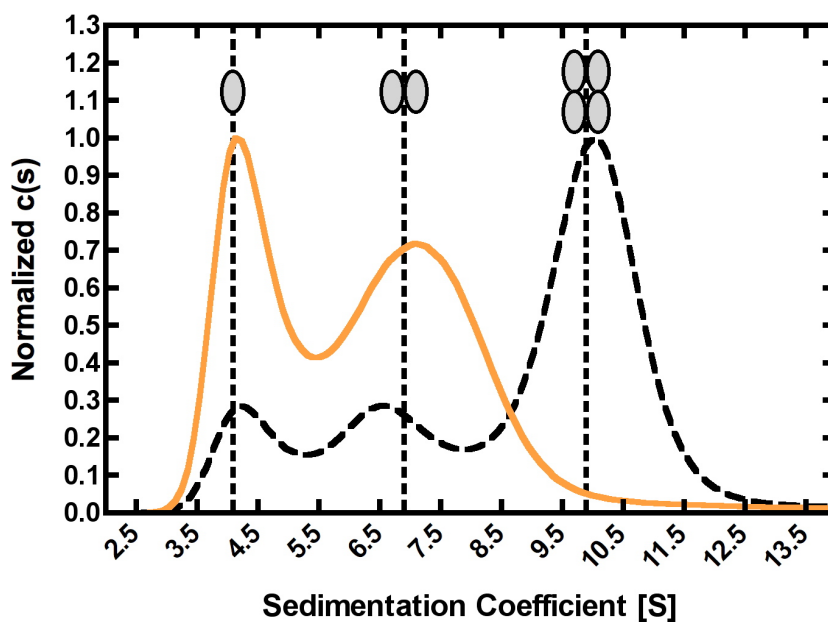
**Figure 2-4. Sedimentation velocity analysis of SecA-N68ΔNC**

The solution behaviour of SecA-N68ΔNC was characterized by sedimentation velocity at a range of concentrations from 1.5 μM (0.1 mg/mL) to 62 μM (4 mg/mL). The dashed line shows the predicted sedimentation coefficient of a 65 kDa globular protein. The oval shapes represent the expected sedimentation coefficients for monomeric, dimeric, and tetrameric SecA-N68. S-values are observed sedimentation coefficient (sobs). The distribution plots are normalized by the amplitude.

### 2.3.3 H<sub>6</sub>-SecA-N68 forms monomers, dimers, and tetramers

SecA must interact with unstructured preprotein substrates during the translocation reaction, and therefore one expects that there will be site(s) on SecA that bind unstructured polypeptides in a manner relatively independent of the specific sequence. On this basis, the observation that oligomerization of H<sub>6</sub>-SecA-N68 is completely dependent on the presence of unstructured segments at its N- and C-termini warranted further investigation.

Previous sedimentation analysis indicated a monomer-tetramer equilibrium for H<sub>6</sub>-SecA-N68 (Dempsey *et al*, 2002). Sedimentation velocity experiments were repeated for the H<sub>6</sub>-SecA-N68 construct, this time incorporating a SedFit analysis (Figure 2-5). At a relatively high concentration (2.6 mg/mL), the monomeric and tetrameric species were evident with sedimentation coefficients of 3.9 S and 10 S, respectively, in line with previous sedimentation coefficients of 4.1 S and 8.9 S (Dempsey *et al*, 2002). In addition, there was a third species present sedimenting at 6.8 S, which likely represents dimers of H<sub>6</sub>-SecA-N68. At a lower concentration of 0.3 mg/mL (4 μM), the 10 S species was no longer present and H<sub>6</sub>-SecA-N68 sedimented as two peaks at 3.9 S and 7.1 S, indicating an equilibrium between monomeric and dimeric forms. This new analysis indicated the presence of H<sub>6</sub>-SecA-N68 dimers, and therefore the tetramer is most likely a dimer of dimers. This is consistent with the SAXS analysis of the tetramer: *ab initio* modeling of the tetramer using SAXS data indicated a particle with D2 rather than C4 symmetry (Dempsey *et al*, 2002).



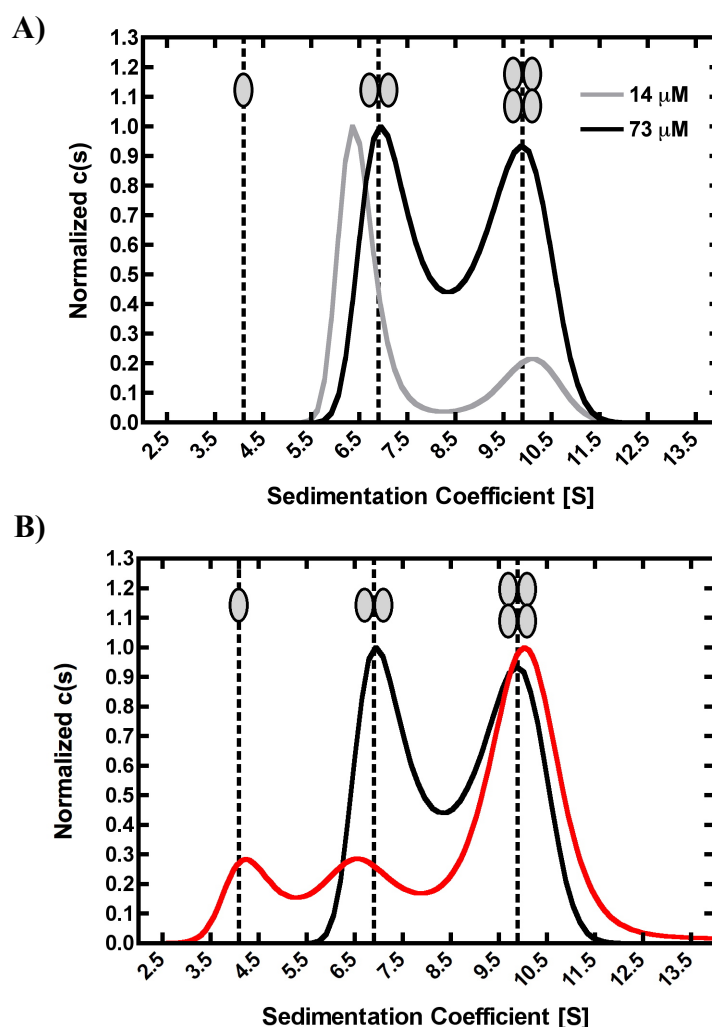
**Figure 2-5.  $H_6$ -SecA-N68 participates in a monomer-dimer-tetramer equilibrium**

The distribution plot for  $H_6$ -SecA-N68 was derived from SedFit analysis of sedimentation velocity experiments conducted at two different  $H_6$ -SecA-N68 concentrations: 38  $\mu$ M (2.6 mg/mL; dashed black curve) and 4  $\mu$ M (0.3 mg/mL; orange curve). The expected position of monomer, dimer, and tetramer are shown with dashed lines and oval shapes. S-values are observed sedimentation coefficient (sobs). The distribution plots have been normalized by the amplitude.

### 2.3.4 Roles of the N- and C-terminal sequences in SecA-N68 oligomerization

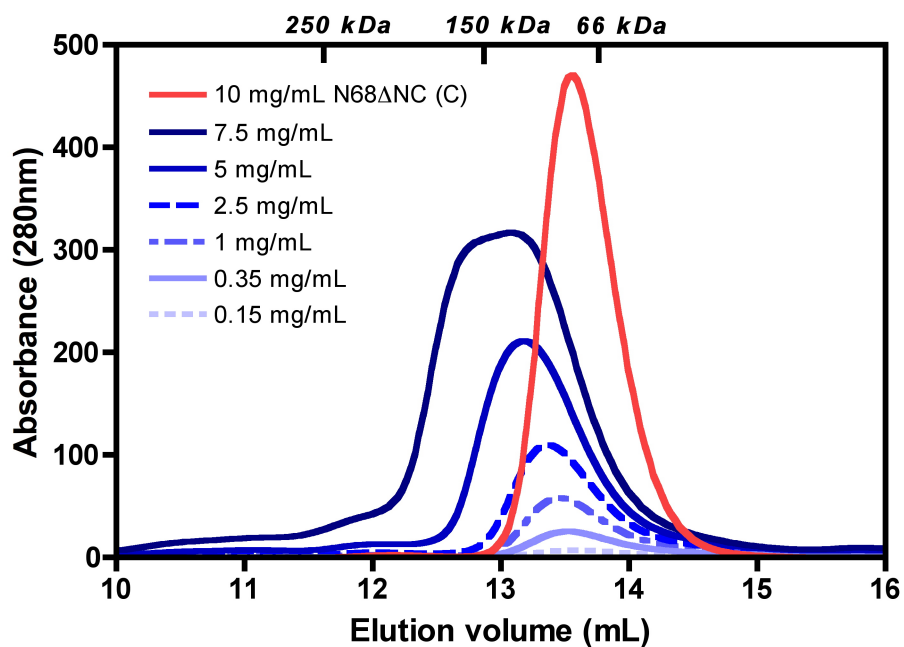
As noted previously, the N-terminus of H<sub>6</sub>-SecA-N68 contains a hexa-histidine tag that has replaced 5 residues of the wild-type SecA sequence. Given the importance of the N-terminus for the self-association of H<sub>6</sub>-SecA-N68, the oligomerization of a construct with a wild-type N-terminal sequence, SecA-N68, was studied. This construct was created by cloning the coding sequence for SecA residues 1 to 610 into an expression vector that includes a cleavable N-terminal hexahistidine affinity tag. The resulting protein carries the wild-type N-terminal sequence GAMLIKLLTKVF along with an extra dipeptide (GA) remaining after proteolytic removal of the affinity tag. When analyzed by sedimentation velocity, SecA-N68 participated in a dimer-tetramer equilibrium (Figure 2-6A), with the dimer predominating at a lower concentration (1 mg/mL), and no evidence of a monomeric species. The absence of a monomeric species is a notable difference from what was observed for H<sub>6</sub>-SecA-N68 (Figure 2-5 and 2-6B), where monomers were evident even at relatively high protein concentrations. The presence of monomeric SecA in solutions of H<sub>6</sub>-SecA-N68, but only dimeric and tetrameric SecA in solutions of SecA-N68, indicates that the wild-type N-terminal sequence is critical for dimer formation, and lends additional support for the idea that the SecA-N68 tetramer is a dimer of dimers.

To further characterize the roles of the N- and C-terminal sequences in oligomerization of SecA-N68, two additional deletion constructs were created: one that lacks the N-terminal unstructured region, SecA-N68 $\Delta$ N (residues 15 to 610), and one that lacks the C-terminal unstructured region, SecA-N68 $\Delta$ C (residues 1 to 590). Analytical size exclusion chromatography of SecA-N68 $\Delta$ C at varying concentrations from 0.15 to 7.5 mg/mL indicated that it eluted at different positions depending on the concentration (Figure 2-7). At the highest concentration tested (7.5 mg/mL, 110  $\mu$ M), SecA-N68 $\Delta$ C eluted as a broad peak with an estimated molecular weight of approximately 120 kDa. It is notable that a tetrameric species is not evident in the chromatogram of SecA-N68 $\Delta$ C, even at 7.5 mg/mL (Figure 2-7), while SecA-N68 clearly forms tetramers at concentrations above 1 mg/mL (Figure 2-3). On this basis, the C-terminal unstructured region is critical for tetramer formation.



**Figure 2-6. Effect of the hexahistidine tag on oligomerization of SecA-N68**

**(A)** Sedimentation velocity analysis was used to characterize the oligomerization of SecA-N68 which lacks the N-terminal hexahistidine affinity tag that is present in H<sub>6</sub>-SecA-N68. The distribution plot from SedFit analysis of two velocity runs at SecA-N68 concentrations of 14 μM (1 mg/mL; light grey curve) and 73 μM (5 mg/mL; black curve) is shown. **(B)** A comparison of the distribution plots of H<sub>6</sub>-SecA-N68 (38 μM; red curve) and SecA-N68 (73 μM; black curve). The two constructs differ only in their extreme N-terminal sequence, which is MHHHHHHLTKVF for H<sub>6</sub>-SecA-N68 and GAMLKLLTKVF for SecA-N68. The expected positions of monomer, dimer, and tetramers are shown with dashed lines and oval shapes. S-values are observed sedimentation coefficient (sobs). The distribution plots have been normalized by amplitude.

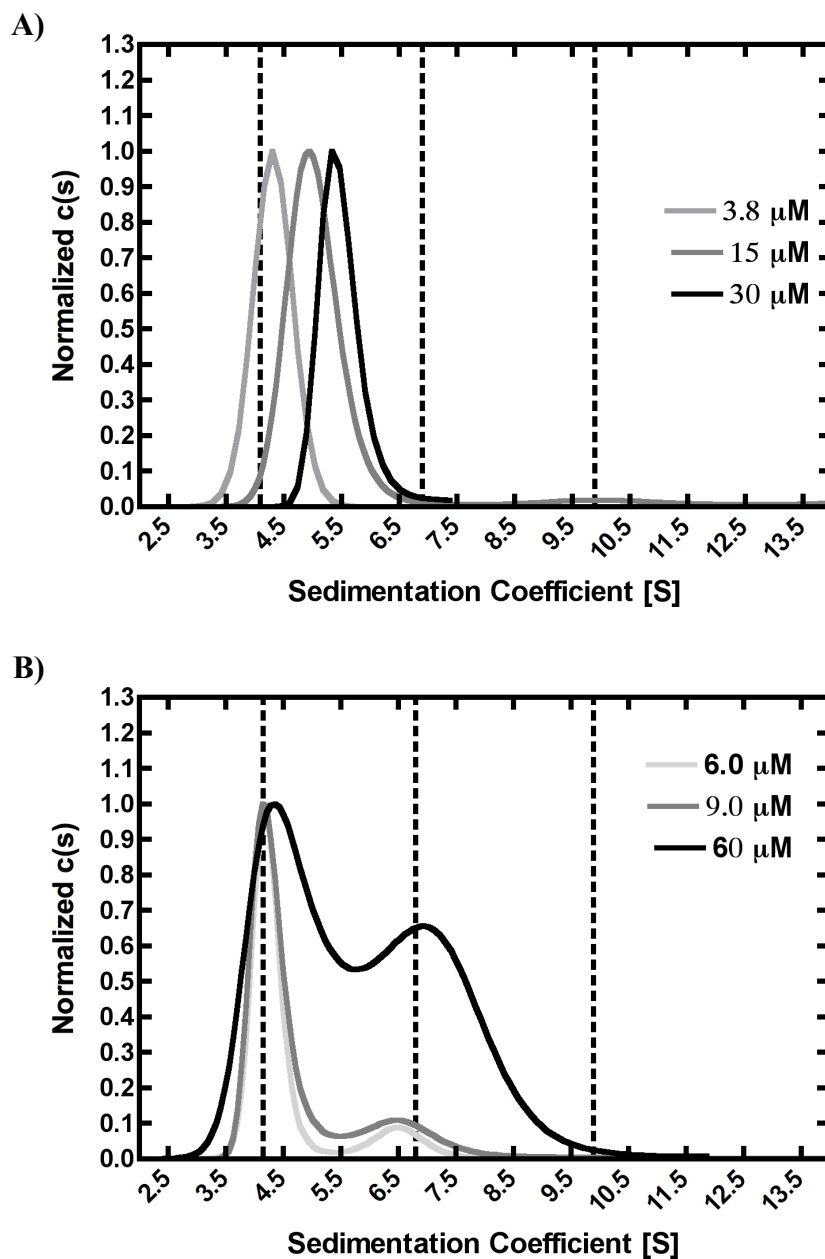


**Figure 2-7. Size exclusion analysis of SecA-N68ΔC**

SecA-N68ΔC lacks the N-terminal unstructured region that is present in SecA-N68. To assess its oligomeric state as a function of concentration, purified SecA-N68ΔC was analyzed by size exclusion chromatography at the indicated concentrations. For comparison, the elution of SecA-N68ΔNC (10 mg/mL) is indicated by the red curve. SecA-N68ΔNC lacks both unstructured termini and is purely monomeric in solution. The elution positions of the molecular weight standards are shown on the top axis of the plot.

Sedimentation velocity analysis of SecA-N68 $\Delta$ C and SecA-N68 $\Delta$ N indicated that both proteins participated in a monomer-dimer equilibrium, but the dimerization behaviour of the two constructs was clearly different (Figure 2-8). In particular, at any given concentration SecA-N68 $\Delta$ C shows a single sharp peak with a molecular weight between that of a monomer and dimer (Figure 2-8A). The observation of a single peak that shows concentration-dependent changes in sedimentation coefficient is indicative of an interaction with fast kinetics (i.e.  $k_{\text{off}} > 10^{-3} \text{ s}^{-1}$ ; (Brown *et al*, 2008; Balbo & Schuck, 2005). In contrast, SecA-N68 $\Delta$ N sediments as two distinct species with coefficients of 4.2 S and 6.6 S. The larger species corresponds to a dimer of SecA-N68 $\Delta$ N, and the relative amount of this species decreases at lower protein concentrations. Thus, the C-terminal domain mediates dimer formation with a relatively slow interconversion between monomer and dimer. Therefore, there appear to be two interaction sites on the SecA-N68 construct, and each one of them mediates dimerization of SecA-N68. One of these interaction sites binds preferentially to the N-terminal unstructured region of SecA and exhibits a fast exchange between monomer and dimer. The C-terminal unstructured region of SecA-N68 interacts with a separate site on SecA to form a dimer that exhibits relatively slow exchange with the monomer.





**Figure 2-8. Roles of unstructured termini in oligomerization of SecA-N68**

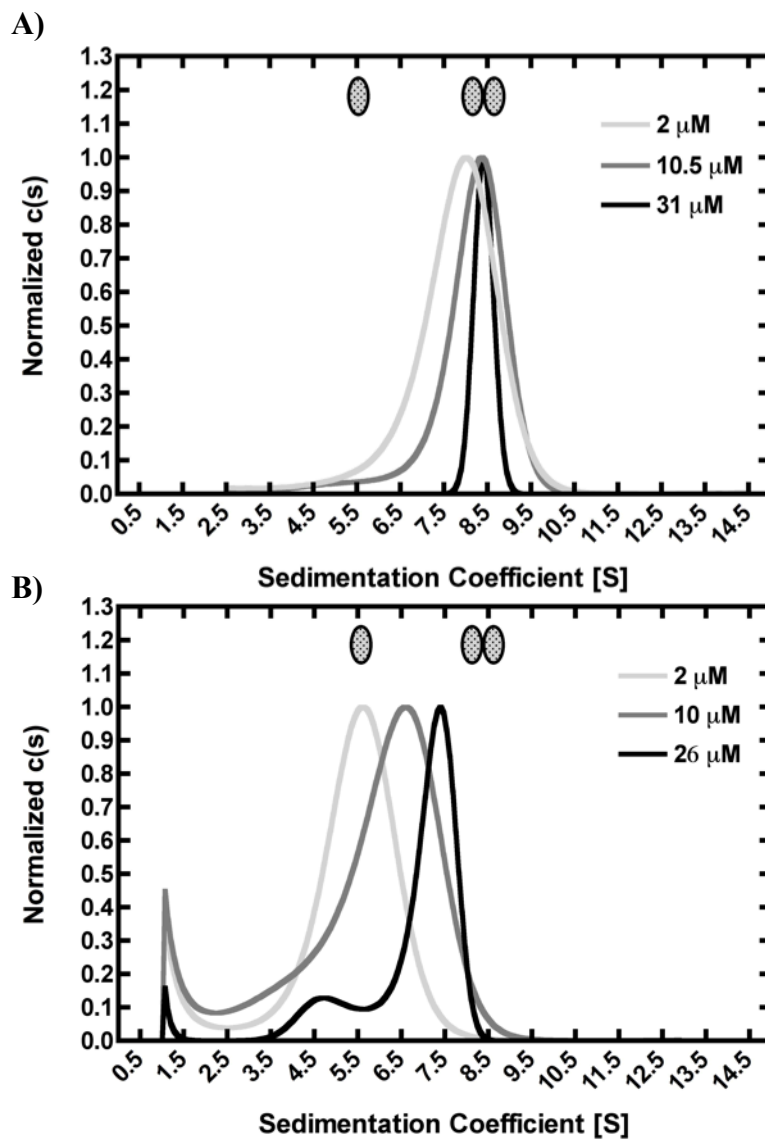
The roles of the extreme N- and C-terminal sequences in the oligomerization of SecA-N68 was assessed by sedimentation velocity analysis of (A) SecA-N68ΔC, which lacks the extreme C-terminal region, and (B) SecA-N68ΔN, which lacks the extreme N-terminal region. Sedimentation velocity experiments were conducted at the indicated protein concentrations, and the resulting distribution plots from SedFit analysis are illustrated. The estimated positions monomeric (4.1 S), dimeric (7 S), and tetrameric (10 S) species of SecA-N68 are indicated with dashed lines. S-values are observed sedimentation coefficient (sobs). The distribution plots have been normalized by amplitude.

### 2.3.5 The role of the extreme N-terminus in dimerization of SecA-N95

SecA is present in solution as a dimer that associates with a  $K_D$  in the low nanomolar range (Gold *et al.*, 2007), depending on the ionic strength of the solution. For example, a  $K_D$  of 0.74 nM was reported when SecA dimerization was analyzed by fluorescence cross-correlation spectroscopy in a low ionic strength buffer (Kusters *et al.*, 2011); sedimentation equilibrium analysis indicated a  $K_D$  of 14 nM in 100 mM KCl, rising to 40.4  $\mu$ M in 500 mM KCl (Wowor *et al.*, 2011). Full-length SecA is 901 residues long, with an unstructured region beginning at residue 835, which links the main body of the SecA protomer to a 22 residue zinc binding domain that interacts with the translocation-specific chaperone SecB (Bechtluft *et al.*, 2010; Zhou & Xu, 2003). Removal of the unstructured C-terminus and zinc-binding domain yields a SecA construct, SecA-N95, comprising residues 1 to 835, that is fully functional for protein translocation *in vivo* (Or *et al.*, 2005).

To assess the effect of the N-terminus on oligomerization of SecA-N95, sedimentation velocity experiments were conducted for SecA-N95 and SecA-N95 $\Delta$ N, which comprises residues 15 to 835 (Figure 2-9). At concentrations above approximately 10  $\mu$ M, SecA-N95 sedimented with a coefficient of 8.4 S, which is consistent with a dimer. At the lowest concentration tested, 2  $\mu$ M (0.2 mg/mL), SecA-N95 sedimented with only a slightly lower coefficient (approximately 8.0 S) indicating that the dimer was still the major species in solution (Figure 2-9A). Thus, SecA-N95, like full length SecA, forms a dimer in solution with high affinity.

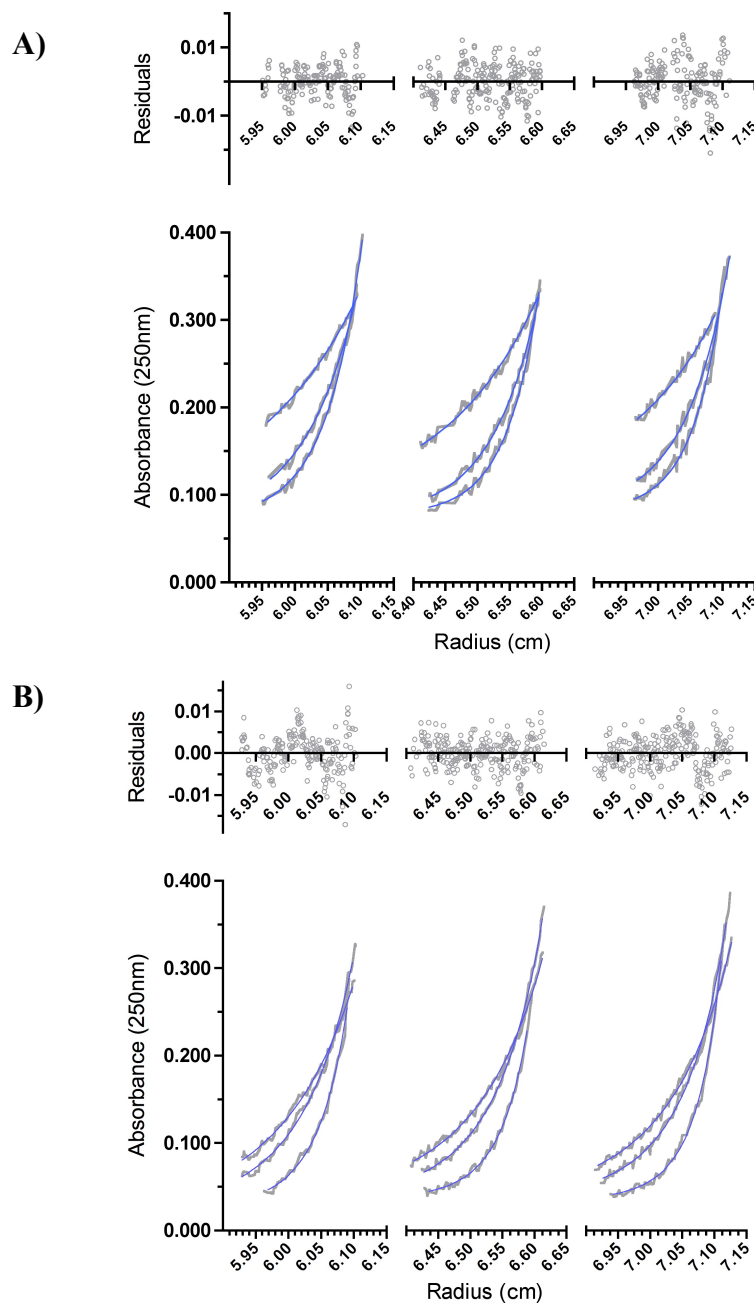
SecA-N95 $\Delta$ N sedimented with lower coefficients than SecA-N95 (Figure 2-9B). For example, at a concentration of 2  $\mu$ M (0.2 mg/mL), SecA-N95 $\Delta$ N sedimented with a coefficient of 5.6 S, which is consistent with a molecular weight of 97 kDa, compared to 8.0 S for SecA-N95. At a higher concentration of 26  $\mu$ M (2.5 mg/mL), the majority of SecA-N95 $\Delta$ N sedimented at 7.5 S, with a smaller species at 4.6 S; in contrast, SecA-N95 which at 31  $\mu$ M sedimented as a single species with coefficient of 8.4 S. These results indicate that deletion of the extreme N-terminus has weakened the SecA-N95 solution dimer.



**Figure 2-9. Effect of the extreme N-terminus on oligomerization of SecA-N95**

Sedimentation velocity analysis of SecA-N95 and SecA-N95ΔN was performed at the indicated protein concentrations. **(A)** Distribution plot for SecA-N95 at 2 μM, 10.5 μM, and 31 μM concentrations. **(B)** Distribution plot of SecA-N95ΔN concentrations of 2 μM, 10 μM, and 26 μM. S-values are observed sedimentation coefficient (sobs). The distribution plots have been normalized by amplitude.

To quantify the changes in the SecA-N95 dimer dissociation constant, sedimentation equilibrium experiments were conducted with SecA-N95 and SecA-N95 $\Delta$ N. These experiments were carried out under the same buffer and experimental conditions used for sedimentation velocity experiments (i.e. 50 mM Tris-HCl, 100 mM KCl, 2 mM EDTA, 5 mM MgCl<sub>2</sub>, and pH 7.5). SecA-N95 $\Delta$ N samples at a concentration of 0.25 mg/mL were brought to equilibrium at three different rotor speeds (10,000 rpm, 12,000 rpm, and 16,000 rpm). Fitting the recorded absorbance data to a single species model gave a best-fit molecular weight of 118.3 kDa, which is between monomer and dimeric species (not shown here). Fitting the same data to a global monomer dimer equilibrium model yielded an excellent fit to the entire data set, with a  $K_D$  of 25.4  $\mu$ M (Figure 2-10A). SecA-N95 was analyzed at the same concentration and rotor speeds of 7,000 rpm, 10,000 rpm, and 12,000 rpm. Data were fitted to a single ideal species model yielding a best-fit molecular weight of 178.6 kDa, relatively close to the molecular weight of the SecA-N95 dimer, 189.2 kDa, indicating that SecA-N95 protein exists chiefly as dimers in solution (Figure 2-10B). The residuals of this fit were distributed randomly, indicating that there was no systematic deviation from the single-species model used. Attempts to fit these data to a monomer-dimer equilibrium were not successful because the monomers of SecA-N95 were in small amounts under these conditions that they could not be detected with sufficient accuracy. Based on this, we expect the  $K_D$  of SecA-N95 to be at or below the high nanomolar range. Thus, absence of the N-terminus weakens the dimerization of SecA-N95 significantly.



**Figure 2-10. Sedimentation equilibrium analysis of SecA-N95 and SecA-N95ΔN**

To compare and quantify the dimerization of SecA-N95 (**A**) and SecA-N95ΔN (**B**), sedimentation equilibrium analysis was conducted under identical buffer conditions. Gray lines represent the experimental absorbance data, while the curves of best fit are shown in blue. The gray circles on the top panels shows the residuals of the curves of best fit, which are an indication of goodness of the applied model to the experimental data. **Panel (A)** shows the global fitting of SecA-N95 equilibrium data to a single ideal species with a molecular weight 178.6 kDa. **Panel (B)** represents the global fitting of SecA-N95ΔN equilibrium data to monomer-dimer equilibrium with a  $K_D$  of 25.4  $\mu\text{M}$ .

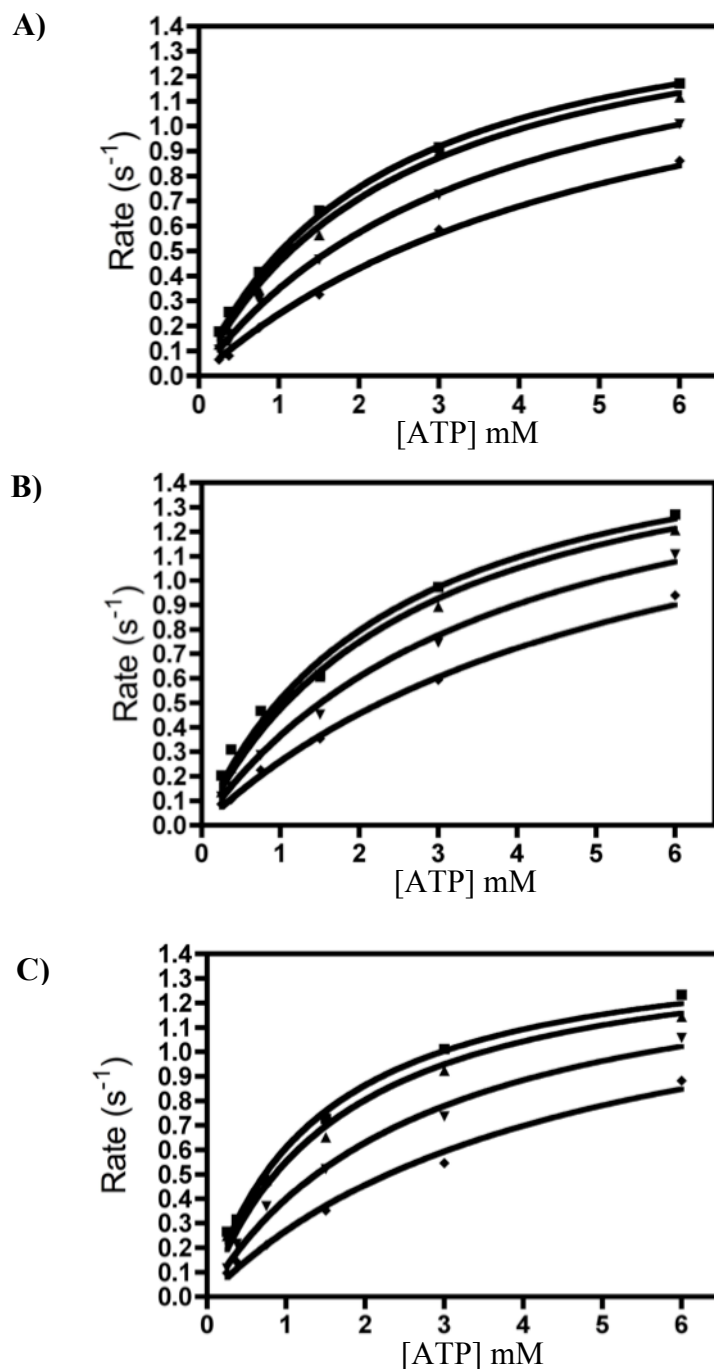
### 2.3.6 Solution ATPase activity of dimeric and monomeric SecA-N95

SecA exists as a dimer in solution, but the significance of this solution dimer is a matter of debate. Previously, by comparing the kinetic parameters of a partially monomeric SecA (His<sub>6</sub>-SecA(T109N)) and a dimeric SecA (SecA(T109N)), it was reported that the protomers of SecA regulate the ATPase activity of each other in solution (Dempsey, B., 2007, PhD. thesis). On this basis, to test whether the dimer formation of SecA due to the presence of the N-terminal segment has an effect on the ATPase activity of this molecule in solution, kinetic analysis was performed on the monomeric (SecA-N95ΔN) and the dimeric proteins (SecA-N95 and full-length SecA) as described previously (Dempsey, 2007). The rationale was that if the two protomers hydrolyze ATP completely independently from each other, then ADP should act as a competitive inhibitor; on the other hand, if the two protomers are somehow allosterically coupled, a different mode of inhibition by ADP would be observed.

A real-time coupled assay was used to measure the kinetic parameters of the SecA constructs. In this assay, phosphate produced by ATP hydrolysis is consumed by purine nucleoside phosphorylase (PNPase) to convert 7-methylinosine (m<sup>7</sup>Ino) into hypoxanthine. Furthermore, since these SecA constructs have a low basal ATPase activity, we used a mutant of SecA, T109N (Mitchell & Oliver, 1993), that has an elevated basal ATPase activity. The SecA(T109N) mutant has approximately 35% of the *in vitro* translocation activity of wild-type SecA, and it has been shown that the overall shape and solution structure of SecA remains unaffected by the mutation (Shilton *et al*, 1998); therefore, the SecA(T109N) retains the essential features of wild-type SecA.

The solution ATPase activity of SecA constructs was measured in the presence of increasing concentrations of ADP. Global fitting of the SecA-N95(T109N) and SecA-N95ΔN(T109N) data at all ADP concentrations using the Michaelis-Menten equation incorporating competitive inhibition yielded excellent fits to each data set (Figure 2-11A and B). Similarly, the kinetics data of SecA(T109N) were globally fitted to a competitive inhibition model (Figure 2-11C). Global fitting of all the data to the non-competitive and uncompetitive models yielded systematic deviations and overall poor fits to the data (data not shown). Kinetics constants derived from

these global fits incorporating competitive inhibition are shown in Table 2-4. Thus, based on these data it is evident that the presence of N-terminus and dimer formation does not significantly affect the kinetics constants and the mode of inhibition of SecA molecules under the assay conditions.



**Figure 2-11. Kinetic analysis of SecA constructs**

Kinetic analyses of SecA-N95(T109N) (**Panel A**), SecA-N95ΔN(T109N) (**Panel B**), and SecA(T109N) (**Panel C**) were performed using a coupled assay to obtain the kinetic parameters listed in Table 2-4. Initial concentrations of ADP were 0 (■), 50 (▲), 250 (▼), and 600 (◆) μM, and concentration of ATP varied as indicated on the figures. Data points are the measured rates, while the solid curves represent the nonlinear regression fits derived from globally fitting of all data to the competitive inhibition model.



**Table 2-4. Kinetic parameters for ATP hydrolysis by SecA constructs**

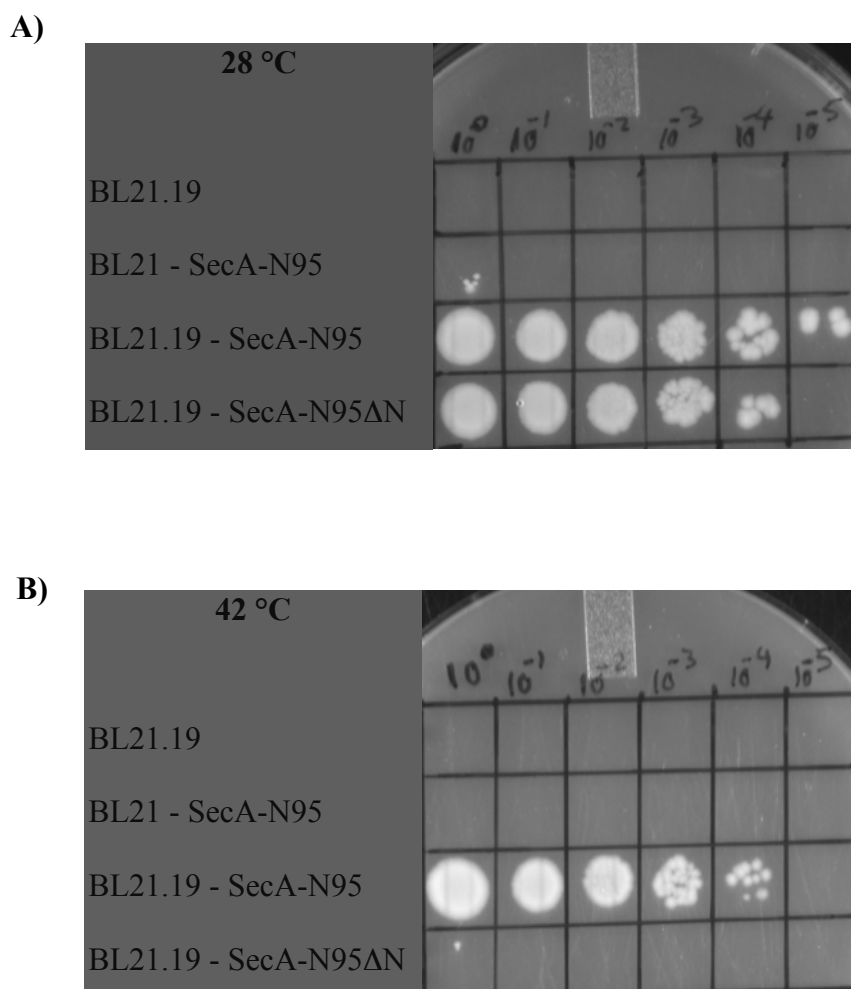
<b>Construct</b>	<b>Inhibition</b>	<b><sup>a</sup>K<sub>M</sub> (mM)</b>	<b><sup>a</sup>K<sub>I</sub> ADP (μM)</b>	<b><sup>a</sup>k<sub>cat</sub> (s<sup>-1</sup>)</b>
SecA-N95ΔN(T109N)	competitive	2.41 ± 0.39	440 ± 94	1.76 ± 0.12
SecA-N95(T109N)	competitive	2.21 ± 0.22	443 ± 55	1.61 ± 0.07
SecA(T109N)	competitive	1.44 ± 0.23	281 ± 59	1.49 ± 0.09

<sup>a</sup> The values (with standard errors) were calculated by globally fitting the complete sets of kinetic data illustrated in Figure 2-11.

### 2.3.7 Functionality of the N-terminally truncated SecA *in vivo*

To further test the importance of the N-terminal segment of SecA *in vivo*, a temperature-sensitive SecA (SecA<sup>ts</sup>) *E. coli* strain, BL21.19(DE3), was used. In this strain, the expression of the endogenous SecA is switched off at elevated temperatures, and therefore it grows only at lower temperatures (e.g 28-30 °C) (Mitchell & Oliver, 1993). Supplementing BL21.19(DE3) cells with a functional *SecA* gene restores its growth at higher temperatures (Jilaveanu *et al*, 2005).

Therefore, both pJZ7-N95 (SecA-N95) and pAK-SecA-N95ΔN14 (SecA-N95ΔN), which does not carry the N-terminal His-tag, were transformed into the BL21.19(DE3) strain. The cells were grown at 28 °C in LB/Amp/Kan, and cultures were adjusted to the same density. Then, they were spotted on two LB/Amp/Kan plates at the indicated dilutions and incubated at 28 °C or 42 °C for 24 hrs. SecA-N95 complemented the SecA<sup>ts</sup> strains at 42 °C, while SecA-N95ΔN, which lacks the 14 N-terminal residues did not (Figure 2-12). At 28 °C, where the endogenous SecA is still expressed by the BL21.19 cell, both cells were able to grow. As negative controls in this assay, BL21.19 without any plasmid and BL21(DE3) carrying pJZ7-N95 plasmid were used. As it can be seen, none of these controls grew at either 28 or 42 °C.



**Figure 2-12. *In vivo* complementation analysis of SecA-N95 and SecA-N95ΔN**

The ability of SecA constructs with the wild-type N-terminus (SecA-N95) and the truncated N-terminus (SecA-N95ΔN) to rescue the SecA<sup>ts</sup> strain (BL21.19) was examined at **(A)** the permissive temperature of 28 °C and **(B)** non permissive temperature of 42 °C. Cells were grown initially in liquid media at 28 °C and after adjusting the cultures to a common OD<sub>600</sub> of 1, they were spotted on two identical LB/Amp/Kan plates at the indicated dilutions. Note that the BL21.19 cells carry the kanamycin marker, while the SecA plasmids carry an ampicillin resistance marker.

## 2.4 Discussion

SecA-N68 self-association is mediated through its N- and C-terminal unstructured polypeptides. It had been shown previously that H<sub>6</sub>-SecA-N68 participates in a monomer-tetramer equilibrium with a  $K_D$  of 63  $\mu$ M for the formation of the tetramer (Dempsey *et al*, 2002); at lower concentrations this molecule is in equilibrium between monomer and dimers, while at higher concentrations it has a tendency to form tetramer. Here, it was revealed that by removal of the unstructured termini, SecA-N68 $\Delta$ NC has no tendency to self-associate in solution. Even at the highest concentrations tested, it appeared that this construct is completely monomeric with a globular shape (Figures 2-3 and 2-4). Therefore, the monomeric behavior of SecA-N68 $\Delta$ NC was a clear indication that these termini are absolutely essential for the oligomerization of SecA-N68. Since the oligomerization of SecA has been suggested to be implicated in the translocation process (Kusters & Driessen, 2011; Sardis & Economou, 2010), this discovery prompted us to further characterize these interactions.

Selective deletion of the unstructured N- or C-terminal regions of SecA-N68 suggested that the tetramer assembly is formed via a “dimer of dimers” and therefore two separate binding sites are involved in the self-association. Investigating the self-association of SecA-N68 $\Delta$ N and SecA-N68 $\Delta$ C, which lack the N- and C-terminal polypeptides of SecA-N68, respectively, showed that they participated in a monomer-dimer equilibrium (Figures 2-7 and 2-8) but with noticeably dissimilar dimerization characteristics (Figure 2-8); SecA-N68 $\Delta$ C contributed in a fast exchange equilibrium while the SecA-N68 $\Delta$ N showed slow kinetics. Therefore, the observed differences in the dimerization behavior of SecA-N68 $\Delta$ N and SecA-N68 $\Delta$ C suggests that these unstructured termini are interacting with two separate sites on the SecA-N68, and each one of them separately mediate the self-association of SecA-N68. Furthermore, the extreme N-terminal polypeptide appears to be involved in the initial dimer formation, and the C-terminal unstructured segment is responsible for interaction between dimers and the appearance of the tetrameric structure. In support of the involvement of the N-terminal segment in the initial dimerization, it was shown that modification of the N-terminus segment of SecA-N68 led to the weakening of dimerization and observation of monomeric species both at lower and higher concentration, while the tetramer formation was not affected (Figure 2-6).

The observed self-association of SecA-N68 provides further evidence regarding the molecular basis of SecA dimerization. Of the two termini involved in the self-association of SecA-N68, the role of the C-terminal unstructured terminus of SecA-N68 in the dimerization of full-length SecA can be ruled out, as this polypeptide is not naturally available on SecA, where it is placed between NBD2 and CTD domains. On the other hand, several studies have pointed towards the importance of the extreme N-terminal segment for self-association of SecA (Randall *et al*, 2005; Or *et al*, 2005; Jilaveanu *et al*, 2005). Therefore, the role of this unstructured N-terminal segment on the dimerization and activity of a functional construct of SecA was investigated.

Subsequent analysis of the role of the N-terminus polypeptide on oligomerization of the functional SecA-N95 construct showed that the SecA-N95 forms a tight dimer in solution with an dissociation constant in the sub-micromolar range (Figures 2-9 and 2-10). This is in agreement with previous results that had shown SecA-N95 behaves similar to the wild-type SecA in terms of dimerization (Or *et al*, 2002). Furthermore, the removal of the first 14 residues in SecA-N95 $\Delta$ N resulted in a dramatic change in the dimerization. SecA-N95 $\Delta$ N was mainly monomer at lower concentrations (Figure 2-9), supporting the previous observations that N-terminus truncation will cause monomerization of SecA monomer (Randall *et al*, 2005; Or *et al*, 2005; Jilaveanu *et al*, 2005). Even at the highest concentration tested, SecA-N95 $\Delta$ N showed a molecular weight between monomer and dimer with a dissociation constant of  $\sim 25 \mu\text{M}$ . It is worth noting that it has been shown that the CTD domains of SecA may play a role in the dimerization of this molecule (Karamanou *et al*, 1999; Hirano *et al*, 1996), therefore, the dimer formation of SecA-N95 $\Delta$ N might be solely caused by the CTD domains, and be independent of the N-terminal polypeptide. Therefore, it seems that a cooperative interaction mediated by the extreme N-terminus and the CTD domains is likely responsible for the high affinity dimerization of full-length SecA.

Next, the possible differences in the functionality of monomeric and dimeric SecA-N95 constructs were investigated by comparing their solution ATPase activity kinetics as well as their ability to restore the growth of a conditional lethal SecA mutant strain. In order to examine the ATPase activity of SecA-N95 constructs, T109N mutants were used, as it has been shown that the wild-type SecA has no significant solution ATPase activity unless the T109N mutation is present (Dempsey, 2007). It was observed that the kinetic parameters of dimeric and monomeric

constructs remained mainly unchanged (Figure 2-11) indicating that the removal of the N-terminal polypeptide does not affect the solution ATPase activity of SecA. In addition, SecA ATPase activity can be inhibited by ADP, and crystallographic data have shown that there is only one nucleotide-binding site on SecA (Hunt *et al*, 2002; Papanikolaou *et al*, 2007b). Therefore, it is expected that ADP will inhibit monomeric SecA competitively. In the case of dimeric SecA, if the ATPase activity of the protomers is independent from each other, a simple competitive inhibition is expected; however, if the binding of ADP to one subunit inhibits the ATPase activity of the other subunit, then a deviation of this mode of inhibition is expected. It was shown that both SecA-N95 proteins were competitively inhibited by ADP (Table 2-4), indicating that the ATPase activity of the protomers are independent of each other in solution. However, in previous work a different result was reported (Dempsey, 2007). It was suggested that the mode of inhibition of a partially monomeric SecA (His<sub>6</sub>-SecA(T109N)) by ADP is different than that of a dimeric SecA (SecA(T109N)), suggesting allosteric regulation between protomers of SecA. The reason for this discrepancy remains unknown. Although it was shown that the presence of N-terminal polypeptide and dimerization does not significantly affect the ATPase activity of SecA, it clearly affects the functionality of SecA *in vivo*, as the removal of the first 14 residues led to the loss of function and the cells carrying this construct did not survive (Figure 2-12). This observation is in line with previous results that showed the constructs lacking the 11 N-terminal residues of SecA are not functional *in vivo* (Randall *et al*, 2005; Jilaveanu *et al*, 2005). In addition, it has been shown that SecA constructs that lack the N-terminal region are compromised in the *in vitro* translocation assay (Jilaveanu *et al*, 2005; Bauer *et al*, 2014).

Although several studies have indicated that the extreme N-terminal segment is critical for both *in vitro* and *in vivo* function of SecA during the translocation reaction (Or *et al*, 2005; Jilaveanu *et al*, 2005; Bauer *et al*, 2014), the interaction site(s) for this N-terminal polypeptide are not identified yet. The first crystal structure of SecA from *B. subtilis* showed that the 10 N-terminal residues are in the dimeric interface and are interacting with the HSD and HWD domains (Hunt *et al*, 2002). However, in the next solved crystal structures of SecA both monomeric and dimeric SecA were observed, and the location of N-terminal segment varied greatly (Sardis & Economou, 2010). In these structures either the electron density for the N-terminus segment is not apparent (Nithianantham & Shilton, 2008; Papanikolaou *et al*, 2007b), the N-terminus is not making any contacts with other domains (Zimmer *et al*, 2008; Vassylyev *et al*, 2006), or it is in

contact with the CTD domains of SecA (Zimmer *et al*, 2006; Hunt *et al*, 2002), which are not present in SecA-N68. Therefore, our results are valuable in terms of localizing the binding site to the N-terminal domains of SecA (i.e NBD1, NBD2, and/or PPXD; the domains available on SecA-N68 protomers) by showing that SecA-N68 $\Delta$ C can form a dimer, and by removal of the extreme N-terminus, the SecA-N68 $\Delta$ NC is not able to self-associate. Evidently, further structural studies are required for identification of the exact binding site(s) for the N-terminal segment on SecA.

Uncovering the molecular basis of self-association of SecA is one of the first steps in understanding the underlying mechanisms of preprotein translocation by the Sec system. There are still many unresolved questions about the oligomeric state and exact molecular role of the SecA dimer and even the dimeric interface(s) of SecA. However, the observation of various dimeric structures of SecA and the functional role of the dimer *in vivo* have led to the development of several translocation models in which a single or multiple copies of SecA are involved in the translocation process (Kusters & Driessen, 2011; Sardis & Economou, 2010; Park & Rapoport, 2012). In the dimeric model, one SecA protomer should be bound to SecYEG at the membrane and the second protomer is interacting with first one. While the first (membrane bound) SecA translocates segments of preprotein through the pore, the second SecA delivers the segments of preprotein to the SecYEG bound SecA (Kusters & Driessen, 2011). In addition, it is believed that SecA dimer may be very dynamic, particularly during the translocation reaction, utilizing alternative dimer interfaces (Kusters & Driessen, 2011; Sardis & Economou, 2010). However, in the dimeric model, the exact mode of interaction of the SecA protomer with each other, preproteins, and the SecYEG channel is not clear yet. Although the results presented here seems to support the role of dimerization in the translocation reaction, these findings do not exclude the possibility that the mere presence of N-terminus segment rather than dimerization of SecA to be responsible for the proper function of SecA during the translocation reaction.

In summary, the results presented here reveal that the oligomerization of SecA constructs is largely dependent on its terminal polypeptides; especially the extreme N-terminal region has a profound effect on the dimerization of this molecule *in vitro* and its functionality *in vivo*. Moreover, it was shown that the N-terminal polypeptide is interacting with the N-terminal domains of SecA. There are still unresolved questions about the oligomeric state and structure of

the SecA translocon components. Understanding the mechanisms of coupling of preprotein translocation to the ATP hydrolysis and the possible role of the N-terminal segment of SecA and the dimerization of this molecule in protein translocation will require additional studies.



## Chapter 3

### 3 Interaction of of ATPase SecA with the Mature Regions of Preprotein Substrates

SecA converts the chemical energy of ATP into a mechanical force to drive unfolded preproteins through the SecYEG channel in a stepwise manner. Besides the possible interactions with signal sequences, SecA is expected to interact with the mature region of preproteins, however, the location of the preprotein binding site and the mechanism of coupling between preprotein binding and ATP binding and hydrolysis are not known. In this chapter, polypeptides corresponding to sequences in the mature regions of preproteins were coupled to a solid support and the interaction between SecA and the polypeptides was characterized. The binding site on SecA for the unfolded polypeptides was mapped to the DEAD motor and/or stem region of the PPXD. Furthermore, interactions between SecA and the polypeptides were affected by both nucleotides and the N-terminal segment of SecA. Finally, the sequences of several high affinity SecA binding peptides were analyzed and a possible “motif” on preproteins for interaction with SecA was identified. These findings have implications for the translocation of preproteins by ATPase SecA.

#### 3.1 Introduction

The essential components of the bacterial Sec system are the pore-forming SecYEG and the ATPase SecA. Biochemical studies of this system have demonstrated that preproteins are transported through the SecYEG channel in an unfolded fashion (Park & Rapoport, 2012; Kusters & Driessen, 2011), and in fact the crystal structure of SecYEG revealed that this pore forming trimeric complex can only accommodate unfolded polypeptides (Zimmer *et al*, 2008; Breyton *et al*, 2002). The motor protein of the Sec system, SecA, is a peripheral multi-domain ATPase associated with the SecYEG, and the minimal biochemical function of this protein is perhaps the ATP dependent translocation of unfolded polypeptide substrates through the SecYEG pore. The mechanism of transport of unfolded preproteins by SecA has been a subject of research for a number of years.

The translocation process must inevitably involve binding and release of the preprotein substrate, and on this basis SecA is expected to possess sites that are capable of interacting with unfolded preproteins. The preprotein substrates are usually composed of a short N-terminal signal sequence and a mature region. Although the exact role of the signal sequence in SecA dependent translocation is not clear, it has been demonstrated that the signal sequence is required for translocation by wild-type SecA and SecYEG, and there is evidence that signal sequences interact with SecA and this interaction is critical for the translocation process (Clérico *et al*, 2008). However, in addition to the signal sequence, SecA must interact with the mature region of preproteins. There are several lines of evidence in support of this idea. First, simply adding a signal sequence to a non-secretory protein does not necessarily lead to its translocation by the Sec system. Second, substrates lacking a signal sequence can still be transported through the Sec system in the presence of the supportive *prl* mutations (Danese & Silhavy, 1998; Emr *et al*, 1981; Oliver & Beckwith, 1981). Third, SecA transports preprotein substrates through the membrane pore in a stepwise manner. In fact, it was shown that multiple ATP molecules are hydrolyzed by SecA during translocation, and it was estimated that approximately 5 kDa of polypeptide was translocated for each ATP molecule hydrolyzed (Kusters & Driessen, 2011; van der Wolk *et al*, 1997). On this basis, one can envision a mechanism in which SecA catalyzes multiple cycles of preprotein binding and release to effect translocation of preprotein substrates through SecYEG pore.

Identification and characterization of the preprotein binding site on SecA are critical for synthesizing a coherent model for SecA-mediated translocation. Full length SecA is a relatively large protein (901 residues) composed of several domains including the NBD1 and NBD2 (“Dead Motor” domains), the “preprotein cross-linking domain” (PPXD), and the C-terminal domains beginning at residue 610, which are followed by an unstructured linker and a 22 residue “Zinc Binding Domain (ZBD) (Figure 1-3). Since removal of the extreme C-terminal ZBD and unstructured linker does not affect the *in vivo* function of SecA, these parts of SecA are likely not involved in the interaction with preproteins. Of the other domains, the “two-helix finger” domain of SecA, which is part of CTDs, has been suggested to be able to interact with the unfolded polypeptides. The crystal structure of SecA-SecYEG complex showed that this domain is deeply inserted into the SecYEG funnel (Zimmer *et al*, 2008). Cross-linking experiments revealed that the translocating polypeptide comes close to the tip of the two-helix finger domain exactly above

the entrance of SecYEG channel (Erlandson *et al*, 2008). However, there is no direct evidence that this segment is responsible for recognizing and interacting with the mature region of preproteins. On the other hand, early cross-linking experiments that involved a preprotein substrate and a collection of deletion constructs of SecA suggested that a region located on NBD1 was the location of cross-links between SecA and a full-length preprotein substrate that contained a non-cleavable signal sequence (Kimura *et al*, 1991). On the basis of these experiments, this domain of SecA was termed the “preprotein cross-linking domain” (PPXD). Consistent with these results, SecA-N68, which is composed of the NBD1/NBD2 DEAD Motor and the PPXD domains, can bind *in vitro* to unfolded polypeptides that do not have any resemblance to signal sequences (Papanikou *et al*, 2005). In this case, the isolated PPXD domain showed some binding affinity towards the unfolded polypeptides. The conclusion from this study was that the “stem” region of the PPXD, which connects the PPXD to NBD1, interacts with the signal sequence while the main PPXD domain (termed the “bulb”) binds unfolded polypeptides that do not have any resemblance to signal sequences (Papanikou *et al*, 2005). However, the crystal structure of SecA-SecYEG complex showed that the PPXD makes major contacts with SecY, therefore it seems that the PPXD domain is responsible for interaction of SecA and SecYEG channel. On this basis, the exact site(s) on SecA for recognition and interaction with unfolded preprotein substrates remains an open question.

The nature of the interaction between SecA and the mature regions of preprotein substrates is difficult to characterize because the unfolded preprotein substrates have a tendency to either refold or aggregate. To explore the direct interaction of SecA with the mature regions of preprotein substrates, a new strategy was adopted that involved the immobilization of large peptides (up to 16 kDa) from maltose binding protein (MBP) to form a heterogeneous population of potential binding sites. Binding to these immobilized peptides was monitored using surface plasmon resonance (SPR) spectroscopy. These studies were extended to a more defined system using an immobilized 20-residue peptide from the mature region of the periplasmic binding protein, FhuD. SecA constructs bound to the immobilized peptides in both cases, and the binding site was mapped to the Dead Motor domains of SecA. Moreover, the interactions between SecA and the peptides were affected by the presence of nucleotides, indicating that the preprotein binding site on SecA is coupled to changes wrought by nucleotide binding and hydrolysis. Surprisingly, the interaction between SecA and the immobilized peptides was almost completely

dependent on the presence of the unstructured N-terminal region of SecA. Finally, in order to study the sequence specificity of the SecA-polypeptides interactions, the sequences of several peptides that are able to bind SecA with relatively high affinity were analyzed to find a common SecA binding “motif”.

## 3.2 Materials and methods

### 3.2.1 Purification of SecA proteins

All the SecA proteins were expressed and purified essentially as described in Section 2-2 of this thesis.

### 3.2.2 Preparation of CNBr peptides from MBP

Mature MBP was purified from a periplasmic extract as previously described (Gould *et al*, 2009). Briefly, wild-type MBP was expressed using *E. coli* strain HS3309 carrying pLH1 (Hor & Shuman, 1993) by overnight growth in LB/Amp at 30°C. Cells were harvested and the periplasmic extract was isolated by osmotic shock. The extract was dialyzed against Buffer A (50 mM Tris-HCl, pH 8.5) and applied to an anion exchange column (Q-Sepharose HP, 1.6 x 13 cm) preequilibrated with Buffer A. The column was washed extensively with Buffer A until absorbance at 280 nm returned to baseline; bound proteins were eluted in a 240 mL linear gradient from 0 % to 100% Buffer B (Buffer A containing 1 M NaCl). Fractions containing MBP were pooled, dialyzed against Buffer A, and chromatographed again by high-resolution anion exchange chromatography (Mono Q HR 16/10). Fractions containing MBP were pooled, dialyzed against Buffer A, flash frozen in liquid N<sub>2</sub> and stored at -80°C.

To digest the MBP, 125 µL of purified MBP (30 mg/mL) was mixed with 375 µL of Trifluoroacetic acid (TFA) to yield a 75% solution of TFA; Cyanogen bromide (CNBr) (Sigma) was added to a final concentration of 10 mM, a 150-fold molar excess over methionine. The reaction mixture was incubated at room temperature in the dark for 72 hrs. To monitor the progress of digestion, samples were taken at various time points, neutralized with NaOH, and analyzed by SDS-PAGE using a 15% gel and a tris-tricine buffer system. Protein bands were visualized with Coomassie blue R-250. At the end of the reaction, the mixture was diluted 15-fold with water and lyophilized. The dried sample was then dissolved in 0.05% TFA and a small fraction analyzed by mass spectrometry. The remaining solution was dialyzed against 50 mM sodium acetate pH 4.5.

### 3.2.3 Surface plasmon resonance analysis

Surface Plasmon Resonance (SPR) analyses were conducted on a BIAcore X instrument using CM5 biosensor chips (GE Healthcare). The experiments were performed at 20°C, and all buffers were degassed prior to analysis. Derivatized sensor chips were stored at 4°C by first washing with water to remove buffer salts, and then placing in 50 mL capped centrifuge tubes containing a few milliliters of the running buffer to prevent drying of the chip.

The CNBr digest of MBP was coupled to a CM5 chip (GE Healthcare) through amine groups as follows. A new sensor chip was equilibrated with 50 mM HEPES, 150 mM NaCl, 0.01% Tween-20 at pH 7.4. Hydration of the dextran matrix was ensured by injection of 50  $\mu$ L of 50 mM HCl and 50  $\mu$ L of 50 mM NaOH at a flow rate of 100  $\mu$ L/min, followed by re-equilibration with the HEPES buffer. Peptide immobilization was carried out at a flow rate of 10  $\mu$ L/min using the same running buffer. Carboxyl groups were activated by a 100  $\mu$ L injection of freshly prepared 1:1 (v:v) solution of 0.4 M 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and 0.1 M N-hydroxysuccinimide (NHS). The CNBr digest of MBP (100  $\mu$ L in 50 mM sodium acetate pH 5.0) was injected, followed by 100  $\mu$ L of ethanolamine-HCl (1 M, pH 8.5) to deactivate the excess NHS ester groups. The chip was washed with several injections of 50  $\mu$ L of 50 mM NaOH at 100  $\mu$ L/min and equilibrated with running buffer until there was no change in the baseline.

Thiol coupling was used to immobilize specific synthetic peptides to CM5 chips as follows. Carboxyl groups of the matrix were activated with a 90  $\mu$ L injection of EDC/NHS as described above. This was followed by an injection (40  $\mu$ L) 2:1 (v:v) mixture of 2-(2-pyridinyldithio)ethaneamine hydrochloride (PDEA; 120 mM) in sodium borate buffer (150 mM, pH 8.5), which introduces a reactive disulfide onto the matrix. To prevent additional amine coupling in the next step, excess NHS groups were removed by reaction with ethanolamine. The sulfhydryl-containing peptide, AQYEDFIRSMKPRFVKRGAR $\{\beta$ -Ala $\}$  $\{\beta$ -Ala $\}$ C, was dissolved at a concentration of 100  $\mu$ M in sodium acetate buffer (50 mM, pH 4.7) and 100  $\mu$ L was injected. Afterwards, 100  $\mu$ L of the cysteine solution (50 mM cysteine and 1 M NaCl in 0.1 M sodium acetate pH 4.3) was injected twice onto the chip to deactivate excess reactive disulfide

groups. The chip was washed with injections of NaOH (50  $\mu$ L, 50 mM) at 100  $\mu$ L/min until a stable baseline was reached.

SPR experiments were conducted at 25  $^{\circ}$ C and a flow rate of 5  $\mu$ L/min; the running buffer consisted of 50 mM HEPES, 150 mM NaCl, 6 mM MgCl<sub>2</sub>, and 0.01% Tween-20 at pH 7.4. Analyte solutions were dialyzed and subsequently diluted in the running buffer to obtain the desired protein concentration in a matched buffer. Binding assays were conducted by equilibrating the chip until a stable baseline was achieved and then injecting 30  $\mu$ L of the analyte; dissociation was monitored for at least 250 seconds after the end of each injection. The chip surface was regenerated after each experiment by injection of 80  $\mu$ L guanidine hydrochloride (6 M) at a flow rate of 10  $\mu$ L/min. The binding signal was recorded and processed using the software accompanying the instrument.

In order to test the effect of nucleotides on binding of SecA to the unfolded polypeptide, a 2  $\mu$ M solution of SecA-N68 protein was supplemented with 4 mM of either ATP and ADP (from 100 mM stock solutions at pH 7.0), or 2 mM of AMP-PNP (100 mM stock solution, pH 7.2) and was then analyzed as described above.

### 3.2.4 Sequence analysis

The sequence specificity of SecA was assessed using the results obtained from the previous *in vitro* oriented peptide library experiments aimed at identifying high affinity SecA binding peptides (B. Shilton, manuscript in preparation). These peptides were ordered based on their relative signal index (RSI), which is a measure of the relative amount of SecA bound to each immobilized peptide. The RSI varies from 0 for no binding, to 1 for the peptide sequence that exhibits the strongest binding. Two groups were used for the analysis, those with an RSI greater than 0.4, and those with an RSI greater than 0.5. The GibbsCluster server (<http://www.cbs.dtu.dk/services/GibbsCluster>; (Andreatta *et al*, 2013)) was used to align and identify a common motif in these two groups. The complete list of these peptides is provided in Appendix B.

The Protein Pattern Find tool (Stothard, 2000) was used to identify sequences similar to the “SecA binding motifs”, which had been found as described above, on the preproteins. First, the sequences of the mature regions of several preprotein substrates of *E. coli* SecA were retrieved from Uniprot, and then by using this tool the locations of [V/I/A/L/F]xx[V/A/I/L/F/M]xx[R/K] were identified.

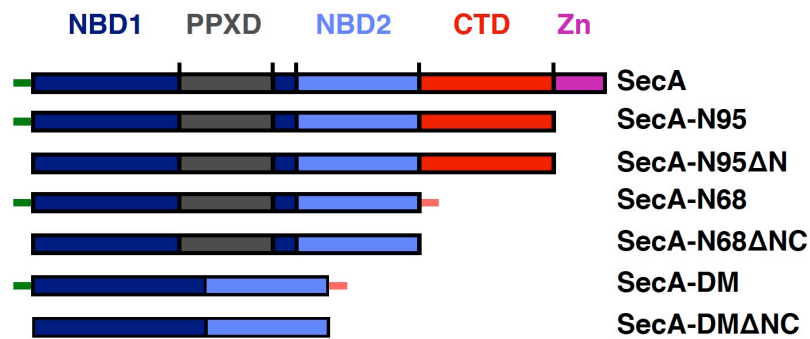


### 3.3 Results

SecA translocates unfolded preproteins through SecYEG in an ATP-dependent manner. To accomplish the translocation of a complete preprotein, SecA undergoes rounds of ATP binding and hydrolysis, each cycle associated with translocation of approximately 5 kDa of polypeptide (Kusters & Driessen, 2011; van der Wolk *et al*, 1997). On this basis, SecA is expected to interact with a number of mature sequences of the preprotein during translocation. The nature of the interaction of SecA with the mature regions of preprotein substrates has not been characterized, likely because of the technical difficulties associated with the tendency of unfolded proteins to aggregate or refold, either of which would prevent interaction with SecA. To overcome this problem, a new strategy was used to explore the interaction of SecA with the unfolded mature regions of preproteins. This strategy involved first the immobilization of large peptides (up to 16 kDa) from maltose binding protein (MBP) to form a heterogeneous population of potential binding sites. The interaction with SecA was monitored using surface plasmon resonance (SPR) spectroscopy, and it was clear that SecA was competent to bind to such immobilized peptides. This was followed by analysis of the sequence specificity of SecA for shorter peptides using immobilized synthetic peptide arrays. The array data provided short peptide sequences (15 residues) that bound to SecA with relatively high affinity. One of the short peptides was used for SPR experiments to characterize in greater detail the binding of SecA to a single peptide sequence.

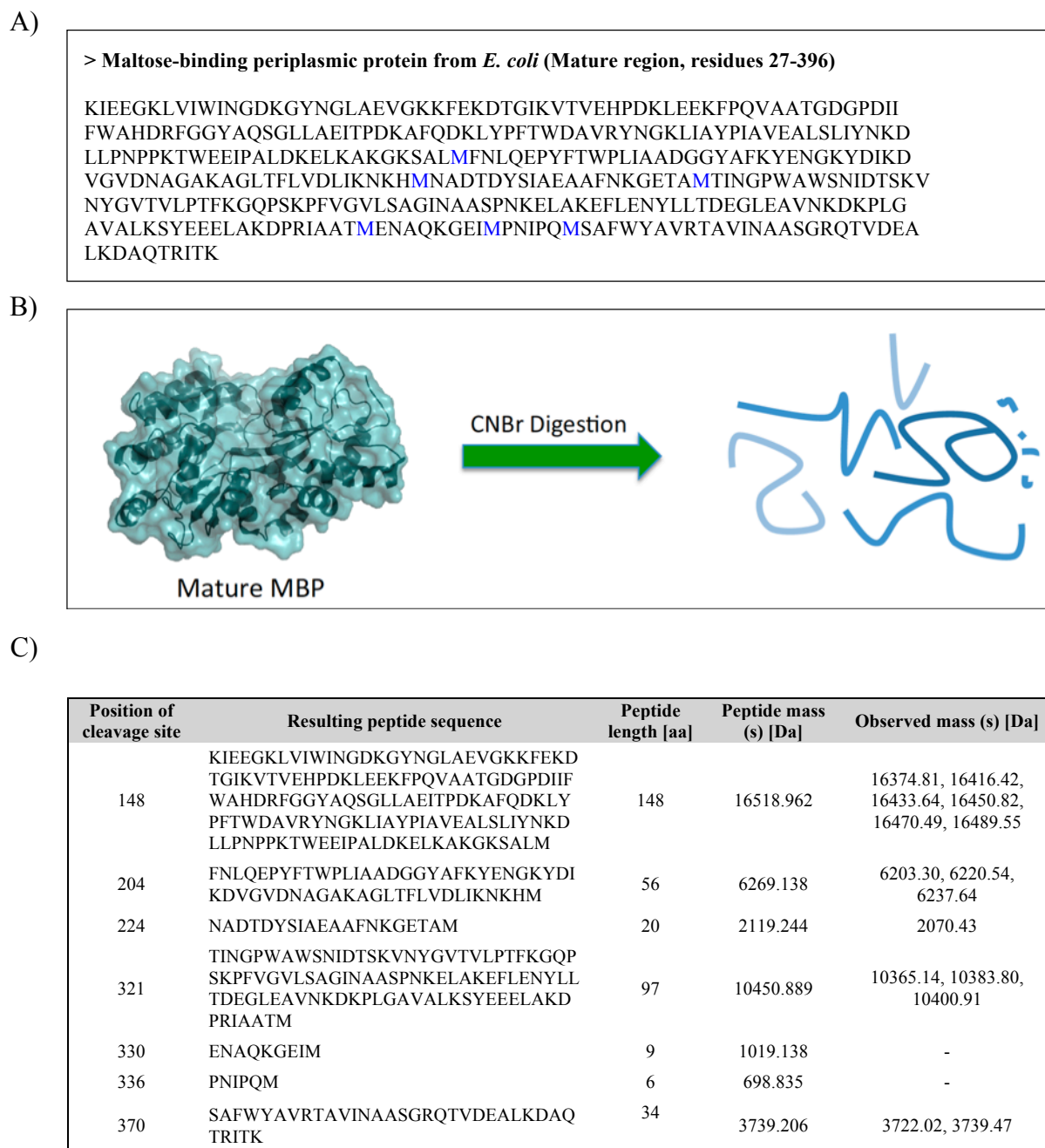
#### 3.3.1 Interaction of SecA with mature regions of MBP

MBP is a natural substrate of SecA that contains 6 methionine residues such that cleavage with cyanogen bromide (CNBr) would lead to the formation of 7 peptides ranging in length from 6 to 148 residues (Figure 3-2). None of the peptides comprise an individual domain of MBP, and therefore none were expected to adopt a stably-folded structure. Such peptides could therefore provide substrates to investigate the ability of SecA to bind the mature region of MBP. The digestion of MBP with CNBr was monitored by SDS-PAGE (Figure 3-3) and the digest was characterized by mass spectrometry (Figure 3-2C).



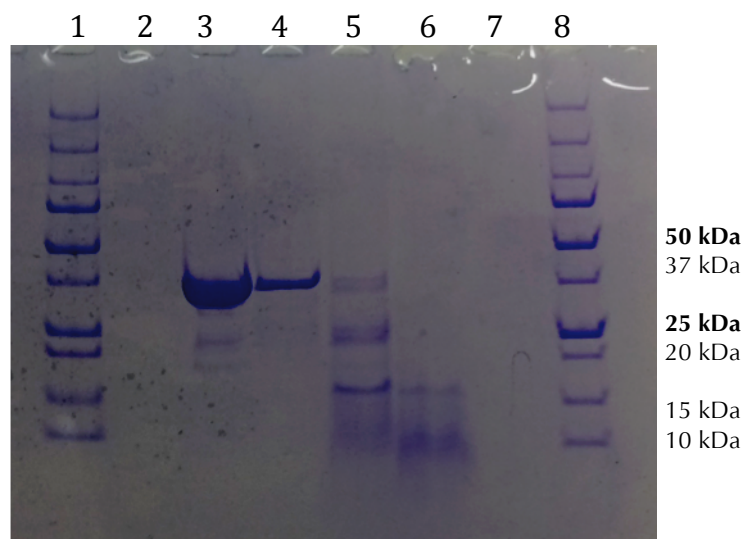
**Figure 3-1. SecA constructs used in this study**

Schematic domain arrangement of various SecA constructs is shown here. SecA-DM (DEAD motor) construct corresponds to SecA-N68 but with the "preprotein cross-linking domain" (PPXD) removed. The unstructured N-terminus of SecA is illustrated by the short green line. The unstructured polypeptide at the C-terminal region of SecA-N68 and SecA-DM constructs is showed by a short orange line.



**Figure 3-2. The sequence of mature MBP and the unfolded polypeptides generated using CNBr digestion**

(A) The sequence of mature MBP (Uniprot accession number: P0AEX9) with the positions of methionine residues highlighted in blue (B) A schematic diagram of MBP digestion with CNBr. (C) The expected sequence and molecular weights masses of the generated peptides, and also the observed molecular weights of the fragments from CNBr digestion of mature MBP by mass spectroscopy are shown in the lower table.

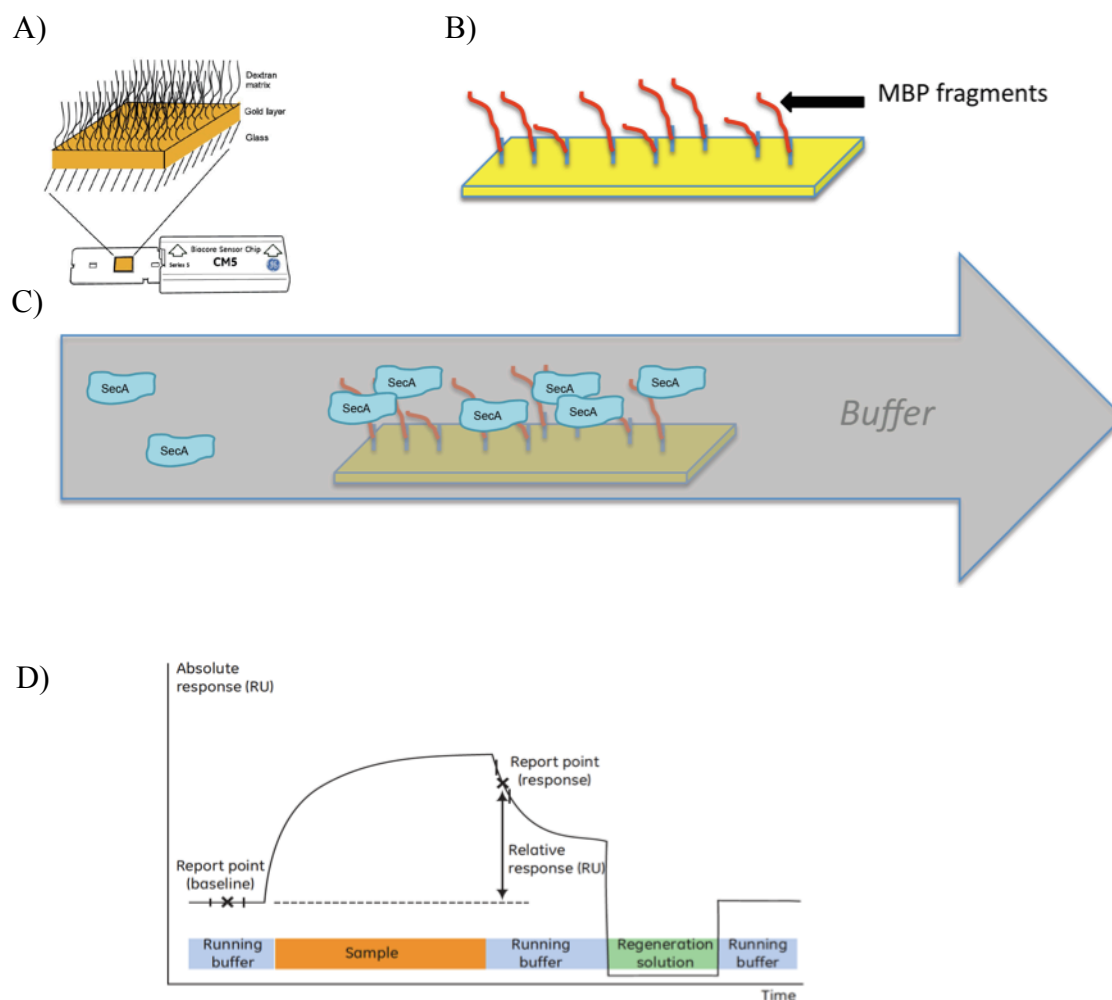


**Figure 3-3. Cyanogen bromide cleavage of MBP**

Mature MBP was digested with cyanogen bromide (CNBr) for 72 hrs, and the reaction monitored using a 15% SDS-PAGE with a Tris-tricine buffer system. Lanes 1 and 8 are molecular weight markers; lane 3 is untreated mature MBP; lane 4 is MBP in 75% TFA without any CNBr; lanes 5 and 6 are the CNBr digest after 24 and 72 hours, respectively.

Surface Plasmon Resonance (SPR) was used to study the interaction between SecA and the peptides representing the mature region of MBP (Figure 3-4). In this technique, a solution of analyte (*prey*; SecA protein) is injected over a ligand molecule (*bait*; unfolded fragments of MBP) that is immobilized on the dextran surface of the SPR chip. The association of the prey molecule, SecA, to the immobilized bait MBP fragments leads to an increase in the SPR signal (Jönsson *et al*, 1991; Johnsson *et al*, 1991; Green *et al*, 2000). At the end of the injection, the complex dissociates resulting in a decrease in the SPR signal (Figure 3-4). A great advantage of this technique is that there is no need to label the interacting SecA molecules, because the SPR method detects only changes in the mass of molecules on the chip's surface. Therefore, it eliminates any effects on the properties of SecA introduced by labeling procedures.

The complete CNBr digest was coupled to a CM5 (carboxymethyl dextran) sensor chip through amine groups on the peptides. Thus, the surface as modified was highly heterogeneous, both in terms of the sequences available for binding and the mode in which they were coupled. Since it was not clear where on MBP SecA would bind, the ability to couple the entire sequence of MBP to the chip ensured that potential binding sites would not be missed. Furthermore, the diversity in coupling (i.e. some peptides would be coupled by their N-terminal amine, others by the epsilon amino group of lysine) would also help to ensure that a particular binding site was not completely lost due coupling to the dextran matrix. In addition, the covalent amine ester linkages meant that the chip could be regenerated with denaturing agents. In summary, this was a stable surface that provided a means to test whether SecA was able to bind to a mature preprotein; however, due to the heterogeneity in the surface, it was not used for quantitative analysis of the binding interactions.

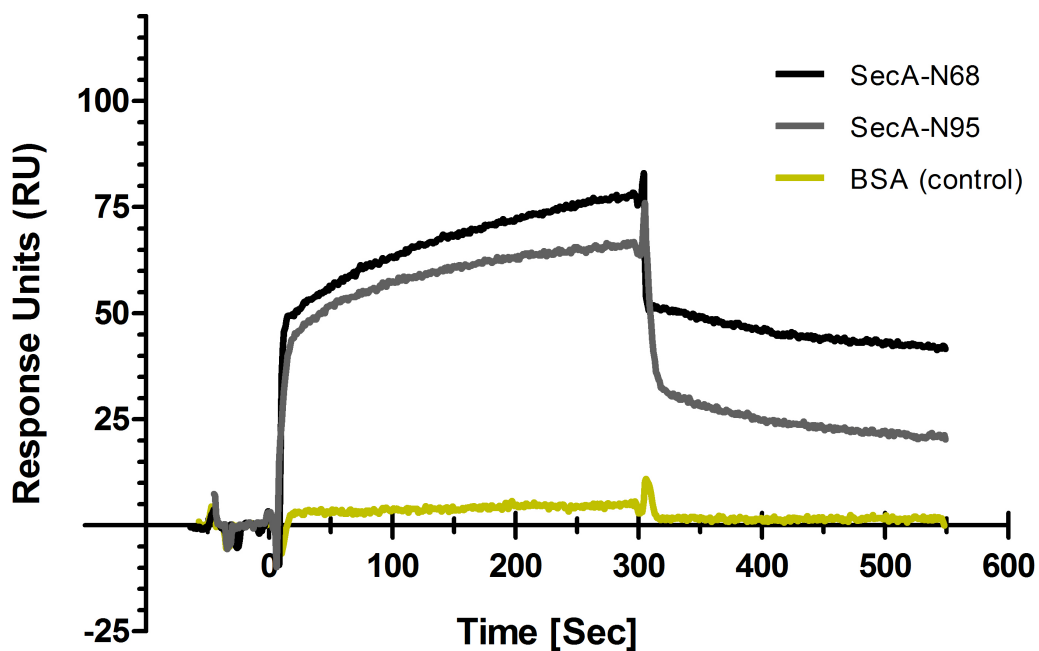


**Figure 3-4 Surface Plasmon Resonance (SPR) to study SecA-peptide interactions.**

SPR spectroscopy was used to analyze the interaction of SecA with mature regions of a preprotein substrate. **(A)** CM5 sensor chips were used to immobilize peptides to a carboxymethyl dextran matrix. **(B)** Preprotein fragments generated by CNBr digestion of mature MBP were covalently coupled to the surface of sensor chip. **(C)** An analyte solution is passed over the immobilized peptides. **(D)** The binding of SecA molecules to the immobilized peptides leads to a change in the refractive index at the surface, which is output as “response units” by the instrument. Panels A and D were taken from Biacore™ Assay Handbook published by GE Healthcare.

Interactions with MBP were first tested with SecA-N95 and SecA-N68, both of which clearly exhibited binding to the derivatized surface (Figure 3-5). Note that for these experiments, the sensor chip has a second flow cell – the “reference” cell – that was not derivatized with peptides, and that non-specific interactions with the surface of the reference flow cell were subtracted from the interactions observed at the derivatized surface. Under the same experimental conditions, bovine serum albumin (BSA) when injected into the system showed little if any binding to the surface (Figure 3-5). These results demonstrate that mature MBP contains sequences to which SecA binds, consistent with previous work indicating that SecA can bind unfolded polypeptides *in vitro* (Papanikou *et al*, 2005). This prompted us to characterize the interactions of SecA molecules with unfolded preprotein substrates in greater detail.

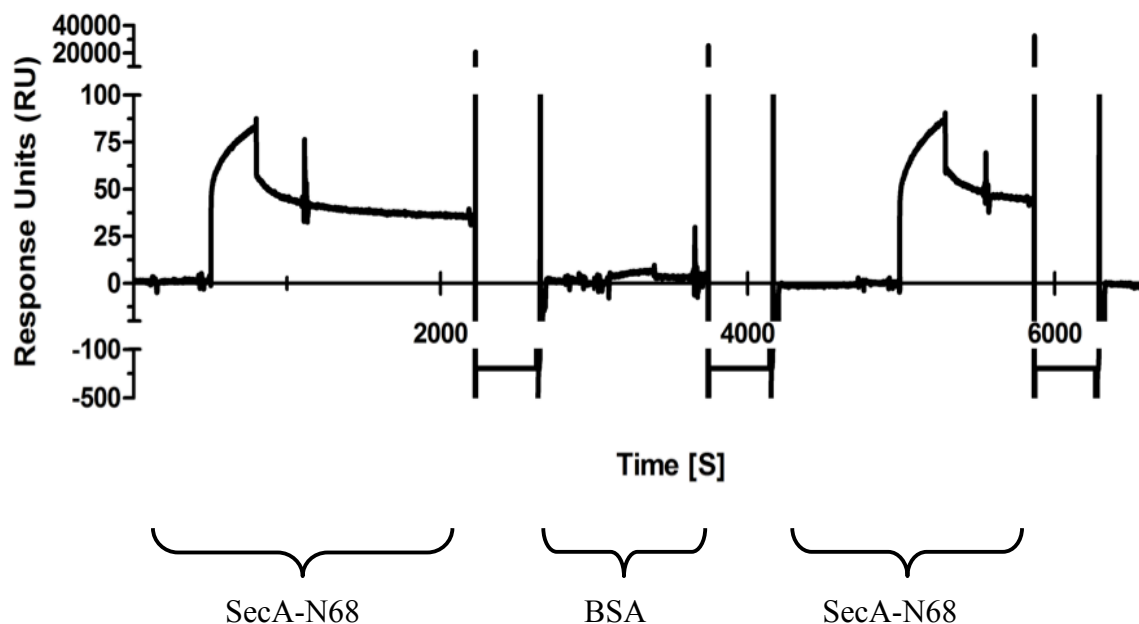
Given the potential for promiscuous interactions between SecA and denatured or unfolded protein associated with the sensor chip, the regeneration of the sensor chip after each injection was critical. To remove non-covalently bound protein molecules, several solutions were tested, including 10 mM glycine-HCl pH 2.0, 50 mM HCl, 50 mM NaOH or a combination of consecutive injections of these solutions. However, the key for regeneration of this chip was found to be an injection of 6 M guanidine hydrochloride (Gn-HCl), typically 80  $\mu$ L over 5 to 10 minutes. This injection returned the sensor signal to baseline, but did not adversely affect the surface because a subsequent injection of the analyte yielded identical results (Figure 3-6).



**Figure 3-5. Binding of SecA-N68 and SecA-N95 to MBP fragments**

To assess the ability of SecA to bind to the mature region of MBP, a CNBr digest of MBP was immobilized on a sensor chip and the interaction of SecA-N68 and SecA-N95 was analyzed using SPR. For both SecA constructs, 30  $\mu\text{L}$  of a 3  $\mu\text{M}$  solution was injected; the flow rate of the running buffer was 5  $\mu\text{L}/\text{min}$ . The same concentration of BSA was injected under identical experimental conditions to provide a negative (non-binding) control.





**Figure 3-6. Regeneration and reproducibility of the MBP biosensor chip**

The sensor chip derivatized with the CNBr digest of MBP was regenerated with injection(s) of 6 M guanidine-HCl. *From left to right:* a solution of SecA-N68 (3  $\mu$ M) was injected on the sensor (at time  $\sim$ 500 Sec); during the dissociation phase of SecA-N68, the chip surface was regenerated with an 8 min injection of 6 M Gn-HCl between 2200 and 2500 s; a solution of BSA (3  $\mu$ M) was injected at  $\sim$ 3000 s; the sensor surface was regenerated again at  $\sim$ 3800 s, followed by a second injection of a SecA-N68 (3  $\mu$ M) at  $\sim$ 5000 s.

### 3.3.2 Domain mapping of SecA-preprotein interactions

In the previous section it was shown that SecA could bind to the mature region of MBP. Both SecA-N95 and SecA-N68 exhibited binding to peptides from MBP, indicating that the C-terminal domains of SecA are not required for the interaction. To further map the site of peptide interaction to particular SecA domains, a SecA “DEAD motor” (SecA-DM) construct was used. SecA-N68 consists of two nucleotide binding domains, NBD1 and NBD2, along with a “preprotein cross-linking domain” (PPXD), which is a 150 residue domain internally fused to NBD1 (Figure 1-3). The SecA-DM construct has the main PPXD (residues 229 to 368) excised, and therefore consists only of NBD1 and NBD2 and a short stem of PPXD. SecA-DM retains the same ATPase activity as SecA-N68, and its crystal structure was solved, showing that deletion of the PPXD had no adverse effects on the structure of NBD1 or NBD2. (Nithianantham & Shilton, 2008).

Using the same SPR chip with the immobilized CNBr digest of MBP, SecA-DM bound to the sensor chip with higher affinity than either SecA-N95 or SecA-N68 (Figure 3-7). This was surprising because SecA-DM lacks the PPXD, which is the very domain expected to be responsible for preprotein binding. From early cross-linking experiments it appeared that the “preprotein cross-linking domain” (PPXD) was the location of cross-links between a preprotein substrate and SecA (Kimura *et al*, 1991). To further investigate a potential role of the PPXD in binding preprotein, a chimeric MBP-PPXD fusion had been constructed. In this protein, the SecA PPXD was fused to a loop in MBP. This construct was expressed, purified, and crystallized. The crystal structure shows that the PPXD adopts a native fold and its location on MBP allows it to sample different conformations, ensuring that all of its surfaces are available for binding (B. Shilton, personal communication). The MBP-PPXD construct showed no propensity to bind to the immobilized MBP peptides (Figure 3-7).

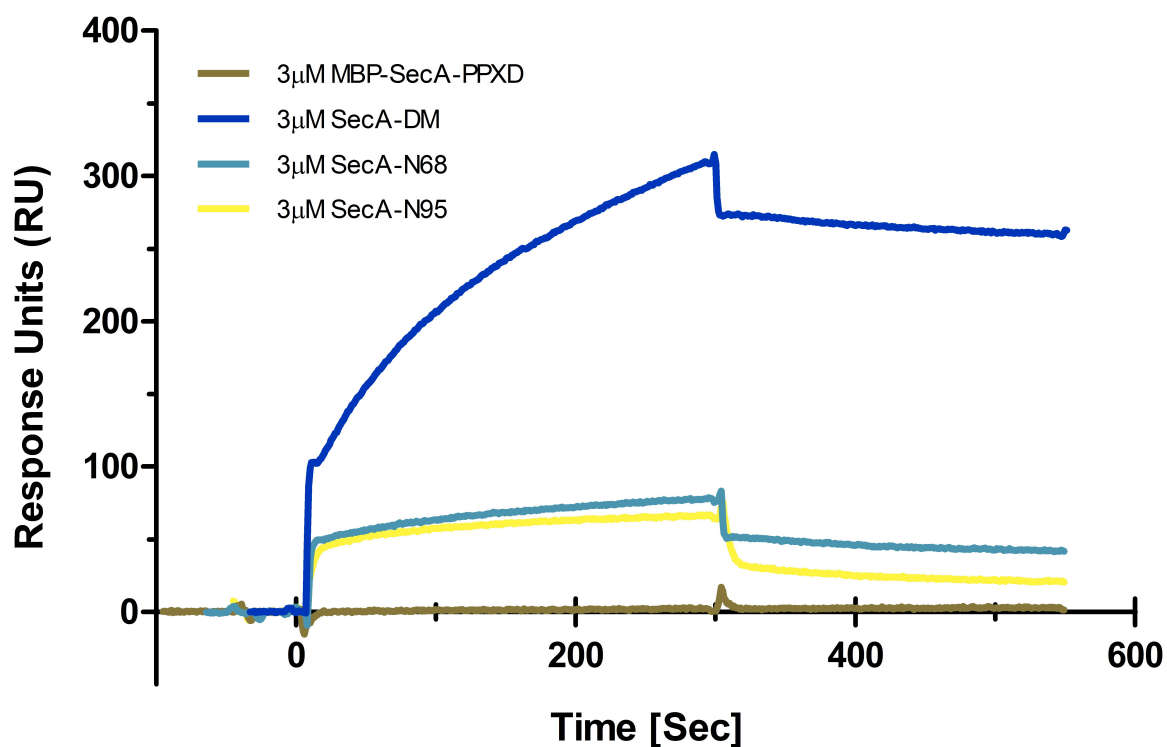
These results were very exciting, and provide valuable insight into the mode of interaction of SecA with its preprotein substrates. As can be seen in Figure 3-5 and 3-7, SecA-N95, SecA-N68, and SecA-DM all bind the immobilized MBP peptides. In these constructs, SecA-N95 shows the lowest binding. In turn, SecA-N68, which is shorter than SecA-N95 and lacks the CTD domain of SecA, shows a relatively higher binding signal than SecA-N95. The highest binding signal is

observed for SecA-DM, which is shorter than SecA-N68 and it lacks the PPXD. Taken together, these results indicate that the preprotein binding site is present on NBD1 and/or NBD2 (i.e. SecA-DM), and that the PPXD has no propensity to bind preproteins.

As an alternative to the heterogeneous surface provided by the immobilized CNBr digest of MBP, immobilized peptide libraries were used to identify short peptide sequences to which SecA would bind with relatively high affinity. From this analysis, the peptide sequence AQYEDFIRSMKPRFVKRGAR, which is part of the sequence of FhuD, a periplasmic protein for iron siderophore transport, was identified as one of strongest SecA interacting sequences (B. Shilton, manuscript in preparation). This peptide was synthesized along with two  $\beta$ -alanines and a cysteine at the C-terminus, to provide a flexible linker and the possibility for immobilizing the peptide by a thiol coupling.

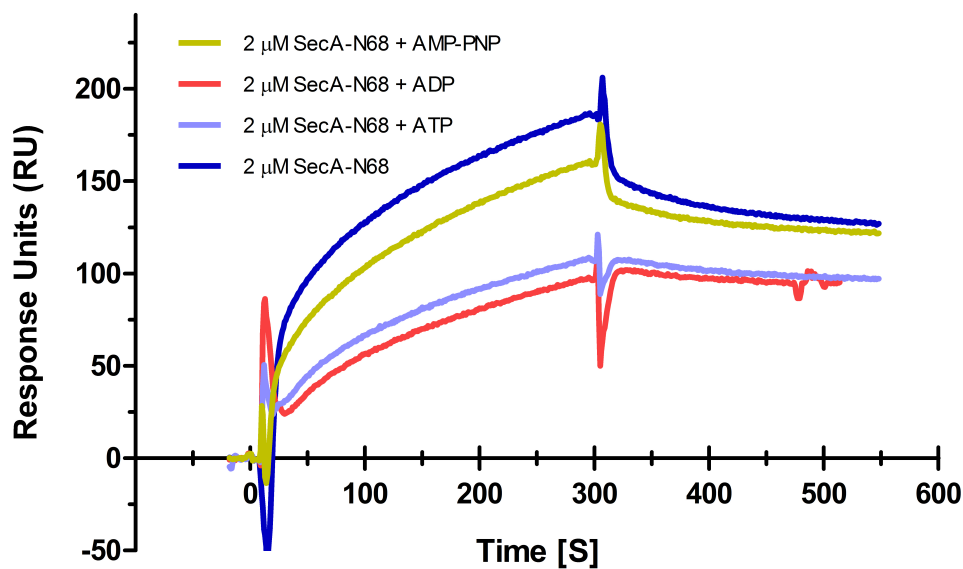
The FhuD peptide was coupled to a CM5 sensor chip through a disulfide linkage, and binding of SecA-N68 to the surface was examined; SecA-N68 clearly bound to the immobilized FhuD peptide (Figure 3-8). Binding to the chip derivatized with the FhuD peptide appeared to be somewhat stronger (175 RU maximum for 2  $\mu$ M SecA-N68) than binding to the chip derivatized with the MBP CNBr digest (75 RU maximum, for a 3  $\mu$ M solution). This is perhaps not surprising since the chip derivatized with the FhuD peptide presents a homogeneous array of a single high-affinity binding site, while the MBP CNBr digest presumably has many sequences for which SecA shows little or no affinity. SecA catalyzes preprotein translocation through SecYEG by consuming the energy from ATP hydrolysis, and it has been shown that binding of nucleotides by SecA will cause conformational changes (Karamanou *et al*, 1999; Shinkai *et al*, 1991). Therefore, the effects of different nucleotides on the binding of SecA-N68 to the FhuD peptide was examined. As can be seen in Figure 3-8, the presence of AMP-PNP decreased binding only slightly, while both ADP and ATP decreased the maximum binding signal of SecA-N68 to 50% of the value observed in the absence of nucleotides. Therefore, the presence of ADP weakens the interaction between SecA-N68 and the immobilized FhuD peptide. Non-hydrolysable ATP – i.e. AMP-PNP – has less of an effect, while ATP causes a similar change as ADP, which is expected since SecA-N68 has a significant ATPase activity. The observed effects of nucleotides on the interaction between SecA-N68 and the FhuD peptide are consistent with the idea that the binding site is formed by NBD1 and/or NBD2, which are subject to

conformational changes during the cycle of ATP binding and hydrolysis.



**Figure 3-7. Domain mapping of SecA-preprotein interactions**

Several SecA constructs were compared for their ability to bind to an immobilized CNBr digest of MBP. Solutions of each SecA construct (3  $\mu$ M) were injected and binding to the immobilized peptides was quantified by the change in refractive index at the sensor surface. SecA-N95 contains the C-terminal domains from residue 610 to 835; SecA-N68 consists of residues 1 to 609; SecA-DM consists of the two nucleotide binding domains, NBD1 and NBD2, with a deletion of the PPXD; while MBP-SecA-PPXD is a chimera of MBP with the isolated PPXD included as an internal fusion.



**Figure 3-8. The effects of different nucleotides on binding of SecA-N68 to the unfolded substrate**

SecA-N68 solution was supplemented with 4 mM ADP, 4 mM ATP, or 2 mM AMP-PNP to test the effect of these nucleotides on the binding SecA to an immobilized peptide from FhuD, with sequence AQYEDFIRSMKPRFVKRGAR-βA-βA-C, which was coupled to the sensor surface by a disulfide at the C-terminal cysteine. The N-terminus of the peptide was acetylated.

### 3.3.3 The extreme N-terminus of SecA is required for peptide binding

Results from Chapter 2 showed that removal of the extreme N-terminus affected oligomerization of SecA-N68 and SecA-N95. In addition, the presence of this segment is essential for efficient *in vivo* function of SecA (see Chapter 2). The exact function of the extreme N-terminal sequence of SecA has been a matter of debate (Kusters & Driessen, 2011; Sardis & Economou, 2010; Or *et al*, 2005; Jilaveanu *et al*, 2005). To further investigate the function of the N-terminus of SecA, we used SPR to compare the binding of SecA molecules with and without the N-terminal sequence.

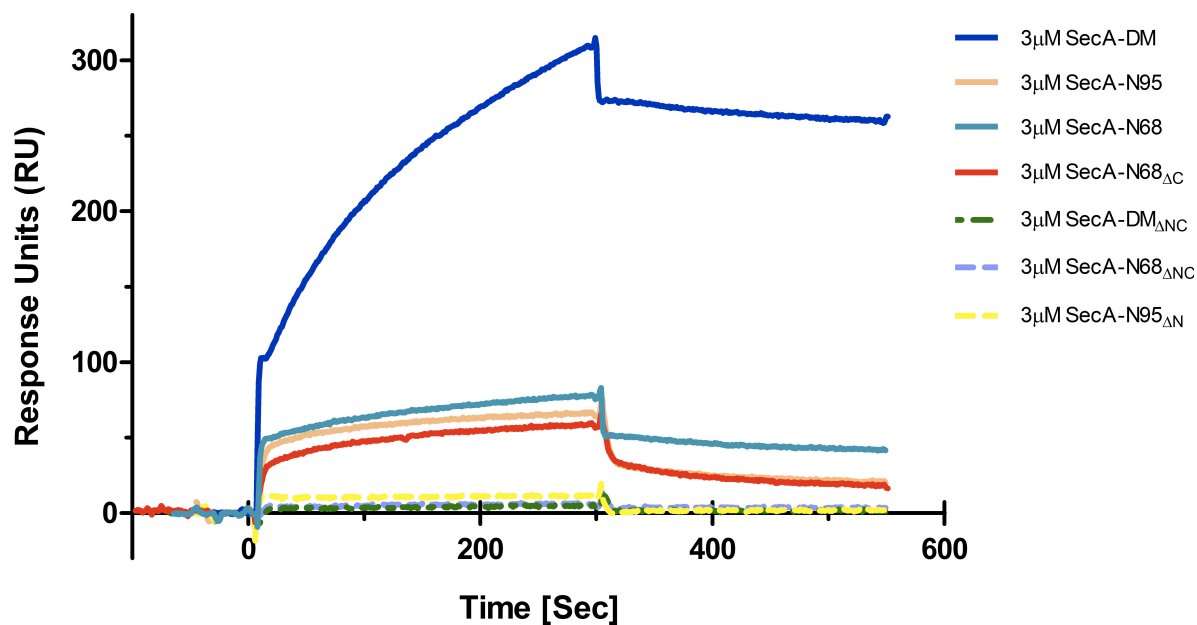
In the previous sections it was demonstrated that SecA-N95, SecA-N68, and SecA-DM interacted with immobilized fragments of mature MBP. To test the effect of the extreme N-terminus of SecA, six SecA constructs, SecA-N95, SecA-N68, SecA-DM, SecA-N95 $\Delta$ N, SecA-N68 $\Delta$ NC, and SecA-DM $\Delta$ NC were separately injected at the same concentration (i.e 3  $\mu$ M) onto the sensor chip derivatized with the MBP CNBr digest. Surprisingly, the SecA constructs that lacked the N-terminal segment, SecA-N95 $\Delta$ N, and SecA-N68 $\Delta$ NC, SecA-DM $\Delta$ NC, were not able to bind to the immobilized MBP peptides (Figure 3-9). On the other hand, constructs that carried the extreme N-terminal sequence, namely SecA-N95, SecA-N68, and SecA-DM, bound to the immobilized peptides. The SecA-N68 and SecA-DM constructs also carry an unstructured C-terminal sequence. To assess the importance of this C-terminal sequence for peptide binding, interactions with the SecA-N68 $\Delta$ C construct were assessed (the SecA-N68 $\Delta$ C construct has the N-terminus, but lacks the C-terminus, of SecA-N68, and was used previously for oligomerization studies). SecA-N68 $\Delta$ C showed only slightly weaker binding compared to SecA-N68, and on this basis it is the extreme N-terminal sequence that is somehow facilitating binding of SecA to the immobilized MBP CNBr digest. The loss of affinity towards the polypeptide substrates upon removal of the extreme N-terminal sequence provides additional insight into the function of this sequence and has implications for the translocation mechanism.

The critical role of the extreme N-terminus in peptide binding was a surprise. To verify this role using a homogeneous and well-defined immobilized peptide, we used the FhuD peptide (AQYEDFIRSMKPRFVKRGAR- $\beta$ A- $\beta$ A-C) that was immobilized to the carboxymethyl dextran

surface by its C-terminal cysteine (Figure 3-10). Both SecA-N95 and SecA-N68 bound to this peptide, consistent with previous experiments. However, both the SecA-N68- $\Delta$ NC and SecA-N95 $\Delta$ N constructs exhibit little, if any binding to the immobilized FhuD peptide (Figure 3-10).

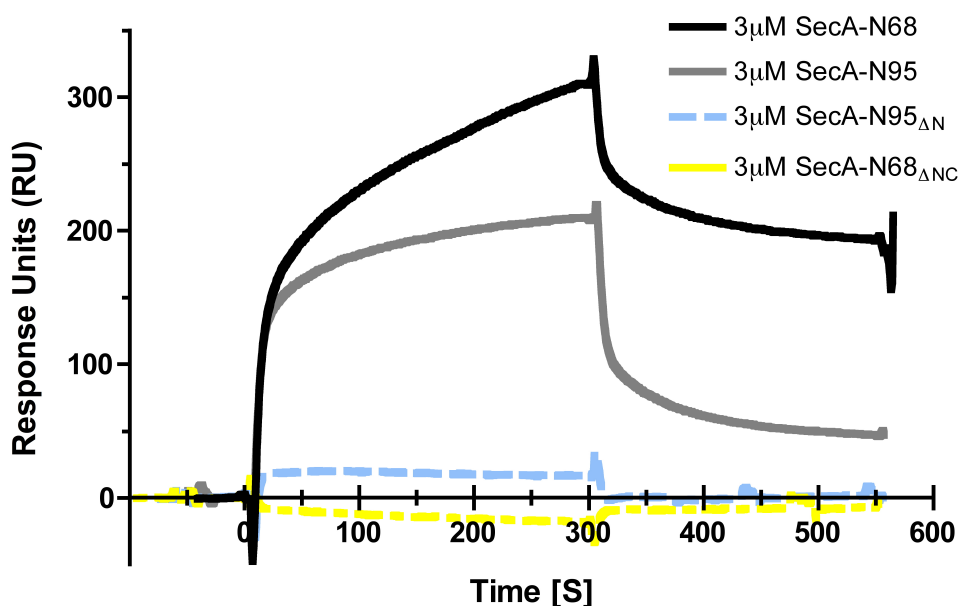
Overall, the experiments in this section have demonstrated that (1) SecA constructs are able to bind to the unfolded preproteins *in vitro*, (2) the binding site(s) for these preproteins is located on the DEAD-motor domains (NBD1 and NBD2) of SecA, (3) peptide binding is affected by nucleotides, and (4) the binding is dependent on the presence of the N-terminal unstructured region of SecA.





**Figure 3-9. The extreme N-terminus of SecA is required for binding to unfolded substrate polypeptides *in vitro***

SecA constructs were analyzed for their binding towards the immobilized CNBr digest of MBP using SPR. SecA-N95, SecA-N68, SecA-DM, and SecA-N68 $\Delta$ C constructs carry the N-terminal unstructured region of SecA, while SecA-N95 $\Delta$ N and SecA-N68 $\Delta$ NC lack this N-terminal segment. Solutions (3  $\mu$ M) of these proteins were analyzed using the same sensor chip under identical running conditions.



**Figure 3-10. Role of the of SecA extreme N-terminal sequence in binding to an immobilized peptide.**

A peptide representing a sequence of FhuD to which SecA binds with relatively high affinity (AQYEDFIRSMKPRFVKRGAR- $\beta$ A- $\beta$ A-C) was coupled to a CM5 chip by the C-terminal cysteine. To assess the role of the extreme N-terminus in binding to this peptide, the binding of two constructs with the N-terminus, SecA-N68 and SecA-N95, was compared to binding of two constructs, SecA-N95 $\Delta$ N and SecA-N68 $\Delta$ NC, in which the unstructured N-terminus had been removed. The constructs were all injected at the same concentration of 3  $\mu$ M.

### 3.3.4 Analysis of the sequence specificity of SecA for peptide binding

To catalyze translocation, SecA must interact with a variety of unfolded polypeptides, and the degree of sequence specificity for the interaction could therefore be relatively low. This was the rationale for using the complete sequence of a translocation substrate (MBP) for the initial SPR experiments. The use of oriented peptide libraries to find specific short sequences that bind to SecA facilitated the creation of a sensor chip that displayed a single peptide sequence from the FhuD protein, immobilized in a defined manner. The results using the immobilized FhuD peptide were fully consistent with the results from the immobilized MBP CNBr digest. This indicated that whatever was happening during the binding reaction with the MBP digest was also happening with a more well-defined system. The results from the oriented peptide libraries also indicate that SecA does exhibit a degree of sequence specificity. We have used the results of these experiments, which were carried out prior to my arrival in the laboratory, to analyze in detail the sequence specificity of the SecA-peptide interaction.

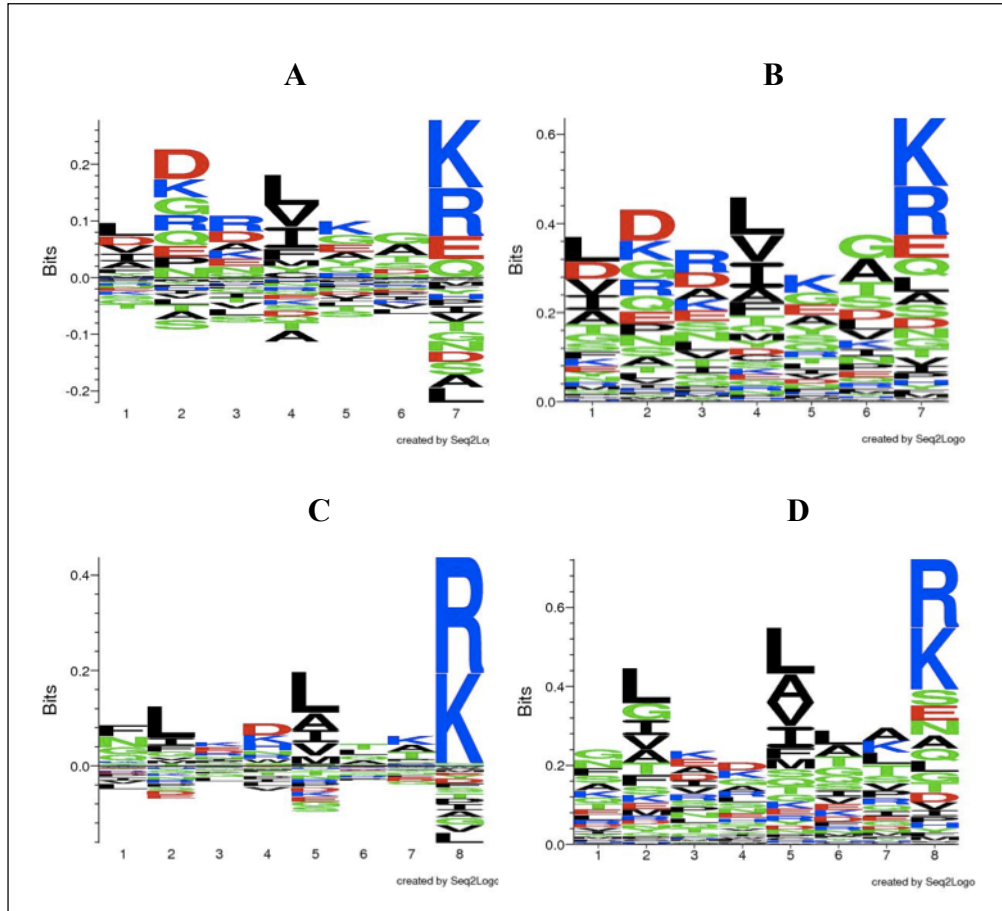
The oriented peptide libraries consisted of 15 mer peptides bound to a cellulose sheet by a flexible linker attached to their C-terminus. The sequences of the peptides corresponded to the sequences of exported preproteins as well as resident cytoplasmic proteins. The arrays were used to “scan” through protein sequences. For example, preMBP (396 residues) was represented by approximately one hundred 15-mer peptides that started at the N-terminal sequence and were shifted by 4 residues until the last peptide that constitutes the extreme C-terminal sequence. A fluorescently labeled SecA molecule was used to measure binding to the immobilized peptides. After removal of the background, each peptide spot was scored for SecA binding in terms of a “relative signal index” (RSI), which ranged from 0 for the weakest binding to 1 for the strongest binding.

To determine the optimal sequence for relatively strong binding to SecA, the peptides (approximately 1500) were ranked according to their RSI and divided into two groups. Group 1 represented peptides that had a  $RSI > 0.4$  (78 members), and Group 2, that had a  $RSI > 0.5$  (47 members). Lists of the peptides in the two groups are provided in the Appendix B.

Alignment of the peptides by common multiple sequence alignment tools did not result in a good alignment (data not shown). Instead, more specialized tools for alignment and clustering of peptide data were used, such as GibbsCluster-1.1. The top-scored clusters resulted by this method are shown in Figure 3-11. The results of the clustering these peptides are shown the two different formats (Shannon style and Kullback-leibler style) for better observation of the patterns. As it can be seen, in the group A, the Lys/Arg residues are common at the C-termini of the polypeptides and the position -3 relative to the Arg/Lys residue is a hydrophobic amino acid (L/V/I/A/F). Analysis of the group 2 (peptides with higher binding strength) revealed a better-resolved pattern of amino acids. In these peptides Arg/Lys is located at the C-terminal of the pattern, and the two hydrophobic residues (L/A/V/I) are located at position -3 and -6 of the Arg/Lys residue. Based on these data it seems that the common sequence between the SecA binder peptides is [V/I/A/L/F/M]xx[V/A/I/L/F/M]xx[R/K].

Furthermore, the alignment of the 25 top peptides that had shown the highest binding strength towards the SecA proteins in the oriented peptide library experiments and share a common pattern is shown in Figure 3-12. It can be seen that binding motif is less than 8 residues in length. The Lys/Arg residues are highly conserved at the C-terminal of this motif followed a relatively conserved hydrophobic residue at -3 position relative to the Arg/Lys residue.

The predicted “SecA binding” sequence ([V/I/A/L/F]xx[V/A/I/L/F/M]xx[R/K]), which was derived from the alignment of top SecA-binder peptides, is relatively simple and can be found in many proteins. For example, on the MBP sequence, this motif has been repeated at least nine times (Figure 3-13). Analysis of other preproteins (such as OmpA, DegP, and GBP) using specialized tools in finding “patterns” in protein sequences (e. g. Protein Pattern Find) showed a similar result. The abundance of this “motif” on the protein sequences is consistent with the fact that SecA has to interact with numerous substrates *in vivo* that do not share a high degree of sequence identity.



**Figure 3-11. Alignment of top SecA-binder peptides**

The peptides that showed the highest SecA binding strength were aligned. Top binders were organized into two groups, the first with relative signal intensity (RSI)  $> 0.4$  and the second with RSI  $> 0.5$ . The alignment of the RSI  $> 0.4$  group is shown in **Panels A and B**. A and B are essentially the results of a single alignment test, which are presented in two different formats for better observation of consensus sequences (A is presented in Kullback-leibler style, while B is in Shannon style). **Panels C and D** show the results of alignment of peptides with RSI  $> 0.5$ , and are represented in two different formats, similar to A and B.

```

-----FIRSMKPRFVKRGAR-----
-----TVDEALKDAQTRITK-----
----LVVSTLNNPFFVSLK-----
----MKKLLKLGFNLLTKS-----
----LLAEITPDKAFQDKL-----
-----FEQAMQTRVFQPLKL-----
----TAEHTQSVLKGFNKF-----
-----TIEKQLARTQRDKKR-----
----KVQAKYPVDLKLVVK-----
----LKSDVLFNFKATLK-----
----LISKGIPADKISARG-----
----EIPALDKELKAKGKS-----
----AQYEDFIRSMKPRFV-----
----LIEWLPGSTIWAGKR-----
----EAVNKDKPLGAVALK-----
----EVSASHASDNVGGK-----
----ANIPVITLDRQATG-----
----ENGAYKAQGVQLTAK-----
----NIDVGFGLSLAATR-----
----ETADKVLKGEKVQAK-----
----ELANVQDLTVRGTKI-----
----LNNPFFVSLKDGAQK-----
----NGSDNKIALAARPVK-----
----SRRERSLRRLEQRKN-----
----SALTTMMFSASALAK-----
      x*xx*xx*

```

**Figure 3-12. Location of the identified motif in the top SecA binder peptides.**

The top 25 peptides that had shown the strongest binding towards SecA and shared a common pattern. The position of the conserved K/R residues, -3 hydrophobic residues, and -6 hydrophobic residues are highlighted with blue, black, and gray respectively.

```

>MBPIP0AEX9I27-396
KIEEGKLVIIWINGDKGYNGLAEVGGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYYAQSG
LLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFN
LQEPYFTWPLIAADGGYAFKYENGYDIKDVGVDNAGAKAGLTFLVDLIKHKHMNADTDYSIAEAAFNKGETAM
TINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKP
LGAVALKSYEEELAKDPRIATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQTRITK

>OmpAIP0A9I0I22-346
APKDNTWYTGAKLGWSQYHDTGFINNNGPTHENQLGAGAFGGYQVNPYPVGFEMGYDWLGRMPYKGSVENG
AYKAQGVQLTAKLGYPIITDDLDIYTRLGGMVWRADTKSNVYGKNHDTGVSPVFAGGVEYAITPEIATRLEYQWT
NNIGDAHTIGTRPDNGMLSLGVSYRFGQGEAAPVVAPAPAPAPEVQTKHFTLKSDVLFNFNKATLKPEGQAAL
DQLYSQLSNLDPKDGSVVVLGYTDRIGSDAYNQLSERRAQSVDYLIISKGIPADKISARGMGESNPVTGNTCD
NVKQRAALIDCLAPDRRVEIEVKGIKDVVTQPQA

>GBPIP0AEE5I24-332
ADTRIGVTIYKYDDNFMSVVRKAIEQDAKAAPDVQLLMNDSQNDQSKQNDQIDVLLAKGVKALAINLVDPAAG
TVIEKARGQNVVFFFNKEPSRKALDSYDKAYYVGTDSKESGIIQGDIAKHWAANQGWDLNKDGQIQFVLLKG
EPGHPDAEARTTYVIKELNDKGIKTEQLQLDTAMWDTAQAKDKMDAWLSGNANKIEVVIANNNDAMAMGAVEA
LKAHNKSSIPVFGVDALPEALALVKSGALAGTVLNDANNQAKATFDLAKNLADGKGAADGTNWKIDNKVVRVP
YVGVDKDNLAEFSSK

```

**Figure 3-13. The SecA-binding motif position along the preprotein substrates**

The sequences of the mature regions of several *E. coli* SecA preprotein substrates were retrieved from Uniprot, and using the Protein Pattern Find tool, the location the “SecA binding motif”, [V/I/A/L/F]xx[V/A/I/L/F/M]xx[R/K], was identified. The identified sites that match this motif are highlighted in yellow. The results for MBP, OmpA, and GBP are provided here.

### 3.4 Discussion

SecA can bind the mature region of unfolded preproteins in solution. Using two different substrates, which include the unfolded fragments of mature regions of MBP (a natural SecA substrate), and an isolated high affinity SecA-binder peptide (FhuD peptide), it was shown that both a functional and a truncated SecA construct (SecA-N95 and SecA-N68, respectively) can bind specifically to the unfolded polypeptides *in vitro*. These results demonstrate that these unstructured polypeptides contain sequences to which SecA can bind. Moreover, it was shown that the polypeptide binding by SecA is affected by different nucleotides (Figure 3-7). In the presence of AMP-PNP, the binding was slightly lowered, while in the presence of ADP the maximum binding signal of SecA-N68 was decreased to around half of the value observed in the absence of nucleotides. ATP caused a similar change as ADP, which was expected given that SecA-N68 has a significant ATPase activity (Price *et al*, 1996). Overall, the observation of interaction of SecA constructs with unfolded polypeptides is consistent with previous works demonstrating that, *in vitro*, SecA constructs can recognize and be cross-linked to unstructured polypeptides (Kimura *et al*, 1991; Papanikou *et al*, 2005). However, the distinctive feature of the current study with the unfolded fragments of MBP is that it showed SecA can bind mature regions of a “natural” substrate of the Sec system *in vitro*.

Sec system, and particularly SecA, must interact with and translocate a vast variety of unfolded preproteins, which presumably do not share a high degree of similarity. Therefore, the degree of sequence specificity for the interaction is expected to be relatively low (Kusters & Driessen, 2011; Vrontou & Economou, 2004). In fact, by analyzing the sequences of the top SecA-binder peptides, which were identified by oriented peptide libraries assay, it was shown that the SecA binding “motif” is relatively simple (Figure 3-11). This simple pattern is frequently found in proteins. However, the exact binding site for these polypeptides on SecA has not been clear.

Several truncated constructs were used for tracing the polypeptide binding site on SecA. Both SecA-N95 and SecA-N68 exhibited binding to the unfolded polypeptides (Figure 3-4), indicating that the preprotein binding site is located on SecA-N68, and the CTD domains as well as the zinc binding domain (residues 610-901) are not critical for the interaction. SecA-N68 consists of two nucleotide binding domains, NBD1 and NBD2, along with a “preprotein cross-linking domain”

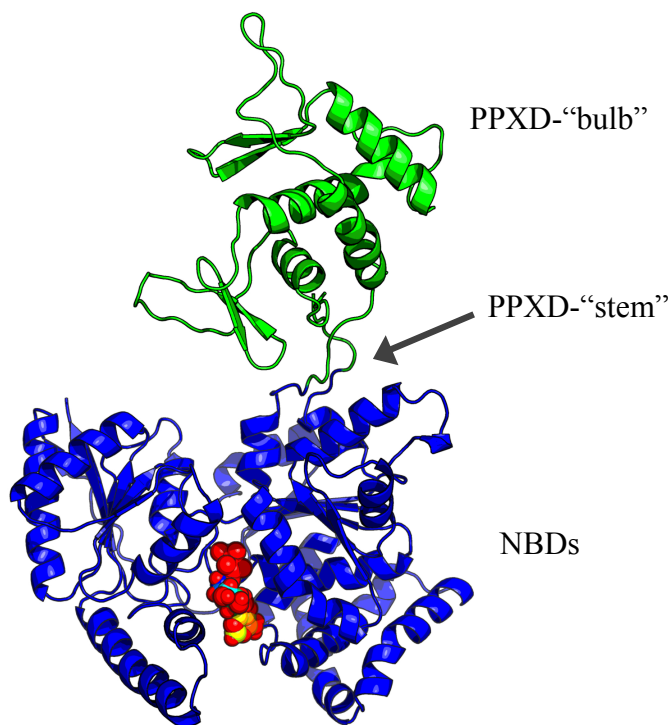


(PPXD), which is comprised of 150 residues, internally fused to NBD1 (Figure 3-14). The PPXD is composed of a “bulb” region (residues 233-362), which is the main part of PPXD, and two short anti-parallel  $\beta$ -strands called the “stem” region (residues 221-233 and 362-377) (Papanikou *et al*, 2005). The stem region connects the PPXD to the NBD1. Therefore, to further map the site of peptide interaction to particular SecA-N68 domains, two shorter constructs were used. The first construct, SecA-DM (composed of the first 227 residues as well as residues 367 to 609), contains the main part of the stem region but has the bulb region of PPXD excised (Nithianantham & Shilton, 2008). The other construct is a chimeric MBP-SecA-PPXD, in which the bulb region of PPXD (residues 229 to 368) was fused to a loop in MBP to ensure that all of its surfaces are available for binding. Subsequent analysis revealed that SecA-DM bound to the unfolded peptides with even higher affinity than either SecA-N95 or SecA-N68, while the MBP-SecA-PPXD showed no binding to the immobilized MBP peptides (Figure 3-6). In fact, the comparison of the binding specificities of various SecA constructs towards the unfolded fragments of MBP suggests that the observed lower binding signal for SecA-N95 and SecA-N68 compared to the SecA-DM may be due to the interference of the PPXD with preprotein binding. This interference may actually happen through covering parts of the “preprotein binding site”. Overall, by using truncated constructs, the polypeptide binding was traced to the SecA-DM while PPXD showed no tendency to bind the unfolded polypeptides.

A critical question is where the unfolded polypeptides bind on the SecA-DM. The observation of preprotein binding by SecA-DM is surprising because this construct lacks the PPXD, which is the very domain expected to be responsible for preprotein binding. That is, the term “preprotein cross-linking domain” (PPXD) originated because of early cross-linking experiments that involved a preprotein substrate and a collection of deletion constructs of SecA (Kimura *et al*, 1991). The specificity of the observed interactions is questioned as a nonspecific cross-linker was used and the cross-linking took place only when SecA proteins were partially “reconstituted” by refolding a mixture of N-terminal and C-terminal fragments of SecA together (Papanikou *et al*, 2005). Therefore, later, in reexamining the functional role the PPXD, it was shown that this domain is essential for translocation of preproteins (Papanikou *et al*, 2005). Moreover, based on proteolytic protection pattern of SecA-N68 in the presence of both a signal sequence and an “artificial” substrate of Sec system, which was used as the mature region of preproteins, it was concluded that the bulb and stem region of PPXD were involved in the

binding to the mature region and signal sequences, respectively (Papanikou *et al*, 2005). However, their direct binding assay revealed that although the isolated PPXD fragment, which contained both bulb and stem regions, retained some binding towards the polypeptides, its binding was several times lower compared to that of SecA-N68 and a construct containing only NBD1 and PPXD domains. Furthermore, the direct polypeptide binding characteristics of the nucleotide binding domains alone was not examined. Therefore, contrary to the presented conclusion, it is conceivable that the binding site for the preproteins is placed on the stem region of PPXD and/or NBDs, and the observed lower polypeptide binding in the isolated PPXD fragment to be likely due to the inability of termini of this construct to completely form the two anti-parallel  $\beta$ -strands of the stem region. In fact, supporting evidence for the presence of a polypeptide binding site on the stem region of PPXD and/or NBDs comes from a solved crystal structure of *B. subtilis* SecA bound to a peptide that does not resemble a signal sequence (Zimmer & Rapoport, 2009). In this complex structure, the short peptide was found to be bound to the base of stem of the PPXD domain, which is immediately adjacent to NBD1, by inducing a  $\beta$ -strand conformation. Also, it seemed that this peptide interacts with the neighboring residues from the NBD1. Interestingly, the exact same residues (corresponding to residues 224-226 of *ec*SecA) are available on the *E. coli* SecA-DM construct (Figure 3-14). Taken all together, it seems that the interaction with the polypeptides takes place through the NBDs and/or stem region of PPXD of SecA-DM.

It is not yet clear how the SecA pushes preproteins through the SecYEG. However, the idea that the DEAD domains of SecA can translocate a polypeptide through the membrane pore is consistent with the ability of a related class of RNA helicases to translocate along nucleic acids. Superfamily 2 RNA helicases (SF2) share the structurally conserved NBD domains with SecA, and they lack the PPXD appendix of SecA. While SecA binds polypeptides, SF2 helicases interact with a different biological polymer, RNA, and their main biochemical function is the ATP-dependent directional translocation on single or double stranded nucleic acids (Singleton *et al*, 2007). For example, in nonstructural protein 3 of hepatitis C virus (HCV NS3) interaction with the single stranded DNA happens through their NBD domains (Kim *et al*, 1998), and it is suggested that the ATP-dependent conformational motions in NBD domains is responsible for the translocation/walking along the nucleic acid. Therefore, due to the high degree of structural conservation, it has been suggested the same mechanism of conformational changes in DEAD



**Figure 3-14 Domain arrangement of *E. coli* SecA-N68**

The crystal structure of *E. coli* SecA-N68 is shown. The residues available in the DEAD motor (SecA-DM) is shown in blue. The “bulb” region of PPXD is shown in green. The “stem” region, which is composed of two  $\beta$ -strands connecting the DEAD motor and the bulb region of PPXD, is showed with an arrow. The nucleotide is shown in red. The coordinates for this structure were obtained from B. Shilton (personal communication).

motor domains may be applied for translocation of a completely unrelated polymer substrate, polypeptides, by SecA ATPase (Singleton *et al*, 2007). In this scenario, the PPXD instead of direct recognition of preproteins plays another critical role(s). For example, the crystal structure of SecA-SecYEG complex revealed that this domain is making major contacts with SecY protein, and consequently, it seems likely that the PPXD domain is responsible for interaction of SecA and SecYEG channel (Zimmer *et al*, 2008).

Furthermore, along with the aforementioned constructs, the binding specificity of other available SecA constructs, which lacked the unfolded termini, towards the unfolded polypeptides was tested. Surprisingly, it was revealed that the constructs that lacked the extreme N-terminal polypeptide were not able to bind the immobilized polypeptides (Figures 3-9 and 3-10). The exact function of the extreme N-terminal sequence of SecA has been a matter of debate (Kusters & Driessen, 2011; Sardis & Economou, 2010; Or *et al*, 2005; Jilaveanu *et al*, 2005). It has been shown previously that upon removal of the first 11 residues SecA turns into a monomer (Randall *et al*, 2005; Or *et al*, 2005; Jilaveanu *et al*, 2005), and is no longer able to support the growth of *E. coli in vivo* (Jilaveanu *et al*, 2005; Randall *et al*, 2005). Our results from Chapter 2 also indicated that the removal of this segment clearly affects the oligomerization of both SecA-N68 and SecA-N95, and its presence is important for *in vivo* function of SecA. On this basis, the observed loss of affinity towards the unfolded substrates upon removal of the extreme N-terminal sequence provides another insight into the function of this important region of SecA. There are two hypotheses for how the presence of the N-terminal sequence affects the preprotein binding by SecA. First, the dimer formation mediated through N-terminus may cause conformational changes in SecA that opens the preprotein binding site(s) for interaction with preproteins. Alternatively, the N-terminus may act in *cis* on each protomer and renders the preprotein binding site(s) accessible.

In summary, our overall results investigating the interaction of unfolded polypeptides with various SecA constructs revealed that preproteins contain sequences that can bind in a nucleotide dependent manner to SecA protein *in vitro* though the binding site(s) located on the NBD domains and/or stem region of PPXD. These interactions are affected by the presence of the N-terminal unstructured region of SecA. As the models for translocation of preproteins by Sec

system are developed, the critical role of these important regions of SecA should be taken into consideration.

## Chapter 4

### 4 Towards High Resolution Analysis of the Sec System's Components

A complete understanding of the translocation mechanism requires high resolution structures of the individual components and the molecular complexes that are formed during translocation. On this basis, structures of *E. coli* SecA-N95, both alone and in complex with a peptide representing a translocation substrate, have been longstanding goals in the field. Crystals of SecA-N95 had been obtained previously, but did not diffract beyond 8 Å resolution. The diffraction quality of SecA-N95 crystals was improved by removal of the unstructured N-terminal segment. A similar strategy was used to improve the crystallization of a SecB-like protein, Rv1957 or “*mtSecB*”, found in *M. tuberculosis*. Rv1957 was cloned, expressed, and shown to form a stable tetramer in solution, in common with *E. coli* SecB. However, crystallization was very slow and difficult to reproduce, problems that were attributed to the N-terminal region of *mtSecB* which is susceptible to proteolysis. A *mtSecB* construct in which the N-terminus was truncated by eight residues yielded large crystals relatively rapidly. Unfortunately these crystals did not diffract, and therefore additional work will be required to obtain a high-resolution structure of *mtSecB*.

#### 4.1 Introduction

SecA, the essential peripheral motor protein of the Sec system, is highly conserved across all the known bacteria (Vrontou & Economou, 2004; Kusters & Driessen, 2011). Many structural and functional aspects of the role of this protein in the translocation reaction have been extensively studied over the past few decades, and also several X-ray structures of SecA from various bacteria such as *B. subtilis*, *T. thermophilus*, *M. tuberculosis*, and *E. coli* are available (Hunt *et al*, 2002; Vassylyev *et al*, 2006; Sharma *et al*, 2003; Papanikolau *et al*, 2007a). *E. coli* SecA is a 204 kDa stable homodimer in solution (Wowor *et al*, 2011; Kusters *et al*, 2011). This protein consists of several distinct domains, including two nucleotide-binding domains (NBD1 and NBD2); the preprotein cross-linking domain (PPXD); the C-terminal domains (CTDs) including several domains that have been shown to be implicated in the inter- and intra-molecular

interactions between SecA protomers and other components of the Sec system (Karamanou *et al*, 1999; Zimmer *et al*, 2008; Hunt *et al*, 2002); and an unstructured zinc binding domain (ZBD) that is involved in the interaction with SecB and is the only “non essential” domain of SecA (Xu *et al*, 2000). SecA in concert with SecYEG is responsible for moving unfolded preproteins in a stepwise fashion through the membrane (Erlandson *et al*, 2008; Papanikolau *et al*, 2007b; Uchida *et al*, 1995). Upon interaction with the SecYEG and preproteins at the membrane, the ATPase activity of SecA is significantly increased, and also it appears that the conformational changes in SecA, which are derived from the binding and hydrolysis of ATP, are responsible for the movement of preproteins through the SecYEG (Economou & Wickner, 1994b). However, the main questions regarding the molecular aspects of the preprotein targeting by SecA and mechanics by which SecA translocates preproteins through the membrane pore remain unanswered. As a result, the high resolution structures of the individual components and the molecular complexes that are formed during translocation, especially in complex with a preprotein substrate, are needed for the complete understanding of the SecA-dependent translocation mechanism.

Despite the fact that SecA protein is most widely studied in *E. coli*, the complete structure of the full-length *E. coli* SecA is not resolved, and a crystal structure of this protein in complex with a translocating polypeptide is not yet available (Nithianantham & Shilton, 2008; Papanikolau *et al*, 2007b). Obtaining the high-resolution structure of *E. coli* SecA has proved to be difficult, as this molecule is refractory to crystallization. In an early investigation, crystals of the full-length *E. coli* SecA did not show detectable diffraction of X-rays. By electron microscopy analysis of thin sections of these crystals, a low resolution ( $\sim 40$  Å) envelope was attained, which did not reveal structural details (Weaver *et al*, 1992). This “misbehavior” was attributed to the large conformational flexibility of both N- and C-termini of this protein or the heterogeneity resulted from proteolysis of these termini (Weaver *et al*, 1992; Weinkauff *et al*, 2001). In fact, unpublished experiments, using various truncated constructs, showed that the presence of the flexible C-terminal ZBD prevented crystallization of full-length *E. coli* SecA (B. Shilton, personal communication). In *E. coli*, the ZBD is very short (only 22 amino acids) and is attached to the main body of SecA by a long unstructured polypeptide linker composed of 59 residues (Dempsey *et al*, 2004; Bechtluft *et al*, 2010; Vrontou & Economou, 2004). The structure of ZBD is stabilized through binding with  $Zn^{2+}$  ion, which is coordinated by 3 cysteines and a histidine

residue (Fekkes *et al*, 1998). By removal of the ZBD and linker, a crystal structure of this protein was obtained, although this structure missed most residues of the critical PPXD domain (Papanikolau *et al*, 2007b). On the other hand, however, X-ray structures of homologous SecA proteins from other bacteria have proved to be helpful (Hunt *et al*, 2002; Sharma *et al*, 2003; Vassylyev *et al*, 2006). In fact, the structure of *B. subtilis* SecA was used for modeling the missing PPXD domain in the *E. coli* SecA structure (Papanikolau *et al*, 2007b). Overall, in order to map the polypeptide binding site of SecA with high precision, obtaining high resolution crystal structures of *E. coli* SecA both alone and in complex with a peptide representing a translocation substrate have been longstanding goals in the field. However, obtaining crystals, that diffract to high resolution has proven difficult and has presented a major hurdle for obtaining a SecA-peptide complex.

Another important factor in SecA-mediated translocation of preproteins by Sec system is the targeting of preproteins to the translocon at the membrane. Given the physical distance between the location of preprotein synthesis by ribosomes in the cytoplasm and the Sec transport apparatus at the membrane, bacteria have devised various ways to maintain preproteins in an unfolded state and prevent their premature folding (Bechtluft *et al*, 2010). Many bacterial species express a dedicated chaperone, SecB, for this task (Randall & Hardy, 2002). Deletion of SecB leads to accumulation of various preprotein aggregations in the cytoplasm. Crystal structures of SecB are solved and have suggested that preproteins would stay bound to SecB by wrapping around this protein (Xu *et al*, 2000; Bechtluft *et al*, 2010). SecB is also believed to deliver preproteins directly to the membrane bound SecA (Zhou & Xu, 2003), and the interactions between *E. coli* SecB and SecA mainly occurs through the ZBD of SecA (Zhou & Xu, 2003; Kimsey *et al*, 1995).

The presence of the SecB chaperone is limited mainly to the Gram-negative bacteria (Bechtluft *et al*, 2010), however, it has been shown that many other prokaryotes have evolved proteins with similar function. For example, it has been proposed that in Gram-positive bacteria another chaperone called CsaA is involved in the targeting of preproteins to the Sec translocon (Müller *et al*, 2000; Kawaguchi *et al*, 2001). Similarly, a relatively recent study revealed that this mycobacterial's genomes contain the coding sequence for a putative SecB (Bordes *et al*, 2011a). Although the *M. tuberculosis* SecB (*mtSecB*) shares only 13% sequence similarity with the *E.*



*coli* SecB, *mtSecB* possesses some of the important structural elements such as the key *E. coli* SecB residues required for oligomerization, along with residues that are important for specific interaction with SecA (Sala *et al*, 2013). *In vivo* and *in vitro* assays showed that this protein exhibits chaperone functions similar to the SecB protein. Interestingly, although the *M. tuberculosis* housekeeping SecA1 lacks the extreme C-terminal cysteine residues involved in Zn<sup>2+</sup> ion binding, it has retained all of the conserved neighboring amino acids involved in the interaction with SecB (Bordes *et al*, 2011b; Zhou & Xu, 2003). Remarkably, *mtSecB* is also involved in a functional stress-responsive toxin-antitoxin (TA) system (Bordes *et al*, 2011b; Gerdes & Maisonneuve, 2012b). This dual function of *mtSecB* in Sec and TA systems has been suggested to be a mechanism for adaptation of mycobacteria to protein export stress conditions (Bordes *et al*, 2011a). However, despite the importance of post-translational secretion for the virulence and survival of *M. tuberculosis*, very little is known about its Sec system and also the mechanisms by which preproteins are kept in the translocation competent state by *mtSecB* or other chaperones are not clear (Feltcher & Braunstein, 2012). Overall, the fact that *mtSecB* possess the main predicted structural elements of *E. coli* SecB, which are essential for the interaction with preproteins and SecA, suggests that this protein plays a role in delivering the preproteins to SecA(s) in the mycobacterial Sec system. Since no atomic structures of *mtSecB* are available, structure determination of this protein can expand our knowledge of interaction of Sec system with unfolded preproteins.

Here, to gain further insight into the molecular mechanisms by which bacterial Sec system interact with unfolded polypeptides, we aimed at obtaining and improving the diffraction-quality crystals of an *E. coli* SecA and an *M. tuberculosis* SecB molecules; namely SecA-N95ΔN(ER) and *mtSecB*ΔN8. The SecA-N95ΔN(ER) lacks the non essential ZBD and the unstructured N-terminal residues of SecA and carries five entropy-reducing mutations. Similarly, the *mtSecB*ΔN8, as its name suggest, has the first eight residues removed. This truncation was carried out in regard to our observations that the extreme N-terminal segment of *mtSecB* is unstructured and interferes with the crystallization process. The overall results presented in this chapter reveal that by removal of the unstructured termini, especially the N-terminal region, of these molecules, larger crystals can be obtained with higher frequency, which in turn will increase the chances of obtaining high-resolution structures of these proteins.

## 4.2 Materials and methods

Trypsin,  $\alpha$ -Chymotrypsin, and proteinase-K were obtained from Sigma. Protein crystallization screening kits were obtained from Qiagen. All other media and reagents were obtained as described in Chapter 2.

### 4.2.1 Molecular cloning

The constructs used in the current study are listed in Table 4-1. All cloning procedures were carried out according to standard protocols (Sambrook & Russell, 2001). The list of primers used for the DNA amplification and sequencing reactions are provided in Table 4-2. The entire coding sequences of the constructs were confirmed by DNA sequencing using the primers listed in Table 4-2.

#### 4.2.1.1 Construction of pAK-SecA-N95 $\Delta$ N(ER) vector

The *EcoRI* fragment from pJH-N68-ER1-ER2 was inserted in the *EcoRI* sites of pAK-SecA-N95 $\Delta$ N plasmid. The pAK-SecA-N95 $\Delta$ N backbone generated by *EcoRI* digestion was treated with Antarctic Phosphatase to reduce the self-ligated vector background. After transformation, the correctly oriented vectors were selected by their ability to express SecA-N95 $\Delta$ N(ER) protein with the expected molecular weight.

#### 4.2.1.2 Cloning of Rv1957 from *M. tuberculosis*

To study the mycobacterial SecB (*mtSecB*) protein (from gene *Rv1957*), two constructs were made: *mtSecB* with an N-terminal hexahistidine tag (pAK-*mtSecB*) and *mtSecB* with an N-terminal hexahistidine tagged MBP (pETM41-AK-*mtSecB*). For construction of pAK-*mtSecB*, the coding sequence of *mtSecB* (*Rv1957*) gene was amplified from the genomic DNA of an attenuated *M. tuberculosis* strain (obtained from ATCC) using primers F-*mtSecB* and R-*mtSecB*

(Table 4-2). To remove some of the non-optimal codons from the *mtSecB* gene, the primers carried silent mutations for improved protein expression in *E. coli*. The PCR reaction was carried out using PfuTurbo DNA polymerase with the following procedure: an initial denaturation step at 96 °C for 5 min was followed by the first cycle (96 °C for 3 min, 68 °C for 5 min, and 72 °C for 5 min); the reaction was continued for 30 cycles (96 °C for 1 min, 51 °C for 1 min, and 72 °C for 4 min) and completed with an extension step for 10 min at 72 °C. The amplified DNA band was gel-purified and ligated between the *Bam*HI and *Ehe*I(*Ssp*DI) cut sites of the expression vector pProEX-HTa (Invitrogen) with T4 DNA ligase. The resulting plasmids were transformed into DH5 $\alpha$  cells.

A restriction enzyme-free cloning strategy was used for creation of the pETM41-AK-*mtSecB* vector (Klock & Lesley, 2009). The coding sequence of the *mtSecB* gene was amplified using primers F-AK-*mtSecB*-i and R-AK-*mtSecB*-i, and the backbone of the pETM41 vector was amplified using F-AK-pETM41-v and R-AK-pETM41-v primers (Table 4-2). After confirming the success of the PCR reactions by gel electrophoresis, the amplified vector and insert were mixed at various ratios and directly transformed into DH5 $\alpha$  cells.

#### 4.2.1.3 Cloning of truncated *mtSecB* constructs

A mutagenesis method based on Klock and Lesley's procedure (Klock & Lesley, 2009) was used for generation of the truncated version of *mtSecB* protein. In this method, for construction of each vector, a pair of mutagenic primers was used to amplify the entire pETM41-AK-*mtSecB* plasmid. For construction of pETM41-AK-*mtSecB* $\Delta$ N15, the primers F-*mtSecB* $\Delta$ N15 and R-*mtSecB* $\Delta$ N15 were used; for pETM41-AK-*mtSecB* $\Delta$ N8, primers F-*mtSecB* $\Delta$ N8 and R-*mtSecB* $\Delta$ N15 were used; and for pETM41-AK-*mtSecB* $\Delta$ N5, primers F-*mtSecB* $\Delta$ N5 and R-*mtSecB* $\Delta$ N15 were used. The PCR products were treated by *Dpn*I for 2 hours at 37 °C to remove the template plasmids and the vectors were transformed into DH5 $\alpha$  cells. DNA sequencing with the MBP primer (Table 4-2) was used to confirm the mutagenesis.

**Table 4-1. List of the constructs used in these studies**

<b>Construct</b>	<b>Plasmid</b>	<b>Features</b>
SecA-N68 $\Delta$ NC-ER	pJH-N68-ER1-ER2	Residues 15-590 in pProEX-Hta; carries entropy-reducing (ER) mutations: E55A, K56A, E58A, E196A, and E197A. Has a cleavable N-terminal His-tag
SecA-N95 $\Delta$ N	pAK-SecA-N95 $\Delta$ N	Residues 15-835 in pProEX-HTa; cleavable N-terminal His-tag
SecA-N95 $\Delta$ N(ER)	pAK-SecA-N95 $\Delta$ N(ER)	Residues 15-835 of <i>E. coli</i> SecA in pProEX-HTa; carries entropy-reducing (ER) mutations: E55A, K56A, E58A, E196A, and E197A. Has a cleavable N-terminal His-tag.
<i>mtSecB</i>	pAK- <i>mtSecB</i>	Residues 1-181 of <i>M. tuberculosis mtSecB</i> (Rv1957) in pProEX-HTa. Has a cleavable N-terminal His <sub>6</sub> -tag.
<i>mtSecB</i>	pETM41-AK- <i>mtSecB</i>	Residues 1-181 of <i>M. tuberculosis mtSecB</i> (Rv1957) in pETM41. Has a cleavable N-terminal His <sub>6</sub> -tagged MBP.
<i>mtSecB</i> $\Delta$ N15	pETM41-AK- <i>mtSecB</i> $\Delta$ N15	Residues 16-181 of <i>M. tuberculosis mtSecB</i> (Rv1957) in pETM41. Has a cleavable N-terminal His <sub>6</sub> -tagged MBP.
<i>mtSecB</i> $\Delta$ N8	pETM41-AK- <i>mtSecB</i> $\Delta$ N8	Residues 9-181 of <i>M. tuberculosis mtSecB</i> (Rv1957) in pETM41. Has a cleavable N-terminal His <sub>6</sub> -tagged MBP.
<i>mtSecB</i> $\Delta$ N5	pETM41-AK- <i>mtSecB</i> $\Delta$ N5	Residues 6-181 of <i>M. tuberculosis mtSecB</i> (Rv1957) in pETM41. Has a cleavable N-terminal His <sub>6</sub> -tagged MBP.

**Table 4-2. List of the primers used in these studies**

<b>Primer</b>	<b>Sequence</b>	<b>Length</b>	<b>Use</b>
F- <i>mtSecB</i>	5'- ATCTGAGGCGCCATGACTGACCGTACCGACGCCGACGACCTTG ACC	46	Cloning
R- <i>mtSecB</i>	5'- CATAATGGATCCTTACGGCGTTCCACGCGTTGCCGGCCATTGG GCACCGGGAGAAACC	58	Cloning
F-AK-pETM41-v	5'- CCACTGAGATTCGGCTGCTAACAAAGC	27	Cloning
R-AK-pETM41-v	5'- CGCCCTGAAAATACAGGTTTTCGCTCATGG	30	Cloning
F-AK- <i>mtSecB</i> -i	5'- CAACAACCCCATGAGCGAAAACCTGTATTTTCAGG	35	Cloning
R-AK- <i>mtSecB</i> -i	5'- AGCCGAATCTCAGTGGTCATCCGCCAAAACAGC	33	Cloning
F- <i>mtSecB</i> ΔN15	5'- TATTTTCAGGGCGCCCGCTGGCAGCCCGCGCAC	34	Cloning
R- <i>mtSecB</i> ΔN15	5'- GCGCCCTGAAAATACAGGTTTTCGCTCATGGGGTTGTTG	40	Cloning
F- <i>mtSecB</i> ΔN8	5'- TATTTTCAGGGCGCCGACCTTGACCTGCAACG	32	Cloning
F- <i>mtSecB</i> ΔN5	5'- TATTTTCAGGGCGCCGACCGACCTTGACC	34	Cloning
MBP	5'- GTCGTCAGACTGTCGATGAAGCC	23	Sequencing
F-AK-pProEX	5'- TAACAATTCACACAGGAAACAGACC	26	Sequencing
R-AK-pProEX	5'- TTCTCTCATCCGCCAAAACAGC	22	Sequencing

#### 4.2.2 Purification of SecA-N95 $\Delta$ N(ER)

For the initial crystallization trials, histidine-tagged SecA-N95 $\Delta$ N(ER) was expressed and purified as described in the Materials and Methods section of Chapter 2. The final purified protein was dialyzed against 20 mM HEPES pH 7.5, 150 mM NaCl, and 1 mM TCEP, aliquoted, flash-frozen in liquid N<sub>2</sub>, and stored at -80°C.

To increase the purity of the SecA-N95 $\Delta$ N(ER) preparation for crystallization, additional purification steps were incorporated. After the initial IMAC affinity purification step, the histidine tag was removed and the protein dialyzed overnight against 0.1 M Tris, 100 mM NaCl, pH 8.0. This initial step was followed by ammonium sulfate precipitation, hydrophobic interaction chromatography, Cibachron Blue affinity chromatography, and ion exchange chromatography. These procedures were used for the purification of non-histidine-tagged SecA constructs and are described in detail in the Materials and Methods section of Chapter 2. Preparative gel filtration was used as the final polishing step with a running buffer that consisted of 20 mM HEPES, 150 mM NaCl, and 1 mM TCEP at pH 7.5. Fractions containing SecA95 $\Delta$ N(ER) were pooled and concentrated to 20 mg/mL using a 15 mL Amicon Ultra Centrifugal Filter Unit (Millipore) with a MWCO of 10 kDa. The concentrated protein was aliquoted, flash frozen in N<sub>2</sub>, and stored in -80°C for crystallization trials.

#### 4.2.3 Purification of *mtSecB* proteins

Protein expression was carried out in a *E. coli* BL21(DE3) background. Cells carrying pAK-*mtSecB* were grown in the media containing ampicillin (50 mg/L), while pETM41 constructs were grown in the presence of of kanamycin (30 mg/L). For recombinant protein expression, overnight cultures were prepared by transferring the freshly grown colonies into LB broth. The overnight cultures were used to inoculate of a few liters of fresh LB broth with 1:1000 dilutions. The cells were grown at 37 °C to an OD<sub>600</sub> of approximately 0.6; protein expression was induced with addition of 0.5 mM IPTG. Thereafter, the cells were allowed to grow overnight at 30 °C while shaking at 220 rpm. The cells were harvested and *mtSecB* was purified by nickel affinity, ion-exchange, and gel filtration chromatography, essentially as described in the methods section

of Chapter 2. Purified *mtSecB* proteins were dialyzed against 10 mM Tris-HCl and 1 mM TCEP at pH 7.5, aliquoted, flash-frozen in liquid N<sub>2</sub>, and stored at -80 °C.

#### 4.2.4 Partial proteolysis and Edman sequencing

For partial proteolysis experiments, a solution of *mtSecB* protein (300 µg/mL) was prepared in digestion buffer (50 mM Tris-HCl and 150 mM NaCl at pH 7.4). Solutions of DPCC-treated trypsin, α-chymotrypsin, and proteinase-K were prepared in the digestion buffer at a concentration of 1 mg/mL. Each protease solution was serially diluted 1:2 (v:v) with the *mtSecB* protein solution to give eight protease concentrations. The solutions were incubated at room temperature for 30 min; proteolysis was stopped by addition of 5X SDS loading buffer (0.5 M DTT, 50% glycerol, 10% SDS, and 0.25% bromophenol blue) and boiling for 6 min. The proteolysis patterns were visualized by SDS-PAGE.

In preparation for N-terminal sequencing by Edman degradation, a 300 µg/mL solution of *mtSecB* was digested with trypsin (10 µg/mL) for 30 min at room temperature; the reaction was stopped by boiling in SDS loading buffer. Samples of partially digested *mtSecB* were loaded into several wells of a 15% SDS-PAGE gel. The bands separated on the acrylamide gel were then blotted onto a Polyvinylidene Difluoride (PVDF) membrane at 30 V overnight in 10 mM 3-(cyclohexylamino)-1-propanesulfonic acid (CAPS), 10% methanol, pH 11. The blotted proteins were visualized by Coomassie blue (0.1% Coomassie blue R-250, 40% methanol, and 10% acetic acid), and submitted to the SPARC BioCentre, Toronto for N-terminal sequencing using an Applied Biosystems PROCISE 491C Protein Sequencing System.

#### 4.2.5 Analytical ultracentrifugation and size exclusion chromatography

Analytical size exclusion chromatography was carried out using a Superdex 200 HR 10/30 column. The running buffer was 50 mM Tris-HCl pH 7.5, 100 mM KCl, 1 mM EDTA, 5 mM MgCl<sub>2</sub>, and 5 mM β-mercaptoethanol, and the column was developed at flow rate of 0.7 mL/minute at room temperature. Samples (50 µL) of the analyte or molecular weight standards

were injected onto the column and absorbance was monitored at 280 nm. The molecular weight standards used were obtained from Sigma and included catalase (250 kDa), alcohol dehydrogenase (150 kDa), BSA (66 kDa), and carbonic anhydrase (29 kDa). Acetone (10 mg/mL) and Blue Dextran 2000 (GE Healthcare; 1 mg/mL) were used to determine the included and void volumes, respectively.

The sedimentation velocity ultracentrifugation experiments were conducted at 20°C as described previously (Chapter 2). Protein samples were extensively dialyzed against 50 mM Tris-HCl pH 7.4, 100 mM KCl, 2 mM EDTA, 5 mM MgCl<sub>2</sub>, and 2 mM TCEP. The final dialysis buffer was used as the “reference solution”. Protein samples and reference solutions were loaded into the sample and reference cells, respectively, and after thermal equilibration of the rotor, samples were centrifuged at 30,000 rpm. Absorbance measurements were collected at 280 or 295 nm with a 0.002 cm radial step and averaged over three readings. A total of 30 scans were collected at 10 minute intervals. The data were processed using software package SedFit. A partial specific volume ( $V\text{-bar}$ ) of 0.729 mL/g for *mtSecB* was calculated from its amino acid composition using the SEDNTERP software.

#### 4.2.6 Protein crystallization

Initial crystallization screens were conducted using sitting drop vapour diffusion in 96 well plates with assorted NeXtal screening kits (Qiagen), including PACT, JCSG+, PEGs, Classics Lite, and Protein Complex Suites. Typically 0.8  $\mu$ L of protein solution (SecA-N95 $\Delta$ N(ER) at 20 mg/mL; *mtSecB* at 20 mg/mL; and *mtSecB* $\Delta$ N8 at 12 mg/mL) was mixed with 0.8  $\mu$ L of the reservoir solution and allowed to equilibrate at either room temperature or 10 °C. SecA-N95 $\Delta$ N(ER) protein was supplemented with 2 mM ADP, 5 mM MgCl<sub>2</sub>, and 2 mM TCEP, while *mtSecB* protein solutions contained 2 mM TCEP.

For the final crystallization of SecA-N95 $\Delta$ N(ER), 0.8  $\mu$ L of protein (10 mg/mL SecA-N95 $\Delta$ N(ER), 2 mM ADP and 5 mM MgCl<sub>2</sub>, and 1 mM TCEP) was mixed with 0.8  $\mu$ L of reservoir solution (100 mM HEPES pH 7.0 and 7-15% PEG 8,000) using a 24-well sitting drop vapor diffusion tray. The trays were stored at room temperature. In preparation for X-ray



diffraction, the crystals were soaked briefly in reservoir solution supplemented with 30% glycerol before flash freezing in liquid N<sub>2</sub>.

For the final crystallization of *mtSecB* $\Delta$ N8, 1  $\mu$ L volumes of *mtSecB* protein (12 mg/mL and 2 mM TCEP) were mixed with an equal volume of reservoir solution (50 mM imidazole, 300 mM potassium thiocyanate, 14-16% PEG 3350 at pH 6.5) using a 24-well sitting drop vapour diffusion tray. The trays were stored at 10 °C and the crystals allowed to grow for three weeks to reach maximum size. In preparation for X-ray diffraction, crystals were soaked in reservoir solution with 30% glycerol and flash-frozen in liquid N<sub>2</sub>.

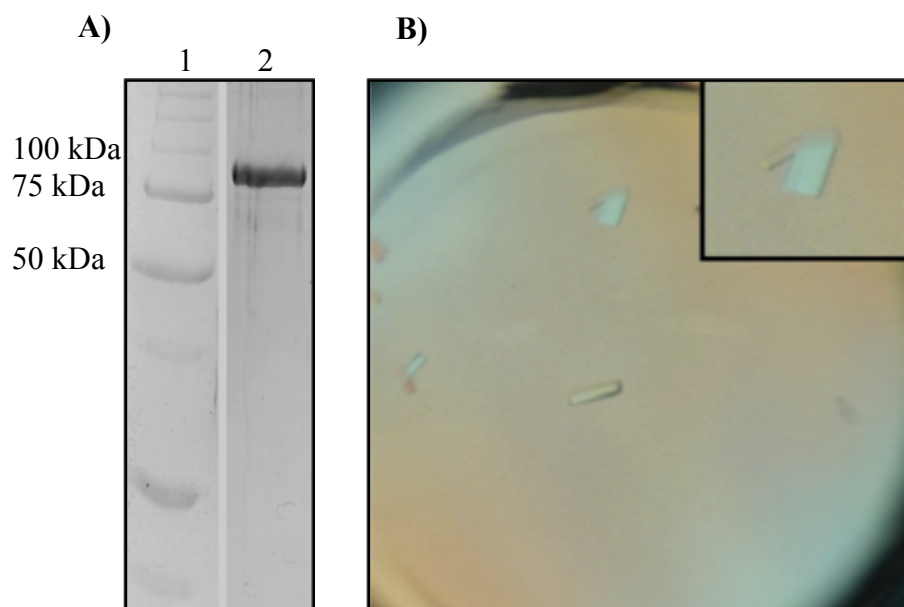
Crystallographic data were collected at the Canadian Light Source CMCF Beamline 08ID-1, and the data analyzed using iMosflm.

## 4.3 Results

### 4.3.1 Purification and initial crystallization of SecA-N95 $\Delta$ N(ER)

SecA-N95 was crystallized previously in the Shilton laboratory, but the crystals did not diffract beyond 8 Å resolution using a synchrotron source (CHESS beamline A-1). In another case, crystals of SecA-DM (Nithianantham & Shilton, 2008) were improved by removal of the unstructured N- and C-termini, and introduction of “entropy reducing” (ER) mutations. These mutations involved the conversion of surface-exposed flexible residues to alanine, specifically E55A, K56A, E58A, which formed a crystal contact to yield a 1.4 Å structure of SecA-DM $\Delta$ NC (B. Shilton, personal communication). Therefore, a similar strategy was used to improve the crystals of SecA-N95. In addition to removal of the unstructured N-terminus and inclusion of the E55A/K56A/E58A ER mutations, two additional residues, E196 and E197, were mutated to alanine. The resulting construct was termed SecA-N95 $\Delta$ N(ER).

SecA-N95 $\Delta$ N(ER) was expressed as a fusion protein with an N-terminal hexahistidine tag connected by a linker with a TEV protease cleavage site. The protein was expressed and purified as described in Chapter 2 for histidine-tagged constructs. The purification included Ni<sup>2+</sup>-affinity, ion-exchange, and gel filtration chromatography. An SDS-PAGE gel of the final preparation is illustrated in Figure 4-1A. Although there appear to be some minor contaminants, the initial crystallization screens yielded SecA-N95 $\Delta$ N(ER) crystals from several solutions. The precipitant in all cases was PEG. Further efforts to optimize the crystallization by varying protein concentration, temperature, buffer concentration, and PEG concentration led to conditions that included 100 mM HEPES, 150 mM magnesium acetate, and 12% PEG 8,000 at pH 7.5 (Figure 4-1B). The crystals appeared both at 10 °C and at room temperature after approximately 24 hrs. Although the crystals were well shaped and diffraction appeared promising, the crystals would stop growing before reaching a suitable size. Furthermore, they would lose their ability to diffract X-rays and eventually disappear from the crystallization drops. Growth of the crystals at 4 °C resulted in the same pattern of events. To address this issue, additional purification steps were included in the preparation of SecA-N95 $\Delta$ N(ER).

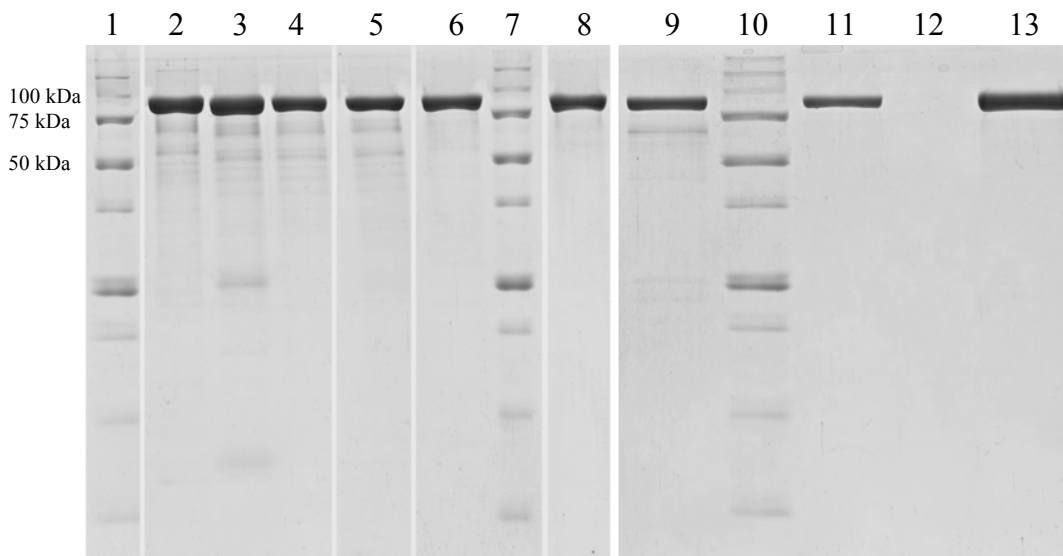


**Figure 4-1 SDS-PAGE analysis and initial crystallization of SecA-N95 $\Delta$ N(ER)**

(A) SDS-PAGE analysis of the initial preparation of SecA-N95 $\Delta$ N(ER) (lane 2), along with molecular weight standards (lane 1). (B) An example of the crystals obtained from this protein, grown at 10°C in a system composed of 100 mM HEPES, 150 mM magnesium acetate, and 12% PEG 8,000 at pH 7.5, whose longest dimension is around 0.2 mm.

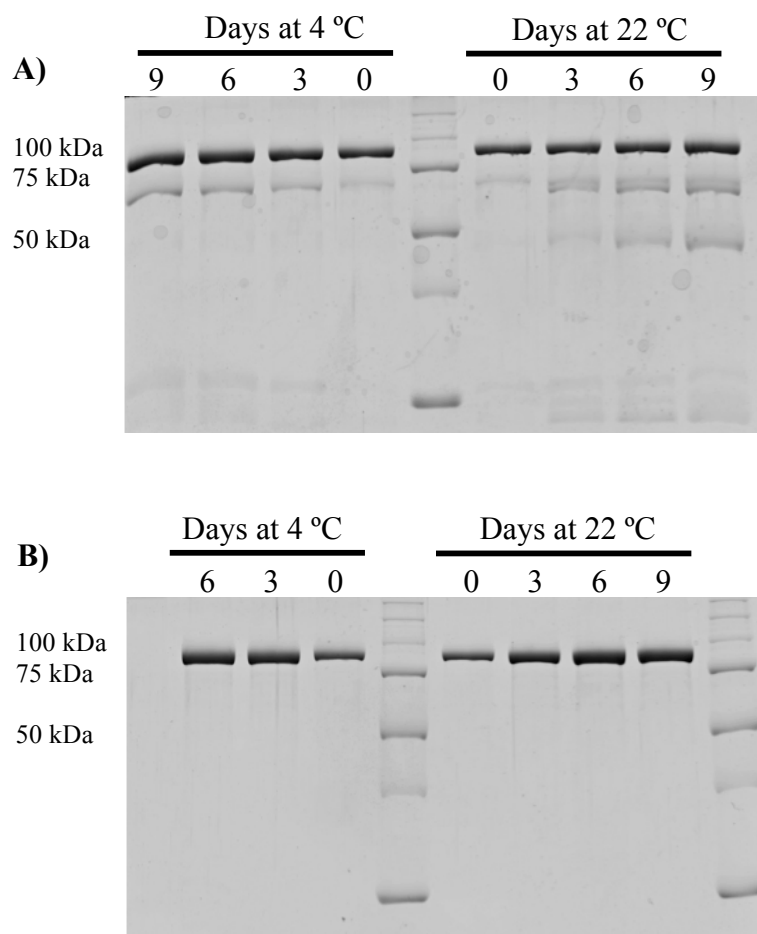
### 4.3.2 Optimization of SecA-N95 $\Delta$ N(ER) purification for greater stability

The minor bands in the SDS-PAGE gel of purified SecA-N95 $\Delta$ N(ER) (Figure 4-1A) increased in intensity over time, indicating that the protein was undergoing degradation. The addition of protease inhibitors did not prevent this degradation, and therefore additional purification steps were included in the preparation of SecA-N95 $\Delta$ N(ER) to ensure that contaminating proteases were completely removed. After the first affinity purification using Ni<sup>2+</sup>-NTA chromatography and removal of the affinity tag using TEV protease, the untagged protein was subjected to the same purification procedure that had initially been developed for untagged SecA, and which is described in detail in the Methods for Chapter 2. The additional purification steps included ammonium sulfate precipitation, Cibacron Blue affinity chromatography, anion-exchange chromatography, and gel filtration chromatography. SDS-PAGE analysis of the SecA-N95 $\Delta$ N(ER) preparation at each stage of the purification is provided in Figure 4-2; the final yield of 60 mg of SecA-N95 $\Delta$ N(ER). The stability of this preparation was compared to the stability of the previous preparation that had been crystallized but failed to yield sufficiently large crystals (Figure 4-3). It can be seen that the original preparation of SecA-N95 $\Delta$ N(ER) undergoes degradation over time even at 4 °C (Figure 4-3A). However, the improved preparation of SecA-N95 $\Delta$ N(ER), with additional purification steps, underwent no apparent degradation even after 9 days of at room temperature (22 °C; Figure 4-3B).



**Figure 4-2. Optimized purification of SecA-N95ΔN(ER) for crystallization.**

SDS-PAGE analysis of SecA-N95ΔN(ER) over the course of an extended purification. Lanes 2 to 4 document the first purification steps using Ni<sup>2+</sup>-affinity chromatography: SecA-N95ΔN(ER) protein was first isolated by Ni<sup>2+</sup>-affinity chromatography (lane 2); the N-terminal affinity tag was removed by digestion with TEV protease (lane 3); tagged proteins and the TEV protease were captured by Ni<sup>2+</sup>-affinity chromatography and the flow through (lane 4) was retained. SecA-N95ΔN(ER) was precipitated using ammonium sulphate (lane 5), and purified using hydrophobic interaction chromatography (lane 6), Cibachron blue affinity chromatography (lane 8), and ion exchange chromatography (lane 9). Gel filtration chromatography was used as the final polishing step: 200 μg and 400 μg of the final preparation is loaded in lanes 11 and 13, respectively. Lanes 1, 7, and 10 are molecular weight standards.



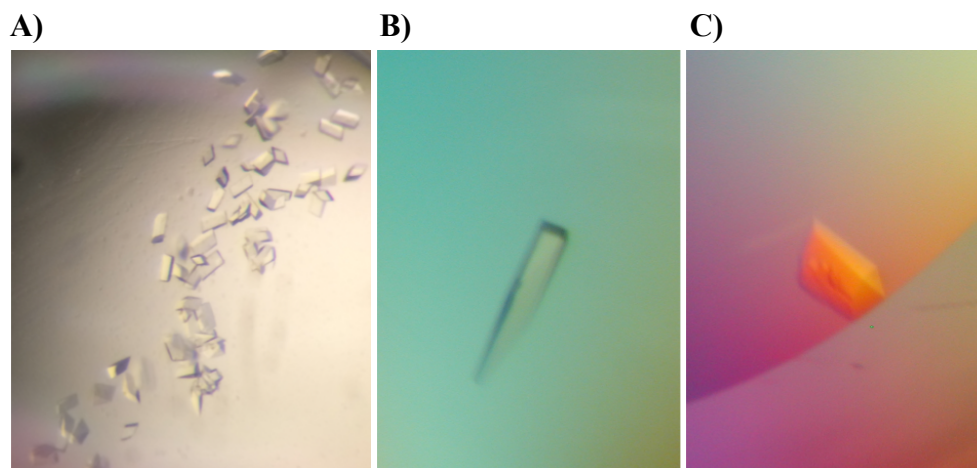
**Figure 4-3. Improved *in vitro* stability of SecA-N95ΔN(ER) protein**

The stability of two SecA-N95ΔN(ER) preparations at 4 °C and 22 °C was followed over 9 days by SDS-PAGE analysis. **Panel A** SecA-N95ΔN(ER) was purified by Ni<sup>2+</sup>-NTA affinity chromatography, followed by ion exchange and gel filtration chromatography. **Panel B** SecA-N95ΔN(ER) was purified using an extended procedure that included Ni<sup>2+</sup>-NTA affinity chromatography followed by ammonium sulfate fractionation and four chromatographic steps: hydrophobic interaction, Cibachron Blue affinity, anion exchange, and gel filtration chromatography.

### 4.3.3 Crystallization of SecA-N95 $\Delta$ N(ER)

The extensively purified and stable SecA-N95 $\Delta$ N(ER) preparation was used for crystallization trials. Three new crystallization conditions were identified by repeating the crystallization screens at room temperature. The new conditions included 0.1 M HEPES, pH 7.0, 20% PEG 8,000; 0.1 M MES pH 6.5, 12% PEG 20,000; and 0.1 M HEPES pH 7.5, 15% PEG 20,000. The initial conditions were further optimized by varying the protein concentration, incubation temperature, buffer concentration, and PEG concentration. Both sitting drop and hanging drop vapour diffusion strategies in combination with micro seeding were successful (Figure 4-4). The best crystals grew using sitting drop vapour diffusion at room temperature with a reservoir solution containing 100 mM HEPES pH 7.0 and 5% to 10% PEG 8,000 both with and without ADP. Crystals would appear overnight and reach their maximum size in a few days. The largest crystal obtained had dimensions of 0.3 mm  $\times$  0.3 mm  $\times$  0.2 mm.

In preparation for X-ray diffraction analysis, crystals of SecA-N95 $\Delta$ N(ER) were soaked in a cryoprotectant consisting of the reservoir solution supplemented with 30% glycerol, and flash frozen in liquid N<sub>2</sub>. Four of the crystals yielded diffraction slightly beyond a resolution of 6 Å. Three of these images were indexed and indicated a monoclinic space group, C2, with average unit cell dimensions of a=159, b=105, and c=77 Å and  $\beta$ =105°. This unit cell has a volume of 1.2  $\times$  10<sup>6</sup> Å<sup>3</sup> and for SecA-N95 $\Delta$ N(ER) with a molecular weight of approximately 95 kDa, one molecule in the asymmetric unit yields solvent content of approximately 63%. Crystals of SecA-N95 obtained previously were in a tetragonal space group, P4, with unit cell dimensions of a=b=259.93 and c=213.52 Å. The very large unit cell of these crystals could contain from 4 to 8 SecA-N95 protomers per asymmetric unit, corresponding to a solvent content between 75% and 50%; these crystals did not diffract beyond 8 Å. On this basis, removal of the unstructured termini and/or introduction of the entropy reducing mutations improved the quality of SecA-N95 crystals.



**Figure 4-4 Crystals of SecA-N95ΔN(ER)**

Crystals of SecA-N95ΔN(ER) were produced by hanging drop (**Panel A**) and sitting drop (**Panels B and C**) vapor diffusion. In Panel A, the drop was mixed with an equal part of reservoir solution (100 mM HEPES pH 7.0 and 5 % PEG 8,000) and allowed to equilibrate; crystallization was initiated by microseeding. In Panels B and C the reservoir solutions contained 100 mM HEPES pH 7.0 and 7% or 8% PEG 8,000, respectively, and crystals appeared spontaneously. SecA-N95ΔN(ER) protein used in the Panel C contained 2 mM ADP.

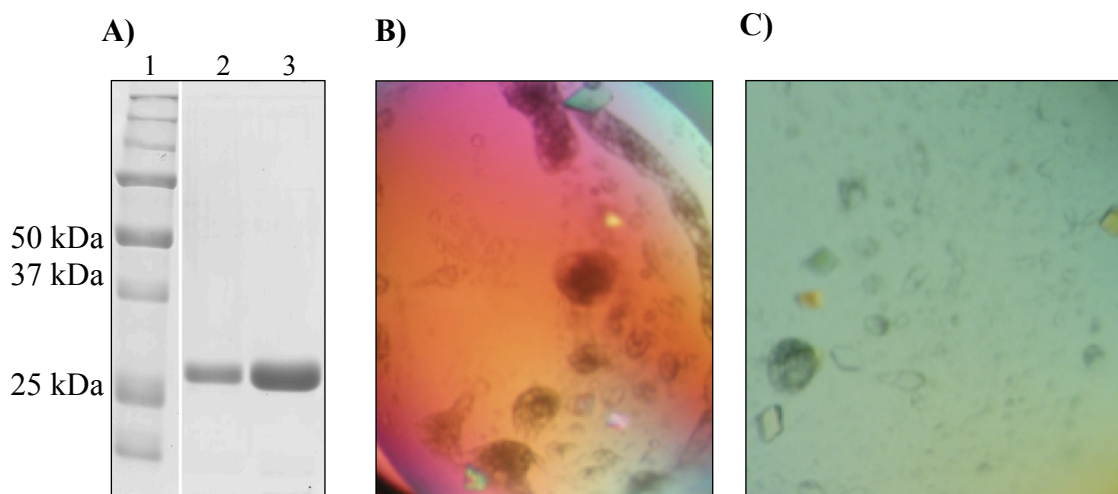


#### 4.3.4 Purification, characterization, and crystallization of *mtSecB*

To obtain a crystal structure of *mtSecB* (Rv1957), the coding region was cloned into two different expression vectors, pProEX-HTa and pETM41. Cloning *mtSecB* into pProEX-HTa yielded pAK-*mtSecB*, which expresses a fusion protein consisting of a 2 kDa N-terminal hexahistidine tag that is connected to *mtSecB* with a linker that can be cleaved by TEV protease. Cloning *mtSecB* into pETM41 resulted in the expression vector pETM41-AK-*mtSecB*, in which the *mtSecB* protein is expressed as a N-terminal fusion with histidine tagged maltose binding protein (MBP), connected by linker with a TEV protease cleavage site. The coding region of *Rv1957* contained two rare Arg codons at the extreme 5' and 3' termini, along with a rare Ile codon at the 3' termini. To ensure robust expression in an *E. coli* background, these rare codons were replaced with corresponding abundant *E. coli* codons in the primers used for amplification of *Rv1957*. Both vectors yielded robust expression of the fusion proteins, and in both cases; over 40 mg of purified protein was obtained by sequential Ni<sup>2+</sup>-affinity, anion exchange, and gel filtration chromatography (Figure 4-5A). The *mtSecB* protein has a molecular weight of 20.2 kDa but migrates anomalously on SDS-PAGE with an apparent molecular weight of 27 kDa SDS-PAGE (Bordes *et al*, 2011b).

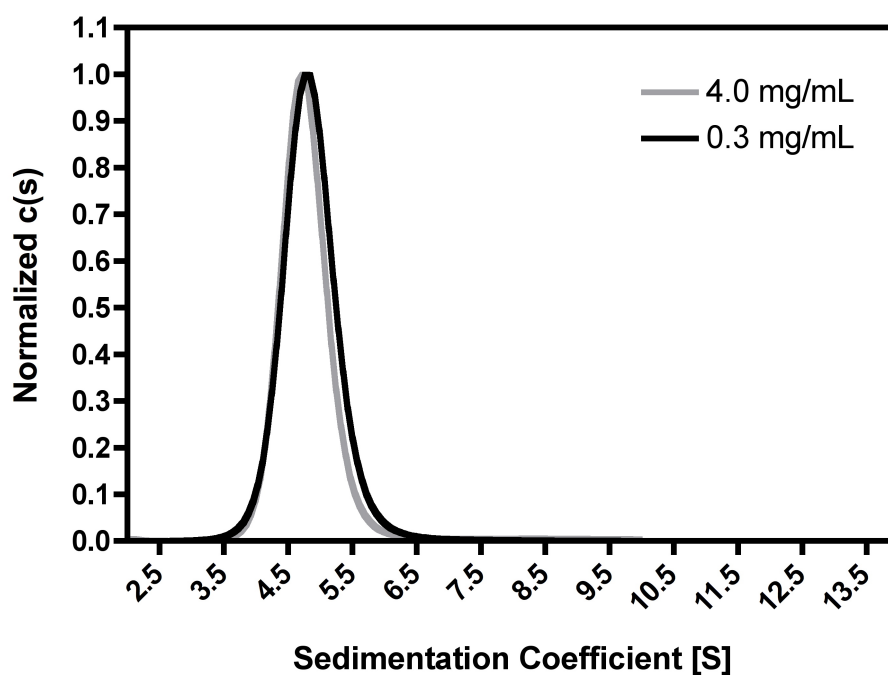
Previous studies with *mtSecB* indicated that it had a solution molecular weight of approximately 115 kDa when analyzed by gel filtration chromatography; therefore, due to its sequence and functional resemblance to *E. coli* SecB, *mtSecB* was predicted to be a tetramer in solution (Bordes *et al*, 2011b). To characterize the solution structure of *mtSecB* in greater detail, sedimentation velocity experiments were performed at protein concentrations of 0.3 mg/mL (15  $\mu$ M) and 4.0 mg/mL (198  $\mu$ M; Figure 4-6). The *mtSecB* protein sedimented as a single species with a coefficient of 4.7 S, which is consistent with a molecular weight of 82 kDa, indicating that *mtSecB* does indeed form a tetramer in solution. Furthermore, *mtSecB* sedimented with the same coefficient at concentrations of 15  $\mu$ M and 198  $\mu$ M, indicating that the tetramer has a dissociation constant of 0.5  $\mu$ M or less. The solution tetramer has a frictional ratio ( $f/f_{min}$ ) of 1.37 indicating a globular shape. No larger molecular weight species could be observed in the sedimentation plot indicating that protein aggregates were not present.

Crystallization trials were initiated using *mtSecB* expressed from plasmid pETM41-AK-*mtSecB*. Crystallization screening was done both at 4 °C and room temperature using a protein concentration of 10 mg/mL and sitting-drop vapour diffusion. The first *mtSecB* crystals appeared after approximately three weeks in 200 mM potassium thiocyanate and 20 % PEG 3350, pH 6.5 and 10 °C (condition 1) and in 200 mM K/Na tartrate, 100 mM Bis-Tris Propane pH 6.5, and 20 % PEG 3350 at 10 °C (condition 2). The crystals were difficult to reproduce: in the shorter term (48 hours) a phase separation was apparent, but crystals did not appear until 2 months later (Figure 4-5C). Further attempts to reproduce the crystallization were not successful.



**Figure 4-5 SDS-PAGE analysis of purified *mtSecB* protein and its initial crystallization**

(A) The SDS-PAGE profile of purified *mtSecB* protein that was expressed as a fusion with either an N-terminal His-tagged MBP (lane 2) or a 2 kDa N-terminal His-tag (lane 3). The *mtSecB* protein has a molecular weight of 20.2 kDa but migrates anomalously on SDS-PAGE. (B) The purified *mtSecB* protein was used for crystallization trials in which an initial hit was identified after approximately three weeks. These crystals appeared in 200 mM potassium thiocyanate and 20 % PEG 3350, pH 6.5. (C) Additional experiments were carried out to optimize the crystallization, but crystallization was extremely slow. In the case pictured here, the *mtSecB* crystals appeared after two months in 200 mM potassium thiocyanate and 18-20 % PEG 3350, pH 6.5; the crystals were approximately 0.05 mm in diameter.

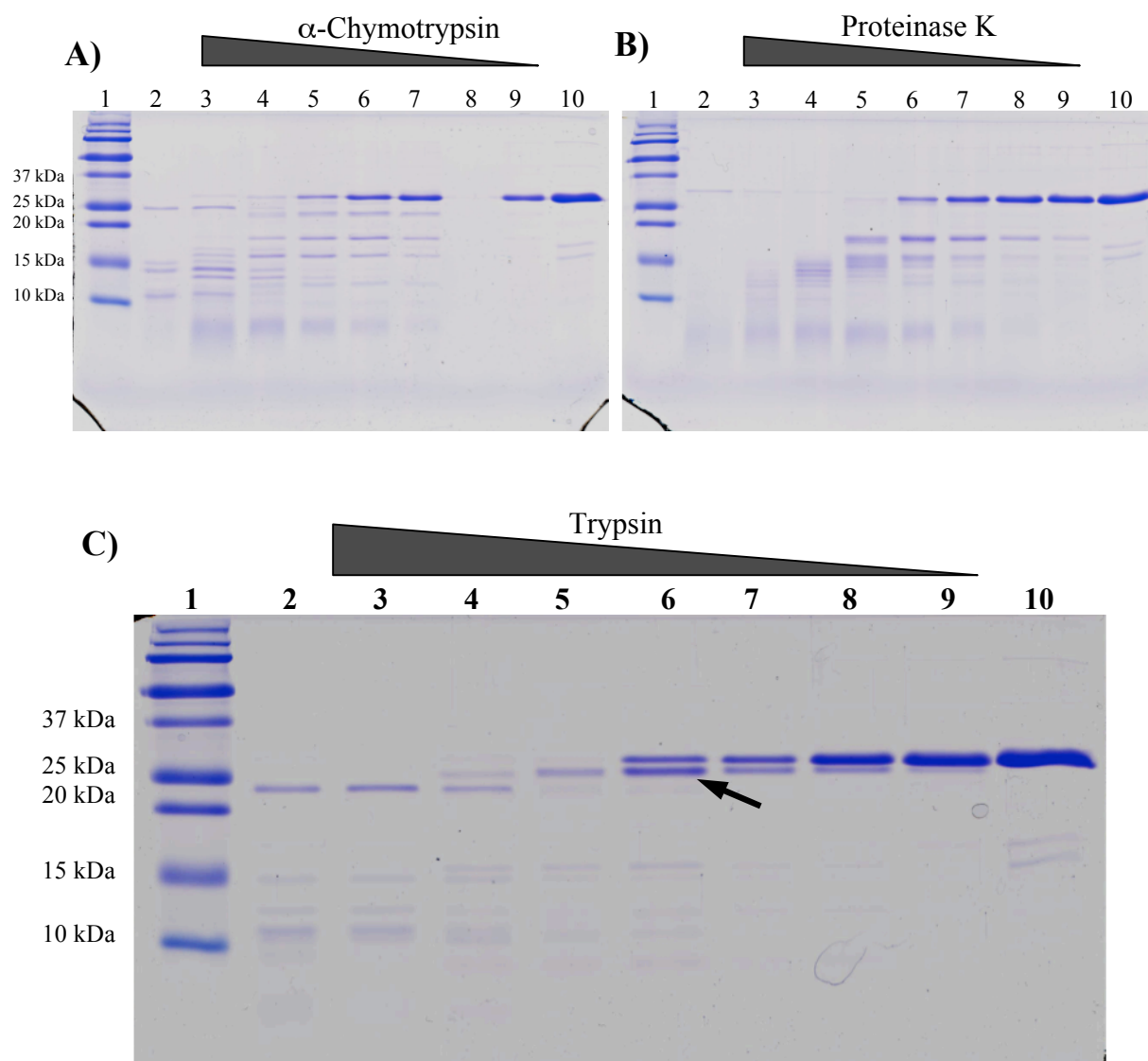


**Figure 4-6 Sedimentation velocity analysis of wild-type *mtSecB* protein**

The solution behavior of the purified mycobacterial *mtSecB* protein was characterized by sedimentation velocity. The distribution plot from SedFit analysis of two velocity runs at *mtSecB* concentrations of 15  $\mu\text{M}$  (0.3 mg/mL; black curve) and 198  $\mu\text{M}$  (4 mg/mL; light grey curve) is shown. At both concentrations, the *mtSecB* protein sedimented with a coefficient of 4.7 S. S-values are observed sedimentation coefficient (sobs). The distribution plots have been normalized by amplitude.

### 4.3.5 Crystallization of an N-terminal truncation of *mtSecB*

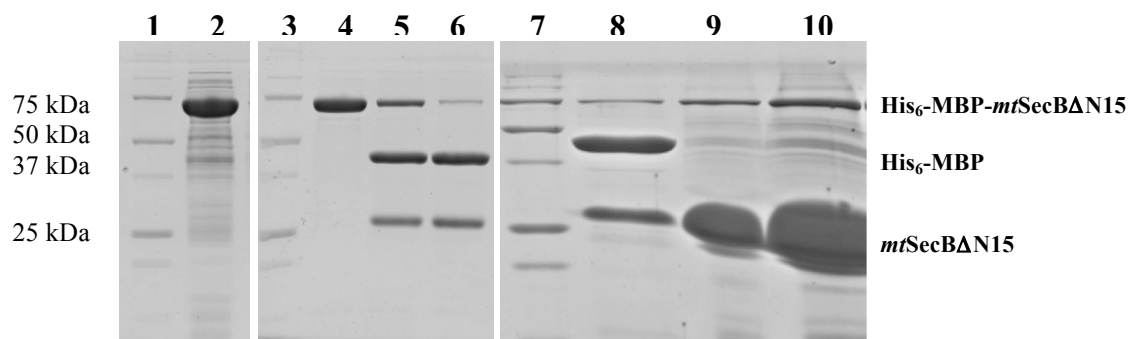
Previous crystallization trials indicated that *mtSecB* formed crystals only after a long incubation. This behaviour could indicate that *mtSecB* was undergoing slow limited proteolysis prior to crystallization. That is, unstructured regions, particularly at the N- and C-termini, could interfere with crystallization, while at the same time these regions are expected to be relatively susceptible to proteolysis. To test this possibility, the purified *mtSecB* protein was subjected to limited proteolysis using  $\alpha$ -chymotrypsin, proteinase-K, and trypsin. The protease solutions were serially diluted from 1 mg/mL stocks with a solution containing *mtSecB* (300  $\mu$ g/mL) to yield protease concentrations from 330  $\mu$ g/mL down to 450 ng/mL in a constant concentration (200  $\mu$ g/mL) of *mtSecB*. The solutions were incubated for 30 minutes, and the reaction stopped by heating aliquots in SDS-PAGE sample buffer. The extent of digestion was assessed by SDS-PAGE (Figure 4-7). When *mtSecB* was digested with  $\alpha$ -chymotrypsin or proteinase-K, a stable proteolytic fragment was not evident (Figure 4-7A and B). In the case of digestion with trypsin, a stable fragment was produced that was approximately 2.1 kDa smaller than full-length *mtSecB* (lane 6 in Figure 4-7B). N-terminal sequencing of this excised band yielded the sequence LAARAQ, corresponding to residues 19-24 of full-length *mtSecB*. The N-terminal sequence of *mtSecB* is GAMTDRTDADDLQRLAARAQ (the initial GA is a remnant from the TEV protease cleavage), and therefore trypsin cleaves the protein after R18.



**Figure 4-7 Partial proteolysis of *mtSecB* protein**

*mtSecB* protein was partially digested using  $\alpha$ -chymotrypsin (**Panel A**), proteinase-K (**Panel B**), and trypsin (**Panel C**). Experiments were performed under the same conditions. For all gels, lane 1 contains molecular weight markers: lane 2 is a sample of protease only, lanes 3 to 9 are the partial proteolysis reactions, and lane 10 contains untreated *mtSecB* protein. The *mtSecB* protein concentrations were constant at 200  $\mu\text{g}/\text{mL}$ , and the increasing concentration of each protease (from 0330  $\mu\text{g}/\text{mL}$  in lane 3 to 450  $\text{ng}/\text{mL}$  in lane 9) is shown as a triangle. Reactions were stopped after 30 minutes by heating in SDS-PAGE sample buffer; the main product of the trypsin partial digestion is shown by an arrow.

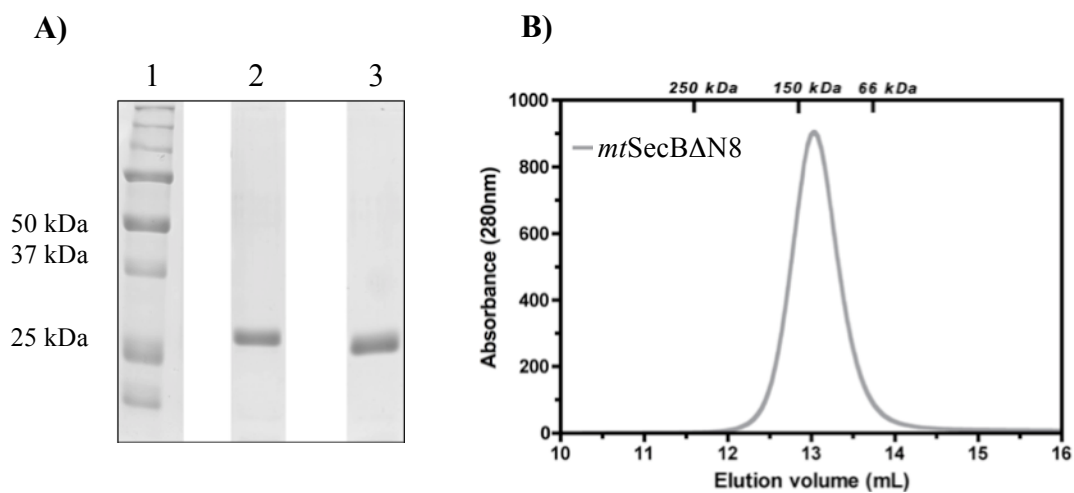
In an attempt to improve the crystallization of *mtSecB*, several N-terminal truncated constructs were made based on the results from the partial proteolysis. The initial construct, His<sub>6</sub>-MBP-*mtSecB*ΔN15, had a deletion of the first 14 residues of the wild-type *mtSecB* N-terminal sequence. Although the construct was robustly expressed, removal of the affinity tag would not proceed to completion; furthermore, the protein was precipitating as a “glassy pellet” during purification (Figure 4-8). Finally, size exclusion chromatography revealed that His<sub>6</sub>-MBP-*mtSecB*ΔN15 eluted as a single peak, but after treatment with TEV protease *mtSecB*ΔN15 appeared in the void volume, indicating that it had aggregated. In summary, removal of N-terminal 14 residues led to destabilization and aggregation of *mtSecB*. On this basis, two other constructs (His<sub>6</sub>-MBP-*mtSecB*-ΔN8 and -ΔN5) with less severe 8- and 5-residue truncations were made. Both *mtSecB* constructs were expressed in *E. coli*. Since the rationale for designing these constructs was to create a *mtSecB* protein with a shorter N-terminus, it was decided to first analyze *mtSecB*ΔN8. In this case, the affinity tag could be completely removed, and *mtSecB*ΔN8 continued to be soluble in subsequent purification steps. Figure 4-9A shows the purified *mtSecB*ΔN8 protein and compares the purity of this protein with that of the purified wild-type *mtSecB*. In order to ensure that the removal of the N-terminal residues of the *mtSecB* did not have an adverse affect on the solution behavior and tetramer formation of *mtSecB*ΔN8 protein, a sample of the purified protein was run at 16 mg/mL on an analytical size exclusion chromatography column (Figure 4-9B). It can be seen that the protein elutes as a single asymmetric peak at around 120 kDa, corresponding to the molecular weight of the tetrameric *mtSecB* protein. These satisfactory results with *mtSecB*ΔN8 meant that the *mtSecB*ΔN5 construct was abandoned and crystallization of *mtSecB*ΔN8 was pursued.



**Figure 4-8 Removal of the N-terminal 15 residues leads to destabilization of *mtSecB***

*mtSecB*ΔN15 was expressed as a fusion protein carrying an N-terminal His<sub>6</sub>-MBP tag (lane 2). The fusion protein was purified by Ni<sup>2+</sup>-NTA chromatography (lane 4) and digested with TEV protease at 4 °C for 48 hrs (lane 5) or 72 hrs (lane 6) to remove the His<sub>6</sub>-MBP. After 72 hrs digestion (lane 8) *mtSecB* was then centrifuged at 35,000 rpm for 45 min. A glassy pellet could be seen at the bottom of the centrifuge tube. The pellet was rinsed briefly with the digestion reaction buffer, and was diluted 20X (lane 9) and 10X (lane 10) with the SDS-PAGE loading buffer. The precipitated *mtSecB*ΔN15 can be seen in lanes 9 and 10. Lanes 1, 3, and 7 contain molecular weight markers.

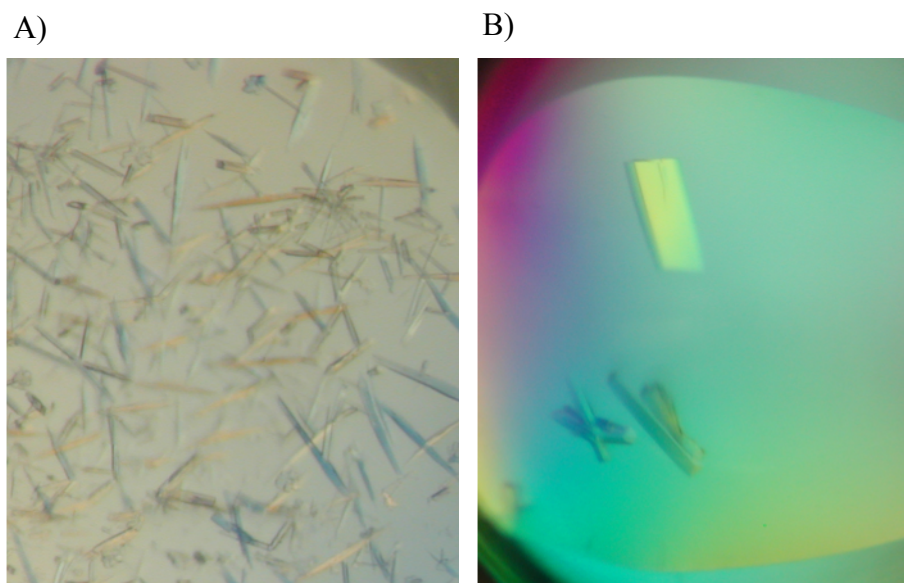




**Figure 4-9 Analysis of purified *mtSecBΔN8* protein**

*mtSecBΔN8* protein was purified for crystallization trials. **(A)** SDS-PAGE was used to compare purified wild-type *mtSecB* (lane 2) with *mtSecBΔN8* (lane 3). **(B)** Analytical size exclusion chromatography of truncated *mtSecBΔN8*. The position of elution for several molecular weight standards is indicated on the top axis of the plots.

The truncated *mtSecBAN8* readily formed crystals under the conditions found previously for the full-length protein (i.e conditions 1 and 2), and there were also crystals under three additional conditions: 0.2 M sodium iodide, 0.1 M Bis-Tris propane pH 6.5, 20% PEG 3,350; 0.2 M potassium thiocyanate, 0.1M Bis-Tris propane pH 6.5, 20% PEG 3,350; 0.2 M sodium nitrate, 0.1 M Bis-Tris propane pH 6.5, 20% PEG 3,350. At room temperature the *mtSecBAN8* crystals tended to have a 1-dimensional needle-like shape, while at lower temperatures (10 and 16 °C) *mtSecBAN8* formed larger 3-dimensional crystals (Figure 4-10). After optimization of crystallization conditions and components, the best crystals were obtained in 300 mM potassium thiocyanate, 50 mM imidazole pH 6.5, and 15% PEG 3,350 (Figure 4-10B). These crystals appear after a week and it takes two additional weeks to reach maximum size. Some of the smaller initial crystals were tested at the Canadian Light Source CMCF beamline 08ID-1 did not diffract. Larger crystals (with a maximum size of approximately 0.4 mm × 0.2 mm × 0.1 mm) have been obtained more recently but not yet tested for diffraction.



**Figure 4-10** *mtSecBΔN8* crystals

*mtSecBΔN8* crystals were grown at room temperature (**A**) and at 10 °C (**B**) in 50 mM imidazole, 300 mM K thiocyanate, 15 % PEG 3350 at pH 6.5. The crystals in Panel B appeared after one week and continued growing for another two weeks; the dimensions of the largest crystal obtained was 0.4 mm × 0.2 mm × 0.1 mm.

## 4.4 Discussion

A longstanding goal in the Shilton laboratory has been to obtain crystals of SecA in complex with a peptide representing a translocation substrate. A more recent goal was to extend structure-function studies of the Sec system to *M. tuberculosis*. For both *E. coli* SecA and the *M. tuberculosis* SecB-like protein (*mtSecB*), protein disorder appeared to have a negative impact on the formation of high quality crystals. It has long been recognized that protein disorder can adversely affect crystallization and crystal quality (Linding *et al*, 2003). On this basis, crystallization of *E. coli* SecA and *mtSecB* was improved by removal of the short unstructured terminal regions of these proteins and, in the case of *E. coli* SecA, introduction of entropy-reducing mutations.

With regard to the *E. coli* SecA, previous unpublished work in the Shilton laboratory showed that the C-terminal Zinc-Binding Domain (ZBD) and unstructured linker prevented crystallization. A construct lacking the linker and ZBD, SecA-N95, formed crystals with a very large unit cell that did not diffract beyond 8 Å resolution. In an effort to improve these crystals, the N-terminal unstructured region of SecA was removed and several entropy-reducing mutations were introduced. The resulting SecA-N95ΔN(ER) construct crystallized in a much smaller unit cell, and diffracted to beyond 6 Å resolution. Therefore, the changes to the construct did improve crystal quality, but not to the extent required for high-resolution structural analysis.

It is noteworthy that deletion of the extreme N-terminal region of SecA not only improved crystallization of SecA-N95, but also facilitated and improved the purification of SecA constructs (Appendix A). These effects are likely due to the ability of the extreme N-terminus to mediate oligomerization of SecA and also facilitate binding of preproteins and corresponding peptides, as described in Chapters 2 and 3. Oligomerization of SecA and promiscuous interactions with unfolded polypeptides would also interfere with crystallization. On this basis, removal of the extreme N-terminus of SecA is likely facilitating the formation of higher quality crystals through a number of different mechanisms. It is clear, however, that additional strategies will need to be employed to obtain crystals that diffract to high resolution and also to obtain crystals of SecA in complex with peptides corresponding to translocation substrates.

The wild-type form of *M. tuberculosis* Rv1957, called “*mtSecB*” because of its sequence and functional resemblance to *E. coli* SecB, was also hindered from crystallization due to the presence of an unstructured N-terminal region. Wild-type *mtSecB* was highly soluble and formed a stable tetramer in solution, but formed crystals only after a prolonged incubation. Removal of the proteolytically sensitive N-terminal region in *mtSecB* $\Delta$ 8 greatly accelerated crystallization of the construct, suggesting that this region was likely being slowly removed during the long incubations of *mtSecB*. Unfortunately, the smaller crystals of *mtSecB* $\Delta$ 8 did not diffract. Larger crystals (0.4 mm  $\times$  0.2 mm  $\times$  0.1 mm) have been obtained more recently but have not yet been tested. Overall, it is clear that removal of the unstructured N-terminus of *mtSecB* represented a valuable step forward, but that obtaining crystals of *mtSecB* that are suitable for high resolution diffraction analysis will require additional work.

Traditionally, protein structure determination has been carried out on a case-by-case basis, usually after years or even decades of investigating individual proteins. Therefore, as a great deal of information regarding the chemical and physical properties of each of these proteins was available, the structure determination of proteins that were “misbehaved” in solution would not be carried out. As a result, the lack of structural knowledge about these ill behaved proteins has been an obstacle in the complete understanding of their underlining biological functions. In addressing this issue, however, it has been shown that the presence of unstructured regions at the N- and C- termini (or even within domains) of some of these proteins is one of the reasons for their misbehavior in solution (Linding *et al*, 2003), which can either affect their solubility and stability or inhibit their crystallization (Reeves & Nissen, 1998; Bandaru *et al*, 2004; Cary *et al*, 1980). Likewise, it has been shown that the removal of these regions can facilitate the crystallization process (Linding *et al*, 2003). The results with crystallization of *E. coli* SecA and *mtSecB* are additional examples in this regard. By removal of the unstructured termini of these molecules, larger crystals were obtained with higher frequency, which in turn will increase the chances of obtaining high-resolution structures of these proteins. Although the high-resolution structures of these proteins were not resolved here, we are hopeful that the advances made in the crystallization of these molecules, which were presented in this chapter, would pave the way for this task.

## Chapter 5

### 5 Summery and conclusions

The central questions in this thesis revolved around two aspects of the bacterial Sec system; namely preprotein recognition and targeting by the Sec system as well as molecular mechanisms by which unfolded preproteins would be moved through the membrane pore. SecA is responsible for moving the unfolded preproteins through the membrane pore and is critical for bacterial post-translational secretion through the Sec system. However, despite the wealth of knowledge available regarding the post-translational translocation system, the fundamental question of how SecA couples ATP binding and hydrolysis to unfolded polypeptide recognition and movement remains largely unknown. In order to fully address this question, details of the inter- and intra-molecular interactions of SecA need to be understood. In particular, the exact preprotein binding site(s) on SecA have not been defined, and also the functional oligomeric state of SecA has been a matter of controversy.

This thesis addressed aspects of preprotein transport by SecA through advancing the molecular understanding of interactions of SecA with unfolded polypeptide substrates as well as investigating the nature of the oligomerization of SecA. In the current chapter an overview of the major findings and their implications are provided. Then, the future directions will be highlighted.

#### 5.1 SecA oligomerization is mediated by unstructured termini

The functional oligomeric state of SecA has been a key question for synthesizing a coherent model of translocation. In Chapter 2, by re-examining the unusual tetramerization of a truncated SecA construct (SecA-N68), we showed that the two unstructured polypeptides at its termini, the naturally occurring N-terminal polypeptide and a byproduct of truncation at C-terminus, are responsible for its self-association. Subsequently, by examining the role of the first 14 N-terminal residues of functional SecA-N95 construct it was shown that the removal of this region drastically weakened dimerization; the dissociation constant was increased from the sub-micromolar range in the case of SecA-N95 to approximately 25  $\mu$ M in SecA-N95 $\Delta$ N. Although

weakened dimerization did not considerably affect the solution ATPase activity of SecA-N95 *in vitro*, it was shown that SecA-N95 $\Delta$ N is not functional *in vivo*.

The functional role of SecA dimer formation both in cytosol and at the membrane, as a part of Sec translocon, is not clear yet. In fact, controversial observations regarding the importance of dimerization of SecA have been reported over the past few years. Beside this, due to the observed different dimeric interfaces in numerous experiments, it has been suggested that SecA dimerization may be a very dynamic process, especially during the translocation process, forming monomers or utilizing alternative dimeric interfaces (Or *et al*, 2005). One of the most controversial regions involved in the dimerization and consequently function of SecA is the extreme N-terminal region of SecA. Our experiments provided further evidence regarding the importance of this region of SecA for its self-association. Although the results presented here seem to support the role of dimerization in the translocation reaction, these findings do not exclude the possibility that the mere presence of N-terminus segment rather than dimerization of SecA to be responsible for the proper function of SecA during the translocation reaction. It is possible that the N-terminal segment, through a *cis* interaction with other domains of SecA or even interaction with other cellular components facilitates the translocation process. In fact, a recent study pointed towards the possible role the extreme N-terminal segment of SecA in interaction with the bacterial lipid membrane (Bauer *et al*, 2014). In addition, our results presented in Chapter 3, which is briefly summarized below, unexpectedly uncovered another possible functional role for this important region of SecA. It was shown that the N-terminal region of SecA is important for interaction with unfolded polypeptide. Despite this, understanding the possible role of the N-terminal segment of SecA and the dimerization of this molecule in coupling of preprotein translocation to the ATP hydrolysis will require additional studies.

## 5.2 Interaction of SecA with unfolded preproteins

There is little known about the mode of interaction of SecA with unfolded polypeptides, presumably due to difficulties associated with characterizing these interactions, mainly because the unfolded preprotein substrates have a tendency to either refold or aggregate *in vitro*. Here, we

have been able to investigate the direct interaction of SecA with the mature regions of preproteins using a new strategy that involved the immobilization of large peptides from maltose binding protein (MBP) to form a heterogeneous population of potential binding sites. Binding to these immobilized peptides was monitored using surface plasmon resonance (SPR). These studies were extended to a more defined system using an immobilized 20-residue peptide from the mature region of the periplasmic protein, FhuD. Our overall results investigating these interactions, which are presented in Chapter 3, revealed that SecA constructs bound to the immobilized peptides in both cases, and the binding site was mapped to the NBD domains and/or stem region of PPXD of SecA.

Our findings are important, because it was shown that SecA is able to directly interact with the “mature region” of preproteins. Also, mapping of the preprotein binding site to the NBDs and/or the adjacent regions together with the observed effects of nucleotides on the interaction between SecA-N68 and the polypeptides provide further clues about the translocation of preproteins by SecA. These are consistent with the idea that the binding site is formed by NBD1 and/or NBD2, which are subject to conformational changes during the cycle of ATP binding and hydrolysis.

Furthermore, our data shed more light on the sequence specificity of the SecA-polypeptides interactions. By analyzing the high affinity SecA-binder peptides a common SecA binding “motif” was identified, which was a simple pattern commonly found in proteins. This finding can explain how SecA is able to recognize different preproteins that do not share any apparent sequence homology.

Another equally important question regarding the function of SecA concerns the exact nature of the interactions between SecA and the mature regions of preproteins, which needs further investigation.

### 5.3 Crystallization of SecA and *mtSecB*

In Chapter 4 of this thesis, in order to gain further insight towards the functional roles of Sec components, we aimed at obtaining the crystal structures of SecA-N95 and *mtSecB*, as a complete understanding of the translocation process requires high resolution structures of the



individual components and the molecular complexes that are formed during translocation. For instance, structures of *E. coli* SecA-N95, both alone and in complex with a peptide representing a translocation substrate, have been longstanding goals. The work presented in this thesis showed that the diffraction quality of SecA-N95 crystals can be improved by removal of the unstructured N-terminal segment. A similar strategy was used to improve the crystallization of Rv1957 or “*mtSecB*”, found in *M. tuberculosis*. Rv1957 was cloned, expressed, and shown to form a stable tetramer in solution, in common with *E. coli* SecB. However, crystallization was very slow and difficult to reproduce, problems that were attributed to the N-terminal region of *mtSecB*, which is susceptible to proteolysis. An *mtSecB* construct in which the N-terminus was truncated by eight residues yielded large crystals relatively rapidly.

Overall, in the Chapter 4, it was shown that presence of unstructured termini have a profound effect on behavior and crystallization of these two members of the Sec system. The following section provides a list of experiments that could provide further insight into preprotein targeting and molecular mechanics in bacterial protein translocation.

## 5.4 Future Directions

### 5.4.1 Structural models for preprotein binding site on SecA

The goal here would be to locate the preprotein binding site on SecA. A structure of SecA in complex with polypeptide may provide invaluable information regarding the mode of interaction of SecA with its preprotein substrates. Given the fact that SecA-DM construct showed the highest binding towards the unfolded polypeptides (Chapter 4) and this construct has been crystallized before with relatively high resolution (Nithianantham & Shilton, 2008), this construct can be used in co-crystallization trials with a high affinity SecA binder polypeptide. Furthermore, this complex structure would also be an invaluable blueprint for studying the physiological roles of the sequence specificities of SecA. In turn, the knowledge of interaction site between SecA and preprotein will likely provide reasons for the observed sequence specificity of SecA.

In addition, it would also be highly valuable to gain knowledge about the possible conformational changes of SecA induced by preprotein binding, as such conformational changes can likely play important roles in the translocation of preproteins by SecA (Duong, 2003). These can be studied by analytical ultracentrifugation, size exclusion chromatography, and SAXS experiments.

Nonetheless, further attempts on crystallization of SecA-N95 $\Delta$ N(ER) may lead to obtaining a high resolution structure, which is valuable in mapping the unfolded polypeptide binding site to SecA. In addition, since our data presented in Chapter 3 showed that the lack of N-terminal segment is detrimental for unfolded polypeptides binding to SecA, this construct can provide further clues regarding why the removal of N-terminal segment leads to the loss of binding to unfolded polypeptides.

#### 5.4.2 Physiological roles of sequence specificity of SecA

The findings presented in Chapter 3 suggested that SecA constructs can bind specifically to polypeptides derived from the mature region of preproteins, which share a simple SecA binding “motif”. By creating MBP constructs that lack this motif, preparation of a CNBr digestion of these proteins, and subsequent testing of their direct binding towards SecA constructs, it will be possible to test the importance of the identified pattern for binding to SecA. Furthermore, since *in vitro* translocation systems are already available (Brundage *et al*, 1990; Akimaru *et al*, 1991), it would be interesting to compare the ability of SecA in translocating the wild-type MBP and these mutated MBP constructs, which lack high affinity SecA binding sites. This would provide a direct indication that SecA indeed uses these binding sites for interaction with and translocation of unfolded preproteins. This can logically be extended to analyzing the ability of SecA to catalyze the translocation of these mutated MBP constructs *in vivo*.

Finally, another equally interesting experiment would involve testing the ability of the isolated high affinity SecA binder peptides to inhibit the normal translocation of preproteins by SecA *in vitro*. This is especially important, because the ability of these peptides to inhibit the function of

SecA *in vitro* would be an indication of potential value of these peptides for development of antibiotics that target the bacterial translocation system.

### 5.4.3 Identification of binding site(s) for the N-terminal segment of SecA

The results presented in Chapter 2 showed the importance of the N-terminal segment of SecA for its function *in vivo*. Data presented in Chapter 3 also pointed towards the importance of this region for the interaction of SecA with its unfolded preprotein substrates. The logical future study of the role of the extreme N-terminal segment of SecA would be to identify its precise binding site(s) on SecA. Protein crystallization is the first option. For this reason, a polypeptide corresponding to the N-terminal segment of SecA has been synthesized. Characterizing the binding specificity of this polypeptide to various SecA constructs is the next step. The shortest construct that retained binding to this synthetic polypeptide will be used for co-crystallization trials. Using the *in vitro* chemical cross-linking technique is an alternative method for identification of regions involved in the interaction with the unfolded N-terminal segment of SecA.

After identification of the “binding site(s)” for the unstructured N-terminal segment, by altering these sites, we expect that the monomer biased SecA molecules to be created. These monomeric constructs, which still carry the N-terminal segment of SecA, can be used to determine their *in vitro* binding towards the unfolded preprotein as well as their *in vivo* function for restoring the growth of *E. coli*. This would reveal whether the mere presence of the N-terminal segment or the dimerization of SecA are important for the function of this molecule in the translocation reaction.

### 5.4.4 Substrate specificities of mycobacterial SecA1 and SecA2

To further probe the interaction of SecA with its unfolded preprotein substrates, it would be interesting to explore and compare the substrate specificities of mycobacterial SecAs. SecA1 is essential and resembles the single, vital SecA protein present in the *E. coli*, whereas SecA2 is non-essential for survival of *M. tuberculosis* and has a slightly altered domain structure, and is crucial for the secretion of only a few specific preproteins (Swanson *et al*, 2015; Rigel &

Braunstein, 2008a; Braunstein *et al*, 2003b). So far, we have purified both proteins. Then, we can try to analyze their binding to the SecA1 and SecA2 substrates on a peptide array. The array has been designed, and will be synthesized shortly. To detect the binding of SecA molecules to the array, we will use His-tagged SecA proteins, and therefore, peptide binding will be traced by immunostaining. Comparing the binding pattern of these proteins may provide us with answers regarding the binding specificities of these proteins. For example, will SecA1 also acts on the SecA2 specific substrates? Are there any consensus sequences for binding of SecA2 to its substrates? Eventually, these SecA-binder peptides can be used for functional and structural studies, essentially in a similar manner as was in Chapter 3.

#### 5.4.5 The role of *mtSecB* in the Sec system

The first logical next step in study of *mtSecB* would include the high-resolution structure determination of this molecule. To this end, recently larger well-formed crystals of this molecule were obtained (Figure 4-10) that have not been tested for diffraction. The diffraction of these crystals can be tested in combination with different cryoprotectant solutions. If the crystals did not diffract well, the quality of *mtSecB* crystals could be further improved by surface entropy reduction which has been shown to be able to effectively improve the quality of several protein crystals in our lab.

There is no direct indication of interaction of *mtSecB* with mycobacterial SecAs. Examining the direct interaction of *mtSecB* with SecA1 and SecA2 would assist in enlightening the function of this protein in the Sec system. It is known that *E. coli* can interact with both free and membrane bound SecA (Hartl *et al*, 1990; Woodbury *et al*, 2000). Interestingly, residues of *E. coli* SecA that are in direct contact with SecB are conserved in mycobacterial SecA1 (Xu *et al*, 2000; Bordes *et al*, 2011b). Therefore, through gel filtration chromatography, analytical ultracentrifugation, or native PAGE, the interaction between these molecules could be evaluated. Furthermore, it has been shown that SecB increases the ATPase activity of *E. coli* SecA in solution (Kim *et al*, 2001; Miller *et al*, 2002). Therefore, it would be interesting to test the effect of presence of *mtSecB* on mycobacterial SecAs' ATPase activity.

Further steps will involve studying the sequence specificity of interaction of *mtSecB* chaperone with other mycobacterial proteins (including preproteins, toxin, and antitoxin proteins), using an oriented peptide library experiment. For this reason, a fluorescently labeled *mtSecB* protein can be used to screen its binding towards mycobacterial substrates.

Identification of other substrates of *mtSecB* protein can be another interesting future work. Mass spectroscopy analysis of the secreted proteins of *mtSecB* mutant strains could lead to identification of proteins that are dependent on *mtSecB* for their secretion.

## 5.5 Conclusion

The work presented in this thesis highlighted the molecular aspects of SecA oligomerization and advanced our knowledge of how SecA interacts with the unfolded substrates. Both of these will help understanding the complex roles of SecA in the preprotein translocation. As the models for the translocation reaction are being developed, our findings regarding the critical role of N-terminal segment of SecA as well as the results concerning the interaction of SecA with preproteins should be taken into consideration.

## 6 References

- Akimaru J, Matsuyama S, Tokuda H & Mizushima S (1991) Reconstitution of a protein translocation system containing purified SecY, SecE, and SecA from *Escherichia coli*. *Proc Natl Acad Sci U S A* 88: 6545–6549
- Akita M, Shinkai A, Matsuyama S & Mizushima S (1991) SecA, an essential component of the secretory machinery of *Escherichia coli*, exists as homodimer. *Biochemical and biophysical research communications* 174: 211–216
- Andreatta M, Lund O & Nielsen M (2013) Simultaneous alignment and clustering of peptide data using a Gibbs sampling approach. *Bioinformatics* 29: 8–14
- Auclair SM, Oliver DB & Mukerji I (2013) Defining the solution state dimer structure of *Escherichia coli* SecA using Förster resonance energy transfer. *Biochemistry* 52: 2388–2401
- Balbo A & Schuck P (2005) Analytical ultracentrifugation in the study of protein self-association and heterogeneous protein-protein interactions. *Protein-Protein Interactions: Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.*: pp. 253–277
- Bandaru V, Cooper W, Wallace SS & Doublé S (2004) Overproduction, crystallization and preliminary crystallographic analysis of a novel human DNA-repair enzyme that recognizes oxidative DNA damage. *Acta Crystallogr. D Biol. Crystallogr.* 60: 1142–1144
- Baud C, Karamanou S, Sianidis G, Vrontou E, Politou AS & Economou A (2002) Allosteric communication between signal peptides and the SecA protein DEAD motor ATPase domain. *J. Biol. Chem.*
- Bauer BW, Shemesh T, Chen Y & Rapoport TA (2014) A ‘push and slide’ mechanism allows sequence-insensitive translocation of secretory proteins by the SecA ATPase. *Cell*
- Bechtluft P, Nouwen N & Tans SJ (2010) SecB---A chaperone dedicated to protein translocation - Molecular BioSystems (RSC Publishing) DOI:10.1039/B915435C. *Molecular BioSystems*
- Bhanu MK, Zhao P & Kendall DA (2013) Mapping of the SecA signal peptide binding site and dimeric interface by using the substituted cysteine accessibility method. *Journal of bacteriology* 195: 4709–4715
- Bordes P, Cirinesi A-M, Ummels R, Sala A, Sakr S, Bitter W & Genevaux P (2011a) SecB-like chaperone controls a toxin-antitoxin stress-responsive system in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America*
- Bordes P, Cirinesi AM & Ummels R (2011b) SecB-like chaperone controls a toxin–antitoxin stress-responsive system in *Mycobacterium tuberculosis*. In
- Braunstein M, Espinosa BJ, Chan J, Belisle JT & Jacobs WR (2003a) SecA2 functions in the secretion of superoxide dismutase A and in the virulence of *Mycobacterium tuberculosis*. *Molecular Microbiology*

- Braunstein M, Espinosa BJ, Chan J, Belisle JT & Jacobs WR (2003b) SecA2 functions in the secretion of superoxide dismutase A and in the virulence of *Mycobacterium tuberculosis*. *Molecular Microbiology* 48: 453–464
- Breukink E, Nouwen N, van Raalte A, Mizushima S, Tommassen J & de Kruijff B (1995) The C terminus of SecA is involved in both lipid binding and SecB binding. *J. Biol. Chem.* 270: 7902–7907
- Breyton C, Haase W, Rapoport TA, Kühlbrandt W & Collinson I (2002) Three-dimensional structure of the bacterial protein-translocation complex SecYEG. *Nature* 418: 662–665
- Briere L-AK & Dunn SD (2006) The periplasmic domains of *Escherichia coli* HflKC oligomerize through right-handed coiled-coil interactions. *Biochemistry* 45: 8607–8616
- Brown PH, Balbo A & Schuck P (2008) Characterizing protein-protein interactions by sedimentation velocity analytical ultracentrifugation. *Curr Protoc Immunol* UNIT 18.15:
- Brundage L, Hendrick JP, Schiebel E, Driessen AJ & Wickner W (1990) The purified *E. coli* integral membrane protein SecY/E is sufficient for reconstitution of SecA-dependent precursor protein translocation. *Cell* 62: 649–657
- Cabelli RJ, Chen L, Tai PC & Oliver DB (1988) SecA protein is required for secretory protein translocation into *E. coli* membrane vesicles. *Cell* 55: 683–692
- Cannon KS, Or E, Clemons WM, Shibata Y & Rapoport TA (2005) Disulfide bridge formation between SecY and a translocating polypeptide localizes the translocation pore to the center of SecY. *J. Cell Biol.* 169: 219–225
- Cary PD, King DS, Crane-Robinson C, Bradbury EM, Rabbani A, Goodwin GH & Johns EW (1980) Structural studies on two high-mobility-group proteins from calf thymus, HMG-14 and HMG-20 (ubiquitin), and their interaction with DNA. *Eur. J. Biochem.* 112: 577–580
- Choi JH & Lee SY (2004) Secretory and extracellular production of recombinant proteins using *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 64: 625–635
- Christie PJ & Cascales E (2005) Structural and dynamic properties of bacterial type IV secretion systems (review). *Mol. Membr. Biol.* 22: 51–61
- Christie PJ & Vogel JP (2000) Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol.* 8: 354–360
- Clérico EM, Maki JL & Gierasch LM (2008) Use of synthetic signal sequences to explore the protein export machinery. *Biopolymers* 90: 307–319
- Collier DN, Bankaitis VA, Weiss JB & Bassford PJ (1988) The antifolding activity of SecB promotes the export of the *E. coli* maltose-binding protein. *Cell*
- Cornelis GR (2006) The type III secretion injectisome. *Nature reviews. Microbiology*
- Dalal K, Chan CS, Sligar SG & Duong F (2012) Two copies of the SecY channel and acidic lipids are necessary to activate the SecA translocation ATPase. *Proc Natl Acad Sci U S A* 109: 4104–4109
- Danese PN & Silhavy TJ (1998) Targeting and assembly of periplasmic and outer-membrane proteins in *Escherichia coli*. *Annu. Rev. Genet.* 32: 59–94

- Das S, Stivison E, Folta-Stogniew E & Oliver D (2008) Reexamination of the role of the amino terminus of SecA in promoting its dimerization and functional state. *Journal of bacteriology* 190: 7302–7307
- De Gier JW, Valent QA, Heijne Von G & Luirink J (1997) The E. coli SRP: preferences of a targeting factor. *FEBS Letters* 408: 1–4
- de Keyzer J, van der Sluis EO, Spelbrink REJ, Nijstad N, Ben de Kruijff, Nouwen N, van der Does C & Driessen AJM (2005) Covalently dimerized SecA is functional in protein translocation. *J. Biol. Chem.* 280: 35255–35260
- Dekker C, de Kruijff B & Gros P (2003) Crystal structure of SecB from Escherichia coli. *J. Struct. Biol.* 144: 313–319
- Delepelaire P (2004) Type I secretion in gram-negative bacteria. *Biochimica et Biophysica Acta (BBA)-Molecular Cell ...*
- Dempsey BR (2007) SecA Protein Translocation: Targeting and Mechanics. Ph.D. thesis. The University of Western Ontario
- Dempsey BR, Economou A, Dunn SD & Shilton BH (2002) The ATPase domain of SecA can form a tetramer in solution. *Journal of Molecular Biology* 315: 831–843
- Dempsey BR, Wrona M, Moulin JM, Gloor GB, Jalilehvand F, Lajoie G, Shaw GS & Shilton BH (2004) Solution NMR structure and X-ray absorption analysis of the C-terminal zinc-binding domain of the SecA ATPase. *Biochemistry* 43: 9361–9371
- Derman AI, Puziss JW, Bassford PJ & Beckwith J (1993) A signal sequence is not required for protein export in prlA mutants of Escherichia coli. *EMBO J* 12: 879–888
- Ding H, Hunt JF, Mukerji I & Oliver D (2003) Bacillus subtilis SecA ATPase exists as an antiparallel dimer in solution. *Biochemistry* 42: 8729–8738
- Driessen AJ (1993) SecA, the peripheral subunit of the Escherichia coli precursor protein translocase, is functional as a dimer. *Biochemistry* 32: 13190–13197
- Driessen AJ (2001) SecB, a molecular chaperone with two faces. *Trends Microbiol.* 9: 193–196
- Duong F (2003) Binding, activation and dissociation of the dimeric SecA ATPase at the dimeric SecYEG translocase. *EMBO J* 22: 4375–4384
- Duong F (2014) Capturing the bacterial holo-complex. *Proceedings of the National Academy of Sciences of the United States of America* 111: 4739–4740
- Duong F & Wickner W (1997) Distinct catalytic roles of the SecYE, SecG and SecDFyajC subunits of preprotein translocase holoenzyme. *EMBO J* 16: 2756–2768
- Dussurget O, Stewart G & Neyrolles O (2001) Role of Mycobacterium tuberculosis Copper-Zinc Superoxide Dismutase. *Infection and ...*
- Economou A (1998) Bacterial preprotein translocase: mechanism and conformational dynamics of a processive enzyme. *Molecular Microbiology* 27: 511–518
- Economou A & Wickner W (1994a) SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion. *Cell*



- Economou A & Wickner W (1994b) SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion. *Cell* 78: 835–843
- Economou A, Pogliano JA, Beckwith J, Oliver DB & Wickner W (1995) SecA membrane cycling at SecYEG is driven by distinct ATP binding and hydrolysis events and is regulated by SecD and SecF. *Cell* 83: 1171–1181
- Emr SD, Hanley-Way S & Silhavy TJ (1981) Suppressor mutations that restore export of a protein with a defective signal sequence. *Cell* 23: 79–88
- Erlanson KJ, Miller SBM, Nam Y, Osborne AR, Zimmer J & Rapoport TA (2008) A role for the two-helix finger of the SecA ATPase in protein translocation. *Nature* 455: 984–987
- Fekkes P, de Wit JG, van der Wolk JP, Kimsey HH, Kumamoto CA & Driessen AJ (1998) Preprotein transfer to the Escherichia coli translocase requires the co-operative binding of SecB and the signal sequence to SecA. *Molecular Microbiology* 29: 1179–1190
- Fekkes P, van der Does C & Driessen AJ (1997) The molecular chaperone SecB is released from the carboxy-terminus of SecA during initiation of precursor protein translocation. *EMBO J* 16: 6105–6113
- Feltcher ME & Braunstein M (2012) Emerging themes in SecA2-mediated protein export. *Nature reviews. Microbiology*
- Fröderberg L, Houben ENG, Baars L, Luirink J & de Gier J-W (2004) Targeting and translocation of two lipoproteins in Escherichia coli via the SRP/Sec/YidC pathway. *J. Biol. Chem.* 279: 31026–31032
- Gannon PM & Kumamoto CA (1993) Mutations of the molecular chaperone protein SecB which alter the interaction between SecB and maltose-binding protein. *Journal of Biological Chemistry*
- Gerdes K & Maisonneuve E (2012a) Bacterial Persistence and Toxin-Antitoxin Loci. *Annu. Rev. Microbiol.*
- Gerdes K & Maisonneuve E (2012b) Bacterial persistence and toxin-antitoxin loci. *Annu. Rev. Microbiol.*
- Gibbons HS, Wolschendorf F & Abshire M (2007) Identification of two Mycobacterium smegmatis lipoproteins exported by a SecA2-dependent pathway. *Journal of ...*
- Gierasch LM (1989) Signal sequences. *Biochemistry* 28: 923–930
- Gold VAM, Robson A, Clarke AR & Collinson I (2007) Allosteric regulation of SecA: magnesium-mediated control of conformation and activity. *J. Biol. Chem.* 282: 17424–17432
- Gould AD, Telmer PG & Shilton BH (2009) Stimulation of the maltose transporter ATPase by unliganded maltose binding protein. *Biochemistry* 48: 8051–8061
- Grady LM, Michtavy J & Oliver DB (2012) Characterization of the Escherichia coli SecA signal peptide-binding site. *Journal of bacteriology* 194: 307–316
- Grant SR, Fisher EJ, Chang JH, Mole BM & Dangl JL (2006) Subterfuge and manipulation: type III effector proteins of phytopathogenic bacteria. *Annu. Rev. Microbiol.* 60: 425–449

- Green RJ, Frazier RA, Shakesheff KM, Davies MC, Roberts CJ & Tendler SJ (2000) Surface plasmon resonance analysis of dynamic biological interactions with biomaterials. *Biomaterials* 21: 1823–1835
- Hardy SJ & Randall LL (1991) A kinetic partitioning model of selective binding of nonnative proteins by the bacterial chaperone SecB. *Science* 251: 439–443
- Harris CR & Silhavy TJ (1999) Mapping an interface of SecY (PrIA) and SecE (PrIG) by using synthetic phenotypes and in vivo cross-linking. *Journal of bacteriology* 181: 3438–3444
- Hartl FU, Lecker S, Schiebel E, Hendrick JP & Wickner W (1990) The binding cascade of SecB to SecA to SecY/E mediates preprotein targeting to the E. coli plasma membrane. *Cell*
- Hartmann E, Sommer T, Prehn S, Görlich D, Jentsch S & Rapoport TA (1994) Evolutionary conservation of components of the protein translocation complex. *Nature* 367: 654–657
- Hirano M, Matsuyama S & Tokuda H (1996) The carboxyl-terminal region is essential for SecA-dimerization. *Biochemical and biophysical research communications* 229: 90–95
- Holland IB (2004) Translocation of bacterial proteins--an overview. *Biochim. Biophys. Acta* 1694: 5–16
- Hor LI & Shuman HA (1993) Genetic analysis of periplasmic binding protein dependent transport in Escherichia coli. Each lobe of maltose-binding protein interacts with a different subunit of the MalFGK2 membrane transport complex. *Journal of Molecular Biology* 233: 659–670
- Huie JL & Silhavy TJ (1995) Suppression of signal sequence defects and azide resistance in Escherichia coli commonly result from the same mutations in secA. *Journal of bacteriology* 177: 3518–3526
- Hunt JF, Weinkauf S, Henry L, Fak JJ, McNicholas P, Oliver DB & Deisenhofer J (2002) Nucleotide control of interdomain interactions in the conformational reaction cycle of SecA. *Science (New York, N.Y.)* 297: 2018–2026
- Ito K, Wittekind M, Nomura M, Shiba K, Yura T, Miura A & Nashimoto H (1983) A temperature-sensitive mutant of E. coli exhibiting slow processing of exported proteins. *Cell* 32: 789–797
- Jilaveanu LB & Oliver D (2006) SecA dimer cross-linked at its subunit interface is functional for protein translocation. *Journal of bacteriology* 188: 335–338
- Jilaveanu LB, Zito CR & Oliver D (2005) Dimeric SecA is essential for protein translocation. *Proc Natl Acad Sci U S A* 102: 7511–7516
- Johnsson B, Löfås S & Lindquist G (1991) Immobilization of proteins to a carboxymethyl-dextran-modified gold surface for biospecific interaction analysis in surface plasmon resonance sensors. *Analytical biochemistry* 198: 268–277
- Jönsson U, Fägerstam L, Ivarsson B, Johnsson B, Karlsson R, Lundh K, Löfås S, Persson B, Roos H & Rönnberg I (1991) Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology. *BioTechniques* 11: 620–627

- Kajava AV, Zolov SN, Kalinin AE & Nesmeyanova MA (2000) The net charge of the first 18 residues of the mature sequence affects protein translocation across the cytoplasmic membrane of gram-negative bacteria. *Journal of bacteriology* 182: 2163–2169
- Karamanou S, Sianidis G, Gouridis G, Pozidis C, Papanikolau Y, Papanikou E & Economou A (2005) Escherichia coli SecA truncated at its termini is functional and dimeric. *FEBS Letters* 579: 1267–1271
- Karamanou S, Vrontou E, Sianidis G, Baud C, Roos T, Kuhn A, Politou AS & Economou A (1999) A molecular switch in SecA protein couples ATP hydrolysis to protein translocation. *Molecular Microbiology* 34: 1133–1145
- Kato Y, Nishiyama K-I & Tokuda H (2003) Depletion of SecDF-YajC causes a decrease in the level of SecE: implication for their functional interaction. *FEBS Letters* 550: 114–118
- Kawaguchi S, Müller J & Linde D (2001) The crystal structure of the ttCsaA protein: an export-related chaperone from Thermus thermophilus. *The EMBO ...*
- Keller R, de Keyser J, Driessen AJM & Palmer T (2012) Co-operation between different targeting pathways during integration of a membrane protein. *J. Cell Biol.* 199: 303–315
- Kim J, Miller A, Wang L, Müller JP & Kendall DA (2001) Evidence That SecB Enhances the Activity of SecA †. *Biochemistry*
- Kim JL, Morgenstern KA, Griffith JP, Dwyer MD, Thomson JA, Murcko MA, Lin C & Caron PR (1998) Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure* 6: 89–100
- Kimsey HH, Dagarag MD & Kumamoto CA (1995) Diverse effects of mutation on the activity of the Escherichia coli export chaperone SecB. *J. Biol. Chem.* 270: 22831–22835
- Kimura E, Akita M, Matsuyama S & Mizushima S (1991) Determination of a region in SecA that interacts with presecretory proteins in Escherichia coli. *J. Biol. Chem.* 266: 6600–6606
- Klock HE & Lesley SA (2009) The Polymerase Incomplete Primer Extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. *Methods Mol. Biol.* 498: 91–103
- Kumamoto CA & Beckwith J (1983) Mutations in a new gene, secB, cause defective protein localization in Escherichia coli. *Journal of bacteriology* 154: 253–260
- Kusters I & Driessen AJM (2011) SecA, a remarkable nanomachine. *Cellular and Molecular Life Sciences*
- Kusters I, van den Bogaart G, Kedrov A, Krasnikov V, Fulyani F, Poolman B & Driessen AJM (2011) Quaternary structure of SecA in solution and bound to SecYEG probed at the single molecule level. *Structure* 19: 430–439
- Lee HC & Bernstein HD (2002) Trigger factor retards protein export in Escherichia coli. *J. Biol. Chem.* 277: 43527–43535
- Ligon LS, Hayden JD & Braunstein M (2012a) The ins and outs of Mycobacterium tuberculosis protein export. *Tuberculosis*

- Ligon LS, Hayden JD & Braunstein M (2012b) The ins and outs of Mycobacterium tuberculosis protein export. *Tuberculosis*
- Lill R, Cunningham K, Brundage LA, Ito K, Oliver D & Wickner W (1989) SecA protein hydrolyzes ATP and is an essential component of the protein translocation ATPase of Escherichia coli. *EMBO J* 8: 961–966
- Linder P (2006) Dead-box proteins: a family affair--active and passive players in RNP-remodeling. *Nucleic Acids Res* 34: 4168–4180
- Linding R, Jensen LJ, Diella F, Bork P & Gibson TJ (2003) Protein disorder prediction: implications for structural proteomics. *Structure*
- Luirink J & Sinning I (2004) SRP-mediated protein targeting: structure and function revisited. *Biochimica et Biophysica Acta (BBA)-Molecular Cell ...*
- Luirink J, Heijne G, Houben E & Gier JW (2005) Biogenesis of inner membrane proteins in Escherichia coli. *Annu. Rev. Microbiol.*
- Mao C, Cheadle CE & Hardy S (2013) Stoichiometry of SecYEG in the active translocase of Escherichia coli varies with precursor species. In
- Miller A, Wang L & Kendall DA (2002) SecB Modulates the Nucleotide-Bound State of SecA and Stimulates ATPase Activity †. *Biochemistry*
- Mitchell C & Oliver D (1993) Two distinct ATP-binding domains are needed to promote protein export by Escherichia coli SecA ATPase. *Molecular Microbiology* 10: 483–497
- Moreno F, Fowler AV, Hall M, Silhavy TJ, Zabin I & Schwartz M (1980) A signal sequence is not sufficient to lead beta-galactosidase out of the cytoplasm. *Nature* 286: 356–359
- Mori H & Ito K (2001) An essential amino acid residue in the protein translocation channel revealed by targeted random mutagenesis of SecY. *Proc Natl Acad Sci U S A* 98: 5128–5133
- Musial-Siwiek M, Rusch SL & Kendall DA (2007) Selective photoaffinity labeling identifies the signal peptide binding domain on SecA. *Journal of Molecular Biology* 365: 637–648
- Müller JP, Ozegowski J, Vettermann S, Swaving J, Van Wely KH & Driessen AJ (2000) Interaction of Bacillus subtilis CsaA with SecA and precursor proteins. *Biochem. J.* 348 Pt 2: 367–373
- Nithianantham S & Shilton BH (2008) Analysis of the Isolated SecA DEAD Motor Suggests a Mechanism for Chemical–Mechanical Coupling. *Journal of Molecular Biology* 383: 380–389
- Oliver DB & Beckwith J (1981) E. coli mutant pleiotropically defective in the export of secreted proteins. *Cell* 25: 765–772
- Oliver DB & Beckwith J (1982) Identification of a new gene (secA) and gene product involved in the secretion of envelope proteins in Escherichia coli. *Journal of bacteriology* 150: 686–691
- Or E & Rapoport T (2007) Cross-linked SecA dimers are not functional in protein translocation. *FEBS Letters*

- Or E, Boyd D, Gon S, Beckwith J & Rapoport T (2005) The bacterial ATPase SecA functions as a monomer in protein translocation. *J. Biol. Chem.* 280: 9097–9105 Available at: <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=15618215&retmode=ref&cmd=prlinks>
- Or E, Navon A & Rapoport T (2002) Dissociation of the dimeric SecA ATPase during protein translocation across the bacterial membrane. *EMBO J* 21: 4470–4479
- Osborne AR, Clemons WM & Rapoport TA (2004) A large conformational change of the translocation ATPase SecA. *Proc Natl Acad Sci U S A* 101: 10937–10942
- Palmer T & Berks BC (2012) The twin-arginine translocation (Tat) protein export pathway. *Nature reviews. Microbiology* 10: 483–496
- Papanikolau Y, Papadovasilaki M, Ravelli RBG, McCarthy AA, Cusack S, Economou A & Petratos K (2007a) Structure of Dimeric SecA, the Escherichia coli Preprotein Translocase Motor. *Journal of Molecular Biology*
- Papanikolau Y, Papadovasilaki M, Ravelli RBG, McCarthy AA, Cusack S, Economou A & Petratos K (2007b) Structure of Dimeric SecA, the Escherichia coli Preprotein Translocase Motor. *Journal of Molecular Biology* 366: 1545–1557
- Papanikou E, Karamanou S & Economou A (2007) Bacterial protein secretion through the translocase nanomachine. *Nature Reviews ...*
- Papanikou E, Karamanou S, Baud C, Frank M, Sianidis G, Keramisanou D, Kalodimos CG, Kuhn A & Economou A (2005) Identification of the preprotein binding domain of SecA. *J. Biol. Chem.* 280: 43209–43217
- Park E & Rapoport TA (2012) Mechanisms of Sec61/SecY-Mediated Protein Translocation Across Membranes. *Annual Review of Biophysics*
- Plath K, Mothes W, Wilkinson BM, Stirling CJ & Rapoport TA (1998) Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. *Cell* 94: 795–807
- Pogliano JA & Beckwith J (1994) SecD and SecF facilitate protein export in Escherichia coli. *EMBO J* 13: 554–561
- Pohlschröder M, Hartmann E, Hand NJ, Dilks K & Haddad A (2005) DIVERSITY AND EVOLUTION OF PROTEIN TRANSLOCATION. *Annu. Rev. Microbiol.* 59: 91–111
- Prabudiansyah I, Kusters I & Driessen AJM (2015) In Vitro Interaction of the Housekeeping SecA1 with the Accessory SecA2 Protein of Mycobacterium tuberculosis. *PLoS One* 10: e0128788–e0128788
- Price A, Economou A, Duong F & Wickner W (1996) Separable ATPase and membrane insertion domains of the SecA subunit of preprotein translocase. *J. Biol. Chem.* 271: 31580–31584
- Prinz A, Behrens C, Rapoport TA, Hartmann E & Kalies KU (2000) Evolutionarily conserved binding of ribosomes to the translocation channel via the large ribosomal RNA. *EMBO J* 19: 1900–1906

- Rajapandi T, Dolan KM & Oliver DB (1991) The first gene in the Escherichia coli secA operon, gene X, encodes a nonessential secretory protein. *Journal of bacteriology* 173: 7092–7097
- Randall LL & Hardy SJ (1986) Correlation of competence for export with lack of tertiary structure of the mature species: a study in vivo of maltose-binding protein in E. coli. *Cell* 46: 921–928
- Randall LL & Hardy SJS (2002) SecB, one small chaperone in the complex milieu of the cell. *Cellular and Molecular Life Sciences* 59: 1617–1623
- Randall LL & Henzl MT (2010) Direct identification of the site of binding on the chaperone SecB for the amino terminus of the translocon motor SecA. *Protein Sci.* 19: 1173–1179
- Randall LL, Crane JM, Lilly AA, Liu GP, Mao CF, Patel CN & Hardy S (2005) Asymmetric binding between SecA and SecB two symmetric proteins: Implications for function in export. *Journal of Molecular Biology* 348: 479–489
- Reeves R & Nissen MS (1998) Purification and assays for high mobility group HMG-I (Y) protein function. *Methods in enzymology*
- Rieger CE, Lee J & Turnbull JL (1997) A continuous spectrophotometric assay for aspartate transcarbamylase and ATPases. *Analytical biochemistry* 246: 86–95
- Rigel NW & Braunstein M (2008a) A new twist on an old pathway – accessory Sec systems. *Molecular Microbiology* 70: 271–271
- Rigel NW & Braunstein M (2008b) A new twist on an old pathway--accessory Sec [corrected] systems. *Molecular Microbiology* 69: 291–302
- Rigel NW, Gibbons HS, McCann JR, McDonough JA, Kurtz S & Braunstein M (2009a) The Accessory SecA2 System of Mycobacteria Requires ATP Binding and the Canonical SecA1. *J. Biol. Chem.*
- Rigel NW, Gibbons HS, McCann JR, McDonough JA, Kurtz S & Braunstein M (2009b) The Accessory SecA2 System of Mycobacteria Requires ATP Binding and the Canonical SecA1. *J. Biol. Chem.* 284: 9927–9936
- Riggs PD, Derman AI & Beckwith J (1988) A mutation affecting the regulation of a secA-lacZ fusion defines a new sec gene. *Genetics* 118: 571–579
- Sala A, Calderon V, Bordes P & Genevaux P (2013) TAC from Mycobacterium tuberculosis: a paradigm for stress-responsive toxin-antitoxin systems controlled by SecB-like chaperones. *Cell Stress Chaperones* 18: 129–135
- Sambrook J & Russell DW (2001) *Molecular Cloning* CSHL Press
- Sardis MF & Economou A (2010) SecA: a tale of two protomers. *Molecular Microbiology*
- Schiebel E, Driessen A, Hartl FU & Wickner W (1991)  $\Delta\mu$  H<sup>+</sup> and ATP function at different steps of the catalytic cycle of preprotein translocase. *Cell*
- Schulze RJ, Komar J, Botte M, Allen WJ, Whitehouse S, Gold VAM, Lycklama A Nijeholt JA, Huard K, Berger I, Schaffitzel C & Collinson I (2014) Membrane protein insertion and proton-motive-force-dependent secretion through the bacterial holo-translocon SecYEG-

- SecDF-YajC-YidC. *Proceedings of the National Academy of Sciences of the United States of America* 111: 4844–4849
- Scott JR & Barnett TC (2006) Surface proteins of gram-positive bacteria and how they get there. *Annu. Rev. Microbiol.* 60: 397–423
- Scotti PA, Urbanus ML, Brunner J, De Gier JW, Heijne Von G, van der Does C, Driessen AJ, Oudega B & Luirink J (2000) YidC, the Escherichia coli homologue of mitochondrial Oxa1p, is a component of the Sec translocase. *EMBO J* 19: 542–549
- Sharma V, Arockiasamy A, Ronning DR, Savva CG, Holzenburg A, Braunstein M, Jacobs WR & Sacchettini JC (2003) Crystal structure of Mycobacterium tuberculosis SecA, a preprotein translocating ATPase. *Proceedings of the National Academy of Sciences of the United States of America*
- Shilton B, Svergun DI, Volkov VV, Koch MH, Cusack S & Economou A (1998) Escherichia coli SecA shape and dimensions. *FEBS Letters* 436: 277–282
- Shinkai A, Mei LH, Tokuda H & Mizushima S (1991) The conformation of SecA, as revealed by its protease sensitivity, is altered upon interaction with ATP, presecretory proteins, everted membrane vesicles, and phospholipids. *J. Biol. Chem.* 266: 5827–5833
- Singleton MR, Dillingham MS & Wigley DB (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* 76: 23–50
- Stader J, Gansheroff LJ & Silhavy TJ (1989) New suppressors of signal-sequence mutations, prIG, are linked tightly to the secE gene of Escherichia coli. *Genes Dev.* 3: 1045–1052
- Stothard P (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques* 28: 1102–1104
- Swanson S, Ioerger TR, Rigel NW, Miller BK, Braunstein M & Sacchettini JC (2015) Structural similarities and differences between two functionally distinct SecA proteins: the Mycobacterium tuberculosis SecA1 and SecA2. *Journal of bacteriology*
- Tanaka Y, Sugano Y, Takemoto M, Mori T, Furukawa A, Kusakizako T, Kumazaki K, Kashima A, Ishitani R, Sugita Y, Nureki O & Tsukazaki T (2015) Crystal Structures of SecYEG in Lipidic Cubic Phase Elucidate a Precise Resting and a Peptide-Bound State. *Cell Rep* 13: 1561–1568
- Thanassi DG & Hultgren SJ (2000) Multiple pathways allow protein secretion across the bacterial outer membrane. *Curr. Opin. Cell Biol.* 12: 420–430
- Tomkiewicz D, Nouwen N, van Leeuwen R, Tans S & Driessen AJM (2006) SecA supports a constant rate of preprotein translocation. *J. Biol. Chem.* 281: 15709–15713
- Tropea JE, Cherry S & Waugh DS (2009) Expression and Purification of Soluble His6-Tagged TEV Protease. In *Methods in Molecular Biology* pp 297–307. Totowa, NJ: Humana Press
- Trueman SF, Mandon EC & Gilmore R (2011) Translocation channel gating kinetics balances protein translocation efficiency with signal sequence recognition fidelity. *Mol. Biol. Cell* 22: 2983–2993
- Tseng T-T, Tyler BM & Setubal JC (2009) Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol.* 9 Suppl 1: S2

- Uchida K, Mori H & Mizushima S (1995) Stepwise movement of preproteins in the process of translocation across the cytoplasmic membrane of *Escherichia coli*. *J. Biol. Chem.* 270: 30862–30868
- Ulbrandt ND, Newitt JA & Bernstein HD (1997) The *E. coli* signal recognition particle is required for the insertion of a subset of inner membrane proteins. *Cell* 88: 187–196
- Valent QA, Scotti PA, High S, De Gier JW, Heijne Von G, Lentzen G, Wintermeyer W, Oudega B & Luirink J (1998) The *Escherichia coli* SRP and SecB targeting pathways converge at the translocon. *EMBO J* 17: 2504–2512
- Van den Berg B, Clemons WM, Collinson I, Modis Y, Hartmann E, Harrison SC & Rapoport TA (2004) X-ray structure of a protein-conducting channel. *Nature* 427: 36–44
- van der Laan M, Bechtluft P, Kol S, Nouwen N & Driessen AJM (2004) F1F0 ATP synthase subunit c is a substrate of the novel YidC pathway for membrane protein biogenesis. *J. Cell Biol.* 165: 213–222
- van der Laan M, Urbanus ML, Hagen-Jongman Ten CM, Nouwen N, Oudega B, Harms N, Driessen AJM & Luirink J (2003) A conserved function of YidC in the biogenesis of respiratory chain complexes. *Proc Natl Acad Sci U S A* 100: 5801–5806
- van der Wolk JP, de Wit JG & Driessen AJ (1997) The catalytic cycle of the *Escherichia coli* SecA ATPase comprises two distinct preprotein translocation events. *EMBO J* 16: 7297–7304
- Vassilyev DG, Mori H, Vassilyeva MN, Tsukazaki T, Kimura Y, Tahirov TH & Ito K (2006) Crystal structure of the translocation ATPase SecA from *Thermus thermophilus* reveals a parallel, head-to-head dimer. *Journal of Molecular Biology* 364: 248–258
- Vrontou E & Economou A (2004) Structure and function of SecA, the preprotein translocase nanomotor. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*
- Wang H, Ma Y, Hsieh Y-H, Yang H, Li M, Wang B & Tai PC (2014) SecAAA trimer is fully functional as SecAA dimer in the membrane: existence of higher oligomers? *Biochemical and biophysical research communications* 447: 250–254
- Weaver AJ, McDowall AW, Oliver DB & Deisenhofer J (1992) Electron microscopy of thin-sectioned three-dimensional crystals of SecA protein from *Escherichia coli*: structure in projection at 40 Å resolution. *J. Struct. Biol.* 109: 87–96
- Weinkauff S, Hunt JF, Scheuring J, Henry L, Fak J, Oliver DB & Deisenhofer J (2001) Conformational stabilization and crystallization of the SecA translocation ATPase from *Bacillus subtilis*. *Acta Crystallogr. D Biol. Crystallogr.* 57: 559–565
- Woodbury RL, Hardy SJS & Randall LL (2002) Complex behavior in solution of homodimeric SecA. *Protein Sci.* 11: 875–882
- Woodbury RL, Topping TB, Diamond DL, Suci D, Kumamoto CA, Hardy SJS & Randall LL (2000) Complexes between Protein Export Chaperone SecB and SecA: EVIDENCE FOR SEPARATE SITES ON SecA PROVIDING BINDING ENERGY AND REGULATORY INTERACTIONS. *Journal of Biological Chemistry*



- Wowor AJ, Yu D, Kendall DA & Cole JL (2011) Energetics of SecA dimerization. *Journal of Molecular Biology* 408: 87–98
- Xu Z, Knafels JD & Yoshino K (2000) Crystal structure of the bacterial protein export chaperone SecB - Nature Structural & Molecular Biology. *Nat. Struct Biol.*
- Yu D, Wowor AJ, Cole JL & Kendall DA (2013) Defining the Escherichia coli SecA Dimer Interface Residues through In Vivo Site-Specific Photo-Cross-Linking. *Journal of bacteriology* 195: 2817–2825
- Yuan J, Zweers JC, van Dijl JM & Dalbey RE (2010) Protein transport across and into cell membranes in bacteria and archaea. *Cell. Mol. Life Sci.* 67: 179–199
- Zhou J & Xu Z (2003) Structural determinants of SecB recognition by SecA in bacterial protein translocation. *Nat. Struct Biol.* 10: 942–947
- Zimmer J & Rapoport TA (2009) Conformational flexibility and peptide interaction of the translocation ATPase SecA. *Journal of Molecular Biology* 394: 606–612
- Zimmer J, Li W & Rapoport TA (2006) A novel dimer interface and conformational changes revealed by an X-ray structure of B-subtilis SecA. *Journal of Molecular Biology* 364: 259–265
- Zimmer J, Nam Y & Rapoport TA (2008) Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* 455: 936-943

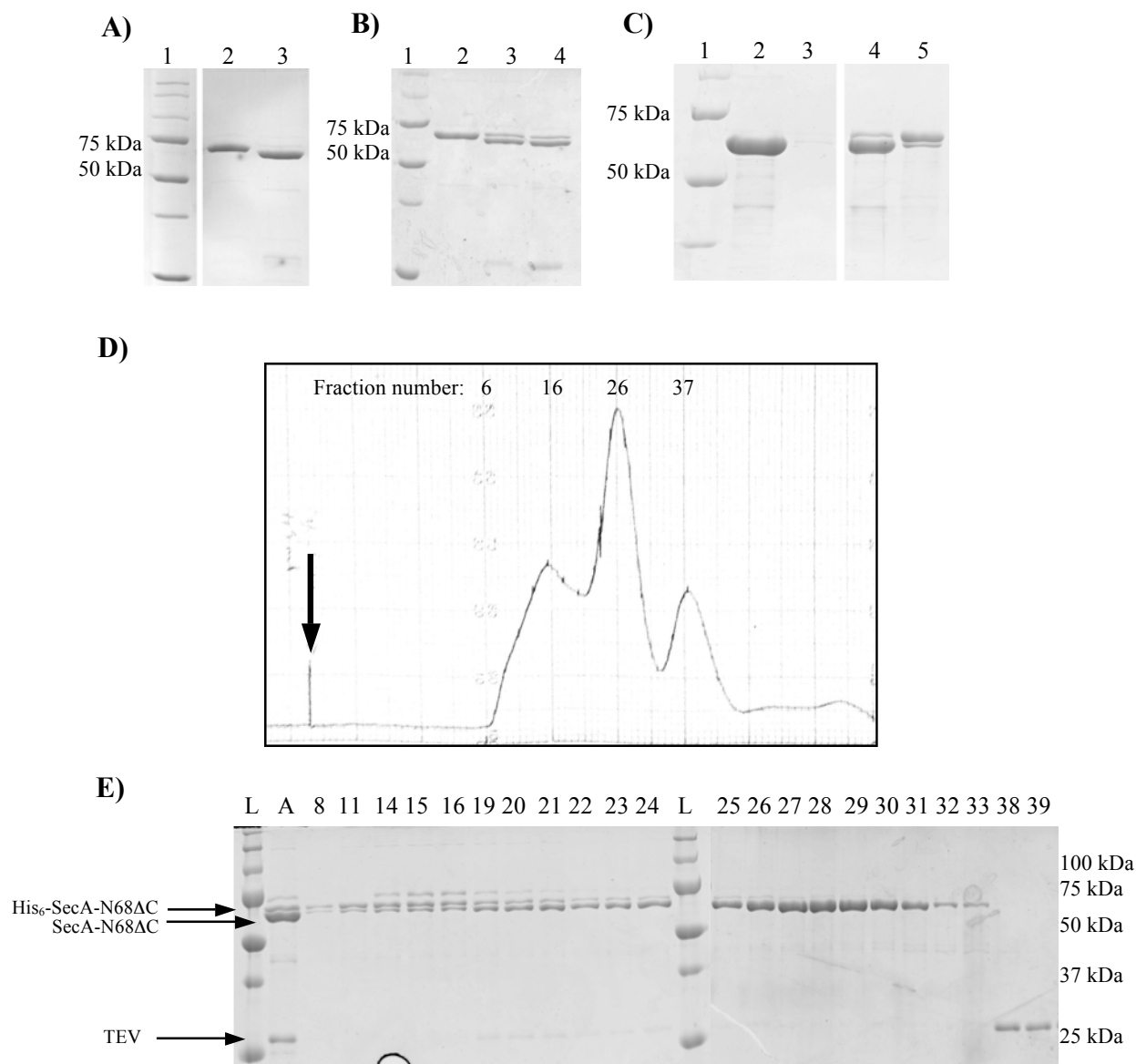
## Appendices

### Appendix A. The effect of the N-terminal segment on purification of SecA constructs

Analysis of SecA constructs with and without the first 14 residues of wild-type SecA, presented in the Chapter 2 and 3, revealed that the extreme N-terminus plays an important role in dimerization and interaction with the unfolded polypeptides. Although the exact roles of the N-terminal sequence of SecA in the translocation mechanism are not clear, previous studies indicated the biological importance of this region of SecA (Jilaveanu *et al*, 2005; Randall *et al*, 2005). During the preparation of SecA constructs, it was noted that the presence of the extreme N-terminus had a major effect on the tendency of SecA to aggregate and/or co-purify with a number of other proteins. To be specific, constructs with the truncated N-terminus were highly soluble and relatively easy to separate from contaminating proteins. On the other hand, constructs with the N-terminus tended to form large complexes and/or insoluble aggregates, and were difficult to resolve from contaminating proteins. As an example, the purification of SecA-N68 $\Delta$ C with SecA-N68 $\Delta$ NC is compared. The details of the procedures are provided in the section 2-2 of this thesis.

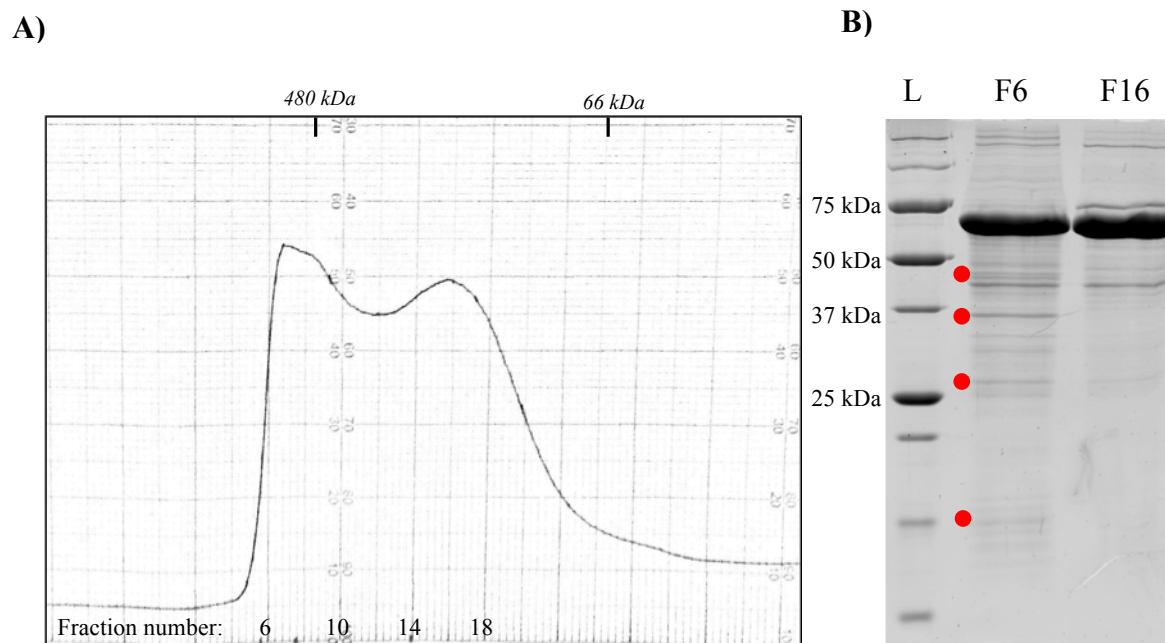
Both SecA-N68 $\Delta$ C and SecA-N68 $\Delta$ NC were expressed from the pProEX-HTa vector to yield SecA constructs with a N-terminal 2 kDa hexahistidine affinity tag, connected via a linker that included a TEV protease cleavage site. SecA-N68 $\Delta$ NC was expressed to high levels and purified by Ni<sup>2+</sup>-NTA affinity chromatography; subsequent removal of the affinity tag by digestion with TEV protease was almost complete within 48 hrs (Figure A-1 A). The resulting untagged construct was purified to homogeneity by ion exchange and size exclusion chromatography (data not shown). SecA-N68 $\Delta$ C, on the other hand, was much more difficult to purify even though the only difference between these constructs was the absence or presence of the SecA N-terminus. In the case of SecA-N68 $\Delta$ C, removal of the affinity tag by TEV protease digestion was highly problematic such that even after several attempts and over 120 hours of incubation, a portion of the protein remained uncleaved (Figure A-1B). In order to separate the digestion products from the proteins that still carried the affinity tag, the digestion reaction was passed over a Ni<sup>2+</sup>-NTA affinity column. Both species of SecA-N68 $\Delta$ C proteins, those that still carried the His-tag and

those with the affinity tag removed, were detected both in the flow-through and in the fraction that bound to the column (Figure A-1C). Subsequent ion-exchange chromatography steps failed to separate these molecules from each other (data not shown). Additional efforts indicated that these molecules could be separated by size exclusion chromatography (Figure A-1D and E), but surprisingly a fraction of the SecA-N68 $\Delta$ C proteins eluted with an apparent molecular weight much larger than expected. Based on these observations, we hypothesized that the Ni<sup>2+</sup>-NTA purified SecA-N68 $\Delta$ C (and other constructs carrying the N-terminal region of SecA) form large protein complexes or soluble aggregates in solution, which prevents complete removal of the affinity tags from these SecA constructs and also affects downstream chromatographic separations. In order to examine this and to separate these protein species, purified SecA-N68 $\Delta$ C, which was not treated with the TEV protease, was injected into the size exclusion chromatography column, and again approximately 50% of the protein eluted with an average molecular weight of several hundred kilodaltons, while the rest of SecA-N68 $\Delta$ C eluted at a volume corresponding to a 150 kDa protein (Figure A-2A). Analysis of these two main peaks by SDS-PAGE indicated that additional protein bands were present in the fractions that eluted with a higher molecular weight (Fraction 16, Figure A-2B), but not in the fractions corresponding to 150 kDa (Fraction 26, Figure A-2B). Based on these results and the effects of the SecA N-terminus on oligomerization and preprotein binding, documented in Chapters 2 and 3 respectively, it seems that the presence of the N-terminal segment in SecA can cause it to interact with other *E. coli* proteins and form large protein complexes in solution. To purify SecA-N68 $\Delta$ C for analysis by analytical ultracentrifugation, after removal of the His-tag, a second round of size exclusion chromatography was used to remove the traces of the large molecular weight species.



**Figure A-1. The effect of the N-terminal segment of SecA on purification of two SecA constructs.**

(A) Ni<sup>2+</sup>-NTA purified SecA-N68 $\Delta$ NC before (lane 2) and after 48 hrs of digestion with TEV protease (lane 3). (B) Ni<sup>2+</sup>-NTA purified SecA-N68 $\Delta$ C before (lane 2) and after TEV digestion for 48 h (lane 3) and 120 h (lane 4). (C) The TEV digested SecA-N68 $\Delta$ NC (lanes 2 and 3) and SecA-N68 $\Delta$ C (lanes 4 and 5) preparations were chromatographed on a Ni<sup>2+</sup>-NTA affinity column. The flow-through is shown in lanes 2 and 4, while proteins eluted with 500 mM imidazole are shown in lanes 3 and 5. (D) After treatment with TEV protease, SecA-N68 $\Delta$ C was chromatographed using a preparative size exclusion column (2.5 x 60 cm, Superdex 200 Prep Grade). The start point of the injection is shown with an arrow and the number of the fractions is indicated at the top of the graph. (E) Fractions obtained from Panel D were analyzed by SDS-PAGE and the number of the fraction is presented above each well. Well A shows the sample applied to the column.



**Figure A-2. Analysis of SecA-N68 $\Delta$ C construct before removal of the affinity tag**

SecA-N68 $\Delta$ C was purified by Ni<sup>2+</sup>-NTA chromatography and concentrated. **(A)** After removal of precipitated protein by centrifugation at 38,000 rpm for 20 min, the SecA-N68 $\Delta$ C preparation was injected onto a preparative size exclusion column (2.5 x 60 cm, Superdex 200 Prep Grade). The absorbance profile of the eluted proteins is shown. The corresponding number of each fraction is presented on the plot. The approximate elution position of the molecular weight standards is shown on top axis of the chromatogram. **(B)** Fractions 6 and 16, obtained from the chromatographic procedure illustrated in Panel A, were analyzed by SDS-PAGE. Protein bands that are evident in Fraction 6 but not present in Fraction 16 are indicated with red dots.

## Appendix B. List of SecA binding peptides identified through oriented peptide library array

Sequence	Protein	RSI	Sequence	Protein	RSI
FIRSMKPRFVKRGAR	FhuD	1.0	TKSLSFCIYDICYAK	SpeD	0.4
SAGINAASPNKELAK	MBP	1.0	KYDIKDVGVNDAGAK	MBP	0.4
TVDEALKDAQTRITK	MBP	1.0	GAGLENIDVGFGLKLS	LamB	0.4
LVVSTLNNPFVSLK	RbsA	1.0	VLFNFKATLKPEGQ	OmpA	0.4
MKKLKLHGfNNLTKS	SpeD	0.9	ADIEVSTCGVISPLK	SpeD	0.4
LLAEITPDKAFQDKL	MBP	0.9	NGTKWWTVGIRPMYK	LamB	0.4
FEQAMQTRVFQPLKL	AmpC	0.9	NIFLVGPMGAGKSTI	AroK	0.4
TAEHTQSVLKGfNKf	LamB	0.8	KVNYGVTVLPTFKGQ	MBP	0.4
TIEQLARTQRDKKR	AroK	0.8	ATPLWQAMPFVRAGR	FhuD	0.4
KVQAKYPVDLKLVVK	RbsA	0.8	VTIRTDQSAKVVAN	AroK	0.4
LKSDVLFNFKATLK	OmpA	0.7	STIEDMARWVQSNLK	AmpC	0.4
LISKGIPADKISARG	OmpA	0.7	YRVRGFTRDINGMKH	SpeD	0.4
VGVLsAGINAASPNK	MBP	0.7	AQSGLLAEITPDKAF	MBP	0.4
EIPALDKELKAKGKS	MBP	0.7	ITPDKAFQDKLYPFT	MBP	0.4
AQYEDFIRSMKPRFV	FhuD	0.7	LPTFKGQPSKPFVGV	MBP	0.4
LIEWLPGSTIWAGKR	LamB	0.7	NNTEMTFQIQRIYTK	SecB	0.4
EAVNKDKPLGAVALK	MBP	0.7	APDRRVEIEVKGIKD	OmpA	0.4
EVVSHIASDNVLGGK	RbsA	0.6	GLTFLVDLIKfNKHMfN	MBP	0.4
ANIPVITLDRQATKG	RbsA	0.6	VGKKFEKDTGIKVTV	MBP	0.4
ENGAYKAQGVQLTAK	OmpA	0.6	EKTLQGGIQLAQsRY	AmpC	0.4
NIDVGFGLSLAATR	LamB	0.6	LREAIQRLEAKGLLL	PdhR	0.4
ETADKVLKGEKVQAK	RbsA	0.6	AMHFVRVLDNAIGGK	FhuD	0.4
ELANVQDLTVRGTKI	RbsA	0.6	YLLTDEGLEAVNKDK	MBP	0.4
LNNPFVSLKDGAQK	RbsA	0.6	VGPMGAGKSTIGRQL	AroK	0.4
NGSDNKIALAARPVK	AmpC	0.6	RREMLPLVSSHRTRI	PdhR	0.4
SRRERSLRLEQRKN	PdhR	0.6	MVWRADTKSNVYGKN	OmpA	0.4
SALTTMMFSASALAK	MBP	0.6	VLASQPADFDRIKGL	RbsA	0.4
LDKELKAKGKSALMF	MBP	0.6	QTRVFQPLKLfNHTWI	AmpC	0.4
QRLEAKGLLLRRQGG	PdhR	0.6	KDVGVDNAGAKAGLT	MBP	0.4
KGYNGLAEVGKKFEK	MBP	0.6	QAMPFVRAGRfQRVP	FhuD	0.4
APAPEVQTKHfTLKS	OmpA	0.6	QLARTQRDKKRPLLH	AroK	0.4
MSGLPLISRRLLTA	FhuD	0.6			
GVTVLPTFKGQPSKP	MBP	0.6			
QCFQTTGAQSKYRLG	LamB	0.6			
ISFEAPNAPHVFQKD	SecB	0.5			
GAKAGLTFLVDLIKfN	MBP	0.5			
YKAQGVQLTAKLGYF	OmpA	0.5			
KHfTLKSDVLFNFK	OmpA	0.5			
VESQRTGDKNfNYKI	LamB	0.5			
VLGGKIAGDYIAKKA	RbsA	0.5			
ALDQLYSQSLNLDPK	OmpA	0.5			
GESNPVTGNTCDNVK	OmpA	0.5			
PGSTIWAGKRfYQRH	LamB	0.5			
MKKTAIAIAVALAGF	OmpA	0.5			
QGIVLATGGGSVKSR	AroK	0.5			
EIADVTIRTDQSAK	AroK	0.5			
IAGDYIAKKAGEGAK	RbsA	0.5			

## Curriculum Vitae

**Name:** Aliakbar Khalili Yazdi

**Post-secondary Education:** PhD in Biochemistry  
The University of Western Ontario  
London, Ontario, Canada  
2012-2016.

MSc in Biotechnology  
Shahid Bahonar University of Kerman  
Kerman, Iran  
2007.

**Honours and Awards:** *Western Graduate Research Scholarship,*  
The University of Western Ontario  
London, Ontario, Canada  
2012-2015

*Full Tuition Waiver,*  
For both BSc & MSc periods  
Ministry of Science Research and Technology  
Iran

**Related Work Experience:** Teaching Assistant  
Biochemistry Laboratory 3380G  
Department of Biochemistry  
The University of Western Ontario  
2013-2015

Research Assistant  
School of Veterinary Medicine  
Shahid Bahonar University of Kerman  
Kerman, Iran  
2008-2011

**Languages:** English: fluent  
Persian/Farsi: native

**Publications:**

- A. Khalili-Yazdi, *et al* (2016) **Preprotein Binding is Linked to Oligomerization of SecA With a Critical Role for the Unstructured SecA N-Terminus**. Manuscript in preparation
- A. Khalili-Yazdi & B. Shilton (2014) **The Role of an Unstructured Region in the Regulation of SecA Activity**. Canadian society for molecular bioscience. 57<sup>th</sup> Annual meeting. *Membrane Proteins in Health and Disease*. Banff, Alberta, Canada.
- A. Khalili-Yazdi & B. Shilton (2013) **Identification of Regions Involved in Self-association of SecA**. *London Health Research Day*. London Ontario, Canada.
- M. Golchin, A. Khalili-Yazdi, *et al.* (2012) **Latex Agglutination Test Based on Single-Chain Fv Recombinant Antibody Fragment**. *Scandinavian Journal of Immunology*, 75: 38–45.
- M. Golchin, F Noori, & A. Khalili-Yazdi (2010) **Staining of an anti-K99 recombinant phage antibody**. *Tehran Univ Med J*, 67: 821-827.